

**PONTIFICIA UNIVERSIDAD JAVERIANA  
FACULTAD DE INGENIERIA  
MAESTRIA EN HIDROSISTEMAS**



**DESARROLLO E IMPLEMENTACIÓN DE PROCESOS DE VIGILANCIA TECNOLÓGICA  
ASOCIADOS AL ESTUDIO DE LOS CICLOS BIOGEOQUÍMICOS E HIDROLÓGICO**

**CARLOS FELIPE GAITÁN OSPINA**

**TRABAJO DE GRADO  
Presentado como requisito parcial  
Para optar al título de**

**MAGISTER EN HIDROSISTEMAS**

**Bogotá, D. C.  
Julio 11 de 2008**

**NOTA DE ADVERTENCIA**  
"La Universidad no se hace responsable por los conceptos emitidos por sus alumnos en sus trabajos de tesis. Solo velará por que no se publique nada contrario al dogma y a la moral católica y por que las tesis no contengan ataques personales contra persona alguna, antes bien se vea en ellas el anhelo de buscar la verdad y la justicia".

Artículo 23 de la Resolución N°13 de julio de 1946.

**DESARROLLO E IMPLEMENTACIÓN DE PROCESOS DE VIGILANCIA TECNOLÓGICA  
ASOCIADOS AL ESTUDIO DE LOS CICLOS BIOGEOQUÍMICOS E HIDROLÓGICO**

**CARLOS FELIPE GAITÁN OSPINA**

**APROBADO**

\_\_\_\_\_  
Nelson Obregón Neira. Ingeniero  
Director

\_\_\_\_\_  
Germán Jiménez. Biólogo  
Jurado 1

\_\_\_\_\_  
Raúl Trujillo Cabezas. Ingeniero  
Jurado 2

\_\_\_\_\_  
Andrés Torres Abello. Ingeniero  
Jurado 3

**DESARROLLO E IMPLEMENTACIÓN DE PROCESOS DE VIGILANCIA TECNOLÓGICA  
ASOCIADOS AL ESTUDIO DE LOS CICLOS BIOGEOQUÍMICOS E HIDROLÓGICO**

**CARLOS FELIPE GAITÁN OSPINA**

---

Francisco Javier Rebolledo Muñoz. Ingeniero  
Decano Académico de la Facultad

---

Nelson Obregón Neira. Ingeniero  
Director de Maestría

*“Nada más difícil de emprender ni más peligroso de conducir que tomar la iniciativa en la introducción de un nuevo orden de cosas, porque la innovación tropieza con la hostilidad de todos aquellos a quienes les sonrió la situación anterior y sólo encuentra tibios defensores en quienes esperan beneficios de la nueva”*  
(Niccoló Machiavelli)

## **AGRADECIMIENTOS**

Gracias al Espíritu Santo y a mis padres por su constante apoyo y compañía. A mi sensei, director, consejero y amigo Nelson Obregón Neira por todas las valiosas enseñanzas y por permitirme soñar con una universidad y una facultad aun más consolidada en el estudio de las ciencias del sistema terrestre. A mis profesores de pregrado por el amor y respeto a la Javeriana y a la ingeniería civil. A todos los docentes de la maestría en Hidrosistemas, por sus valiosos aportes y por la visión holística con la que abordan diferentes casos de estudio. Finalmente a todos mis compañeros de estudio por contribuir con diversos puntos de vista y enriquecer todos los debates.

## **I. Tabla de contenido**

II.	Tablas en el Documento.....	9
III.	Tabla de Figuras.....	12
IV.	Introducción .....	16
V.	Objetivos .....	19
A.	General.....	19
B.	Específicos .....	19
VI.	Estructura del Documento .....	19
VII.	Marco Conceptual .....	21
A.	Antecedentes .....	21
B.	¿Qué es Vigilancia Tecnológica? .....	22
1.	Tipos de Vigilancia.....	29
2.	Tipos de Vigilancia Tecnológica.....	30
3.	Patentes .....	35
VIII.	Marco Teórico .....	50
A.	Antecedentes históricos.....	50
B.	Ciclos Biogeoquímicos.....	55
1.	Introducción .....	55
2.	Descripción .....	56
3.	Conceptos básicos sobre circulación de nutrientes .....	60
4.	Relación entre el Ciclo Hidrológico y el Ciclo Sedimentológico .....	61
5.	CICLO HIDROLÓGICO .....	61
6.	CICLO DEL CARBONO.....	66
7.	CICLO DEL FÓSFORO.....	75
8.	CICLO DEL NITRÓGENO .....	77
9.	CICLO DEL AZUFRE.....	82

10.	CICLO DEL OXIGENO .....	85
11.	CICLO DEL POTASIO .....	87
C.	Introducción al Algoritmo de Redes Neuronales de Kohonen (RNK).....	89
1.	Estructura de la RNK.....	89
2.	Entrada de datos .....	89
3.	Salida de datos .....	90
4.	Procesamiento de la información .....	90
5.	Normalización la Entrada .....	90
6.	Calculo de la salida de cada neurona .....	91
7.	Mapeo de números en formato Bipolar .....	91
8.	Selección de la neurona ganadora .....	92
9.	Entrenamiento de la red .....	92
10.	Razón de aprendizaje .....	92
11.	Ajuste de los pesos .....	93
D.	Mapas Auto – Organizados de Kohonen.....	94
1.	Topologías neuronales .....	96
2.	Arquitectura de la red .....	99
3.	Proceso de entrenamiento.....	101
E.	Diagrama de Voronoi .....	105
F.	Teselaciones de Delaunay .....	105
IX.	Delimitación del problema .....	106
A.	Tipo de Vigilancia Tecnológica a Emplear .....	106
B.	Herramientas de Vigilancia Tecnológica a Incorporar .....	108
X.	Implementación de la Herramienta para la Toma de Decisiones .....	110
A.	Características de la herramienta desarrollada para la creación de SOMs. ....	110
B.	Lenguaje de programación usado .....	110
C.	Entrada de datos .....	110
D.	Pre-tratamiento de la información .....	112

E.	Representación grafica de la información de entrada .....	112
F.	Inicialización y Entrenamiento del SOM .....	114
G.	Determinación de la calidad del Mapa Auto Organizado .....	115
H.	Visualización de Celdas .....	117
XI.	Vigilancia Científica sobre los Ciclos Biogeoquímicos y el Ciclo Hidrológico .....	119
A.	Seguimiento de Publicaciones Científicas y Técnicas.....	120
1.	Pre tratamiento .....	122
2.	Proceso de entrada de datos .....	133
B.	Análisis de patentes .....	138
XII.	Resultados y Discusiones.....	138
XIII.	Conclusiones y comentarios finales .....	143
A.	Conclusiones generales de la investigación .....	144
B.	Conclusiones y comentarios finales desde el trabajo investigativo respecto a la Vigilancia Tecnológica. ....	145
C.	Conclusiones y comentarios finales desde el trabajo investigativo respecto al algoritmo de Kohonen. ....	146
D.	Conclusiones respecto a la aplicación del algoritmo de Kohonen en la Vigilancia Tecnológica. ....	146
E.	Conclusiones respecto a la aplicación del algoritmo de Kohonen en la Vigilancia Tecnológica de los ciclos biogeoquímicos e hidrológico.....	147
XIV.	Recomendaciones generales.....	148
XV.	Bibliografía .....	150

## II. Tablas en el Documento

Tabla 1. Clases Tecnológicas asociadas a los ciclos biogeoquímicos e hidrológico. Clasificación USPTO.....	40
Tabla 2 Total de patentes en las clases tecnológicas asociadas a los ciclos biogeoquímicos e hidrológico.....	47
Tabla 3 Publicaciones por campo tecnológico. Fuente WIPO 2007.....	49

Tabla 4. Elementos de interés y su abundancia en diferentes ambientes (Garcia, Mary Ruth , 2005) .....	59
Tabla 5. Reservorios de agua en el planeta (University of Colorado, Geological Sciences, 2007) .....	63
Tabla 6. Flujos de agua en el planeta (University of Colorado, Geological Sciences, 2007) .....	64
Tabla 7. Tiempos de residencia de agua en el planeta (University of Colorado, Geological Sciences, 2007).....	64
Tabla 8. Reservorios mundiales de Carbono (University of Colorado, Geological Sciences, 2007) .....	69
Tabla 9. Procesos y Flujos del Ciclo del Carbono (University of Colorado, Geological Sciences, 2007) .....	69
Tabla 10. Tiempos de residencia promedio en los reservorios de carbono. (University of Colorado, Geological Sciences, 2007) .....	71
Tabla 11. Reservorios de Fósforo en el planeta (University of Colorado, Geological Sciences, 2007) .....	75
Tabla 12. Procesos y flujos de Fósforo en el planeta (University of Colorado, Geological Sciences, 2007).....	76
Tabla 13. Tiempos de residencia. Ciclo del Fósforo (University of Colorado, Geological Sciences, 2007) .....	77
Tabla 14 Reservorios planetarios de Nitrógeno (University of Colorado, Geological Sciences, 2007) .....	79
Tabla 15. Flujos y Procesos. Ciclo del Nitrógeno (University of Colorado, Geological Sciences, 2007) .....	80
Tabla 16. Tiempos de Residencia en los reservorios. Ciclo del Nitrógeno (University of Colorado, Geological Sciences, 2007) .....	81
Tabla 17. Abreviaturas de procesos relacionados al ciclo del nitrógeno (University of Colorado, Geological Sciences, 2007) .....	82
Tabla 18. Reservorios y tiempos de residencia. Ciclo del Oxígeno (University of Colorado, Geological Sciences, 2007) .....	86
Tabla 19. Códigos RGB de los colores. Ejemplo .....	102
Tabla 20. Características de un archivo .DATA para un vector de entrada dado (ejemplo) .....	111
Tabla 21. Tabla de características estadísticas. (Ejemplo) .....	114
Tabla 22. Indicadores de Calidad del Mapa Auto Organizado. (Datos ejemplo) .....	117
Tabla 23 Registros obtenidos para el criterio BIOGEOCHEMISTRY AND HYDROLOGY, en las diferentes bases de datos consultadas .....	121

Tabla 24. Categorías y Elementos usados para la clasificación de los textos. ....	125
Tabla 25. Valor dimensional para la partícula NUTRIENT .....	127
Tabla 26. Valores dimensionales máximos si se encuentran los caracteres de la entrada de datos en cualquier posición de una palabra .....	127
Tabla 27 Valores dimensionales máximos posibles .....	129
Tabla 28 Familia predominante en cada celda del mapa auto organizado .....	139
Tabla 29 Número de ocurrencias por celda del SOM generado .....	140
Tabla 30 Porcentaje de ocurrencias por celda del SOM generado .....	142
Tabla 31 Distribución de celdas dentro de la clasificación obtenida por el SOM .....	143
Tabla 32 Porcentaje de títulos catalogados en cada familia usando el SOM .....	143

### III. Tabla de Figuras

Figura VII-1 Diagrama de actividades relacionadas a la vigilancia tecnológica y la inteligencia. Adaptada (Fundacion COTEC, 1999) .....	23
Figura VII-2 Qué es la Vigilancia? Mapa conceptual (EAFIT, 2008) .....	24
Figura VII-3 ¿Cuáles son los Tipos de Vigilancia? Mapa Conceptual (EAFIT, 2008) .....	32
Figura VII-4 Frentes de medición de Capacidad tecnológica (Sanchez, 2008).....	33
Figura VII-5 Patentes Tecnológicas en la clase 405 - Ingeniería Hidráulica y de la Tierra - Países con el 1% o más del total de patentes. Periodo 2003-2007. (USPTO, 2008).....	42
Figura VII-6 Número de patentes en la clase tecnológica 405 - Ingeniería Hidráulica y de la Tierra - Primeros tres países por número de patentes periodo 2003-2007. (USPTO, 2008).....	43
Figura VII-7 Número de patentes en la clase tecnológica 405 - Ingeniería Hidráulica y de la Tierra - Países con menos de 20 patentes anuales. Periodo 2003-2007. (USPTO, 2008).....	43
Figura VII-8 Patentes Tecnológicas en la clase 405 - Ingeniería Hidráulica y de la Tierra - Porcentaje de patentes en América del Norte por país. Periodo 2003-2007. (USPTO, 2008)....	44
Figura VII-9 Patentes Tecnológicas en la clase 405 - Ingeniería Hidráulica y de la Tierra - Porcentaje de patentes en Asia por país. Periodo 2003-2007. (USPTO, 2008). .....	44
Figura VII-10 Patentes Tecnológicas en la clase 405 - Ingeniería Hidráulica y de la Tierra - Porcentaje de patentes en Europa por país. Periodo 2003-2007. (USPTO, 2008).....	45
Figura VII-11 Patentes Tecnológicas en la clase 405 - Ingeniería Hidráulica y de la Tierra - Porcentaje de patentes en Oceanía por país. Periodo 2003-2007. (USPTO, 2008).....	45
Figura VII-12 Patentes Tecnológicas en la clase 405 - Ingeniería Hidráulica y de la Tierra - Porcentaje de patentes en Centro y Sur América por país. Periodo 2003-2007. (USPTO, 2008). .....	46
Figura VII-13 Patentes Tecnológicas en la clase 405 - Ingeniería Hidráulica y de la Tierra - Porcentaje de patentes en el Lejano y Medio Oriente por país. Periodo 2003-2007. (USPTO, 2008). .....	46
Figura VII-14 Patentes Tecnológicas en la clase 405 - Ingeniería Hidráulica y de la Tierra - Porcentaje de patentes por región. Periodo 2003-2007. (USPTO, 2008).....	47
Figura VII-15 Número de patentes por región. Periodo 2003-2007. (USPTO, 2008).....	48
Figura VIII-1 Componentes del Ciclo Geológico .....	56
Figura VIII-2 Clasificación de ciclos biogeoquímicos según el tipo de nutrientes .....	57
Figura VIII-3 Interacción entre las diferentes esferas terrestres (esquema) .....	58
Figura VIII-4 Circulación de nutrientes. Adaptado de (University of Colorado, Geological Sciences, 2007).....	60

Figura VIII-5 Interacción de cuerpos de agua en el Ciclo Hidrológico (University of Colorado, Geological Sciences, 2007).....	62
Figura VIII-6 Modelo de Caja. Interacción de Procesos Hidrológicos (University of Colorado, Geological Sciences, 2007).....	63
Figura VIII-7. Ciclo Hidrológico. Se muestra la transferencia de agua entre sus componentes. Adaptado (University of Colorado, Geological Sciences, 2007).....	65
Figura VIII-8 Cambios medios anuales en precipitación, humedad del suelo, escorrentía y evaporación, para el periodo 2080-2099, relativos al periodo 1980-1999. (IPCC, 2007).....	66
Figura VIII-9 Modelo del ciclo del carbono. Adaptado. ( University of Colorado, Geological Sciences, 2007)).....	68
Figura VIII-10 Modelo de caja, Ciclo del Carbono. Adaptado. ( University of Colorado, Geological Sciences, 2007)).....	69
Figura VIII-11 Curva Keeling. Mediciones de CO <sub>2</sub> atmosférico en el monte Mauna Loa, Hawaii. Basada en datos de la NOAA y la UCSD (University Corporation for Atmospheric Research (UCAR), 2007).....	72
Figura VIII-12 Escenarios de concentración de CO <sub>2</sub> y calentamiento de la superficie terrestre, para el siglo 21. (IPCC, 2007).....	73
Figura VIII-13 Estimaciones de la liberación de CH <sub>4</sub> a la Atmosfera. (Instituto de Microbiología Bioquímica, Universidad de Salamanca, 2005).....	74
Figura VIII-14 Reservas globales de metano, flujos y tiempos de residencia. Adaptado. (Universidad de California, Irvine, 2007).....	75
Figura VIII-15 Modelo de caja. Ciclo del fósforo. Adaptado (University of Colorado, Geological Sciences).....	76
Figura VIII-16 Ciclo global de Nitrógeno. Tomado de ( University of Colorado, Geological Sciences, 2007)).....	79
Figura VIII-17 Modelo de Caja. Ciclo del Nitrógeno. Adaptado. (University of Colorado, Geological Sciences, 2007).....	80
Figura VIII-18 Ciclo químico del nitrógeno. Adaptado ( University of Colorado, Geological Sciences, 2007)).....	82
Figura VIII-19 Transformaciones bióticas entre los estados de oxidación del azufre. (García, Mary Ruth , 2005).....	83
Figura VIII-20 Modelo de Caja. Flujos de azufre entre los diferentes reservorios Tamaño de los depósitos en 10 <sup>9</sup> kg y tiempo aproximado de recambio del azufre. Adaptado de (García, Mary Ruth, 2005).....	84

Figura VIII-21 Reservas globales de Azufre, Flujos y Tiempos de Residencia. Adaptado de (Universidad de California, Irvine, 2007).....	85
Figura VIII-22 Reservas globales de oxígeno, flujos y tiempos de residencia. Adaptado de (Universidad de California, Irvine, 2007).....	86
Figura VIII-23 Formas más importantes de potasio en el suelo y procesos relacionados. (Conti, Marta Elvira. Facultad de Agronomía - Universidad de Buenos Aires) .....	88
Figura VIII-24 Estructura típica de una Red Neuronal de Kohonen .....	90
Figura VIII-25 Patrones de Color usados en el ejemplo de Mapas Auto Organizados (Germano, 1999) .....	95
Figura VIII-26 Estado intermedio de organización. SOMs. Usando Self Organizing Map Demo. Iteración 41 .....	95
Figura VIII-27 Clasificación final de colores del SOM, usando SOM demo, luego de mil iteraciones.....	96
Figura VIII-28 Topología rectangular. Ejemplo usando la función gridtop de Matlab 7.1 Var= gridtop (4,5) ) .....	97
Figura VIII-29 Topología rectangular. Usando la función GRIDTOP de Matlab 7.1. (Var= gridtop (5,4) ) .....	98
Figura VIII-30 Topología Hexagonal. Función HEXTOP usando Matlab 7.1 (var=hextop(4,5)) ...	98
Figura VIII-31 Topología Aleatoria. Función RANDTOP usando Matlab 7.1 (var=randtop(4,5)).	99
Figura VIII-32 Arquitectura típica de una RNK. ( <a href="http://www.generation5.org/content/2004/aiSomPic.asp">http://www.generation5.org/content/2004/aiSomPic.asp</a> ) .....	99
Figura VIII-33 Asociación de colores con vectores RGB. ( <a href="http://davis.wpi.edu/~matt/courses/soms/">http://davis.wpi.edu/~matt/courses/soms/</a> ) .....	100
Figura VIII-34 . Grilla de 40 x 40. Ejemplo usando SOM Demo (Buckland, 2005) .....	100
Figura VIII-35 Inicialización de pesos. Ejemplo. ( <a href="http://davis.wpi.edu/~matt/courses/soms/">http://davis.wpi.edu/~matt/courses/soms/</a> )	101
Figura VIII-36 Radios variables de la RNK. (Buckland, 2005).....	103
Figura VIII-37 Distancia entre cuadrados concéntricos. Función boxdist Matlab 7.1. (The MathWorks Inc., 2008).....	104
Figura VIII-38 Ajuste de pesos, según el decaimiento gaussiano, valores en colores cálidos implican un mayor ajuste de pesos en la vecindad MUC.....	104
Figura VIII-39. Teselación de Delaunay. (Wikipedia 2006).....	105
Figura IX-1 Aparte de la Figura VI-3. Tipos de Vigilancia Tecnológica. (EAFIT, 2008) .....	107
Figura IX-2 Ramas de la Vigilancia Tecnológica Científica viables para este estudio. Aparte de la Figura IV-4. (EAFIT, 2008) .....	108
Figura X-1 . Visualización de los datos de entrada. Programa de creación de SOM.....	113

Figura X-2 Graficas auxiliares. Columna 1: Histogramas de Ocurrencias, Columnas 2 y 3: Diagramas de dispersión de las variables. ....	114
Figura X-3 Mapas de inicialización y entrenamiento. De izquierda a derecha: 1) Valores RGB de entrada al SOM, 2) Mapa de Inicialización lineal, 3) Mapa de Inicialización aleatoria, 4) Mapa de entrenamiento con el algoritmo Batch y 5) Mapa de entrenamiento secuencial.....	116
Figura X-4 Representación de Celdas. En la esquina superior izquierda Matriz - U, las otras tres graficas corresponden a los planos de componentes R, G y B. Ejemplo .....	118
Figura X-5 Identificación de la MUC en los planos de componentes y en la Matriz - U .....	119
Figura XI-1. Ejemplo arreglo de caracteres en Matlab.....	123
Figura XI-2. Ejemplo función REGEXP en Matlab .....	124
Figura XI-3. Matriz AUXBIO, con los valores dimensionales para cada una de las palabras de la familia ECO, de los primeros cuatro registros de entrada de datos. ....	131
Figura XI-4. Matrices auxiliares con valores dimensionales para cada elemento de las familias ECO, HYDRO, CHEM y GEO. Primeros 4 elementos analizados .....	132
XI-5. Matriz de pesos para los cuatro primeros elementos analizados, cada columna representa una dimensión vectorial.....	132
Figura XI-6 Graficas relacionadas a la entrada de datos. Se muestra un diagrama de dispersión tridimensional y tres Teselaciones tridimensionales de Delaunay.....	133
Figura XI-7 Histogramas y Diagramas de Dispersión de las cuatro dimensiones utilizadas.....	134
Figura XI-8 Información estadística de los datos de entrada.....	135
Figura XI-9 Inicialización de pesos para el SOM, usando procesos lineales (centro) y aleatorios (derecha) .....	136
Figura XI-10. Mapas de entrenamiento utilizando el algoritmo BATCH (centro) y el SECUENCIAL (derecha). Valores de entrada (izquierda) .....	136
Figura XI-11 SOMs y Diagramas de Ocurrencias para el caso de estudio. ....	137
Figura XII-1 Matriz - U para el caso de estudio. ....	139
Figura XII-2 Histograma de ocurrencias en la Matriz - U. Grandes hexágonos blancos representan mayores ocurrencias como MUC para una celda.....	141
Figura XII-3. Matriz de distancias vectoriales para los vectores organizados por el SOM.....	142
Figura XIV-1. Interfaz gráfica del programa Kvisu. En el ejemplo mostrado se buscó HYDROLOGY se encontraron 704.280 resultados y se creó un mapa temático en la parte derecha de la figura. (KVISU, 2007).....	148

## IV. Introducción

A través de la historia, la Tierra ha soportado una constante variabilidad de sus cuatro esferas fundamentales: la biósfera, la hidrósfera, la litósfera y la atmósfera. El hombre ha vivido algunos de estos cambios naturales, como periodos glaciares e interglaciares, movimientos de masas tectónicas, y extinciones de especies; sin embargo, a pesar de estos cambios, ha logrado adaptarse al medio, y superar los inconvenientes generados por los mismos. No obstante, en la actualidad, la especie humana, se encuentra en una posición en la que el cambio global, generado por el hombre al sistema terrestre, está afectando directamente a la especie y a las diferentes esferas fundamentales, en especial las interacciones de los ciclos biogeoquímicos (carbono, fósforo, nitrógeno, oxígeno, azufre y potasio, entre otros), con el ambiente y con el ciclo hidrológico.

Ante este panorama y dadas las interacciones existentes entre los sistemas físicos, biogeoquímicos y socio económicos, la comunidad científica internacional desarrolló en la última década del siglo pasado, los primeros modelos integrados que tenían en consideración dichas interacciones. Sin embargo, estos modelos no cuentan con el detalle y la complejidad necesaria, para ser considerados globales; motivo por el cual, se inició el desarrollo de modelos que funcionaran a menor escala, pero que consideraran la globalidad del sistema terrestre. Estos modelos, según el Programa Internacional Geósfera-Biósfera<sup>1</sup>, estarán disponibles en la próxima década, y tienen como finalidad, procurar adelantarse a los cambios y predecir el comportamiento del sistema terrestre, y las interacciones entre sus elementos

De esta forma, con el fin de estudiar los cambios en las diferentes esferas, es necesario crear y validar modelos acoplados, así como desarrollar escenarios de simulación, basados en información obtenida mediante sensores remotos, y el estudio y monitoreo de los ciclos biogeoquímicos e hidrológico. En estos modelos, se busca determinar posibles respuestas del sistema ante perturbaciones de origen natural, como erupciones volcánicas o de origen antrópico, como las crecientes emisiones de dióxido de carbono a la atmósfera.

El deseo de estudiar el sistema como un todo, creó la necesidad de obtener información, sobre los diferentes componentes de los ciclos biogeoquímicos e hidrológico, y de desarrollar la instrumentación adecuada para realizar mediciones sobre los mismos; esta tendencia ha crecido desde los años setenta, y actualmente se cuenta con instrumentos especializados que miden casi todos los procesos y variables involucradas en cada uno de los ciclos; es así como

---

<sup>1</sup> Programa desarrollado por el Concejo Internacional para la Ciencia en 1986 <http://www.icsu.org/index.php>

se han desarrollado aparatos que miden variables tan disímiles como la radiación solar, la fotosíntesis, los campos de precipitación sobre la tierra, el contenido de carbono, fósforo o nitrógeno en el aire, agua o tierra, la altura de la capa límite atmosférica y las temperaturas y circulaciones oceánicas, entre otras.

Sin embargo, la invención de todos estos dispositivos y el desarrollo de los modelos de circulación planetaria, requieren la asignación de grandes recursos humanos y financieros, que permitan avanzar constantemente en el estado del conocimiento. Es así como Estados Unidos, Japón, China y Alemania, invirtieron en el 2001, en Investigación y Desarrollo (I+D), 282.000 millones, 104.000 millones, 60.000 millones y 54.000 millones de dólares respectivamente, según información de la Organización de Cooperación y Desarrollo Económicos (OCDE)<sup>2</sup>.

Ante este escenario, y dada la limitación de recursos económicos en los países en vía de desarrollo, para invertir en Investigación y Desarrollo de tecnología, es necesario optimizar el uso de los recursos, y no invertir en “inventar” tecnologías y desarrollos que ya se han creado, ya que, sólo por citar un ejemplo, la Oficina Europea de Patentes<sup>3</sup> manifiesta que se registran pérdidas de veinte mil millones de dólares en la Unión Europea por desarrollos y procesos ya inventados. Para reducir estas pérdidas, la solución más simple consiste en investigar el estado actual de la tecnología o el desarrollo de interés a nivel mundial, siendo clave el análisis de la información, obtenida en bases de datos especializadas en publicaciones científicas y el análisis de información en las diferentes oficinas de patentes. Estas actividades, las trata la **Vigilancia Tecnológica**, herramienta de gestión asociada con acciones de observación, captación y análisis de información, para convertir señales dispersas en tendencias y recomendaciones para tomar decisiones.

Una vez expuestos los anteriores argumentos, y dado que Colombia sólo ha registrado 35 patentes en los últimos cinco años<sup>4</sup>, es necesario preguntarse si actualmente el país, o los grupos de investigación en las universidades, realizan estas actividades o si se están invirtiendo los pocos recursos en desarrollos ya realizados. Este cuestionamiento no busca limitar la inventiva local, sino direccionarla hacia campos y desarrollos en los que sea más competitiva. De otra parte, el sistema de patentes a nivel mundial está diseñado de forma tal, que la investigación sea uno de sus pilares; motivo por el cual la legislación establece ciertas excepciones, en las que no se puede ejercer el derecho sobre la patente, como cuando la invención se utiliza en el ámbito privado y con fines no comerciales; cuando su uso es con

---

<sup>2</sup> <http://www.oecd.org/>

<sup>3</sup> [www.epo.org](http://www.epo.org)

<sup>4</sup> según la Oficina de Patentes y Marcas de Estados Unidos USPTO [www.uspto.gov/](http://www.uspto.gov/)

fin de experimentación; o cuando se usa con fines de enseñanza o investigación científica o académica. Esto permite a otros actores diferentes a los inventores y desarrolladores de la patente, conocer, analizar y comprender la innovación; siendo justamente esas excepciones, las que deben aprovechar el país y los grupos de investigación, para avanzar en el conocimiento y comprender los últimos desarrollos realizados por países con presupuestos mucho más generosos para el desarrollo de nuevas tecnologías.

Lo anterior, sumado a la inexistencia en nuestro medio de trabajos sobre el estado de las patentes relacionadas a los campos tecnológicos afines a los ciclos biogeoquímicos e hidrológico, así como la ausencia de una herramienta que facilite la agrupación y visualización de información proveniente de bases de datos, llevó a plantear la necesidad de realizar un desarrollo que permitiera, tanto al país como a la comunidad de investigadores, avanzar en el estado del conocimiento de los ciclos biogeoquímicos e hidrológico, mediante la incorporación de una herramienta que permitiera catalogar las publicaciones en el tema, existentes en bases de datos especializadas, así como realizar una primera contribución, sobre las tendencias mundiales de patentes, en los campos tecnológicos relacionados.

Respecto a la herramienta desarrollada, se optó por emplear en la misma, características del naciente campo de Inteligencia Artificial, específicamente en el área de las Redes Neuronales Artificiales, y nuevas aplicaciones encontradas en dicha área, como el aprendizaje no supervisado, y la creación de mapas auto organizados; aplicaciones que son abordadas utilizando el algoritmo de Kohonen. Este algoritmo ha sido usado en el país, para la detección de arritmias cardiacas, la clasificación de aminoácidos, segmentación psicográfica, identificación de patrones económicos, y la identificación de descargas atmosféricas. No obstante, este trabajo es el primero en el ámbito nacional, que implementa el algoritmo en clasificación de datos multidimensionales, no numéricos. Específicamente, se utilizó el algoritmo de Kohonen, para la clasificación de publicaciones técnicas, relacionadas con el ciclo hidrológico y los ciclos biogeoquímicos, en cuatro grandes clústeres de información.

Vale la pena notar que el aporte sobre el conocimiento de los ciclos biogeoquímicos, es realizado desde la óptica de un ingeniero civil, quien no recibió información relacionada a los mismos, en ningún curso a nivel de pregrado o de posgrado. Es sin embargo, un primer aporte a quienes como el autor, no contaron con dicha formación, y sienten una creciente necesidad no sólo, de conocer más sobre las ciencias del sistema terrestre, en especial sobre las interacciones del hombre con las esferas que lo rodean, sino también busquen conocer los conceptos básicos, y las actividades relacionadas a la Vigilancia Tecnológica, así como obtener

herramientas que les permitan comprender y desarrollar en el futuro mapas auto organizados, usando el algoritmo de Kohonen.

## **V. Objetivos**

Este capítulo, tiene como fin describir los objetivos del presente trabajo.

Dada la naturaleza de esta investigación, donde se trata temas tan variados como Vigilancia Tecnológica (VT), los ciclos biogeoquímicos e hidrológico (CBGQH) e Inteligencia Artificial (IA), es necesario desarrollar una estructura de objetivos generales al trabajo de investigación, y específicos a cada uno de estos temas.

### **A. General**

- Contribuir a los grupos de investigación y entes dedicados al estudio de los ciclos biogeoquímicos e hidrológico, con una herramienta de Vigilancia Tecnológica que permita la clasificación de información de Bases de Datos especializadas.

### **B. Específicos**

- Implementar el algoritmo de Kohonen para la creación de clústeres de información relacionada a los ciclos biogeoquímicos e hidrológico.
- Desarrollar mapas auto organizados de la información obtenida en las bases de datos.
- Visualizar el comportamiento de las patentes en los campos tecnológicos asociados a los ciclos biogeoquímicos e hidrológico.

## **VI. Estructura del Documento**

El presente documento, tiene como finalidad mostrar los aspectos conceptuales y funcionales relacionados al desarrollo e implementación de procesos de vigilancia tecnológica asociados al estudio de los ciclos biogeoquímicos y el ciclo hidrológico.

En este orden de ideas, se presentará un marco conceptual, que oriente al lector sobre el alcance y definición de vigilancia tecnológica y buscará resolver inquietudes comunes como ¿Qué tipos de vigilancia tecnología existen? , ¿Qué herramientas son usadas en los procesos de vigilancia tecnológica?, y ¿En qué se diferencia la vigilancia tecnológica de otras actividades, como el Benchmarking y la Prospectiva Empresarial? Posteriormente, en el marco teórico, se expondrá el concepto de ciclo biogeoquímico, su importancia y relación con el ciclo geológico,

así como las principales características de los seis grandes ciclos biogeoquímicos y del ciclo hidrológico.

El capítulo nueve, define el caso de estudio, los tipos de vigilancia tecnológica a emplear, las herramientas que se usaran para tal fin y los ciclos biogeoquímicos que se estudiarán; luego en el capítulo diez, se enseñará la herramienta escogida para la toma de decisiones, sus bases conceptuales y características generales; adicionalmente, en este capítulo se desarrolla un ejemplo de aplicación, como muestra de la versatilidad de la herramienta. A continuación, en el capítulo once, se aplica la herramienta desarrollada, en el estudio de los ciclos biogeoquímicos y el ciclo hidrológico. Finalmente, se presentan las conclusiones y comentarios finales sobre este trabajo de investigación.

Adicionalmente, se presenta el tomo 2 que incluye los anexos relacionados a la información obtenida de la USPTO, sobre los campos tecnológicos asociados al estudio de los ciclos biogeoquímicos e hidrológico, así como un documento de patente completo sobre un aparato y un método para analizar un gas, sujeto a ser estudiado y a determinar su posible fabricación en el medio. Por último, dicho tomo presenta un listado de patentes obtenidas de la base de datos INvenciones EN ESpañol (INVENES)<sup>5</sup>, de la Oficina Española de Patentes y Marcas.

Se espera que este trabajo, facilite el proceso de adopción de tecnologías asociadas al estudio de los ciclos biogeoquímicos e hidrológicos, al suministrar una herramienta valiosa para catalogar grandes volúmenes de información.

---

<sup>5</sup> <http://www.oepm.es/>

## VII. Marco Conceptual

Con el fin de determinar los procesos de Vigilancia Tecnológica que serán involucrados en el estudio de los ciclos biogeoquímicos e hidrológico, se desarrollará a continuación el Marco Conceptual, cuyo fin es clarificar conceptos relacionados a la actividad de Vigilancia Tecnológica, como : ¿Qué es?, ¿Cual es su finalidad?, ¿Quién debe practicarla? y ¿Qué tipos de Vigilancia Tecnológica existen?; para luego en el capítulo siguiente, introducir las bases teóricas relacionadas a los ciclos biogeoquímicos.

### A. Antecedentes

La Vigilancia Tecnológica (VT), es una práctica común que la humanidad ha realizado desde siempre, en función de las herramientas disponibles en cada momento de la historia; sin embargo, la Vigilancia Tecnológica presentó un auge en la última década del siglo pasado a causa de la aceleración presentada en todas las áreas de la tecnología y de los avances logrados en las tecnologías de información y comunicaciones.

Esta práctica, aparece como una de las seis funciones claves para la gestión de tecnología, según el trabajo de 1985, del ingeniero francés Morin, donde para mantener la vanguardia tecnológica es esencial la detección de cambios y nuevas tecnologías, con la suficiente antelación para poder evaluarlas, prepararse, adoptarlas y explotarlas.

Según Peter Drucker, experto en gerencia empresarial, *“la mayoría de innovaciones con éxito para el innovador explotan el cambio, y es dicho cambio lo que siempre proporciona la oportunidad o lo nuevo y diferente”*; más concretamente, menciona que la innovación sistemática está ligada a la vigilancia en cuanto la primera requiere la búsqueda organizada y con un objetivo, de cambios en el entorno y del análisis de sus implicaciones. Todos ellos elementos característicos de la vigilancia.

A nivel mundial, se tiene que en Japón, el presupuesto medio destinado por las empresas para captación y valorización de información, tanto externa como interna, alcanza el 1,5% del volumen de facturación; en Francia, varios estudios hablan de una dedicación en torno al 2% de la cifra de facturación, a la búsqueda de información y para muchos directivos de pymes francesas la no identificación de personas dedicadas en mayor o menor medida a la vigilancia, así como el no reconocimiento de un cierto presupuesto dedicado a la misma es sinónimo de una vigilancia desorganizada. (Fundacion COTEC, 1999)

El sistema de vigilancia de una empresa, puede ir desde la organización de pequeñas reuniones en las que se vierten todos los hechos detectados y con posible significado para la empresa,

hasta sistemas con fuentes de información sofisticadas y sistemas de comunicación e intercambio de información de estructura bien definida.

## **B. ¿Qué es Vigilancia Tecnológica?**

En términos generales, la vigilancia tecnológica es una herramienta de gestión que permite reducir el riesgo en la toma de decisiones. La vigilancia, debe basarse en la captación, análisis, síntesis y utilización de la información pública existente, se encuentre o no formalizada en algún medio. Pública en este contexto, se entiende como toda aquella información que puede ser obtenida por medios legítimos. (Combs, 2006) Es importante notar, que la vigilancia tecnológica no implica labores de espionaje o el desarrollo de prácticas para obtener información clasificada o de uso privado. (Fundacion COTEC, 1999)

La Universidad de Navarra, en la implementación de su plan tecnológico, resume los objetivos generales de la vigilancia tecnológica, en dos (2) puntos:

- Captar e informar puntualmente de ciertas cuestiones, en cuanto ocurran ciertos eventos, o en cuanto se publiquen ciertos documentos. (Plan Tecnológico de Navarra, 2007)
- Analizar las informaciones que van apareciendo de modo continuo con criterios y objetivos definidos. (Plan Tecnológico de Navarra, 2007)

En la Figura VII- 1, se aprecia cómo la vigilancia tecnológica está asociada con acciones de observación, captación y análisis de información, para convertir señales dispersas en tendencias y recomendaciones para tomar decisiones. A diferencia del término inteligencia corporativa, la vigilancia tecnológica no se centra en las fases de difusión del análisis y su uso en el proceso de decisión (Fundacion COTEC, 1999).

En algunos países francófonos, se implementa el término vigilancia económica, como la práctica de la vigilancia en todos los niveles de la empresa, ya sea a nivel tecnológico, comercial, jurídico, financiero y estratégico, con carácter estratégico para la empresa.

Los procesos relacionados a la actividad de vigilancia tecnológica, son principalmente los siguientes:

- Aceleración en todas las áreas de la tecnología: La aplicación de avances desarrollados para otro sector y adaptarlos a los intereses de la organización, es cada vez más

relevante y económico que desarrollar tecnologías originales. Análogamente, se debe garantizar que las tecnologías utilizadas por la organización no sean obsoletas.

- Avances en las tecnologías de información y comunicaciones: Actualmente, existen muchas formas de automatizar y captar información, así como herramientas potentes de filtrado y análisis de datos; herramientas que facilitan los procesos de vigilancia tecnológica.



Figura VII-1 Diagrama de actividades relacionadas a la vigilancia tecnológica y la inteligencia. Adaptada (Fundación COTEC, 1999)

Entre los objetivos concretos de la vigilancia tecnológica, se tienen los siguientes (Plan Tecnológico de Navarra, 2007):

1. Detectar los cambios en las tecnologías, mercados, competencia y legislación.
2. Reducir riesgos y detectar amenazas.
3. Detectar las necesidades del cliente y como suplirlas con un determinado producto
4. Determinar brechas entre las capacidades propias y las de los competidores
5. Innovar
6. Integrar (clientes, expertos, socios, desarrolladores, etc.)

Esquemáticamente, la actividad de Vigilancia se representa en la Figura VII-2:

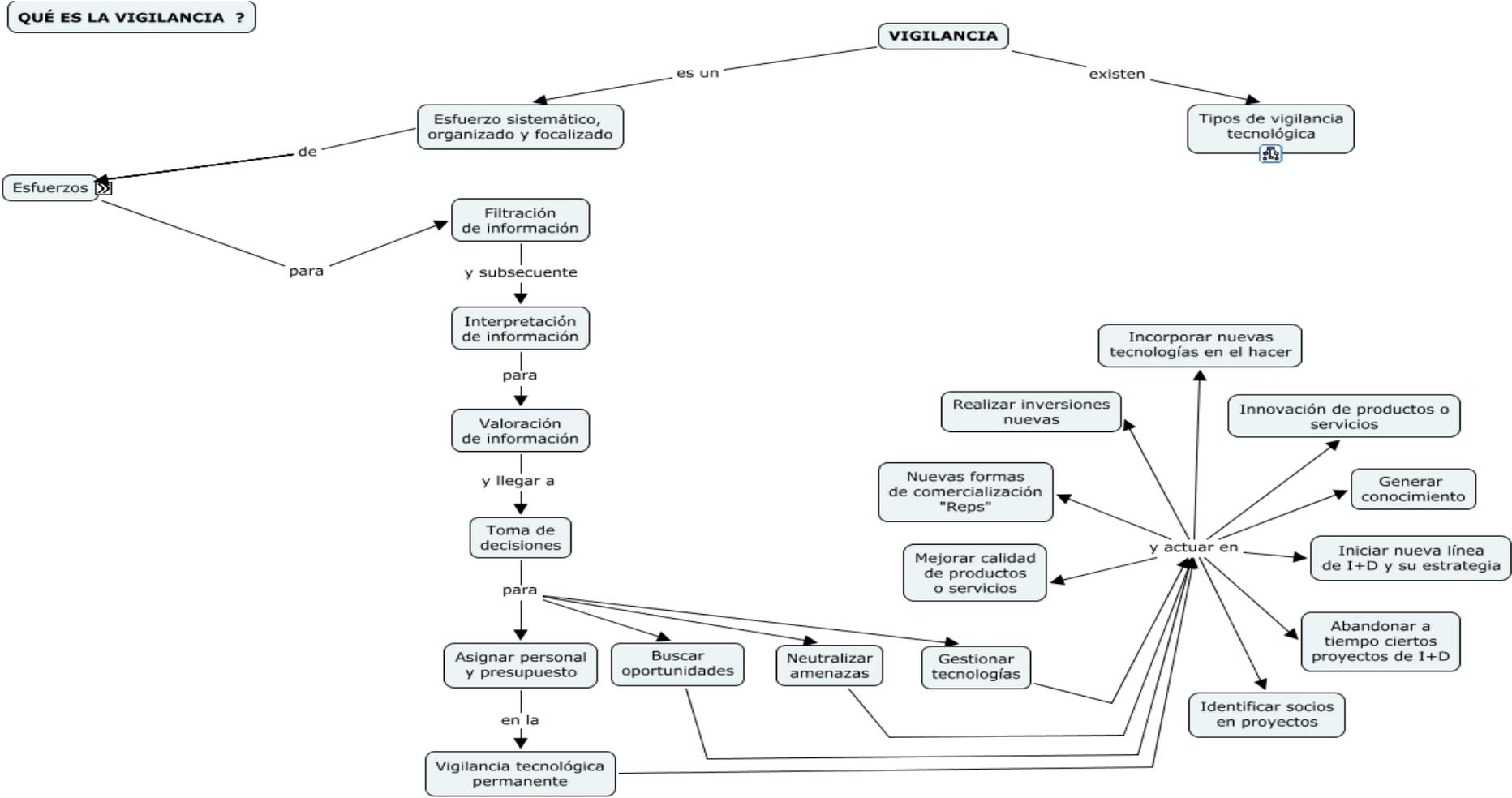


Figura VII-2 Qué es la Vigilancia? Mapa conceptual (EAFIT, 2008)

A continuación, se responderá en base al **Plan Tecnológico de la Universidad de Navarra**<sup>6</sup>, una serie de interrogantes relacionados a la vigilancia tecnológica y otras herramientas de gestión de conocimiento.

### **¿Quién debe practicar la Vigilancia Tecnológica?**

La vigilancia tecnológica es realizada por cualquier organismo interesado en mantener un producto o tecnología en el mercado; esta práctica no necesariamente se efectúa de manera consciente y sistémica; sin embargo, implica tomar decisiones sobre las prestaciones, características y precio del producto, en aras de posicionarlo frente a otros productos o tecnologías con las que compite. En el ámbito empresarial, la vigilancia tecnológica garantiza la existencia de métodos y herramientas, para captar y analizar los cambios del entorno, otorgándole más posibilidades de adaptarse y reaccionar ante entornos cambiantes.

### **¿Se debe realizar la Vigilancia Tecnológica a corto o mediano plazo?**

Realmente, no existen restricciones sobre el plazo en el que debe realizarse la vigilancia tecnológica, siendo efectiva en cualquier horizonte de tiempo, que permita maximizar sus ventajas, esto significa que la vigilancia tecnológica, aunque puede emplearse en el largo plazo, no sería tan efectiva como al realizarse en horizontes menores.

Bajo este supuesto, la vigilancia tecnológica se divide en dos tipos principalmente:

- Vigilancia Tecnológica de tipo operativo: Es la encargada de los estudios a corto plazo.
- Vigilancia Tecnológica de tipo estratégico: Encargada de los estudios a mediano plazo.

### **¿Existen otros tipos de vigilancia?**

Desde el interior de una empresa u organización, los procesos de vigilancia pueden estar compuestos de varios componentes, entre los que se distinguen los siguientes:

- Vigilancia Tecnológica
- Vigilancia de la Competencia
- Vigilancia Comercial
- Vigilancia de Regulaciones

Como se mencionó anteriormente, la vigilancia tecnológica se ocupa de nuevos desarrollos tecnológicos que puedan afectar a la organización; de otra parte, la vigilancia de competencia

---

<sup>6</sup> <http://www.plantecnologico.com/es/herramientas/vigilancia/documentacion/112.php>

se encarga de analizar las adquisiciones y fusiones empresariales, inversiones y nuevas plantas de producción de la competencia; la vigilancia comercial, trata sobre la detección de cambios en las estrategias de distribución y comerciales, por último la vigilancia de regulaciones, se refiere a la evolución de las normas y regulaciones de obligado cumplimiento por los productos de la empresa (Plan Tecnológico de Navarra, 2007).

### **¿En qué se diferencia la Inteligencia Tecnológica y Económica de la Vigilancia Tecnológica y Económica?**

La expresión “*Inteligencia*” es un término anglosajón, que abarca conceptos más modernos y desarrollados que los comprendidos por el término de uso europeo “*vigilancia*”; usualmente, se relaciona la inteligencia al ámbito empresarial en general, sin hacer énfasis en la tecnología, esta relación, llevó a la utilización del término Inteligencia Empresarial.

Dados los alcances de la Inteligencia tecnológica, esta supera a la Vigilancia Tecnológica en tres aspectos fundamentales:

- La inteligencia tecnológica, abarca todo tipo de contenidos, siempre que sean de interés para la organización.
- Se acepta la captación de información de fuentes formales y no formales, como una conversación.
- Incluye tratamientos adicionales de la información, como asimilación de contenidos, toma de decisiones y protección de resultados.

Citando a Michael E. Porter, de la Escuela de Negocios de Harvard, la Inteligencia Empresarial, *"Es la metodología que tiene como objetivo dar la información correcta a la persona correcta en el momento correcto para tomar la decisión correcta"*. (Combs, 2006)

### **¿Cuál es la diferencia entre Vigilancia Tecnológica y la Prospectiva Tecnológica?**

Según el primer documento COTEC, sobre oportunidades tecnológicas, la Prospectiva Tecnológica, estudia métodos y estrategias para intentar predecir con cierto nivel de confianza posibles estados futuros de la tecnología y su influencia en la organización, sector industrial o en la sociedad en general.

Existen tres diferencias fundamentales entre estas herramientas (Fundacion COTEC, 1999):

- El plazo: La Prospectiva Tecnológica se centra en el análisis del mediano y largo plazo, mientras que la Vigilancia Tecnológica está enfocada al corto y mediano plazo.

- La Metodología: La Vigilancia Tecnológica se basa en la cooperación de un grupo de personas dentro de una organización, con soportes puntuales de expertos externos a la misma; adicionalmente, se basa en un procedimiento sistemático de búsqueda, filtrado, evaluación y análisis de información para soportar la toma de decisiones. De otro lado, la Prospectiva Tecnológica, usa varias metodologías, siendo las más común, lograr consensos entre un grupo multidisciplinario de expertos, sobre los escenarios más probables en relación con un tema específico.
- Su Aplicación o uso: La Vigilancia Tecnológica, tiene un mayor interés por las empresas, mientras que la Prospectiva Tecnológica, es más utilizada para establecer políticas tecnológicas; sin embargo, ambas pueden utilizarse indistintamente en cualquier uso.

Los resultados obtenidos luego de aplicar Prospectiva Tecnológica, constituyen una fuente de información muy valiosa para alimentar el sistema de vigilancia de una organización.

La Prospectiva constituye un ejercicio sólidamente fundamentado para analizar el curso esperable de los desarrollos científico- tecnológicos, interpretar las oportunidades que ofrecen y confrontarlos con las preferencias de la sociedad y las empresas, para apoyar la toma de decisiones. (Cabezas, et al., 2007)

### **¿Cuál es la relación entre Vigilancia Tecnológica y Gestión del Conocimiento?**

Estas dos actividades están interrelacionadas, al estar ambas soportadas en documentos que coexisten en la Intranet; teniendo la Vigilancia Tecnológica, un enfoque complementario a la Gestión de Conocimiento, ya que los documentos estudiados por la Vigilancia Tecnológica, corresponden a señales externas sin analizar o en fase de análisis, mientras que los documentos de la Gestión del Conocimiento, corresponden a documentos que incluyen procedimientos y detalles internos de la empresa; documentos que son el resultado del análisis realizado a la información obtenida en la Vigilancia Tecnológica y de haber tomado las decisiones correspondientes (Plan Tecnológico de Navarra, 2007).

### **¿Qué relación tiene la Vigilancia Tecnológica con las búsquedas en las bases de datos?**

Las Bases de Datos, solo son *herramientas* en servicio de la Vigilancia Tecnológica, metodología que utiliza todas las herramientas posibles para llevar a cabo sus objetivos.

### **¿Qué tipo de información es analizada en la Vigilancia Tecnológica?**

Generalmente, se analizan patentes, bibliografía técnica, proyectos de Investigación y Desarrollo, ofertas tecnológicas, tesis doctorales, congresos etc.

### **¿Cuáles son las ventajas de la Vigilancia Tecnológica?**

Las ventajas más notorias son (Plan Tecnológico de Navarra, 2007):

- Mejora del conocimiento de la propia organización (capacidades tecnológicas, relaciones de dependencia entre tecnologías/normas/productos, debilidades y fortalezas respecto a clientes y proveedores)
- Permite identificar nuevos competidores potenciales.
- Analiza fortalezas y debilidades de los competidores conocidos.
- Identifica posibles socios para cooperar, detectar nuevos productos o desarrollos y crear redes de colaboración.
- Mejora la gestión de Investigación y Desarrollo al:
  - Definir socios tecnológicos en la negociación de licencias.
  - Mejorar la selección y comparación de tecnologías para realizar una inversión.
  - Definir las líneas de I+D, en las que hay que seleccionar socios estratégicos.
  - Definir las líneas de I+D, para las que cuenta la organización con recursos propios suficientes.

### **¿Cuál es la diferencia entre Vigilancia Tecnológica y Benchmarking?**

El Benchmarking, generalmente está orientado a un aspecto o función determinada, como la mejora continua de los procesos de calidad de una empresa, a su vez, está centrado en un periodo de tiempo determinado; a diferencia de la Vigilancia Tecnológica, que es continua en el tiempo y está ligada a los aspectos estratégicos de de la organización; permitiendo incluso, en determinado tiempo, establecer cuales empresas u organizaciones son objetivos potenciales para realizar benchmarking (Plan Tecnológico de Navarra, 2007).

### **¿Qué es Inteligencia Competitiva?**

La inteligencia competitiva es el proceso por el cual las organizaciones recopilan y utilizan la información sobre los productos, clientes, y los competidores, para su planificación a corto y largo plazo. Realizando la colección sistemática de la información abierta, que una vez compaginada y analizada proporciona una mejor comprensión de la estructura, de la cultura, del comportamiento, de las capacidades, y de las debilidades de una firma del competidor (Van der Reis, 2007).

La inteligencia competitiva es una actividad de suma importancia para porque le ayuda a las empresas a entender mejor cómo funciona el negocio. De esta manera puede aprender a ser mejor que sus competidores. (Van der Reis, 2007)

### **¿Qué relación tiene la Vigilancia Tecnológica con la Inteligencia Competitiva?**

A finales de los años ochenta, empezó a desarrollarse una actividad llamada vigilancia tecnológica, que consiste en el análisis del entorno tecnológico de la empresa. En conjunto, dicha actividad de vigilancia intenta sistematizar la recogida de información. A mediados de los años noventa, sin embargo, se abandona el concepto de vigilancia y se empieza a hablar de inteligencia competitiva: se trata del proceso de obtención, análisis, interpretación y difusión de la información con valor estratégico sobre la industria y los competidores, que se transmite a los responsables de la toma de decisiones en el momento oportuno. (Maspons, et al., 2007)

### **¿Qué diferencia existe entre la Vigilancia Tecnológica y la Inteligencia Competitiva?**

Básicamente, aunque a veces es difícil marcar fronteras, podemos decir que a la inteligencia se le atribuye un carácter mucho más activo. Así como la vigilancia podría ser un control bastante pasivo de lo que está sucediendo en el entorno, cuando hablamos de inteligencia queda clara la idea de que la información sobre cómo está evolucionando nuestro entorno se utilizará en la toma de decisiones. Por lo tanto, debe tratarse, validarse en términos de calidad y de fiabilidad, para ayudar en la toma de decisiones. (Maspons, et al., 2007)

El concepto americano va un poco más allá, porque en Estados Unidos el concepto de inteligencia también se ha desarrollado en algunos sectores mediante el concepto *lobby*. En cierta medida, una de las actividades del *lobby* es el desarrollo de actividades y de inteligencias, estar atento a todo aquello que lo pueda afectar y utilizar esta información para seguir ejerciendo la actividad su actividad. En Europa, en la medida en que los lobbies funcionan de forma diferente y no tienen el mismo poder que en Estados Unidos, es común pensar que el concepto inteligencia tiene este sentido mucho más activo que el concepto vigilancia. **“A pesar de ello, en ocasiones sólo se trata de un problema de terminología”.** (Maspons, et al., 2007)

## **1. Tipos de Vigilancia**

La actividad de vigilancia, se divide en cuatro (4) grandes ramas, la Vigilancia Tecnológica (VT), Vigilancia Competitiva (VComp.), Vigilancia Comercial (VC) y Vigilancia del Entorno (VE). (EAFIT, 2008); Estas ramas, están muy bien diferenciadas en sus alcances y objetivos, estando la

vigilancia tecnológica dividida en vigilancia científica y vigilancia estratégica, sub ramas que hacen seguimiento a patentes, publicaciones científicas, capacidad tecnológica de la competencia y las relaciones entre las empresas del sector de estudio; de otro lado, la Vigilancia Competitiva busca determinar posibles competidores, la Vigilancia Comercial está dirigida a clientes, proveedores y canales de distribución, y la Vigilancia del Entorno estudia las fuentes, características geográficas, legislación, regulación ambiental, estructura social y política de la locación en la que se lleva a cabo una actividad productiva. La Figura VII-3, esquematiza los diferentes tipos de Vigilancia. (EAFIT, 2008).

## 2. Tipos de Vigilancia Tecnológica

Como se mencionó anteriormente, la Vigilancia Tecnológica se divide en dos (2) grandes ramas, la Vigilancia Científica y la Vigilancia Estratégica; en forma de lista multinivel se puede representar de la siguiente manera:

### **Vigilancia Tecnológica**

- a. Vigilancia Científica
  - i. Ingeniería Inversa
  - ii. Análisis de Patentes
  - iii. Seguimiento de Publicaciones Científicas
- b. Vigilancia Estratégica
  - i. Análisis de la Capacidad Tecnológica de la Competencia
  - ii. Seguimiento de artículos de la competencia
  - iii. Relaciones Económico-Financieras y de trabajo entre las empresas del sector.

Respecto a los componentes de la Vigilancia Científica, el objetivo de la ingeniería inversa es obtener información a partir de un producto accesible al público, con el fin de determinar de qué está hecho, qué lo hace funcionar y cómo fue fabricado. Los productos más comunes que son sometidos a la ingeniería inversa son los programas de computadoras y los componentes electrónicos, pero básicamente casi cualquier proceso puede ser sometido a un análisis de Ingeniería Inversa. Se denomina ingeniería inversa, porque avanza en contravía a las tareas habituales de ingeniería, como utilizar datos técnicos para elaborar un producto determinado. (Wikipedia, 2008) En general si el producto sometido a la ingeniería inversa, fue obtenido en forma apropiada, entonces el proceso es legítimo y legal.

Sobre el análisis de patentes, este procedimiento se usa como apoyo al diagnóstico tecnológico y puede permitir (Escuela Técnica Superior de Ingenieros de Telecomunicaciones, Universidad Politécnica de Madrid, 2007):

1. Apoyar a las técnicas de prospectiva tecnológica dando información sobre las tendencias tecnológicas tras un análisis de las fechas de las patentes.
2. Apoyar a las técnicas de benchmarking dando información sobre prácticas innovadoras de otras empresas.
3. Obtener información sobre disponibilidad de tecnologías útiles adquiribles. Por lo general patentar no implica que la tecnología se use.

El seguimiento de las publicaciones científicas, como su nombre lo indica, trata sobre revisar la información contenida en artículos técnicos que puedan ser de interés a la compañía, relacionados generalmente a un tema específico.

Sobre las subdivisiones de la Vigilancia Estratégica, el análisis de la capacidad tecnológica está constituido por el conjunto de conocimientos y habilidades que dan sustento al proceso de producción. Abarcan desde los conocimientos acumulados de las fuentes de energías empleadas, las formas de extracción de reservas naturales, su procesamiento, transformación y reciclaje, hasta la configuración y desempeño de los productos finales resultantes. Por tanto, se trata de un factor de producción transversal al proceso productivo. (Sanchez, 2008)

La capacidad tecnológica, se puede medir en tres frentes:

1. Capacidad para la investigación y el desarrollo (I y D)
2. Capacidad para desarrollar e implementar proyectos de nuevos procesos o nuevos productos, pasando del proceso de descubrimiento a la innovación.
3. Capacidad para realizar las actividades de producción propiamente dichas.

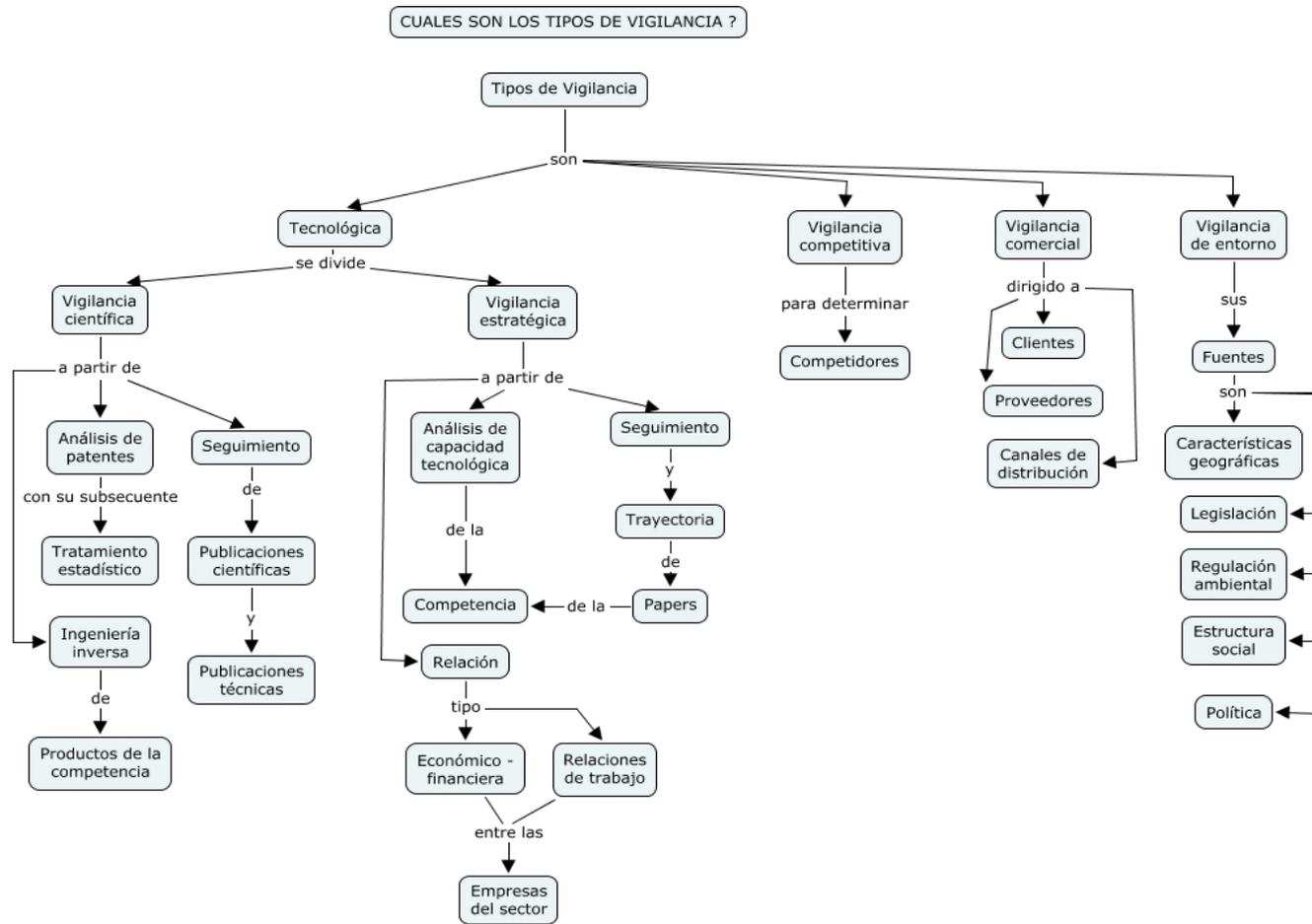


Figura VII-3 ¿Cuáles son los Tipos de Vigilancia? Mapa Conceptual (EAFIT, 2008)

Acerca del seguimiento de artículos de la competencia, consiste en monitorear la producción científica y técnica de documentos, desarrollados por la competencia. Sobre las relaciones Económico-Financieras y de trabajo entre las empresas del sector, mide el grado de cooperación entre diferentes actores del sector productivo de interés, estudia las fusiones, Joint-Ventures y Spin - Offs, realizadas por compañías de la competencia y analiza posibles líneas de inversión de las mismas, basados en información económica de libre acceso.

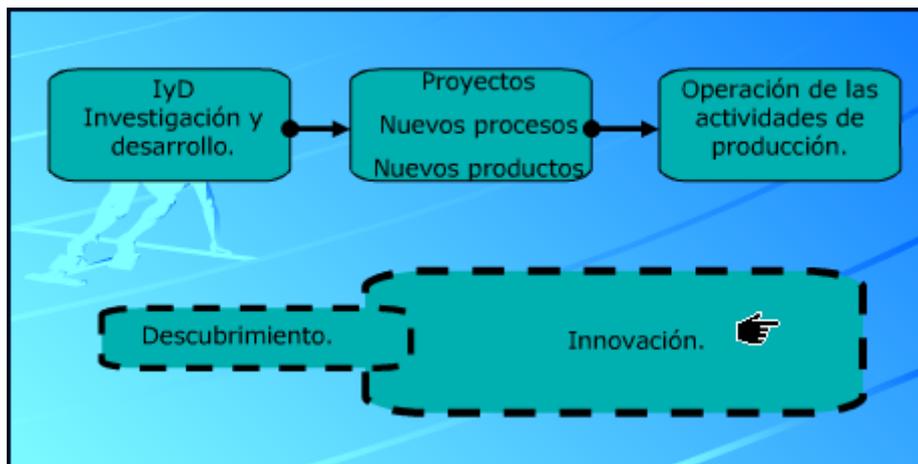


Figura VII-4 Frentes de medición de Capacidad tecnológica (Sanchez, 2008)

#### a) *Herramientas de Vigilancia Tecnológica*

Dado el alto número de registros, datos y variables, involucradas en un proceso de Vigilancia Tecnológica, es necesario desarrollar herramientas que faciliten la captura y el procesamiento de información y su posterior análisis. En este orden de ideas, el presente capítulo diferenciara las herramientas según su función, ya sean orientadas a las primeras etapas del proceso (recopilación y procesamiento de información), o al análisis de los datos.

En el primer grupo se ubican herramientas como:

- **Minería de Datos:** Se define como el proceso de aplicar metodologías basadas en conceptos estadísticos y de administración, para extraer conocimiento de grandes volúmenes de información, usando herramientas computacionales y algoritmos de optimización que agilizan el proceso. En el mercado existen diversos programas de computador que realizan minería de datos, entre los que se encuentran: KDNuggets, STATISTICA, Angoss, DataAnalyst y Weka 3.
- **Redes Neuronales Artificiales:** Las redes neuronales artificiales, en adelante RNA, se dicen “neuronales”, en el sentido que fueron inspiradas por la neurociencia, pero no necesariamente representan cabalmente fenómenos cognitivos o neuronales en su

aspecto biológico; son modelos de computación en paralelo, comprendidos de unidades de procesamiento adaptativas, densamente interconectadas (Hecht-Nielsen, 1990). Estas redes pueden ser implementaciones de sistemas no lineales estáticos o de sistemas dinámicos, siendo su característica principal, la naturaleza adaptativa, en la que el aprendizaje de la red, y su capacidad de resolver problemas, se logra mediante ejemplos y no con programación estructurada. Esta característica, es en particular interesante al aplicarse en dominios, donde se tiene un conocimiento limitado de la física o matemática, del problema a resolver, pero se tienen datos de entrenamiento.

Entre las aplicaciones viables de las RNA, se encuentran: clasificación de patrones, síntesis y reconocimiento de voz, interfaces entre humanos y sistemas físicos complejos, aproximación de funciones, compresión de imágenes, memoria asociativa, clústeres, pronóstico y predicción, optimización de combinatorias, modelación de sistemas no lineales y control. (Hassoun, 1995)

En relación con el la utilización de esta herramienta para practicar VT a los ciclos biogeoquímicos e hidrológico, se destaca la capacidad para clasificar patrones y realizar clústeres de información.

- **Bases de datos:** Es un conjunto de datos, en formato digital o físico, con elementos que comparten un mismo contexto y que son almacenados para su consulta posterior; en el contexto de la vigilancia tecnológica, estas bases de datos pueden ser archivos de patentes o de producción técnica y académica en forma de artículos publicados.

Para el presente trabajo las bases de datos relacionadas con los ciclos biogeoquímicos e hidrológicos son: ProQuest, ScienceDirect, Springer Link, y las usadas por diferentes oficinas de patentes como INVENES, en España.

De otra parte, este trabajo ha identificado las siguientes herramientas relacionadas al análisis de datos:

- **Indicadores de Patentes:** esta herramienta permite a las organizaciones realizar procesos de Vigilancia Tecnológica, con mayor orientación y veracidad. Según cifras publicadas por la Oficina Europea de Patentes, se registran 20.000 millones de dólares de pérdidas en la Unión Europea, por desarrollos de productos y procesos ya existentes. Estos indicadores como su nombre lo indica trabajan con patentes, y por lo tanto con las bases de datos nacionales que contienen esta información.

- **Niveles de actividad tecnológica:** permiten determinar el liderazgo de una u otra institución, frente a la investigación o desarrollo de un tema dado; se pueden obtener indicadores del nivel de actividad tecnológica, estudiando el número de patentes registradas por entidad, los productos y publicaciones realizadas en un determinado horizonte de tiempo.
- **Campos tecnológicos:** Los campos tecnológicos son herramientas que facilitan la ubicación de las actividades investigativas o productivas de una empresa, ayudan a identificar clústeres de tecnologías asociadas y desarrollos poco competitivos en términos de tecnología, respecto a los líderes del sector.
- **Dinámica y visibilidad de la tecnología:** Esta herramienta permite ver las tecnologías asociadas a un determinado desarrollo y su evolución temporal, así como la aplicación de dicha tecnología en otros desarrollos de la empresa o de terceros; puede valerse del número de citas a publicaciones de la empresa o del personal de la misma, realizadas en documentos científicos de terceros.

### 3. Patentes

Es importante suministrarle al lector algunos conceptos relacionados con la legislación sobre propiedad intelectual y patentes. Inicialmente se expondrán las consideraciones relacionadas al tema, en el ámbito nacional y luego se procederá a realizar una breve descripción de las principales oficinas de patentes a nivel mundial.

En el marco de este trabajo, se asistió al *Seminario de Capacitación sobre Patentes, en el marco del proyecto "Redes de Investigación y Propiedad Intelectual"* (2007), coordinado por la Organización Mundial de Propiedad Intelectual (OMPI), organizado por el Servicio Compartido de Propiedad Intelectual (SECOPI) conjuntamente con la Superintendencia de Industria y Comercio (SIC).

A continuación, se expondrán algunos conceptos impartidos en este seminario y extractos de la información contenida en las memorias de dicho evento.

#### a) Propiedad Intelectual

La propiedad Intelectual está conformada por todas las creaciones derivadas del talento y esfuerzo humano; se divide en dos (2) ramas, la Propiedad Industrial y los Derechos de Autor. Dentro de la Propiedad Industrial se encuentran las invenciones, diseños industriales, esquemas de trazado de los circuitos integrados, y las marcas comerciales. Los Derechos de Autor, comprenden las obras literarias, artísticas, musicales y los programas de computador

(Seminario de Capacitación sobre Patentes, en el marco del proyecto “Redes de Investigación y Propiedad Intelectual” , 2007).

### ***b) Invención y la Patente***

Una invención puede considerarse como la solución que se le va a dar a un problema técnico específico; así como permite la producción de productos nuevos, o la mejora de productos existentes. Las invenciones pueden ser de producto (objetos tangibles) o procedimiento (secuencias de etapas para obtener un producto final o un efecto específico), y pertenecer a cualquier campo tecnológico (Seminario de Capacitación sobre Patentes, en el marco del proyecto “Redes de Investigación y Propiedad Intelectual” , 2007).

Generalmente, las invenciones se obtienen luego de un proceso investigativo, que implica grandes gastos de tiempo y dinero; motivo por el cual, la legislación otorga al inventor o empresa para la cual se realizó la investigación, un privilegio mediante el cual le da un derecho exclusivo para el empleo de su invención. El derecho otorgado por el gobierno al inventor, se denomina Patente, y en este consta que se ha realizado una invención y que es propiedad de “El Titular de la Patente” (Seminario de Capacitación sobre Patentes, en el marco del proyecto “Redes de Investigación y Propiedad Intelectual” , 2007).

Las patentes, además de proteger la creatividad de los inventores, representan una fuente importante de información técnica, ya que contienen descripciones de conceptos científicos y técnicos, así como detalles prácticos de procesos y aparatos.

#### **(1) ¿Quién concede el derecho?**

En Colombia, es la Superintendencia de Industria y Comercio

#### **(2) ¿Cuál es la ley que rige la propiedad industrial?**

Para los países miembros del Pacto Andino como Bolivia, Colombia, Ecuador, Perú y Venezuela, el régimen común sobre propiedad industrial es la Decisión 486 de la Comisión de la Comunidad Andina, que rige a partir de Diciembre 1 de 2000. (Seminario de Capacitación sobre Patentes, en el marco del proyecto “Redes de Investigación y Propiedad Intelectual” , 2007)

En Colombia estas disposiciones se encuentran incorporadas en la circular única 010 de 2001 de la Superintendencia de Industria y Comercio.

### (3) ¿Qué se puede patentar?

Para que se pueda patentar una invención, se debe tener en cuenta que existen tres (3) requisitos básicos de patentabilidad (Seminario de Capacitación sobre Patentes, en el marco del proyecto “Redes de Investigación y Propiedad Intelectual” , 2007):

- 1) Ser novedosa: la invención debe ser la primera en el mundo, es decir no puede encontrarse dentro del estado de la técnica, al presentar la solicitud. Siendo el estado de la técnica, toda la información que se haya hecho accesible al público, en cualquier lado del mundo, a través de cualquier medio, ya sea por descripción escrita u oral, por utilización o comercialización.
- 2) Tener nivel inventivo: este requisito establece que no es suficiente que una invención sea diferente a lo existente en el estado de la técnica, sino que esa diferencia tenga dos (2) características: que sea inventiva o el resultado de una idea creativa, y que la diferencia con el estado de la técnica implique un avance o progreso significativo.
- 3) Ser susceptible de aplicación industrial: Si el invento se puede utilizar o fabricar en cualquier tipo de industria, no fue producto del azar y no fue realizada en el ámbito privado para sus propias necesidades o de su familia, se puede patentar.

### (4) ¿Qué no se puede patentar?

La legislación ha considerado algunos campos en los que no se pueden obtener patentes, por no considerarse invenciones o aun siendo invenciones, porque no se les debe otorgar dicho privilegio.

Dentro de los campos que la legislación considera que no son invenciones se pueden incluir los siguientes (Seminario de Capacitación sobre Patentes, en el marco del proyecto “Redes de Investigación y Propiedad Intelectual” , 2007):

- 1) Los descubrimientos, incluyendo la materia existente en la naturaleza
- 2) Las teorías científicas y métodos matemáticos
- 3) Los planes, reglas y métodos para el ejercicio de actividades intelectuales, juegos o actividades económico-comerciales
- 4) Los programas de computador (software)
- 5) Las formas de presentar la información.
- 6) El todo o parte de seres vivos, tal como se encuentran en la naturaleza; los procesos biológicos naturales, el material biológico existente en la naturaleza, o aquel que pueda ser aislado, inclusive el genoma o germoplasma de cualquier ser vivo natural.

Dentro de las invenciones no patentables están (Seminario de Capacitación sobre Patentes, en el marco del proyecto “Redes de Investigación y Propiedad Intelectual” , 2007):

- 1) Las invenciones cuya explotación comercial deba impedirse necesariamente para los siguientes fines, siempre que la exclusión no se haga solamente porque su explotación está prohibida por la legislación:
  - a. Proteger el orden público, o la moral
  - b. Proteger la salud, o la vida de las personas o de los animales
  - c. Preservar los vegetales
  - d. Evitar daños al medio ambiente
- 2) Los métodos de diagnóstico, terapéuticos y quirúrgicos para el tratamiento de personas o animales
- 3) Las plantas y los animales, y los procedimientos esencialmente biológicos para la producción de plantas y animales
- 4) Los procedimientos esencialmente biológicos para la producción de plantas o animales que no sean procedimientos no biológicos o microbiológicos.

(5) ¿Qué derecho otorga la patente de invención y de modelo de utilidad?

La patente confiere el derecho de impedir que terceras personas, sin el consentimiento del Titular de la Patente, fabriquen, comercialicen, utilicen o importen el producto objeto de invención; de tratarse de un procedimiento, se puede impedir su empleo o ejecución de los actos anteriores.

Este derecho **no se puede ejercer en determinadas circunstancias** establecidas por la legislación; por ejemplo, cuando la invención **se utiliza en el ámbito privado y con fines no comerciales**, es decir, **su uso es sin ánimo de lucro** o cuando **su uso es con fines de experimentación**. Con esto se da la oportunidad a terceras personas de conocer, analizar y comprender la invención; también se considera el **uso con fines de enseñanza o investigación científica o académica**. Esta excepción es importante en la medida que permite alcanzar uno de los objetivos principales del sistema de patentes, como es el desarrollo y promoción de la ciencia y la tecnología (Seminario de Capacitación sobre Patentes, en el marco del proyecto “Redes de Investigación y Propiedad Intelectual” , 2007).

(6) ¿Cuánto es el tiempo de duración de una patente?

Una patente de invención se concede por un término de veinte (20) años, a partir de la fecha de presentación de la solicitud, y en el caso de un modelo de utilidad diez (10) años a partir de esta misma fecha.

*c) Las patentes en el mundo*

En el ámbito de las oficinas de patentes a nivel mundial, se maneja el término de “las *Tres Grandes*” u oficinas trilaterales, siendo esta denominación usada para identificar a la Oficina Europea de Patentes, la Oficina Estadounidense de Patentes y Marcas Registradas y la Oficina Japonesa de Patentes. Vale la pena notar, que estas agencias sostienen una relación de cooperación técnica muy estrecha, y son el referente en la materia a nivel mundial.

A continuación se presenta una breve descripción de cada una de las agencias pertenecientes a las Tres Grandes:

- Oficina Europea de Patentes (EPO): Establecida en 1977 por la convención europea de patentes, con el fin de crear un sistema unificado de aplicación de patentes y de subvenciones de investigación, para todos los estados miembros. Su misión principal es apoyar la innovación, competitividad y crecimiento económico de los ciudadanos de Europa. A marzo de 2007, la EPO tiene jurisdicción en 32 países europeos y sobre un mercado de cerca de 600 millones de personas. (EPO, 2008).
- Oficina Estadounidense de Patentes y Marcas Registradas (USPTO): Esta oficina fue creada por el Artículo 1, Sección 8 de la Constitución de Estados Unidos, con el fin de promover el progreso de la ciencia y de las artes útiles, al garantizarle a los inventores por un tiempo limitado los derechos exclusivos sobre sus descubrimientos. Modelo de protección que ha permitido a la industria americana, invertir en nuevos desarrollos y crear millones de empleos. La USPTO, con sede en Alexandria, Virginia, emplea más de 7000 personas que apoyan las funciones de otorgar patentes, examinar y registrar marcas. Para el 2006 la oficina estaba atendiendo cerca de 400.000 solicitudes. (USPTO, 2008)
- Oficina Japonesa de Patentes (JPO): La JPO, es una agencia federal que cuenta con 1358 examinadores de un total de 2651 empleados; la agencia cuenta con un presupuesto cercano a los 800 millones de Euros, y atiende un promedio de 400.000 solicitudes al año desde 1998 (EPO, 2008)

**d) Análisis de Patentes en los Campos Tecnológicos  
Relacionados a los Ciclos Biogeoquímicos y el Ciclo Hidrológico**

Para realizar el análisis de patentes de los campos tecnológicos relacionados, se usó información del catálogo en línea de la USPTO (*United States Patent and Trademark Office*), sobre las clases tecnológicas en las que se presentan los desarrollos, e inventos afines con los ciclos biogeoquímicos y el ciclo hidrológico, en cualquiera de sus componentes.

La consulta del catálogo en línea de la USPTO, garantiza el acceso a los desarrollos tecnológicos elaborados y registrados por la EPO y la JPO, las tres grandes oficinas de patentes a nivel mundial.

Respecto a las clases investigadas, se buscó información de patentes en las siguientes clases tecnológicas, sin importar el origen geográfico y para el periodo 2003-2007. Basados en la clasificación primaria de la patente.

**Tabla 1. Clases Tecnológicas asociadas a los ciclos biogeoquímicos e hidrológico. Clasificación USPTO**

<b>Clase No. (USPTO)</b>	<b>Nombre de la Clase (USPTO)</b>	<b>Nombre de la Clase (español)</b>
016	Miscellaneous Hardware	Hardware varios
023	Chemistry: Physical Processes	Química: Procesos Físicos
165	Heat Exchange	Intercambio de Calor
171	Unearthing Plants or Buried Objects	Desenterrar plantas u objetos de la corteza terrestre
172	Earth Working	Trabajos en Tierra
204	Electrical and Wave Energy	Energía Eléctrica y de Onda
250	Radiant Energy	Energía Radiante
260	Chemistry of Carbon Compounds	Química de Compuestos del Carbono
342	Communications Directive Radiowave Systems & Devices	Sistemas y dispositivos direccionales de radio ondas
343	Communications: Radio Wave Antennas	Comunicaciones: antenas de ondas de radio
345	Computer Graphics Processing and Selective Visual Display Systems	Procesamiento de graficas de computador y sistemas selectivos de visualización de datos.
367	Communication Electrical: Acoustic Wave systems & Devices	Sistemas y dispositivos de ondas acústicas
374	Thermal Measuring and Testing	Pruebas y Mediciones Termiales
405	Hydraulic and Earth Engineering	Ingeniería Hidráulica y de la Tierra

A manera de ejemplo, se presenta a continuación la información relacionada al comportamiento de las patentes en la clase tecnológica Ingeniería Hidráulica y de la Tierra. La información recopilada, para cada una de las clases contenidas en la tabla anterior, se presenta bajo el título *Patentes en Clases Tecnológicas*, en el Tomo 2 de este documento. Adicionalmente se incluyen figuras que muestran el comportamiento de las patentes por zonas geográficas, para cada uno de los campos tecnológicos.

#### (1) Ingeniería Hidráulica y de la Tierra - Clase 405

En la clase 405, de no estar contemplado en otra clase se incluyen: a) El control y tratamiento del agua, b) control y tratamiento de materiales de la tierra in-situ, c) almacenamiento subterráneo de fluidos y disposición de desechos en la tierra, d) aparatos y métodos para operaciones debajo del agua, entre otros.

Se observa en las figuras relacionadas (VII-5 a VII-14), cuales países son los líderes mundiales en patentar bajo esta categoría, así como el número de patentes en los últimos 5 años para los tres países líderes, también se muestra una comparación porcentual entre países de diferentes zonas geográficas.

Para la clase 405, se obtienen las siguientes características en el periodo 2003-2007:

1. Estados Unidos registra el 71 % de las patentes en esta clase tecnológica, los demás países patentan menos del 5 % cada uno. Japón y Canadá, son los siguientes dos países por número de patentes anuales, siendo su actividad muy similar (18 y 16 patentes en el 2003 y 7 y 6 patentes en el 2007).
2. En Estados Unidos, se presentó una disminución del 29% en el número de patentes respecto al 2003, en el 2007; sin embargo, en el 2006, se presentó un aumento del 16% respecto al 2003.
3. Respecto al análisis zonal, se tiene que para el periodo 2003-2007:
  - a. En Norte América (EE.UU y Canadá), Estados Unidos registró el 95% de las patentes y Canadá el 5% restante.
  - b. En los países asiáticos, Japón lideró la actividad inventiva en esta categoría con el 61% de los registros, seguido por Corea del Sur con el 20 %, Taiwán con el 10 %, China con el 5 %, y participaciones menores de Singapur y Hong Kong.
  - c. En Europa, el líder es Reino Unido con el 21 %, le siguen Alemania con el 17%, Francia con el 13 %, Holanda con el 12 %, Noruega con el 10%, Italia con el 8%, Suiza, Austria y Suecia con el 4% cada uno, y luego otros países con menos del porcentaje como Finlandia y España.

- d. En Oceanía, Australia lideró con el 94 % de los registros y Nueva Zelandia realizo el 6 % de patentes.
- e. En Centro y Sur América, Brasil e Islas Caimán realizaron las únicas patentes de la zona, liderando Brasil con el 78% de los registros e Islas Caimán con el 22% restante.
- f. En el Lejano y Medio Oriente, Israel realizó el 56% de las patentes, seguido por Arabia Saudita con el 33% y Turquía con el 11 %, para el periodo de registro.
- g. Por zonas en América del Norte se registró el 72% del total de patentes, seguida por Europa con el 18 %, y Asia con el 6%, luego aparece Oceanía con el 2 %, y participaciones menores de Latinoamérica, África, y del Lejano y Medio Oriente.

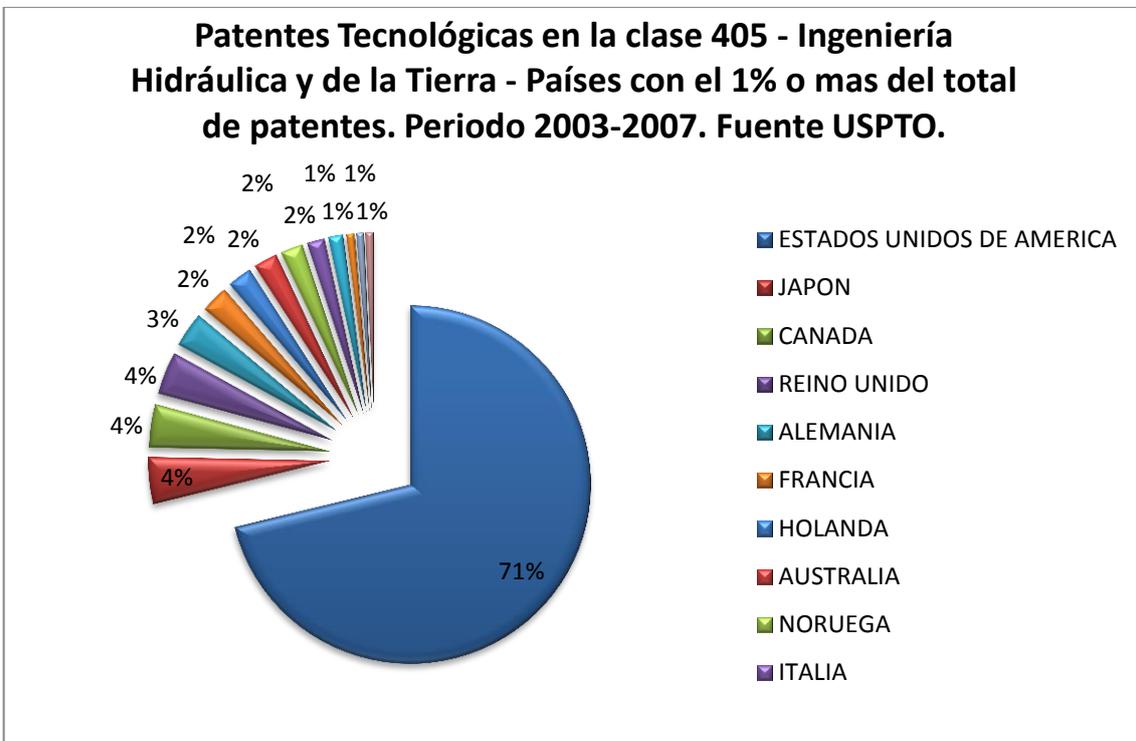


Figura VII-5 Patentes Tecnológicas en la clase 405 - Ingeniería Hidráulica y de la Tierra - Países con el 1% o más del total de patentes. Periodo 2003-2007. (USPTO, 2008).

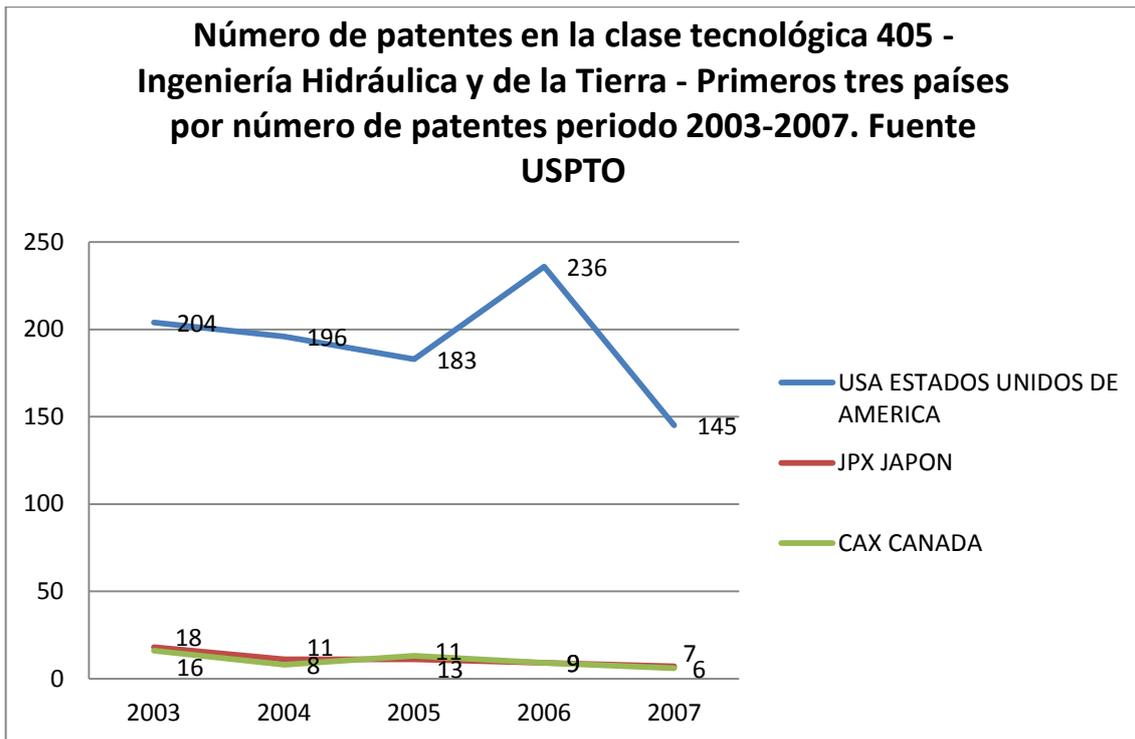


Figura VII-6 Número de patentes en la clase tecnológica 405 - Ingeniería Hidráulica y de la Tierra - Primeros tres países por número de patentes periodo 2003-2007. (USPTO, 2008)

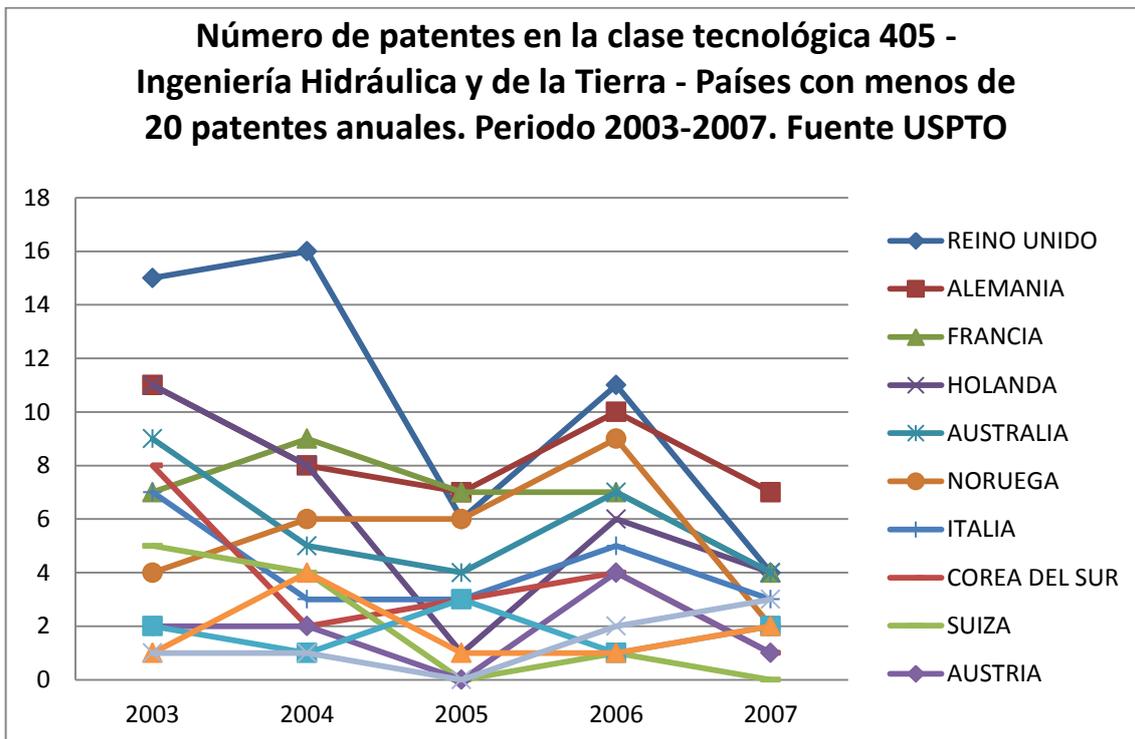


Figura VII-7 Número de patentes en la clase tecnológica 405 - Ingeniería Hidráulica y de la Tierra - Países con menos de 20 patentes anuales. Periodo 2003-2007. (USPTO, 2008)

**Patentes Tecnológicas en la clase 405 - Ingeniería Hidráulica y de la Tierra - Porcentaje de patentes en America del Norte por país. Periodo 2003-2007. Fuente USPTO.**

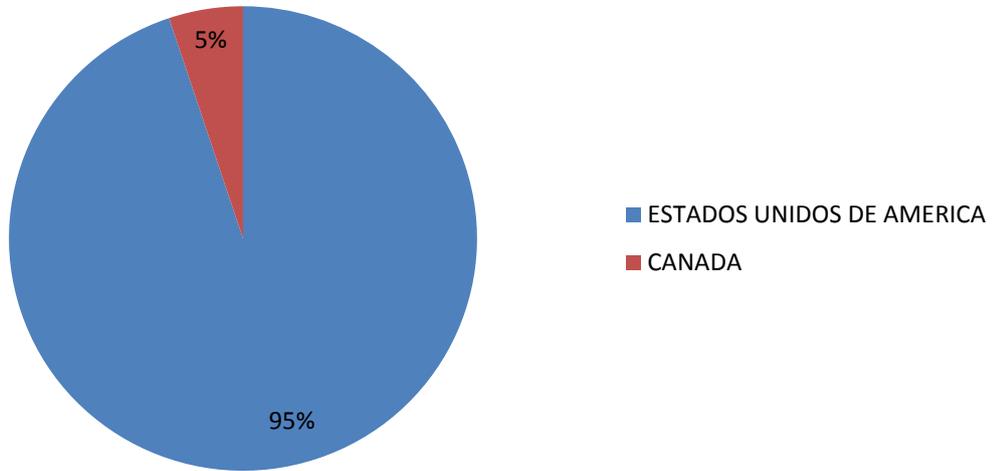


Figura VII-8 Patentes Tecnológicas en la clase 405 - Ingeniería Hidráulica y de la Tierra - Porcentaje de patentes en América del Norte por país. Periodo 2003-2007. (USPTO, 2008).

**Patentes Tecnológicas en la clase 405 - Ingeniería Hidráulica y de la Tierra - Porcentaje de patentes en Asia por país. Periodo 2003-2007. Fuente USPTO.**

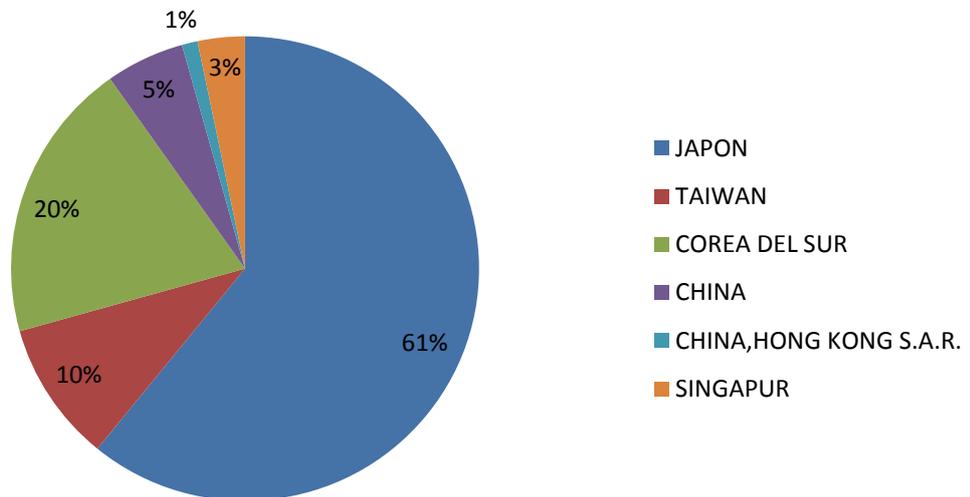


Figura VII-9 Patentes Tecnológicas en la clase 405 - Ingeniería Hidráulica y de la Tierra - Porcentaje de patentes en Asia por país. Periodo 2003-2007. (USPTO, 2008).

**Patentes Tecnológicas en la clase 405 - Ingeniería Hidráulica y de la Tierra - Porcentaje de patentes en Europa por país. Periodo 2003-2007. Fuente USPTO.**

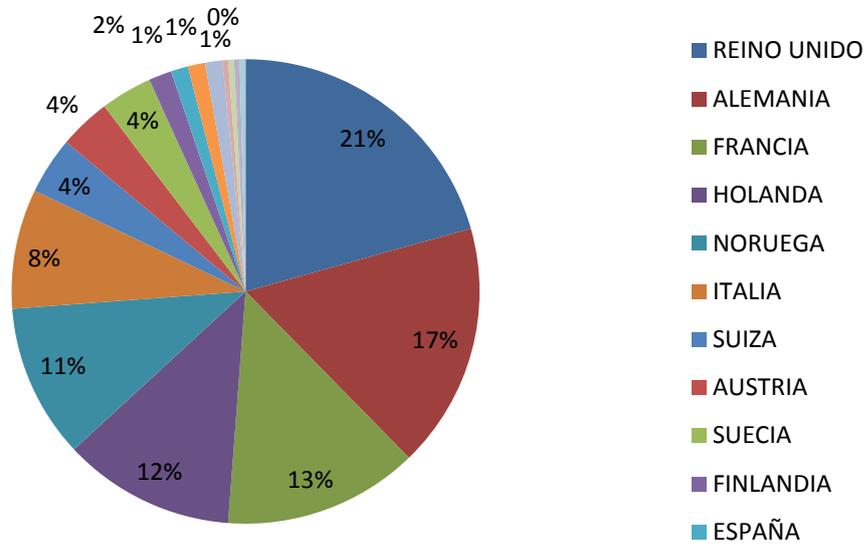


Figura VII-10 Patentes Tecnológicas en la clase 405 - Ingeniería Hidráulica y de la Tierra - Porcentaje de patentes en Europa por país. Periodo 2003-2007. (USPTO, 2008).

**Patentes Tecnológicas en la clase 405 - Ingeniería Hidráulica y de la Tierra - Porcentaje de patentes en Oceanía por país. Periodo 2003-2007. Fuente USPTO.**

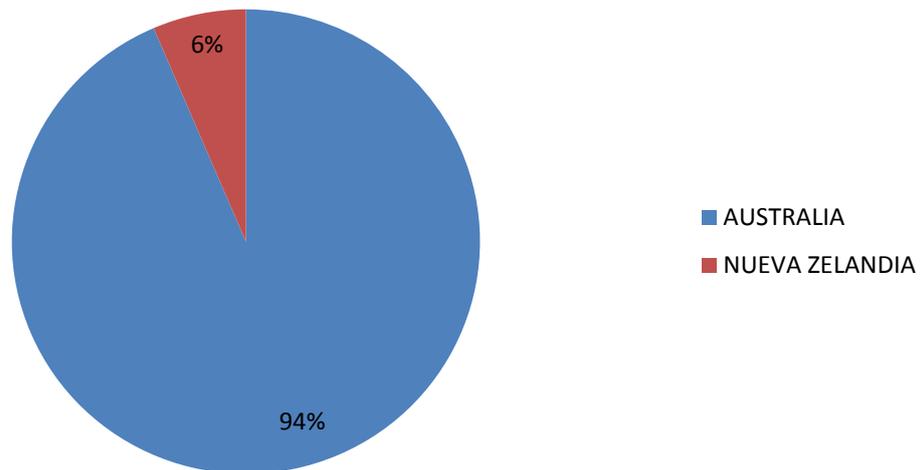


Figura VII-11 Patentes Tecnológicas en la clase 405 - Ingeniería Hidráulica y de la Tierra - Porcentaje de patentes en Oceanía por país. Periodo 2003-2007. (USPTO, 2008).

**Patentes Tecnológicas en la clase 405 - Ingeniería Hidráulica y de la Tierra - Porcentaje de patentes en Centro y Sur América por país. Periodo 2003-2007. Fuente USPTO.**

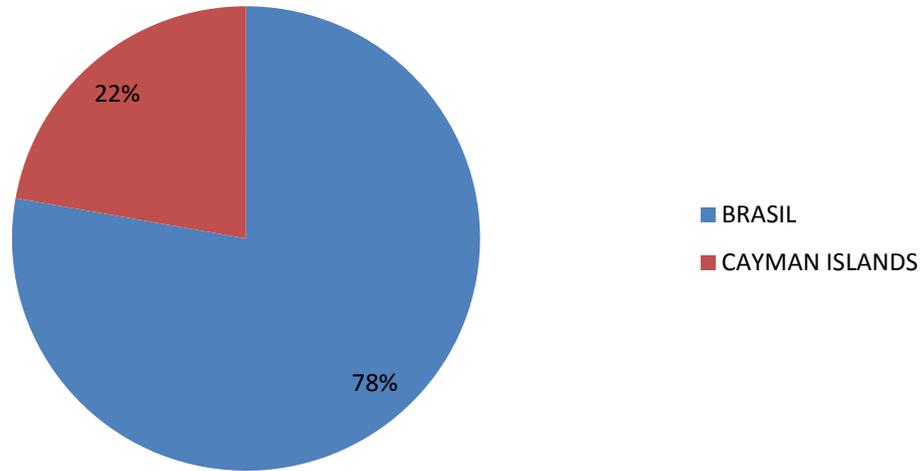


Figura VII-12 Patentes Tecnológicas en la clase 405 - Ingeniería Hidráulica y de la Tierra - Porcentaje de patentes en Centro y Sur América por país. Periodo 2003-2007. (USPTO, 2008).

**Patentes Tecnológicas en la clase 405 - Ingeniería Hidráulica y de la Tierra - Porcentaje de patentes en el Lejano y Medio Oriente por país. Periodo 2003-2007. Fuente USPTO.**

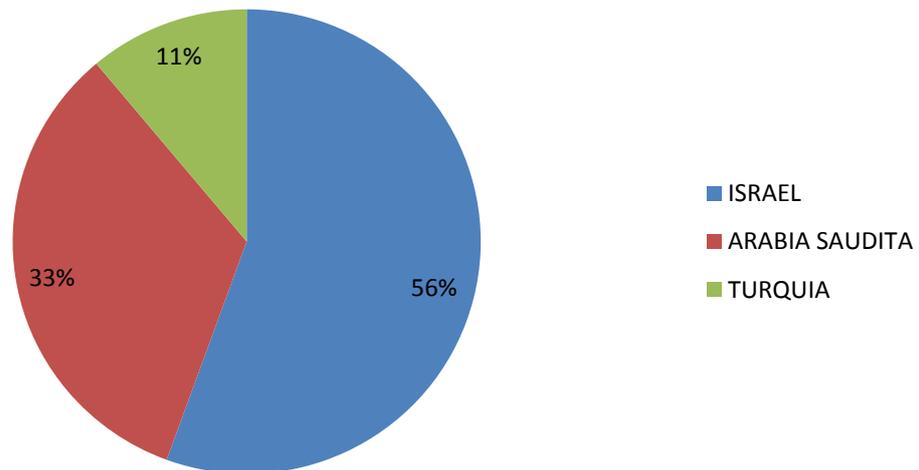


Figura VII-13 Patentes Tecnológicas en la clase 405 - Ingeniería Hidráulica y de la Tierra - Porcentaje de patentes en el Lejano y Medio Oriente por país. Periodo 2003-2007. (USPTO, 2008).

**Patentes Tecnológicas en la clase 405 - Ingeniería Hidráulica y de la Tierra - Porcentaje de patentes por región. Periodo 2003-2007.**  
Fuente USPTO.

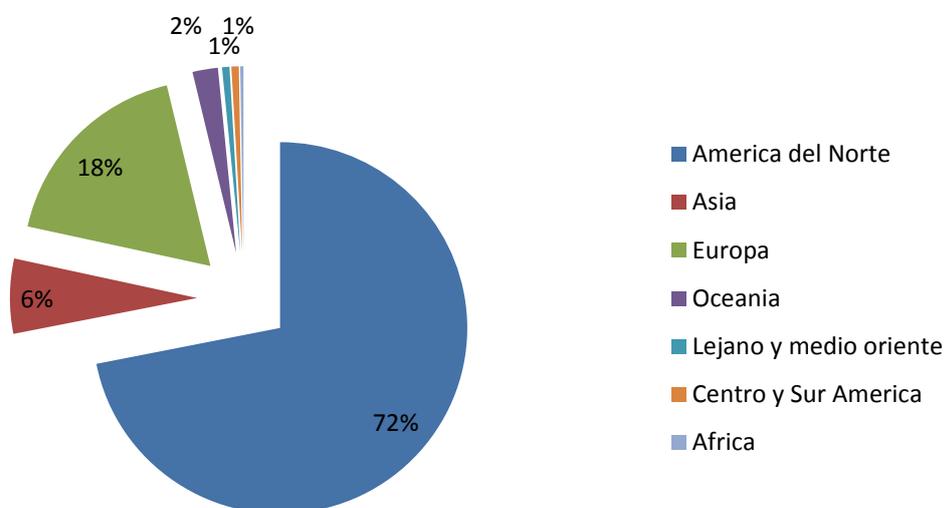


Figura VII-14 Patentes Tecnológicas en la clase 405 - Ingeniería Hidráulica y de la Tierra - Porcentaje de patentes por región. Periodo 2003-2007. (USPTO, 2008).

(2) Acumulado Global

Luego de realizar análisis similares al usado para la clase tecnológica 405 en todas las clases tecnológicas mencionadas en la Tabla 1, se procedió a determinar un acumulado global por regiones, de las patentes registradas en los últimos 5 años.

Tabla 2 Total de patentes en las clases tecnológicas asociadas a los ciclos biogeoquímicos e hidrológico

Total patentes en clases tecnológicas asociadas a los ciclos biogeoquímicos e hidrológico						
	2003	2004	2005	2006	2007	Total
<b>América del Norte</b>	3486	3632	3311	4140	3358	17927
<b>Asia</b>	2102	2313	2143	3124	2689	12371
<b>Europa</b>	1037	1064	942	1101	1003	5147
<b>Oceanía</b>	56	46	46	60	57	265
<b>Lejano y medio oriente</b>	57	64	42	81	77	321
<b>Centro y Sur América</b>	7	5	4	7	7	30
<b>África</b>	8	4	1	4	8	25
<b>TOTAL ANUAL</b>	<b>6753</b>	<b>7128</b>	<b>6489</b>	<b>8517</b>	<b>7199</b>	<b>36086</b>

Como se observa en la Tabla 2, en el periodo 2003-2007 América del Norte registró 17927 patentes en las clases tecnológicas asociadas a los ciclos biogeoquímicos e hidrológico,

mientras que Centro y Sur América, sólo registraron 30 patentes en el mismo periodo, para las estas clases tecnológicas. De esas 30 Colombia registró UNA (1) patente, en la clase tecnológica 374 Pruebas y Mediciones Termales. (En el Tomo 2 se presenta la información relacionada a las clases tecnológicas asociadas)

El escenario observado en estos últimos 5 años para el registro de innovaciones relacionadas a los ciclos biogeoquímicos e hidrológico, no puede ser más preocupante para la región, dado que sólo aportó 30 de 36086 patentes para este periodo (0.08%). Superando únicamente a África, y patentando 10 veces menos que el Lejano y Medio Oriente.

Respecto a los líderes mundiales, Estados Unidos, Japón y Alemania se ubican en las tres primeras posiciones en cuanto al número de patentes en las clases tecnológicas asociadas a los ciclos biogeoquímicos e hidrológico.

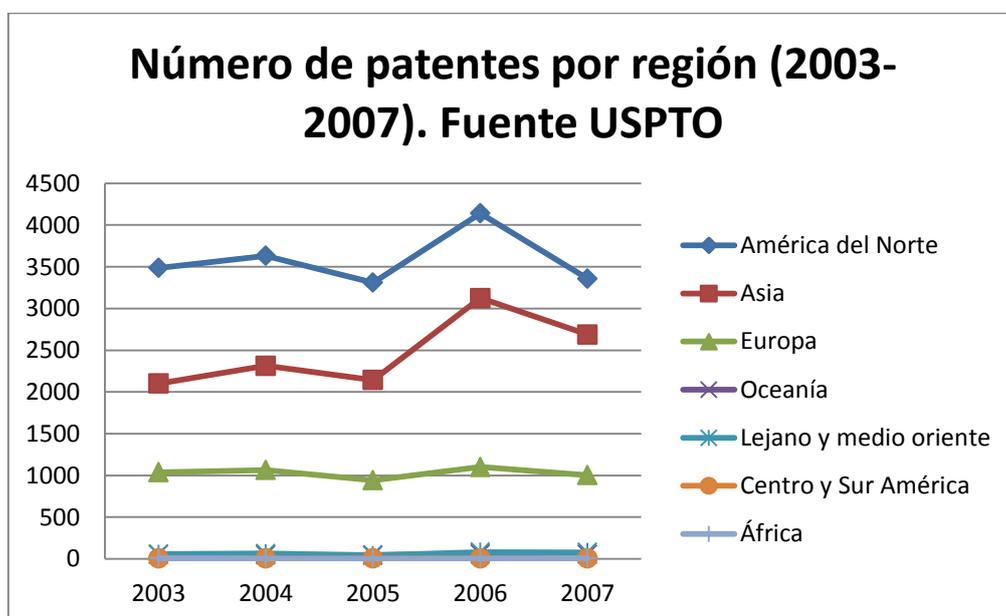


Figura VII-15 Número de patentes por región. Periodo 2003-2007. (USPTO, 2008)

La Figura V-15 presenta gráficamente el número de patentes registrado en los campos tecnológicos asociados a los ciclos biogeoquímicos e hidrológico en los últimos cinco años. Se observa como Asia reduce la brecha con el número de patentes registradas por Estados Unidos, a pesar de registrar comportamientos similares, subiendo su actividad de patentes en el 2004 y el 2006, y presentando una disminución respecto al año inmediatamente anterior en el 2005 y 2007.

Adicionalmente al estudio de la información suministrada por la USPTO, se realizó una búsqueda en la base de datos INVENES, de la Oficina Española de Patentes y Marcas; con el fin

de obtener resúmenes de patentes relacionadas con los ciclos biogeoquímicos e hidrológico. Estas patentes, se anexan en el Tomo 2 de este documento, y representan información muy valiosa, para los interesados en apropiar y adaptar desarrollos tecnológicos en esta línea de investigación, a su medio. En el Tomo 2, se presenta la interfaz de búsqueda de la base de datos INVENES, incluyendo los campos de su formulario interactivo, así como los datos mostrados en pantalla relacionados con las patentes que incluyen la palabra OXIGENO es su título, a manera de ejemplo.

La tabla 3 muestra los registros realizados usando el *Patent Cooperation Treaty (PCT)*, en el que se hace un sólo registro de patente internacional con jurisdicción en los 138 países miembros del PCT. Los números totales son superiores a los registros realizados ya que una misma patente puede registrarse en varios campos tecnológicos.

**Tabla 3 Publicaciones por campo tecnológico. Fuente WIPO 2007**

CAMPOS TECNOLOGICOS		2003	2004	2005	2006	2007	Cambio respecto al 2006 (%)
I	<b>ELECTRICIDAD-ELECTRONICA</b>						
1	Dispositivos eléctricos, ingeniería eléctrica, energía eléctrica	7365	7568	8768	10069	11035	9.6
2	Tecnología Audio-Visual	6057	6074	6713	7453	7759	4.1
3	Telecomunicaciones	10821	10441	11670	13634	15751	15.5
4	Tecnologías de Información	9917	9531	10992	13791	15109	9.6
II	<b>INSTRUMENTOS</b>						
5	Semiconductores	4051	4109	4719	6111	6587	7.8
6	Óptica	2616	2563	3215	5898	5960	1.1
7	Análisis, medición, tecnologías de control	11447	10869	11867	13225	13531	2.3
8	Tecnología Médica	8600	8877	9554	11249	11890	5.7
9	Ingeniería Nuclear	517	496	499	572	712	24.5
III	<b>QUIMICOS - MATERIALES</b>						
10	Química orgánica fina	5225	5652	6112	6512	6082	-6.6
11	Química macromolecular, polímeros	3984	4002	4532	5904	5946	0.7
12	Ingeniería química	3879	3702	4266	5680	5863	3.2
13	Tecnología de superficies, cubiertas	3293	3326	3642	4361	4247	-2.6
14	Materiales y Metalurgia	3037	3031	3252	3836	4045	5.4
IV	<b>FARMACEUTICOS-BIOTECNOLOGIA</b>						
15	Biotecnología	8604	7609	7312	7413	7228	-2.5
16	Farmacéuticos, cosméticos	9976	9436	11090	13920	13936	0.1
17	Agricultura y comida	1660	1840	1949	2336	2309	-1.2
V	<b>INGENIERIA DE PROCESOS</b>						
18	Procesos Industriales	5365	4909	4911	5010	5295	5.7
19	Manejo e impresión	4540	4555	5401	6234	6261	0.4
20	Procesos Agrícolas y de Alimentos, maquinaria y aparatos	1274	1334	1521	1504	1478	-1.7
21	Procesamiento de materiales, textiles y papel	4780	4285	4763	5484	5312	-3.1
22	Tecnología Ambiental	1314	1249	1380	1585	1780	12.3
VI	<b>MAQUINARIA-MECANICA-TRANSPORTE</b>						
23	Herramientas de Maquinaria	2485	2323	2773	3009	3132	4.1
24	Motores, bombas y turbinas	2820	2975	3205	3700	4170	12.7
25	Procesos Térmicos y aparatos	1580	1542	1825	2062	2297	11.4
26	Componentes Mecánicos	3567	3720	4108	4748	5084	7.1
27	Transporte	4597	4883	5542	6078	6696	10.2
28	Tecnología Espacial y Armamento	494	436	536	513	507	-1.2
VII	<b>BIENES DE CONSUMO-INGENIERIA CIVIL</b>						
29	Bienes de consumo y equipamiento	5757	6040	7228	8310	8629	3.8

Como se observa en la tabla anterior, todos los campos tecnológicos han presentado un aumento en el número de registros desde el 2003; en el 2007 sólo se presentaron siete variaciones negativas en las subcategorías respecto al 2006.

## VIII. Marco Teórico

Una vez vistos los conceptos de Vigilancia Tecnológica en el Marco Conceptual; se profundizará en este capítulo, en la evolución del conocimiento entre las interacciones de diferentes elementos químicos, desde la antigüedad y en el concepto de ciclo biogeoquímico, sus componentes principales, fuentes y sumideros de elementos, tiempos de residencia e importancia de cada ciclo en la vida en el planeta.

### A. Antecedentes históricos

Desde la prehistoria, el hombre siempre ha estudiado el comportamiento del mundo que lo rodea, sus elementos e interacciones. En la antigua Grecia, Tales de Mileto (aproximadamente 640-546 antes de Cristo), inició el estudio sobre el significado de los cambios en la naturaleza de la materia, al preguntarse si una sustancia puede transformarse en otra, y realizase los siguientes cuestionamientos: ¿como un trozo de mineral azulado puede transformarse en cobre rojo? (refiriéndose al proceso en el que luego de encender fuego de leña sobre estos minerales, se obtiene cobre), ¿Cuál es la naturaleza de la sustancia?, ¿es piedra, cobre o ambas cosas a la vez?, ¿puede cualquier sustancia transformarse en otra mediante un determinado número de pasos? (Suárez Menéndez, 2008); posteriormente Aristóteles (384-322 antes de Cristo), concibió a los elementos como combinaciones de dos pares de propiedades opuestas, que no se pueden combinar entre si, como Frio y Calor y Humedad y Sequedad; de esta forma, se forman cuatro posibles parejas, cada una originando un elemento, así calor y humedad forman Aire, frio y sequedad forman Tierra, frio y humedad forman Agua y calor y sequedad forman Fuego, adicional a estos cuatro elementos cambiantes, Aristóteles supuso la existencia de un quinto elemento: el Éter, que es eterno, perfecto e incorruptible, elemento que debía constituir el cielo (Asimov, 1980).

Posteriormente en Alejandría, se unió la teoría griega de los elementos, con la técnica egipcia en química aplicada, siendo el primer practicante conocido de esta *khemeia*<sup>7</sup> greco-egipcia

---

<sup>7</sup> De acuerdo con cierta teoría, la palabra *khemeia* deriva del nombre que los egipcios daban a su propio país: *Kham*, por consiguiente, *khemeia* puede ser el arte egipcio.

Bolos de Mendes (aproximadamente 200 antes de Cristo); quien, utilizó el nombre de Demócrito en sus escritos y adaptó la teoría de los cuatro elementos, que consideraba que las sustancias en el Universo, solo diferían en la mezcla elemental (Suárez Menéndez, 2008); esta hipótesis, no descartaba la posibilidad que los elementos podrían mezclarse como átomos o como sustancia continua; esta teoría llevo a pensar que los elementos eran intercambiables entre sí, y en lo que a mi juicio, constituye la primera aproximación conocida al ciclo hidrológico, ya que aparentemente, el agua se convertía en aire al evaporarse y retomaba su forma original cuando llovía sobre la tierra, así mismo describió como la leña al calentarla se transformaba en fuego y aire (vapor).

Luego, desde el año 650 después de Cristo, los árabes mantuvieron y extendieron el uso de la *khemeia* greco-egipcia, y la estudiaron como *al-kimiya* (siendo el prefijo *al*, equivalente a *la*) realizando importantes aportes a los aspectos médicos de la *al-kimiya*, siendo el persa Ibs Sina (979-1037d.c) conocido como Avicena, el médico más importante entre la época del Imperio Romano y los orígenes de la ciencia moderna (Suárez Menéndez, 2008). El estudio de la *al-kimiya*, declinó al sufrir invasiones turcas y mongoles en el siglo X, así como las cruzadas emprendidas desde Europa en el siglo XI, siendo precisamente estas cruzadas las encargadas de llevar este conocimiento a los europeos, donde la *al-kimiya* se adaptó como alquimia y luego de dos siglos de fusión cultural, se registró cierta apreciación de la ciencia árabe al conocer que los árabes poseían libros de gran contenido científico ya sean traducidos de los originales griegos, o productos originales de su cultura como los de Ibs Sina. (Bensaude-Vincent, y otros, 1997)

El primer alquimista europeo importante fue Alberto de Bollstadt (1200-1280), quien además de descubrir el arsénico en forma pura, estudió la obra aristotélica y la llevó a ser tema primario de estudio para quienes buscaban la erudición, a finales de la edad media; sin embargo, en 1317 comenzó el fin de la alquimia, cuando el papa Juan XXII condenó su práctica al mismo tiempo que los europeos comenzaban a dudar sobre los hallazgos del mundo antiguo y la superioridad científica de los griegos, ya que posteriormente al descubrimiento de la brújula, probaron sus habilidades superiores de navegación, al explorar la costa de África (1497) y emprender Cristóbal Colon viajes entre 1492 y 1504, que abrieron nuevas rutas de comercio, con partes del mundo desconocidas por Europa occidental hasta ese entonces. (Asimov, 1980; Suárez Menéndez, 2008)

La duda sobre la veracidad de las teorías desarrolladas en Grecia y Egipto, llevó a Galileo Galilei (1564-1642) a estudiar el comportamiento de los cuerpos durante su caída, aplicando las

matemáticas y las mediciones cuidadosas a la física, así como a los primeros indicios de la futura revolución química (Suárez Menéndez, 2008), entre estos indicios, se destacan los trabajos del médico flamenco Van Helmont (1574-1644), quien cultivó un árbol en una cantidad de tierra dada y lo pesaba periódicamente conforme le agregaba agua y crecía, aplicando así la medición a problemas de biología y química; Van Helmont también fue pionero en el estudio de los vapores (gases), observando su similitud al aire en su apariencia física pero no en todas sus propiedades; en especial estudió los vapores de la madera al arder (Dióxido de carbono), que parecían aire, pero no se comportaban como este, estas sustancias parecidas al aire, según Von Helmont, eran semejantes al *chaos* griego, la materia original, sin forma y desordenada a partir de la que se creó el universo, según la mitología griega; otro aporte importante en la época, fue la publicación en 1630 de un Ensayo sobre la investigación de la causa por la cual el estaño y el plomo aumentan de peso cuando se les calienta, por Jean Rey (1583-1645), en este ensayo, se atribuye el aumento de peso a la absorción de aire y se enuncia el principio de conservación de la materia de manera particular “ el peso que cada trozo de materia toma en la cuna, lo lleva hasta el ataúd” (Suárez Menéndez, 2008), principio que demostró Lavoisier en 1774 en su obra *Opúsculos químicos y físicos*, sobre experimentos de calcinación del plomo y de su estado en un recipiente cerrado, en esta publicación, Lavoisier demuestra gracias a la medición sistemática de pesos, que el aumento en el peso del material calcinado es igual al peso de aire absorbido y que la masa no se creaba ni destruía solo cambiaba de unas sustancias a otras (Ley de conservación de masa). (Suárez Menéndez, 2008)

La nueva concepción de los elementos, se apartó de la teoría greco-egipcia, con los trabajos de Robert Boyle (1627-1691) y la publicación de su libro *El Químico Escéptico* en 1661, en donde definía los elementos de una forma real y practica y no a través de simples razonamientos sobre la composición del universo. Posteriormente, Georg Ernest Stahl (1660-1734), desarrolló la teoría del flogisto, basado en el principio de inflamabilidad de la madera, en donde el aire resultaba útil en la combustión de manera indirecta, sirviendo únicamente como transportador del flogisto que abandonaba la materia, la transferencia del flogisto explicaba la reducción del mismo, uniendo por primera vez en la historia de la química, la oxidación y la reducción en un fenómeno de transferencia, muy similar al aceptado actualmente sobre transferencia de electrones (Suárez Menéndez, 2008), incluso Kant llegó a afirmar que era una ley tan importante como la ley de Galileo sobre caída de los cuerpos; fue precisamente la ley de caída de los cuerpos de Galileo, la que condujo a las conclusiones publicadas por Isaac Newton (1642-1742) en su libro *Principia Matemática* de 1687, en esta obra, Newton introdujo

sus tres leyes de movimiento, que sirvieron como base para el estudio de la ciencia de la mecánica. (Bensaude-Vincent, y otros, 1997)

En 1754, Joseph Black (1728-1799) en su tesis de medicina, trató sobre el estudio del carbonato cálcico y su descomposición al ser calentado, en gas y óxido de calcio, el gas lo recombinó con el óxido de calcio formando de nuevo carbonato cálcico; en su obra, mostró que este gas (dióxido de carbono), podía formarse calentando un mineral o quemando madera, estableciendo una conexión entre los minerales y el reino vegetal, así mismo demostró que las sustancias gaseosas pueden combinarse con sólidos y líquidos para producir cambios químicos y que el óxido de calcio al abandonarse el aire se transforma en carbonato cálcico, deduciendo correctamente que hay pequeñas cantidades de dióxido de carbono en la atmósfera (Suárez Menéndez, 2008). Black introdujo a Daniel Rutherford (1749-1819), en el estudio de las propiedades del dióxido de carbono, quien luego a una serie de estudios llamó a un gas aislado que no permitía vivir a los ratones, ni a las velas arder, aire flogistizado (Nitrógeno). En 1766 Cavendish (1731-1810), estudió sistemáticamente las propiedades del gas que se formaba al reaccionar ciertos metales con ácidos y fue el primero en medir el peso de volúmenes determinados de distintos gases, posteriormente en 1783, Cavendish comprobó que los vapores producidos al arder su gas, se condensaban para formar un líquido que resulto ser agua; este experimento llevó a Lavoisier a deducir que el gas de Cavendish era productor de agua (Hidrogeno) y que ardía por combinación con el oxígeno, por tanto el agua era una combinación de ambos elementos, demostrando que el agua no era una sustancia simple como afirmaban los griegos, sino la combinación de dos gases. (Asimov, 1980)

Retomando la obra de Lavoisier (Bensaude-Vincent, y otros, 1997), en 1774 se concluye que solo una parte del aire se puede combinar con los metales, dejando el interrogante de ¿Qué porción de aire se puede absorber?, Lavoisier contesta afirmando que se puede respirar el aire “puro” o “inminentemente respirable”, posteriormente llamado “principio oxígeno” y oxígeno, realizando así primer trabajo conocido sobre el oxígeno, aunque no lo caracterizó; simultáneamente Robert Priestley (1733-1804), obtiene dos vapores al calentar monóxido de mercurio, uno se condensó en gotas de Mercurio y el otro (Oxígeno) lo llamó aire desflogistizado; este hecho indujo a Lavoisier a retomar los experimentos de Priestley, usando su conocida práctica cuantitativa y calentando el óxido de mercurio en presencia de carbono; en este experimento, descubre que una parte del gas es soluble en agua ( $\text{CO}_2$ ) y que la otra parte activa la combustión (Leicester, 1967); las conclusiones de este trabajo las presentó en 1777 en su obra *Sobre la combustión en general*, posteriormente en su tesis de 1783 *Reflexiones sobre el flogisto para servir de continuación a la teoría de la combustión y de la*

*calcinación*, retoma los argumentos de Stahl y los refuta uno a uno; también consideró que los tejidos vivos y los alimentos contenían una combinación de carbono e hidrogeno, así al respirar, el oxígeno se consumía formando agua a partir de hidrogeno y dióxido de carbono a partir del carbono de los alimentos. (Suárez Menéndez, 2008)

Luego en el siglo XIX, Proust (1754-1826) formuló la ley de proporciones definidas o ley de Proust, en la que explica que un compuesto dado siempre tiene la misma proporción de elementos constituyentes, independientemente del método usado para obtenerlo; este hecho sería explicado a partir de la teoría atomista de Demócrito, que consideraba a la materia como la unión de partículas indivisibles (átomos); la ley de Proust, sería complementada por John Dalton(1766-1844), al formular la ley de proporciones múltiples, en la que afirmaba que diferentes proporciones de dos o más elementos formaban diferentes compuestos, descubrimiento que sirvió de base para formular la primera tabla de pesos atómicos; posteriormente, en 1840 el químico alemán Justus von Liebig, investigó el papel de los constituyentes del suelo en la nutrición de los cultivos. (Leicester, 1967)

En el siglo XX, se estableció que los grandes ciclos de Carbono, Hidrogeno, Oxígeno, Nitrógeno, Fósforo, Azufre y quizás de otros 25 elementos, sostienen la vida en la Tierra; estos ciclos interactúan y se retroalimentan entre sí, conforme los elementos se mueven en el medio ambiente; estas interacciones pueden ser directas, como cuando un elemento dado inmoviliza o cataliza reacciones que involucran otros elementos, o indirectas, como en los casos en que las limitaciones de un elemento dado, influyen en los ciclos de otros elementos en el ecosistema (Scientific Committee on Problems of the Environment (SCOPE) Series , 2003); por ejemplo la oferta de nitrógeno puede limitar los procesos vitales en los océanos; estos acontecimientos, determinaron que la comprensión de los ciclos biogeoquímicos era esencial para entender el funcionamiento de la tierra como sistema.

Recientemente, ha crecido la preocupación por conocer en qué medida actividades humanas, como la deforestación, la intensificación de la agricultura, la urbanización, la industrialización y el uso de combustibles fósiles, han afectado directa o indirectamente los procesos biogeoquímicos y físicos de la Tierra, procesos que pueden afectar un gran número de servicios ofrecidos por los ecosistemas a la humanidad, como mantener la relativa estabilidad climática, la auto depuración del ambiente, la protección de los rayos ultravioletas del sol y la regulación de los ciclos de nutrientes, entre otros.

## B. Ciclos Biogeoquímicos

### 1. Introducción

El estudio de los ciclos globales de los elementos biológicamente activos es una parte importante de diferentes cursos básicos y avanzados de Ciencias de la Tierra, Ecología y Biogeoquímica. Así como el entendimiento de los ciclos biogeoquímicos y los impactos antropogénicos sobre los mismos, es fundamental para el estudio de cambio climático a nivel global; sin embargo, la mayoría de investigadores se especializan en facetas determinadas de estos ciclos, dejando de lado el estudio de las interacciones entre los mismos y su análisis usando una perspectiva más amplia.

Citando a Lavoisier y su trabajo de 1774 (Corsi, 2008), sobre la formulación de la ley de conservación de masa, donde establece que *“En toda reacción química la masa se conserva, es decir, la masa consumida de los reactivos es igual a la masa obtenida de los productos”*, y partiendo de la suposición que esencialmente la Tierra es un sistema cerrado, respecto al balance de materia, se puede afirmar que toda la materia dentro del planeta pertenece a un ciclo; sin embargo la Tierra es un sistema abierto al tratarse de energía, ya que permite la entrada y salida de radiación en la atmosfera (Suárez Menéndez, 2008).

Estos ciclos globales incluyen la circulación de elementos o nutrientes, que son fundamentales para la vida y el sistema climático de la Tierra, siendo el vapor de agua crítico para la supervivencia de los sistemas ecológicos, al condensarse para formar nubes, se reflejan más rayos de sol a la atmósfera, enfriando así el clima, de otra parte, el mismo el vapor de agua, constituye un gas de efecto de invernadero, atrapando calor en la parte infrarroja del espectro electromagnético. De estos ciclos, sobresale el ciclo del agua, ya que interseca la mayoría de los otros ciclos biogeoquímicos, incluyendo los ciclos del carbono, nitrógeno, azufre, fósforo y el ciclo sedimentológico.

Las figuras de modelos de caja, presentadas en este capítulo fueron desarrolladas por la Universidad de Colorado (2007), como complemento a un curso de posgrado en Sistemas de la Tierra, y permiten comparar diferentes ciclos biogeoquímicos (C, O, N, S, P, Si, H<sub>2</sub>O y CH<sub>4</sub>), dado que el formato usado muestra los tamaños de los reservorios o piscinas, las transferencias o flujos entre piscinas, ya sean antropogénicos o naturales y los tiempos de residencia estimados, usando una paleta de colores estándar que identifica las piscinas, flujos y tiempos de residencia.

## 2. Descripción

El concepto de ciclo biogeoquímico se usa para describir la distribución, el transporte de materiales y la transformación de sustancias en el ambiente, sujetas a reacciones químicas, en presencia de energía, mediante interacciones con formas de vida (*bios* = vida) y suelos (*geo* = tierra). Algunos autores, ubican a los ciclos biogeoquímicos como uno de los cuatro componentes del ciclo geológico, junto con el ciclo tectónico, el ciclo hidrológico y el ciclo sedimentológico. (University of Colorado, Geological Sciences, 2007)



Figura VIII-1 Componentes del Ciclo Geológico

Como se explicó anteriormente, existen los grandes ciclos biogeoquímicos de macronutrientes, siendo usual su clasificación entre “los seis grandes”, o elementos que constituyen la materia orgánica y “otros” macronutrientes que cumplen la función de matrices iónicas o estructuras de soporte (Ca, Mg, Si, K, Na, Cl, F), de otro lado, entre los 80 elementos que se encuentran en el suelo, existen aproximadamente 25 ciclos de micronutrientes esenciales en plantas y animales. Siendo el carbono, el azufre y el mercurio son elementos que han experimentado perturbaciones significativas de sus ciclos en las últimas 10 generaciones del hombre.

A continuación, se presenta en la Figura VIII-2, una clasificación entre macronutrientes y micronutrientes; estando los macronutrientes divididos en dos categorías, según la explicación anterior: Los Seis Grandes y los Otros.

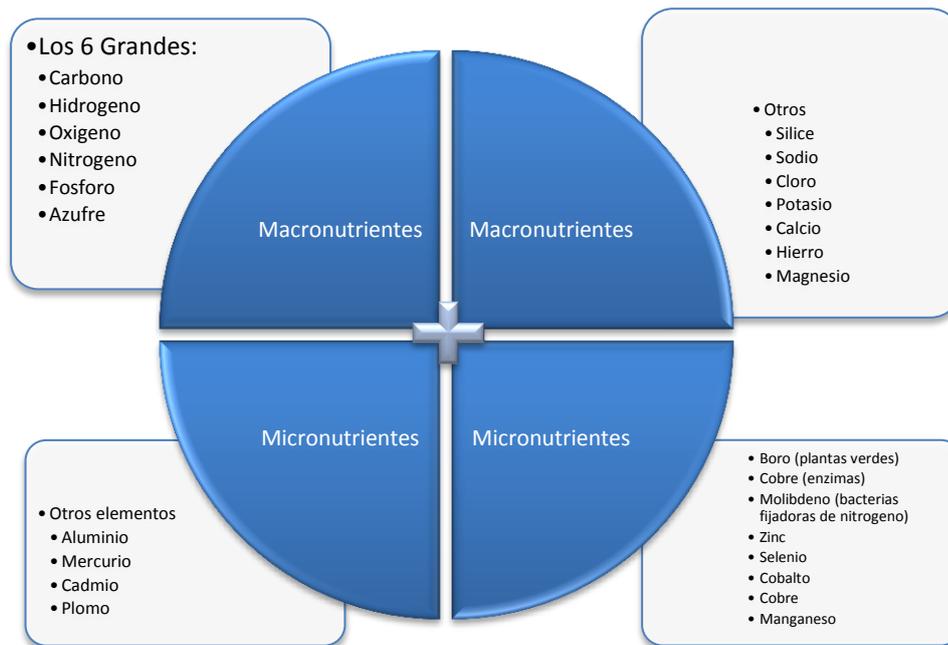


Figura VIII-2 Clasificación de ciclos biogeoquímicos según el tipo de nutrientes

Usualmente los ciclos miden la dinámica del recambio de materia, comparando las magnitudes en el depósito y sus flujos en los diferentes compartimentos del ecosistema, siendo de particular interés el estudio de las escalas espaciales y temporales de las transformaciones y las fases de transición.

Los aspectos básicos de los ciclos biogeoquímicos son (Garcia, Mary Ruth , 2005):

- La distribución de materiales- localización y tamaño del depósito.
- El transporte - patrones y ratas de flujo.
- La transformación- rata de flujo del depósito a otro componente.
- Tiempo de residencia- Tiempo de almacenamiento.

Adicionalmente, el sistema terrestre se puede subdividir en atmósfera, hidrósfera, litósfera y biósfera; en donde la materia ya sea en forma de moléculas como agua, o elementos como carbono, fósforo y oxígeno, se moviliza entre los diversos subsistemas en diferentes fases de la misma (sólido, líquido y gas), constituyendo un ciclo biogeoquímico (ver Figura VIII - 3).

En un ciclo biogeoquímico generalizado, la biosfera está conectada por el flujo de elementos y compuestos químicos, donde en la mayoría de los casos, la biota desempeña un papel primordial, de otra parte, la materia del interior de la Tierra es liberada por los volcanes y la atmosfera intercambia algunos elementos y compuestos rápidamente con la biota y los

océanos. Los intercambios de material entre suelos, rocas y océanos ocurren, pero requieren de ventanas de tiempo mayores.

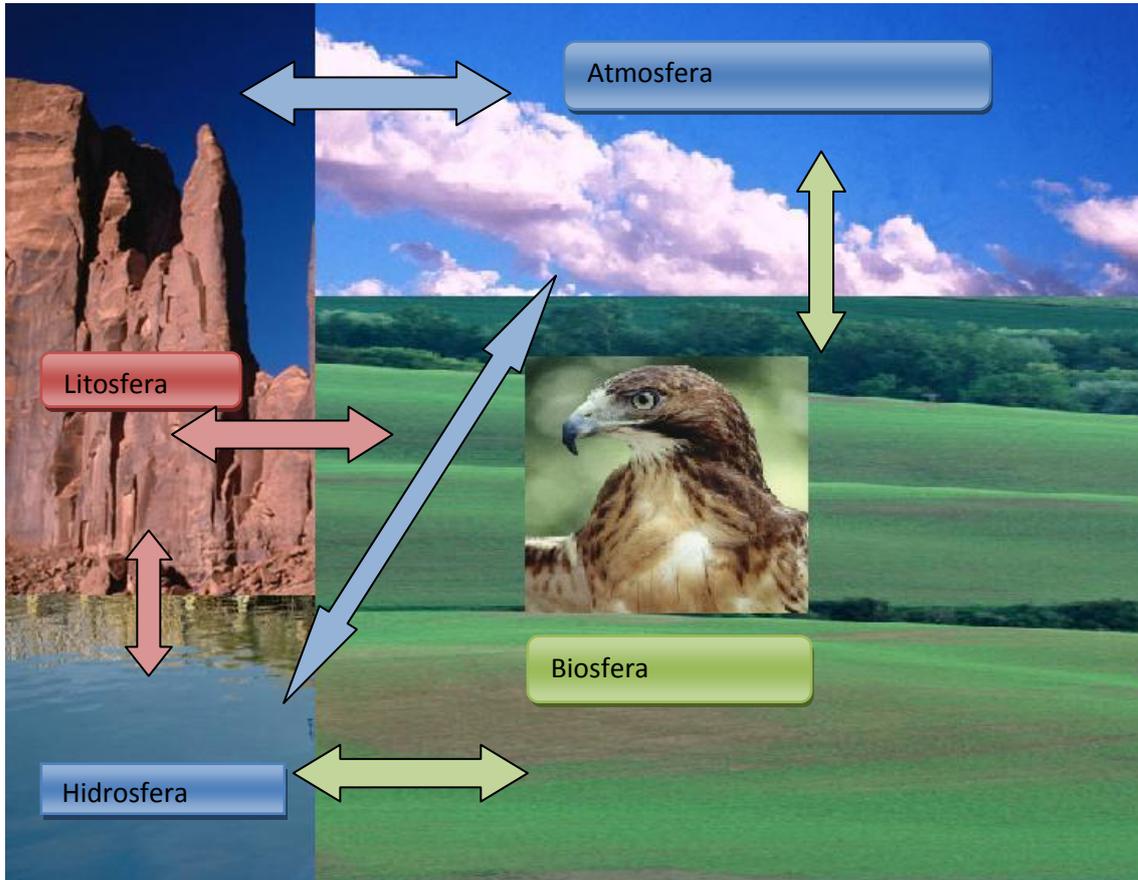


Figura VIII-3 Interacción entre las diferentes esferas terrestres (esquema)

De los elementos en la biomasa, el agua de mar y la corteza terrestre (Tabla 4), se resalta el contenido de los siguientes elementos:

1- Hidrógeno (H)	2- Carbono (C)	3- Nitrógeno (N)
4- Oxígeno (O)	5- Sodio (Na)	6- Magnesio (Mg)
7- Aluminio (Al)	8- Silicio (Si)	9- Fósforo (P)
10- Azufre (S)	11- Potasio (K)	12- Calcio (Ca)
13- Titanio (Ti)	14- Manganeso (Mn)	15- Hierro (Fe)

Tabla 4. Elementos de interés y su abundancia en diferentes ambientes (García, Mary Ruth , 2005)

Elemento			Biomasa			Corteza terrestre		Mar	
Símbolo	No Atómico	Masa Atómica	Masa ppm	Átomos ppm	Esencial biota	Masa ppm	Átomos ppm	Masa ppm	Átomos ppm
H	1	1	65.900	496.800	*	1400	28.800	111.000	110.000
B	5	10.8			*			4.5	0.410
C	6	12	393.460	248.900	*	200	350	27.8	2387
N	7	14	5020	2720	*			0.4	0.030
O	8	16	524.290	249.00	*	466.000	604.000	883.000	55.200
F	9	19			*	625	660	1.3	0.068
Na	11	23	190	63	*	28.300	25.500	10.700	468
Mg	12	24.3	980	307	*	20.900	17.800	1290	53.2
Al	13	27	560	157		81.300	62.500		
Si	14	28.1	1210	327	*	277.200	205.000	2.9	0.103
P	15	31.0	520	128	*	1050	700	0.07	0.002
S	16	32.1	710	169	*	260	170	904	28.2
Cl	17	35.5	500	106	*			19.353	546
K	19	39.1	2290	444	*	25.900	13.700	399	10.2
Ca	20	40.1	3780	717	*	36.300	18.800	412	10.2
Ti	22	47.9				4.400	1900		
Mn	25	54.9	210	29	*	950	360		
Fe	26	55.9	390	553	*	50.000	18.600	0.03	
Co	27	58.9			*				
Cu	29	63.6			*				
Zn	30	65.4			*				
Se	34	79			*				
Br	35	79.9						67	0.084
Sr	38	87.6				375	89	8	0.091
Mo	42	95.9			*				
Ba	56	137.3				425	64		

### 3. Conceptos básicos sobre circulación de nutrientes

En un esquema general sobre la circulación de nutrientes en un ecosistema, los elementos químicos circulan en el ecosistema o son intercambiados entre el ecosistema y la biosfera; los organismos intercambian elementos con los componentes inanimados del medio ambiente en donde algunos elementos son tomados y luego liberados a la atmosfera y otros son intercambiados con agua, suelo o sedimentos.

Los elementos se mueven entre piscinas de almacenamiento, ubicadas dentro del ecosistema a diferentes razones de transferencia y en cada piscina tienen diferentes tiempos de residencia. Por ejemplo el suelo en un bosque tiene una parte activa, que cambia elementos rápidamente con organismos vivos y otra parte inactiva, que intercambia elementos lentamente. Como característica general de la circulación de nutrientes, se observa que el movimiento a través de la atmosfera es usualmente más rápido que por otros medios, siendo el movimiento por los suelos particularmente lento. (Universidad de California, Irvine, 2007)

Generalmente, la vida se beneficia si los elementos son conservados dentro del ecosistema, y no son eliminados por procesos geológicos, como la erosión. (University of Colorado, Geological Sciences, 2007)

A continuación, a manera de guía para el lector, se muestra el diagrama usado por la Universidad de Colorado (Figura VIII - 4), en su curso de Geología, sobre conceptos básicos de circulación de nutrientes.

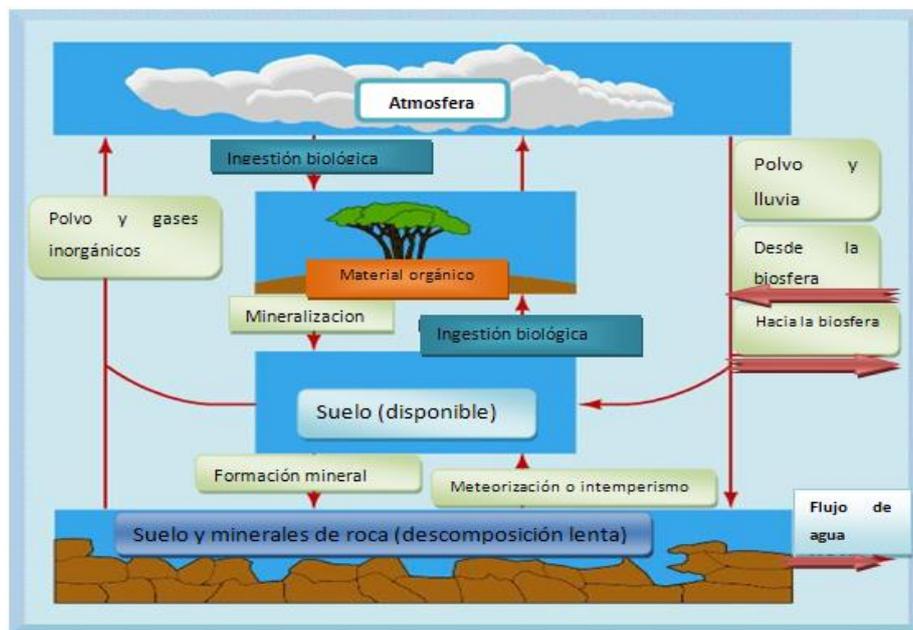


Figura VIII-4 Circulación de nutrientes. Adaptado de (University of Colorado, Geological Sciences, 2007)

#### 4. Relación entre el Ciclo Hidrológico y el Ciclo Sedimentológico

El ciclo hidrológico está unido al ciclo sedimentológico a través de la precipitación, ya que el agua arrastra materiales desde tierra hasta los océanos, donde se depositan como sedimentos; en una escala de tiempo corta, el ciclo sedimentológico incluye los procesos de erosión física o química, transporte de nutrientes y formación de sedimentos, procesos que están relacionados a los flujos de agua. En escalas de tiempo mayores, operan los procesos de sedimentación, transformación química, deriva continental y de cambios geológicos en el fondo del mar.

Tanto el ciclo hidrológico como el sedimentológico, están afectados por la distribución y los flujos de los *seis grandes* macronutrientes mencionados anteriormente; estos elementos se combinan en diversas formas para comprender más del 95% de todas las formas de vida, (University of Colorado, Geological Sciences, 2007) en donde cantidades adecuadas de estos macronutrientes, en las concentraciones correctas y en los sitios indicados son necesarias para mantener la vida.

#### 5. CICLO HIDROLÓGICO

Tratar el ciclo hidrológico de primero en este trabajo, no es cuestión de azar, por el contrario esta posición es más que merecida, ya que la base para la comprensión de la mayor parte de los ciclos biogeoquímicos está en el ciclo del agua y es esencial para la formación del sistema climático; además no se puede desconocer la importancia del abastecimiento de agua, que junto con la radiación solar son elementos esenciales para asegurar el crecimiento biótico, sirviendo como solvente, como reactante químico y como agente físico; siendo el abastecimiento de agua renovable continuamente distribuido por el ciclo hidrológico (California State University Monterey Bay, 2006).

En términos generales el ciclo se comporta así: El vapor de agua en la atmósfera se condensa, distribuye y se deposita en forma de precipitación sobre la superficie terrestre a través de la lluvia, la nieve o el granizo; la evaporación procedente tanto de los lagos, como de ríos, plantas y del suelo húmedo convierte el agua líquida en vapor de agua, gracias a la radiación solar. El agua es transferida al aire desde las hojas de las plantas, principalmente mediante un proceso llamado transpiración, este proceso combinado con la evaporación de cuerpos de agua y del suelo, se conoce como evapotranspiración, siendo la evaporación oceánica seis veces mayor a la registrada en los suelos continentales, el ciclo se cierra con la relación entre los océanos y la

corteza terrestre, donde el agua de las precipitaciones pasa a formar parte de los ríos, lagos, glaciares y aguas subterráneas y gracias a diferentes escalas temporales, todas estas fuentes acaban aportando su flujo de agua a los océanos (Universidad de California, Irvine, 2007).

La evaporación del agua permite la concentración de compuestos disueltos y eventualmente la precipitación de sólidos disueltos; mientras que la precipitación permite la dilución de materiales. Ambos procesos permiten la formación y meteorización de las rocas (University of Colorado, Geological Sciences, 2007).

A continuación se mostraran dos diagramas de interacción, el primero de ellos, enseña los flujos dentro de los diferentes componentes del ciclo hidrológico, el segundo, muestra los procesos que lo componen, como se dijo anteriormente el diagrama de cajas correspondiente a los procesos, fue adaptado del elaborado por la Universidad de Colorado (2007)

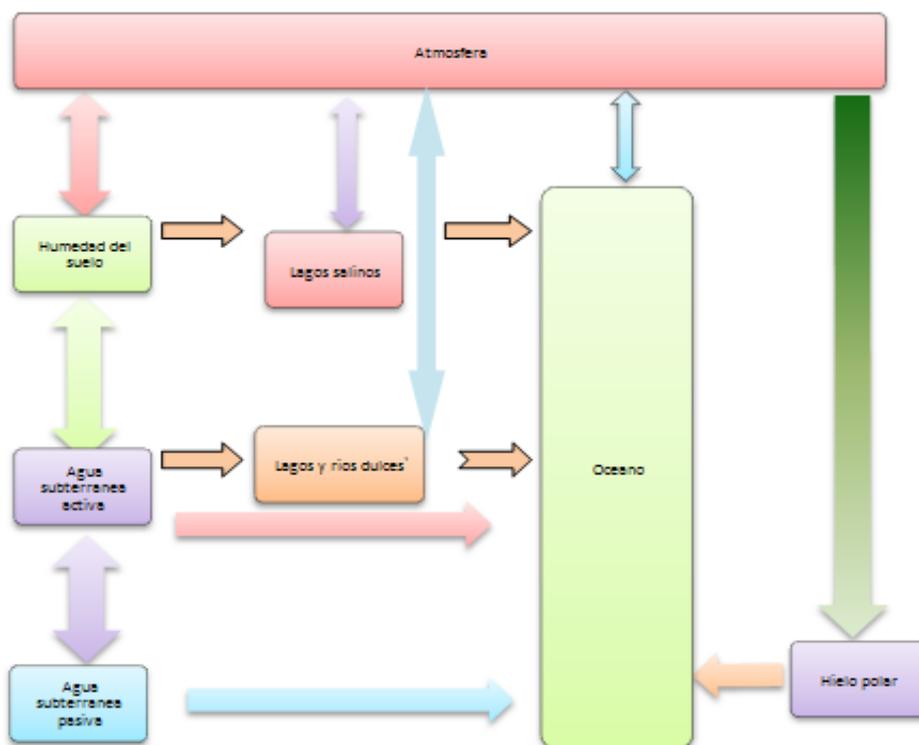


Figura VIII-5 Interacción de cuerpos de agua en el Ciclo Hidrológico (University of Colorado, Geological Sciences, 2007)

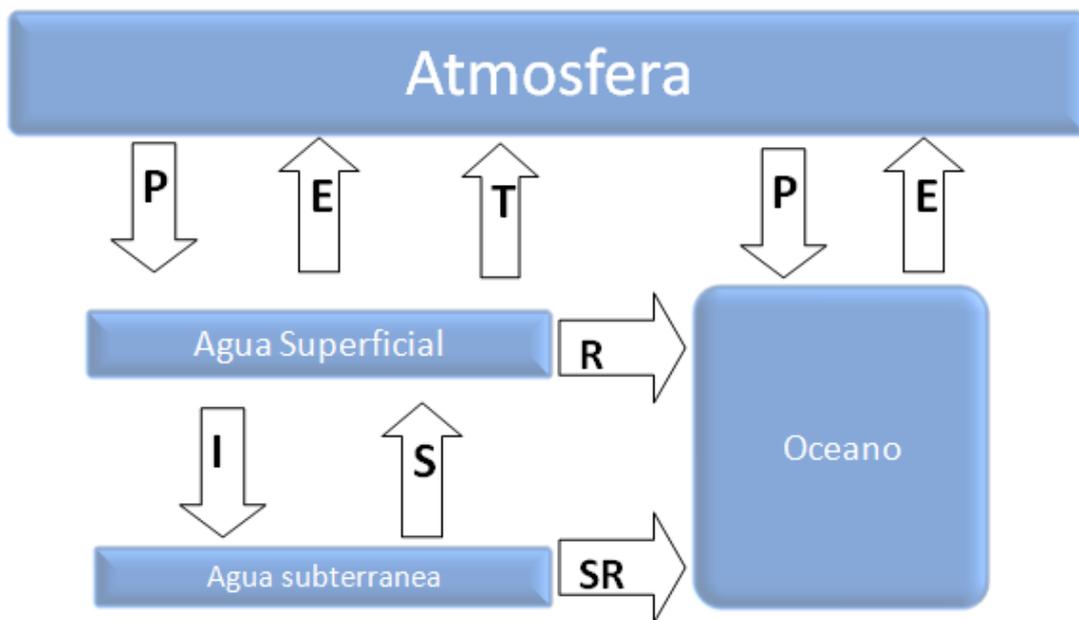


Figura VIII-6 Modelo de Caja. Interacción de Procesos Hidrológicos (University of Colorado, Geological Sciences, 2007)

a) *Reservorios, Flujos y Tiempos de Residencia*

**Reservorios:**

Mientras que la cantidad de agua en el planeta pareciera ser inmensa, la cantidad de agua fresca precipitable es una pequeña fracción de este total.

Tabla 5. Reservorios de agua en el planeta (University of Colorado, Geological Sciences, 2007)

Reservorio	km <sup>3</sup>	%
Atmosfera	12,700	0.001
Océano	1,230,000,000	97.200
Lagos Superficiales	123,000	0.009
Ríos y cursos de agua	1,200	0.0001
Agua Subterránea	4,000,000	0.310
Criósfera (glaciares)	28,600,000	2.150

## Flujos:

Tabla 6. Flujos de agua en el planeta (University of Colorado, Geological Sciences, 2007)

Proceso	Componente	Flujos: km <sup>3</sup> /año
P: precipitación total		496,000
	suelo	111,000
	océano	385,000
E, T: evapotranspiración total		496,000
	suelo	71,000
	océano	425,000
R: escorrentía superficial		26,000
SR: escorrentía sub superficial		
	líquido	12,000
	hielo	2,000
I: infiltración		14,000
S: manantiales		2,000

## Tiempos de Residencia<sup>8</sup>:

Tabla 7. Tiempos de residencia de agua en el planeta (University of Colorado, Geological Sciences, 2007)

Reservorio	Volumen (km <sup>3</sup> )	Flujo (km <sup>3</sup> /año)	Tiempo de Residencia (años)
Atmósfera	12.700	496.000	0.03 – 9 días
Océano:	1230.000.000	425.000	2900
Ríos:	1200	26.000	0.05 -17 días
Agua subterránea:	4.000.000	12.000	330

El tiempo de renovación del agua es del casi 9 días en la atmósfera y de 2-3 semanas en los ríos grandes. El tiempo de renovación de los océanos está entre 2900 y 37000 años, lo cual depende del tamaño y de la evaporación neta. El hombre ha incrementado la rata de evaporación natural, con actividades como la irrigación y la deforestación, en cerca de 3%, pero se estima que esta rata alcanzará valores entre 10-50% en el futuro cercano. (Garcia, Mary Ruth , 2005)

<sup>8</sup> Conforme a la definición de Klieber (Nature Publishing Group, 1955), donde considera el tiempo de residencia (*turnover time*), como la razón en la que el reservorio de la sustancia es remplazado.

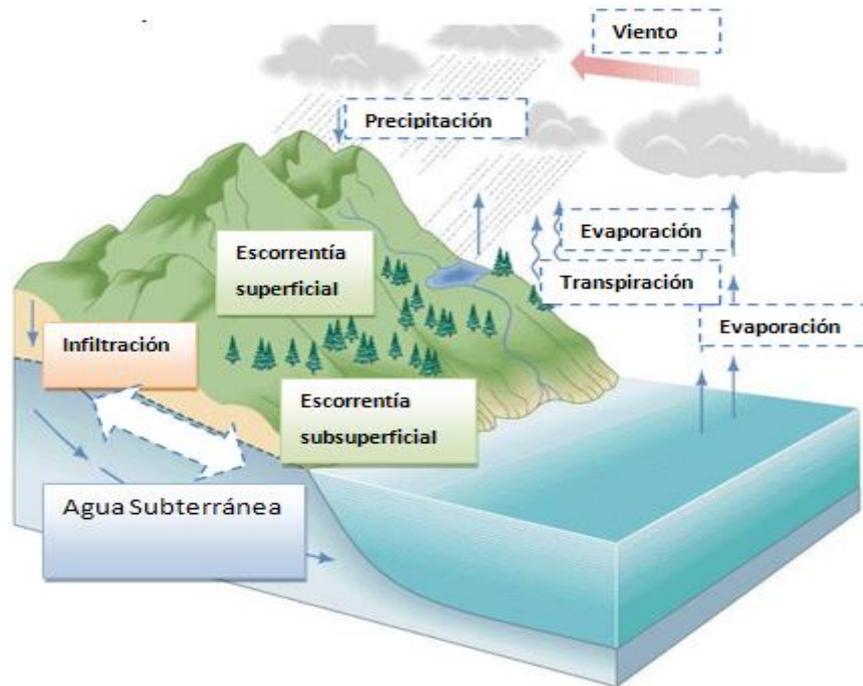


Figura VIII-7. Ciclo Hidrológico. Se muestra la transferencia de agua entre sus componentes. Adaptado (University of Colorado, Geological Sciences, 2007)

El ciclo de agua desempeña un papel fundamental en el funcionamiento tanto del sistema climático como del conjunto de ciclos biogeoquímicos, conectando un sistema con el otro. Cambios en el clima global, pueden causar cambios en el ciclo hidrológico, se estima que aumentos en la temperatura y en la evaporación, crearían un aumento en la precipitación, cambio que afectaría los la escorrentía y la humedad en los suelos, con una posible afectación de los patrones de vegetación y la agricultura a nivel global. (Ver Figura VIII-7).

Los modelos y observaciones estudiados por el Panel Intergubernamental sobre Cambio Climático, indican que los cambios en el flujo radiativo de la superficie de la Tierra afectan al balance de calor y humedad de la superficie y por tanto, el ciclo hidrológico se ve implicado. En concreto, los cambios en los aerosoles pueden afectar con más fuerza a las precipitaciones y a otros aspectos del ciclo hidrológico que otros agentes de forzamiento antropogénico. (IPCC, 2007)

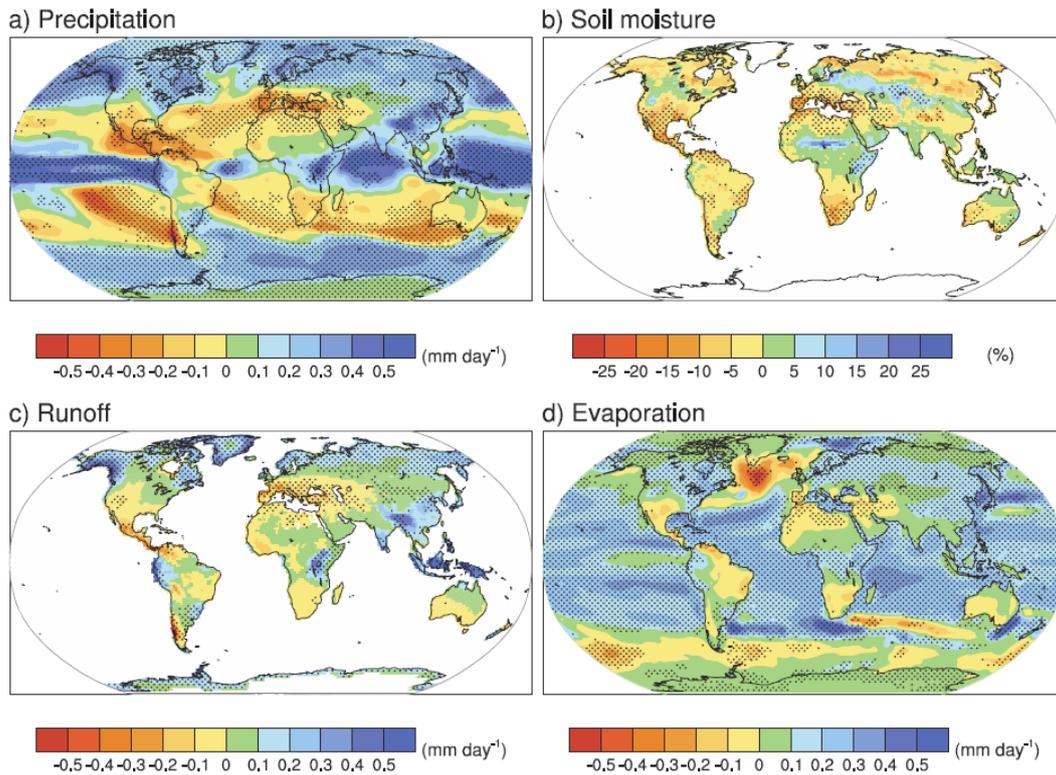


Figura VIII-8 Cambios medios anuales en precipitación, humedad del suelo, escorrentía y evaporación, para el periodo 2080-2099, relativos al periodo 1980-1999. (IPCC, 2007)

## 6. CICLO DEL CARBONO

El ciclo biogeoquímico del carbono es de mayor complejidad que el del agua y tiene gran importancia para el estudio del cambio global, a su vez, el carbono es un elemento clave para la vida en el planeta, y su ciclo incluye cuatro grandes reservorios : el  $\text{CO}_2$  en la atmosfera, los compuestos orgánicos en seres vivos o en organismos que estaban vivos, como  $\text{CO}_2$  disuelto en los océanos y en otros cuerpos de agua y como carbonato de calcio en rocas y en materia orgánica enterrada, como gas natural, carbón y petróleo. El ciclo del carbono por lo tanto está estrechamente relacionado con organismos vivos. (Universidad de California, Irvine, 2007)

Las plantas extraen continuamente carbono de la de atmosfera y lo usan para formar hidratos de carbono y azúcares, para construir sus tejidos gracias al proceso de fotosíntesis; los animales consumen plantas y usan estos compuestos orgánicos en su metabolismo, al morir las plantas y los animales, se forma  $\text{CO}_2$  de nuevo, al combinarse los compuestos orgánicos con el oxígeno durante la descomposición; aunque no todos los compuestos se oxidan, una pequeña fracción es transportada y re depositada como sedimentos en lugares donde es

atrapada para formar depósitos de carbón y petróleo. El dióxido de carbono en la atmósfera, también se disuelve en los océanos y en los cuerpos de agua, donde las plantas acuáticas lo usan para fotosíntesis y algunos animales acuáticos lo emplean para construir conchas de carbonato de calcio ( $\text{CaCO}_3$ ), estas conchas de organismos muertos como fitoplancton y arrecifes de coral, se acumulan en el lecho marino y pueden formar calizas que hacen parte del ciclo sedimentológico. (University of Colorado, Geological Sciences, 2007)

El dióxido de carbono, es un gas atmosférico que tiene un efecto sustancial en el balance de calor del planeta, al absorber radiación infrarroja; este gas así como el vapor de agua, el metano ( $\text{CH}_4$ ), y el  $\text{N}_2\text{O}$ , tiene una gran incidencia en el efecto invernadero, y sin ellos el planeta estaría congelado, sin embargo la vida en el planeta puede alterar la concentración global de  $\text{CO}_2$  en periodos de tiempo muy cortos, durante las etapas de crecimiento, el  $\text{CO}_2$  disminuye en la atmósfera, cerca a latitudes ecuatoriales, dado que la mayor radiación del sol y la temperatura facilitan el crecimiento de las plantas y por lo tanto su razón de ingestión de  $\text{CO}_2$ . En el invierno, entra más dióxido de carbono que el que es removido por las plantas, por eso su concentración crece al aumentar la respiración de las plantas y porque los procesos de descomposición de plantas y animales, ocurren más rápidamente que la fotosíntesis. De otro lado, al ser la masa continental del hemisferio norte mayor que en el hemisferio sur, la concentración global de  $\text{CO}_2$ , varía con la estacionalidad de la vegetación terrestre en dicho hemisferio. (University of Colorado, Geological Sciences, 2007)

La fotosíntesis, es un proceso químico complejo, en el que las plantas y otros organismos autótrofos, convierten agua, dióxido de carbono y luz solar en azúcares y oxígeno, necesarios para la vida en la Tierra, siendo la clorofila encontrada dentro de los cloroplastos de las plantas la encargada de capturar la luz solar. (Villegas, 1990)



La ecuación anterior muestra la producción de formaldehído ( $\text{CH}_2\text{O}$ ), uno de los carbohidratos formados en el proceso, como se aprecia en la ecuación, el oxígeno es un subproducto de la reacción; este oxígeno abandona las plantas por los estomas de las mismas.

El proceso inverso a la fotosíntesis es llamado respiración, mediante este proceso, los organismos vivos (animales, plantas, hongos y la mayoría de bacterias) transforman azúcares y

oxígeno en energía bioquímica. Durante la respiración los carbohidratos se rompen y la energía almacenada en los enlaces es utilizada para producir adenosintrifosfato (ATP) y otros compuestos con altas energías, que pueden usarse en las células para promover los diferentes procesos necesarios para soportar la vida. Adicional a la producción de energía, la respiración produce dióxido de carbono y agua como subproductos, teniendo la ecuación de la reacción la siguiente forma (Villem, 1990):



Oxígeno y carbohidratos

La Figura VIII -9, muestra los procesos involucrados en el ciclo del carbono básico, y sus tiempos de residencia (University of Colorado, Geological Sciences, 2007):



Figura VIII-9 Modelo del ciclo del carbono. Adaptado. ( University of Colorado, Geological Sciences, 2007))

### a) Reservorios, Flujos y Tiempos de Residencia

**Reservorios:**

La mayoría del carbono se encuentra en las rocas en forma de carbonatos y otros sedimentos, del carbono que no está en rocas, la mayor parte está en el océano, respecto a la cantidad de carbono en los suelos, es aproximadamente tres veces la cantidad encontrada en las plantas.

Tabla 8. Reservorios mundiales de Carbono (University of Colorado, Geological Sciences, 2007)

Reservorio	Miles de millones de toneladas métricas
Atmosfera	720
Océano	39.000
Carbonatos	100.000.000
Combustibles fósiles	4.000
Plantas terrestres	560
Suelos	1500

**Flujos:**

En el siguiente modelo de caja (Figura V-10), se muestran los flujos dentro del ciclo del carbono y los componentes del mismo.

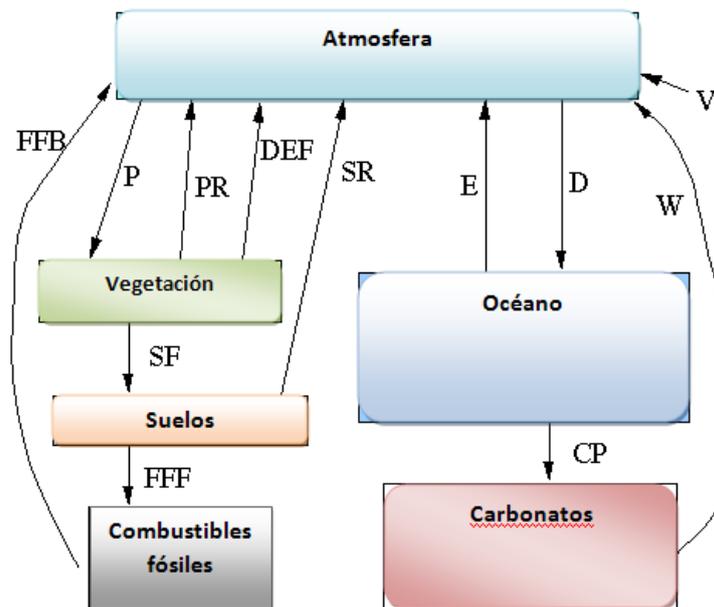


Figura VIII-10 Modelo de caja, Ciclo del Carbono. Adaptado. ( University of Colorado, Geological Sciences, 2007))

Los procesos y flujos se muestran a continuación:

Tabla 9. Procesos y Flujos del Ciclo del Carbono (University of Colorado, Geological Sciences, 2007)

Reservorio	Procesos	Sigla según Figura 10	Flujos en miles de millones de toneladas /año
Plantas	Fotosíntesis	P	120
	Respiración Vegetal	PR	60
	Respiración del suelo	SR	60
	Plantas a suelos	SF	60
	Formación de combustibles fósiles	FFF	0.0001
	Combustión de combustibles fósiles	FFB	6
	Deforestación	DEF	2
	Océanos	Disolución	D
Exsolving <sup>9</sup>		E	103
Formación de Carbonato		CP	4
Meteorización		W	0.6
Volcanes		Volcanes	V

De la tabla anterior (6), se puede inferir que según los flujos actuales, hay un aumento de CO<sub>2</sub> en la atmosfera; al realizar un balance de flujos entrantes (**I**), como respiración vegetal, respiración del suelo, combustión de combustibles fósiles, deforestación, exsolving y meteorización y salientes (**O**) como fotosíntesis y disolución, se obtiene una diferencia aproximada de cuatro mil seiscientos millones de toneladas entrando a la atmosfera, a pesar que el flujo hacia la atmosfera de FFB y DEF es cercano a los ocho mil millones de toneladas al año; este fenómeno se debe a que el intercambio oceánico se realiza por difusión. (University of Colorado, Geological Sciences, 2007)

$$I - O = \Delta\phi$$

$$231.6 - 221 \approx 4.6 \text{ miles de millones de toneladas/año}$$

Se observa que los flujos causados por la actividad humana, son pequeños pero persistentes, así como que los mayores flujos se encuentran entre la vegetación y la atmosfera y entre el océano y la atmosfera, de otra parte, el flujo de carbono por FFB (combustión de combustibles fósiles) es sesenta mil veces mayor que el flujo FFF (formación de combustibles fósiles).

<sup>9</sup> Es el proceso en el que una solución homogénea en estado sólido, se separa en al menos dos moléculas sin la adición o remoción de material.

**Tiempos de Residencia:**

Las escalas de tiempo para los diferentes procesos, difieren en varios órdenes de magnitud, desde millones de años para los procesos que involucran la formación y liberación de carbonatos así como días o segundos para procesos como fotosíntesis e intercambios entre la atmosfera y el océano.

**Tabla 10. Tiempos de residencia promedio en los reservorios de carbono. (University of Colorado, Geological Sciences, 2007)**

Reservorio	Tiempo de Residencia promedio (años)
Atmosfera	3
Vegetación:	5
Suelos:	25
Combustibles fósiles	650
Océanos	350 millones
Carbonatos	150 millones

***b) Cambios en el Ciclo del Carbono***

Desde 1958 Charles Keeling y otros han medido las cantidades de dióxido de carbono, en la atmosfera del monte Mauna Loa en Hawaii, E.E.U.U. En la Figura VIII - 11, se observan las fluctuaciones anuales del ciclo, debidas al crecimiento de la vegetación, como se explicó anteriormente, al mismo tiempo, es posible observar la tendencia creciente de las emisiones conforme pasa el tiempo, atribuida a la quema de combustibles fósiles, la deforestación y la producción de cemento, entre otros factores. (University Corporation for Atmospheric Research (UCAR), 2007)

**Curva Keeling. Mediciones de CO<sub>2</sub> atmosférico en Mauna Loa, Hawaii**

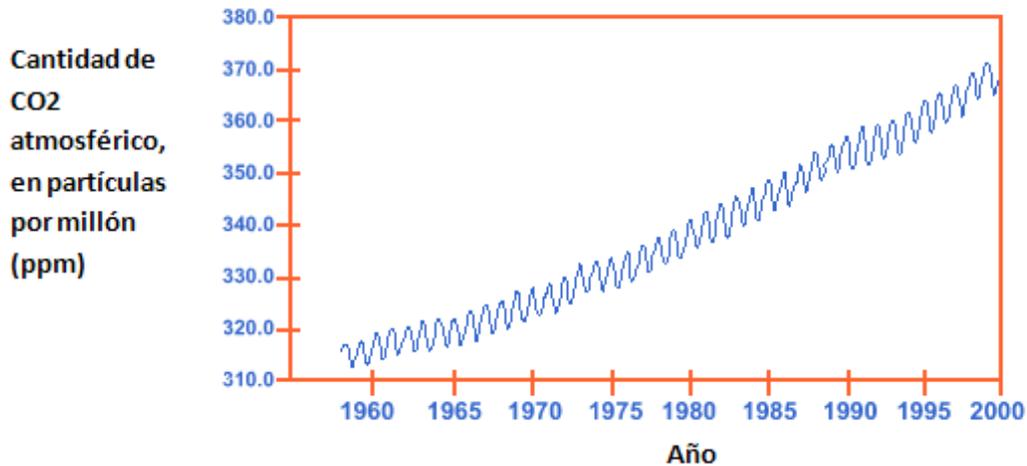


Figura VIII-11 Curva Keeling. Mediciones de CO<sub>2</sub> atmosférico en el monte Mauna Loa, Hawaii. Basada en datos de la NOAA y la UCSD (University Corporation for Atmospheric Research (UCAR), 2007)

No es la primera vez que el planeta tiene esta cantidad de dióxido de carbono en la atmosfera; durante el Paleozoico, gracias a la cantidad de erupciones volcánicas, se emitieron toneladas de lava, cenizas y CO<sub>2</sub>, el dióxido de carbono que permaneció por cientos de años en la atmosfera. (IPCC, 2007)

En la Figura VIII-12, se observa en el cuadro a) la concentración de CO<sub>2</sub> atmosférico, para el siglo XXI, conforme a los nuevos modelos planteados por el IPCC (rojo), respecto a la concentración estándar calculada con el modelo BERNCC<sup>10</sup> (negro). En el cuadro b) Temperatura superficial media global, relativa al 2000, simulada por los modelos C4MIP (usan emisiones de CO<sub>2</sub>), comparada con el calentamiento global simulado por los modelos AR4 del IPCC (usan concentraciones de CO<sub>2</sub>). (IPCC, 2007)

<sup>10</sup> Bern Climate Cycle Model, desarrollado por la Universidad de Bern, Suiza

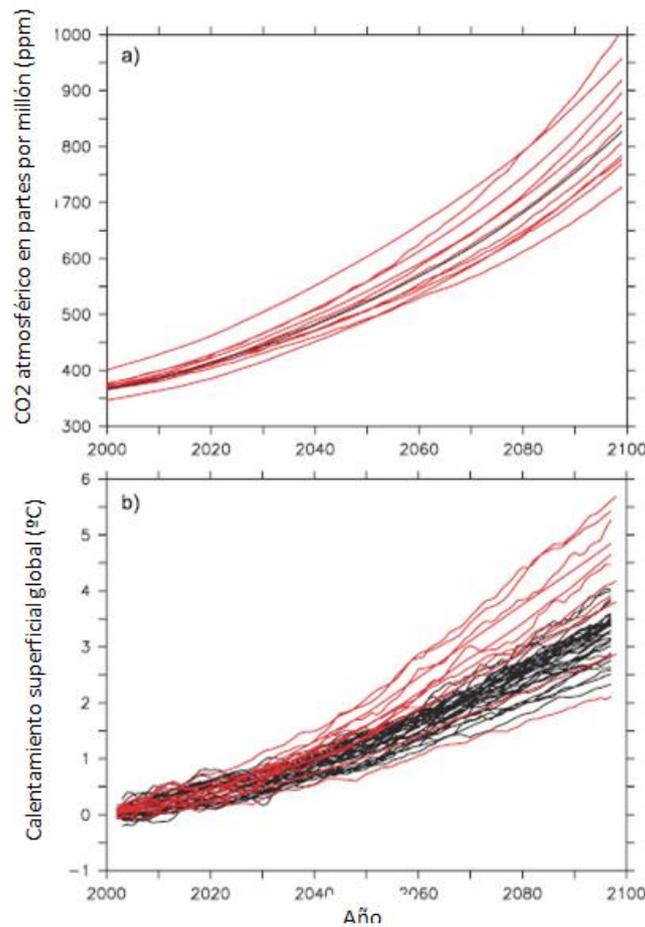


Figura VIII-12 Escenarios de concentración de CO<sub>2</sub> y calentamiento de la superficie terrestre, para el siglo 21. (IPCC, 2007)

### c) *El Metano*

El metano está presente en tres (3) reservorios en el planeta: a) como gas natural asociado a las reservas de combustibles fósiles. b) como hidratos de gas o clathratos (estructura de cristales de hielo que contiene metano), y en c) la atmosfera, donde es fotoquímicamente oxidado, y en el que el aumento en su concentración reciente, es un resultado del desbalance entre las fuentes y los sumideros de metano. (Universidad de California, Irvine, 2007)

Usualmente el estudio del metano (CH<sub>4</sub>), como gas de invernadero se ha centrado en los flujos que afectan la atmosfera. En la siguiente tabla se presentan las estimaciones, realizadas por el Instituto de Microbiología y Bioquímica de la Universidad de Salamanca (Instituto de Microbiología Bioquímica, Universidad de Salamanca, 2005).

Estimaciones de la liberación de CH <sub>4</sub> a la Atmósfera		
Fuente	Emisión de CH <sub>4</sub> (10 <sup>6</sup> Ton./año)	
<b>Biogénica</b>		
<i>Rumiantes</i>	80-100	
<i>Termes</i>	25-150	
<i>Zonas encharcadas</i>	70-120	
<i>Humedales naturales</i>	120-200	
<i>Vertederos</i>	5-70	
<i>Océanos y Lagos</i>	1-20	
<i>Tundra</i>	1-5	
<b>Abiogénica</b>		
Minería del Carbón	10-35	
Emisiones de Gas natural	10-30	
Escapes industriales y de gaseoductos	15-45	
Combustión de Biomasa	10-40	
Hidratos de metano	2-4	
Volcanes	0,5	
Automóviles	0,5	
<b>Total Biogénico</b>	<b>302-665</b>	<b>81-86% del Total</b>
<b>Total Abiogénico</b>	<b>48-155</b>	<b>14-19% del Total</b>
<b>Total</b>	<b>349-820</b>	

Figura VIII-13 Estimaciones de la liberación de CH<sub>4</sub> a la Atmósfera. (Instituto de Microbiología Bioquímica, Universidad de Salamanca, 2005)

#### d) *Reservorios, Flujos y Tiempos de Residencia*

En la siguiente Figura (VIII-14), los nombres de los reservorios están subrayados, los tamaños y los flujos están dados en Tg<sup>11</sup>(10<sup>12</sup> g) de CH<sub>4</sub> y en Tg CH<sub>4</sub>/año. Los tiempos de residencia corresponden a la división del tamaño del reservorio entre el mayor flujo entrante o saliente del mismo y se encuentran entre paréntesis. (Universidad de California, Irvine, 2007)

<sup>11</sup> Tg: Tera gramos de material. Para convertir Tg de CH<sub>4</sub> a moles de Carbono, se debe multiplicar por 6.25 x 10<sup>10</sup>.

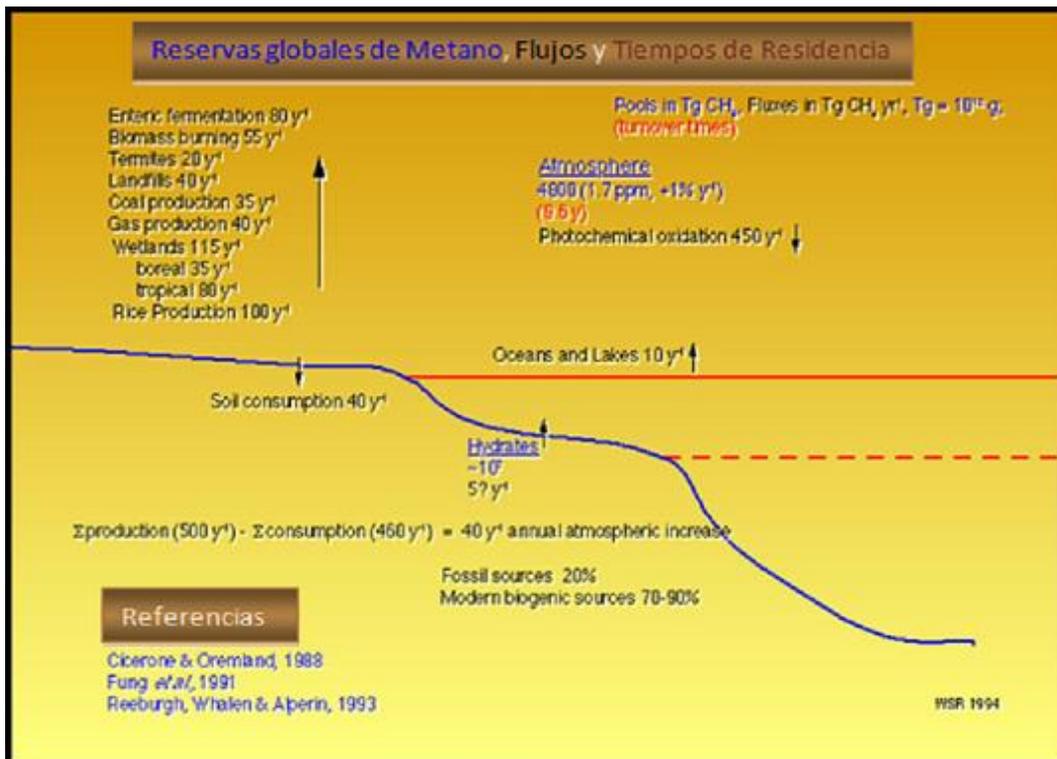


Figura VIII-14 Reservas globales de metano, flujos y tiempos de residencia. Adaptado. (Universidad de California, Irvine, 2007)

## 7. CICLO DEL FÓSFORO

Este ciclo es importante porque el Fósforo, es necesario para limitar la cantidad de nutrientes y porque la escorrentía de fosfato causa procesos de eutrofización (Vilsee, 1990).

### a) Reservorios, Flujos y Tiempos de Residencia

#### Reservorios:

Como se observa en la tabla 8, la mayoría del fósforo en la corteza terrestre es irre recuperable con los medios actuales.

Tabla 11. Reservorios de Fósforo en el planeta (University of Colorado, Geological Sciences, 2007)

Reservorio	Millones de toneladas métricas
<b>Corteza terrestre</b>	
<b>Total</b>	20000 millones
<b>Recobrable</b>	20000
<b>Océano</b>	100000
<b>Vegetación</b>	3000
<b>Agua dulce</b>	100
<b>Suelos</b>	10000

**Flujos:**

En el siguiente modelo de caja (Figura VIII-15), se describen los flujos dentro del ciclo del carbono y los componentes del mismo. (University of Colorado, Geological Sciences, 2007)

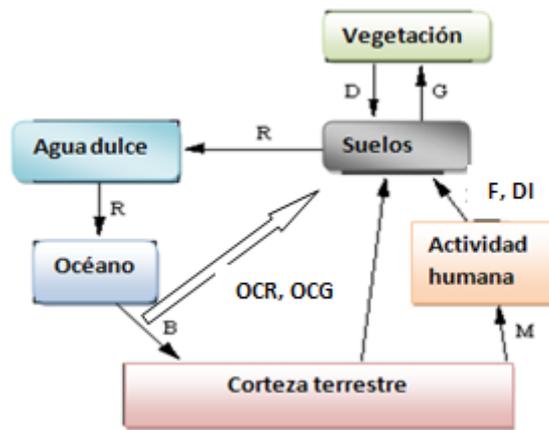


Figura VIII-15 Modelo de caja. Ciclo del fósforo. Adaptado (University of Colorado, Geological Sciences)

Los procesos, nombres y flujos se muestran en la siguiente tabla:

Tabla 12. Procesos y flujos de Fósforo en el planeta (University of Colorado, Geological Sciences, 2007)

Flujos	Sigla usada en la Figura 15	Millones de toneladas métricas/año
Minería (humana)	M	50
Fertilización (humana)	F	50
Meteorización	W	10
Escorrentía	R	20
Por Entierros	B	13
Descomposición	D	200
Crecimiento	G	200
Océano – Continentes (por rocío del mar)	OCR	0.03
Océano – Continentes (por guano)	OCG	0.01
Desechos industriales	DI	2

El fósforo, no posee una fase gaseosa estable, motivo por el cual la adición de este elemento al suelo es lenta; la mayoría del fósforo en la vegetación circula entre plantas vivas y muertas, las adiciones por meteorización son muy bajas respecto al ciclo presentado entre las plantas; de otra parte, los humanos han aumentado la transferencia de fósforo a las plantas y al suelo,

siendo esta razón, cinco veces más rápida que la meteorización. Es importante resaltar que el flujo desde el océano hacia los continentes, ya sea por rocío del mar o por guano, es muy bajo respecto a los otros flujos representados y que su disponibilidad a los organismos marinos está limitada porque la mayoría de los depósitos de fósforo en el océano, se encuentran en aguas profundas. (University of Colorado, Geological Sciences, 2007)

**Tiempos de residencia:**

**Tabla 13. Tiempos de residencia. Ciclo del Fósforo (University of Colorado, Geological Sciences, 2007)**

Reservorio	Años
Depósitos de fosfato en la corteza terrestre	44-175
Océano	5000

**8. CICLO DEL NITRÓGENO**

El nitrógeno existe en una variedad de formas en los sistemas naturales y sus compuestos están involucrados en numerosos procesos abióticos y biológicos. El nitrógeno en su forma de gas ( $N_2$ ), representa más del 80% de la atmosfera, constituyéndose en el mayor reservorio de este elemento. Este gas se convierte en los suelos o en presencia de agua en amonía ( $NH_3$ ), amonium ( $NH_4^+$ ) u otros compuestos nitrogenados. El proceso es conocido como fijación de nitrógeno, y en ausencia de fertilizantes industriales, es la mayor fuente de nitrógeno para los seres vivos; la fijación de nitrógeno en la biosfera, se realiza gracias a bacterias y algas especializadas en la fijación de este elemento. En el suelo se encuentran estas bacterias, en los nódulos de las raíces de las legumbres, donde toman la energía de las plantas para cumplir su objetivo; en ecosistemas de agua dulce y marinos, esta labor la realizan las cianobacterias. (Garcia, Mary Ruth , 2005)

Una vez el nitrógeno ha sido fijado, puede tomar dos diferentes caminos, se puede oxidar por energía en un proceso llamado nitrificación, o puede ser asimilado por un organismo, en su biomasa, en un proceso llamado asimilación de amonio. (California State University Monterey Bay, 2006)

Las plantas incorporan el nitrógeno en sus tejidos, mediante su sistema de raíces, luego es usado para crear aminoácidos y convertirlos en proteínas. El nitrógeno fijado en los cuerpos de seres vivos, regresa al ciclo eventualmente en la forma de gas de nitrógeno. El proceso de denitrificación (proceso químico que transforma nitrato  $NO_3$  en  $N_2$ ) comienza al morir las plantas o al ser ingeridas. Generalmente, el  $N_2$ , es el producto final de este proceso, pero el

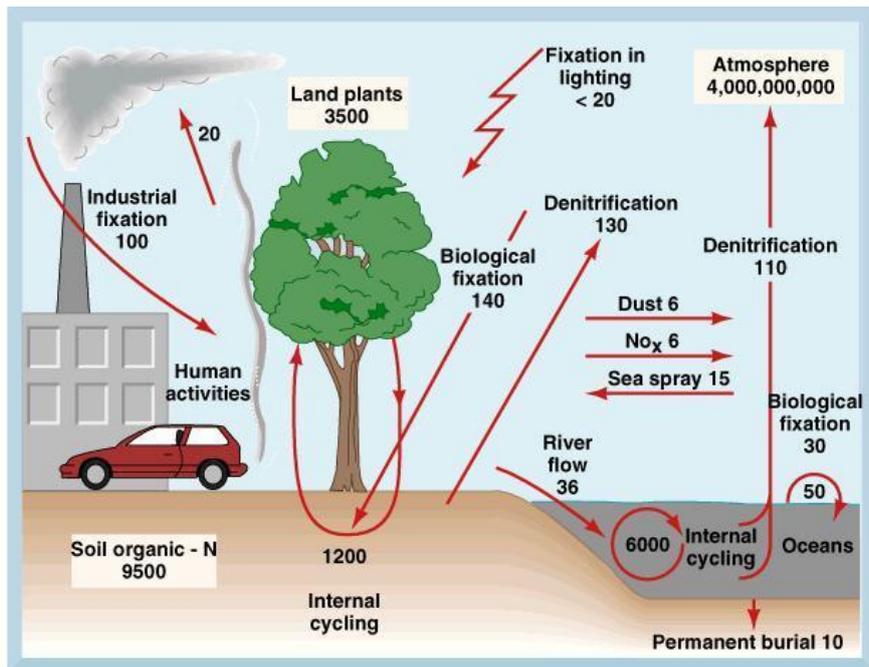
óxido nítrico ( $\text{NO}$ ) también es producido en cantidades menores. (Conti, Marta Elvira. Facultad de Agronomía - Universidad de Buenos Aires, 2005)

La interrupción del ciclo del nitrógeno (Figura V-16), a causa de la actividad humana, produce problemas ambientales tan variados como la producción en la troposfera de smog, la perturbación del océano estratosférico y la contaminación de acuíferos confinados. El óxido nítrico, por ejemplo, es un gas de invernadero y puede destruir el ozono en la estratosfera; eventualmente el óxido nítrico es roto en la por la radiación ultravioleta, produciendo dióxido de nitrógeno ( $\text{NO}_2$ ) y óxido nítrico, compuesto que reduce el ozono. (University of Colorado, Geological Sciences, 2007)

Los óxidos del nitrógeno, se transforman químicamente en  $\text{N}_2$ , nitrato o en compuestos de nitritos, estos compuestos pueden ser usados por la vegetación, una vez sean arrastrados por la lluvia; sin embargo, la lluvia de nitratos es ácida y puede causar problemas ecológicos así como servir de fertilizante a la vegetación. Un átomo de nitrógeno que forma parte de un compuesto formador de smog, puede depositarse en lagos y bosques, como ácido nítrico, capaz de matar peces e insectos; una vez arrastrado a la costa, este átomo de nitrógeno contribuye a la formación de *mareas rojas*; al regresar el átomo a la atmósfera, puede formar parte del óxido nítrico, compuesto capaz de destruir el ozono atmosférico. (García, Mary Ruth, 2005)

En resumen la importancia de este ciclo es debida a que el nitrógeno es un nutriente esencial, así como parte de la lluvia ácida.

**Ciclo del Nitrógeno global:**



The global nitrogen cycle. Pools (□) and annual (→) flux in  $10^{12} \text{gN}_2$ . Note that the industrial fixation of nitrogen is nearly equal to the global biological fixation. (SOURCE: Data from Söderlund, and T. Rosswall, 1982, O. Hutzinger (ed.), *The Handbook of Environmental Chemistry*, Vol 1, Pt. B., Springer-Verlag New York, Inc., New York).

Figura VIII-16 Ciclo global de Nitrógeno. Tomado de (University of Colorado, Geological Sciences, 2007))

a) *Reservorios, Flujos y Tiempos de Residencia*

Reservorios:

Reservorio	Millones de toneladas métricas
Atmosfera	4,000,000,000
Vegetación	3500
Suelos	9500
Océanos	23,000,000
Sedimentos y Rocas	200,000,000,000

Tabla 14 Reservorios planetarios de Nitrógeno (University of Colorado, Geological Sciences, 2007)

Los sedimentos enterrados y las rocas, representan el mayor reservorio de nitrógeno, sin embargo, este reservorio representa tan solo una parte menor del ciclo; a pesar que la atmosfera está conformada en su mayoría por nitrógeno, esta forma del elemento, no puede ser usada por las plantas y necesitan del proceso de fijación de nitrógeno.

Modelo de caja

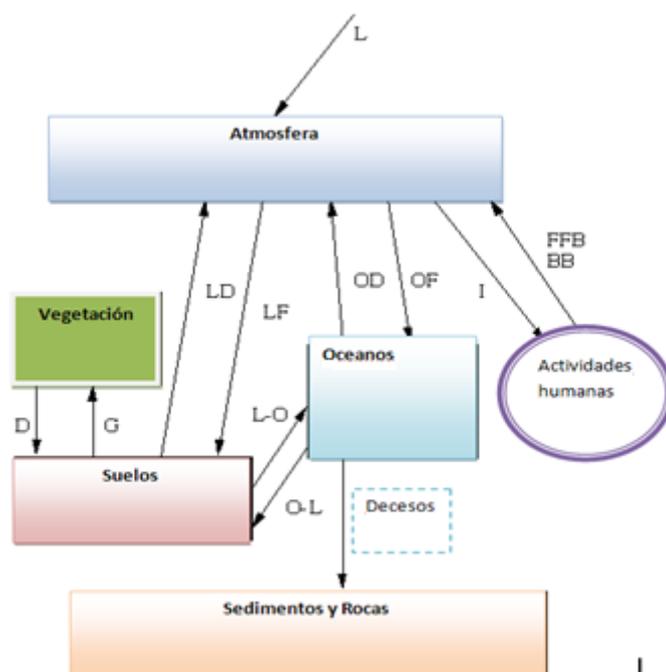


Figura VIII-17 Modelo de Caja. Ciclo del Nitrógeno. Adaptado. (University of Colorado, Geological Sciences, 2007)

**Flujos:**

La fijación industrial es usada para que los fertilizantes provean de nitrógeno a los cultivos; la mayoría de los flujos ocurren entre plantas y suelos ya que las plantas reciclan el nitrógeno, porque este es un nutriente limitado. Solo mediante la generación de luz y las bacterias especializadas, se puede fijar el nitrógeno naturalmente.

Tabla 15. Flujos y Procesos. Ciclo del Nitrógeno (University of Colorado, Geological Sciences, 2007)

Abreviatura, según Figura 17	Proceso	Flujo en millones de toneladas métricas/año
LF	Fijación en suelos	140
LD	Denitrificación de suelos	130
OF	Fijación Oceánica	50
OD	Denitrificación oceánica	110
I	Fijación Industrial	100
FFB	Quema de combustibles fósiles	20
BB	Quema de Biomasa	10

L	Generación de luz	20
Otros		
D	Descomposición	1200
G	Crecimiento	1200
L-O	Suelo-Océano	48
	Ríos	(36)
	Polvo	(6)
	NOx	(6)
O-L(Rocío de mar)	Océano-Suelo	15
	Entierro	10

### Tiempos de Residencia

Tabla 16. Tiempos de Residencia en los reservorios. Ciclo del Nitrógeno (University of Colorado, Geological Sciences, 2007)

Reservorios	Tiempo de residencia (aproximado) en años
<b>Mayores</b>	
Atmosfera	14 millones.
Vegetación	~ 3
Océanos	~ 20,000
Suelos	~ 9
<b>Polucionantes atmosféricos</b>	
NO <sub>x</sub>	~ 4 días
N <sub>2</sub> O	120

Los reservorios en donde el N<sub>2</sub>, es la forma dominante de nitrógeno, como la atmosfera y el océano tienen largos tiempos de residencia; por el contrario, donde el nitrógeno fijado es dominante, se hallan tiempos de residencia cortos. (Garcia, Mary Ruth, 2005)

### Ciclo químico del nitrógeno

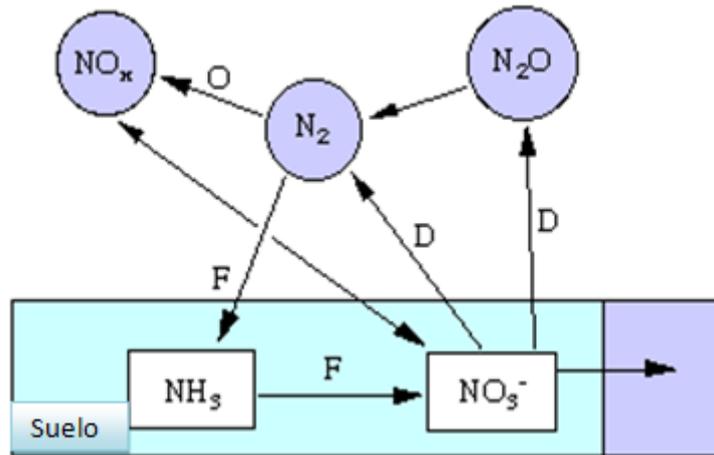


Figura VIII-18 Ciclo químico del nitrógeno. Adaptado ( (University of Colorado, Geological Sciences, 2007))

Tabla 17. Abreviaturas de procesos relacionados al ciclo del nitrógeno (University of Colorado, Geological Sciences, 2007)

Abreviaturas usadas en la Figura VI-18	Proceso
D	Denitrificación
O	Oxidación
F	Fijación

## 9. CICLO DEL AZUFRE

El ciclo del azufre, es importante porque los organismos requieren bajos niveles de este nutriente para cubrir los requerimientos para la síntesis de proteínas, la cual es limitada por lo regular por la disponibilidad de nitrógeno, usualmente la cantidad de azufre varía desde 0.02 a 5% en algunas bacterias que oxidan azufre, pero en general constituye el 0.25% de peso seco; en la figura siguiente, se identifican algunas transformaciones bióticas entre los diferentes estados de oxidación del azufre. (Garcia, Mary Ruth , 2005)

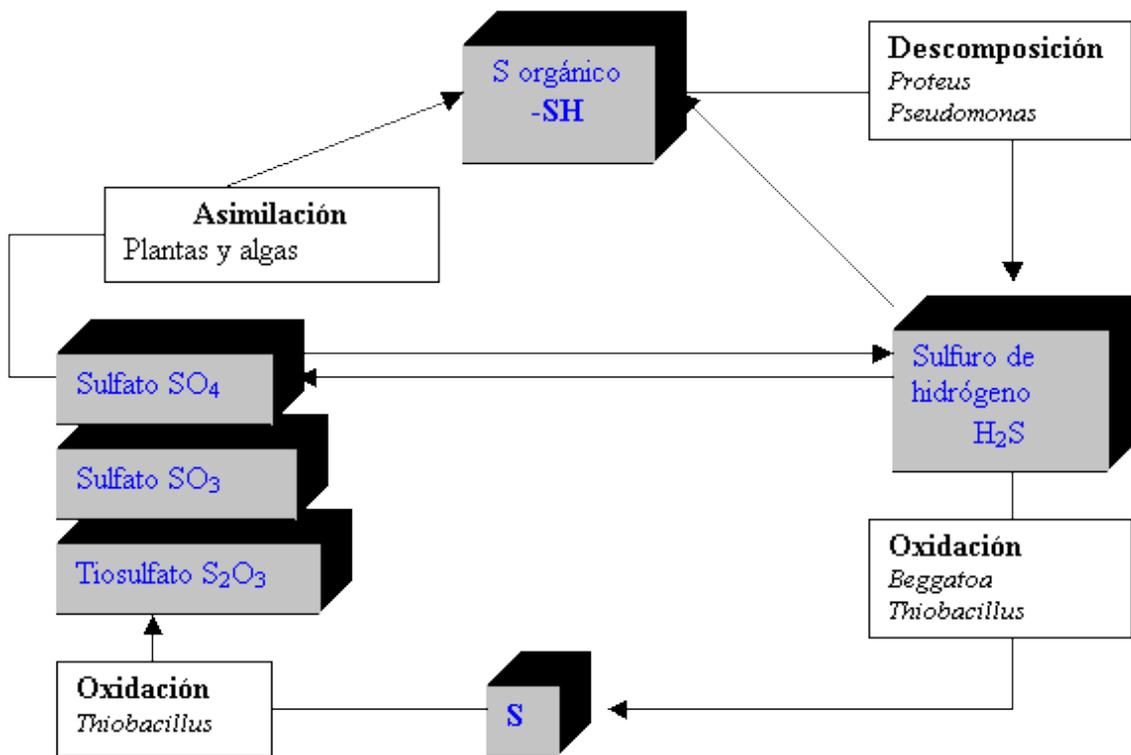


Figura VIII-19 Transformaciones bióticas entre los estados de oxidación del azufre. (García, Mary Ruth , 2005)

Se puede considerar el comienzo del ciclo con la formación de dióxido de azufre (SO<sub>2</sub>) o con compuestos de sulfatos (SO<sub>4</sub>=) en el aire, estos compuestos se precipitan y son lavados de la atmosfera, luego las plantas toman algunos de estos compuestos y los incorporan en sus tejidos; al morir las plantas, estos compuestos son transferidos a los animales o al suelo y cuerpos de agua; posteriormente, las bacterias transforman el sulfuro orgánico en H<sub>2</sub>S gaseoso, mientras que en los océanos, algunos tipos de fitoplancton, producen un químico que se transforma en SO<sub>2</sub>, liberado a la atmosfera, reiniciando así el ciclo. (Garcia, Mary Ruth , 2005)

En su estado reducido de oxidación, el azufre desempeña un papel importante en la estructura y función de las proteínas; en su estado completo de oxidación, el azufre existe como sulfato, siendo este el segundo anión más abundante en agua dulce (después del bicarbonato) y en el agua salada (después del cloruro), y el mayor causante de la acides en la lluvia; esta acides, es importante en diversos procesos biogeoquímicos como meteorización de rocas, lluvia acida y tasas de denitrificación. (Garcia, Mary Ruth , 2005)

El del azufre es uno de principales ciclos biogeoquímicos perturbados por el hombre; se estima que las emisiones a la atmosfera por la actividad humana son iguales a las causadas por procesos naturales, de otra parte, el sulfato en la atmosfera influencia el ciclo hidrológico y

constituye el componente dominante del núcleo de condensación igualmente en zonas no contaminadas. (California State University Monterey Bay, 2006)

En este ciclo aunque el tiempo de vida de la mayoría de los compuestos de azufre en el aire, es del orden de días, los tiempos de residencia en la corteza terrestre son muy lentos, y los fenómenos de meteorización, erosión y sedimentación de rocas con azufre se pueden demorar miles de años. (University of Colorado, Geological Sciences, 2007)

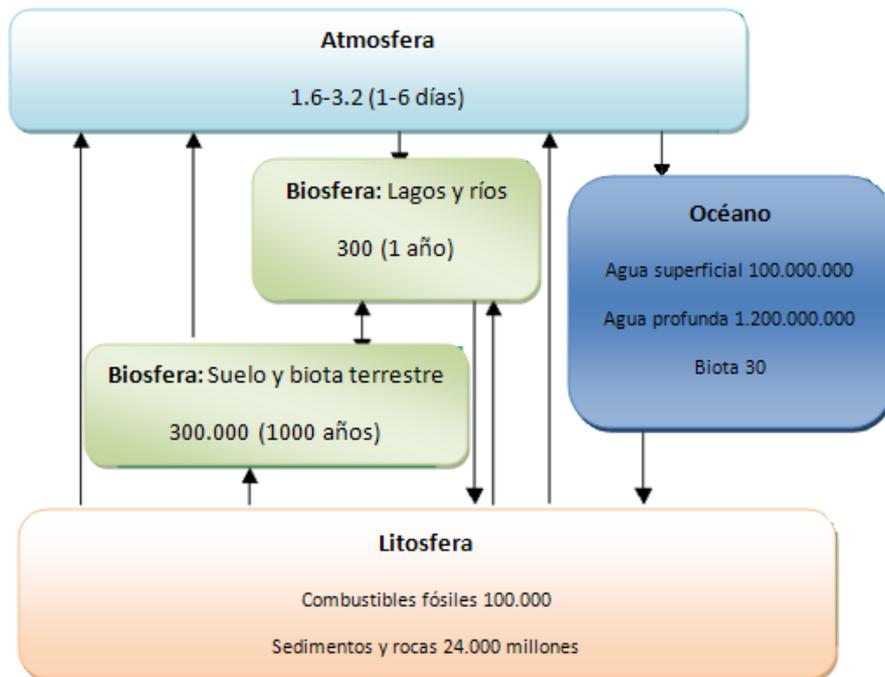


Figura VIII-20 Modelo de Caja. Flujos de azufre entre los diferentes reservorios Tamaño de los depósitos en 10<sup>9</sup> kg y tiempo aproximado de recambio del azufre. Adaptado de (García, Mary Ruth, 2005)

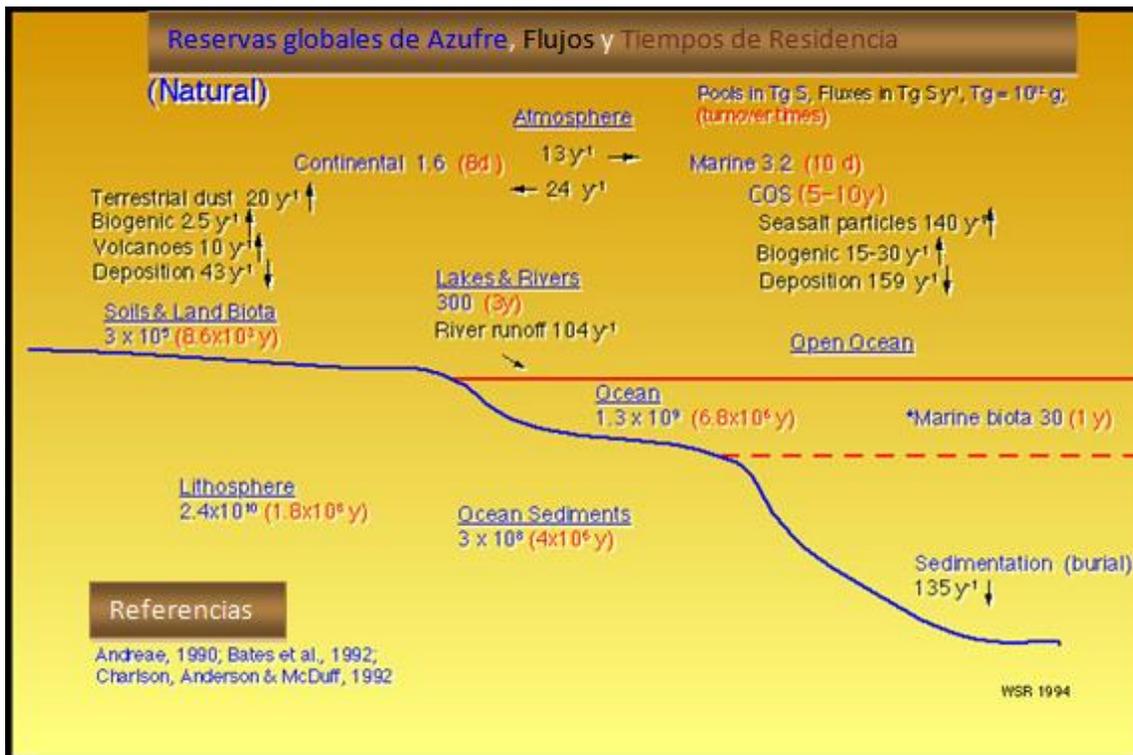


Figura VIII-21 Reservas globales de Azufre, Flujos y Tiempos de Residencia. Adaptado de (Universidad de California, Irvine, 2007)

## 10. CICLO DEL OXIGENO

El oxígeno es uno de los principales constituyentes de la materia viva y gracias a su capacidad de combinación química, que le hace presentarse bajo múltiples formas, es más complicado de analizar que el ciclo del carbono al presentar varios sub ciclos, como los encontrados entre la litosfera y la atmósfera y entre la hidrosfera y las dos fases anteriores. (University Corporation for Atmospheric Research (UCAR), 2007)

El oxígeno molecular puede ser formado por disociación de las moléculas de agua en las capas altas de la atmósfera, bajo actuar las radiaciones de alta energía; sin embargo, el oxígeno presente, tanto en la atmósfera como en las rocas superficiales, es de origen biológico, es decir producido por organismos autótrofos. La formación de la ozonofera, impidió la penetración excesiva de las radiaciones ultravioletas, y favoreció el desarrollo de organismos fotosintéticos que produjeron mayor cantidad de oxígeno. (Universidad de California, Irvine, 2007)

La relación entre los gases atmosféricos y los sistemas terrestre y marino se da principalmente, a través de la fotosíntesis, donde el dióxido de carbono atmosférico se transforma en oxígeno para los seres vivos. Siendo esta la principal fuente de formación de oxígeno (400 mil millones de toneladas año). (University of Colorado, Geological Sciences, 2007)

Por otro lado, el oxígeno atmosférico facilita la oxidación de sustancias inorgánicas, colabora en la meteorización de sedimentos orgánicos fósiles, y en otros procesos de oxidación muy importantes que producen CO<sub>2</sub>, sulfatos y nitritos, entre otros. (Universidad de California, Irvine, 2007)

**Reservorios, Flujos y Tiempos de Residencia**

En la siguiente Figura (VIII - 22), los nombres de los reservorios están subrayados, los tamaños y los flujos están dados en Tg (10<sup>12</sup> g) de CH<sub>4</sub> y en Tg CH<sub>4</sub>/año. Los tiempos de residencia corresponden a la división del tamaño del reservorio entre el mayor flujo entrante o saliente del mismo y se encuentran entre paréntesis. (University of Colorado, Geological Sciences, 2007)

Para convertir moles O<sub>2</sub> a Tg O<sub>2</sub>, se debe multiplicar por 3.2 x 10<sup>11</sup>

Reservorio	Masa (10 <sup>15</sup> moles)	Tiempo de residencia
Rocas sedimentarias	10 <sup>6</sup>	
Atmosfera	37000	3x10 <sup>6</sup> años
Biota	180-10	1000 – 50 años
Océano	219	500 años
Océano superficial	6	22 días

Tabla 18. Reservorios y tiempos de residencia. Ciclo del Oxígeno (University of Colorado, Geological Sciences, 2007)

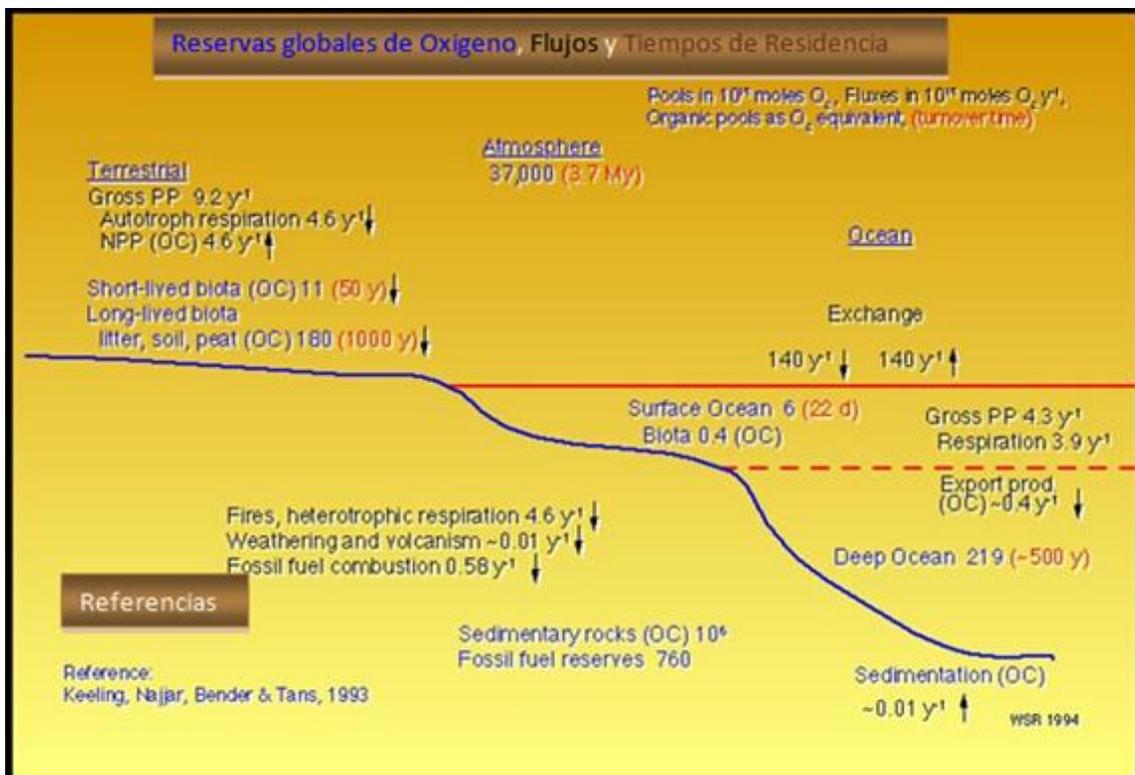


Figura VIII-22 Reservas globales de oxígeno, flujos y tiempos de residencia. Adaptado de (Universidad de California, Irvine, 2007)

## 11. CICLO DEL POTASIO

Aunque el potasio es un elemento nutritivo esencial para todos los organismos vivos, entre el 90 % y el 98 % del potasio encontrado en los suelos, se encuentra en minerales primarios insolubles como los feldspatos y las micas, minerales que contienen silicatos de potasio-aluminio, resistentes a las rupturas químicas. En los suelos, el contenido de este mineral está estrechamente relacionado con el tipo de material parental y la pedogénesis<sup>12</sup> (Mengel y Rahmatullah, 1994).

Los vegetales necesitan cantidades elevadas de este nutriente, que obtienen de la meteorización de minerales, la mineralización de los residuos orgánicos o de abonos y fertilizantes, siendo semejante al requerimiento de nitrógeno (Mengel y Kirkby, 1987). El potasio cumple un rol importante en la activación de enzimas (conociéndose más de 60 activadas por este catión), que actúan en diversos procesos metabólicos tales como fotosíntesis, síntesis de proteínas y carbohidratos; también tiene incidencia en el balance de agua y en el crecimiento meristemático de las plantas (Mengel y Kirby, 1987). Al participar de estos procesos metabólicos el potasio actúa favoreciendo el crecimiento vegetativo, la fructificación, la maduración y la calidad de los frutos.

El potasio en el suelo se encuentra principalmente en dos formas: disponibilidad lenta y disponibilidad inmediata; el primer tipo, comprende del 1 al 10 % del total de potasio disponible y se puede originar de la disolución de minerales primarios o de fertilizantes; en este caso, el potasio es atraído a la superficie de materiales arcillosos, donde es firmemente fijado entre las capas de arcilla, siendo lenta su disponibilidad para las plantas. De otro lado, el potasio de disponibilidad inmediata, comprende solo del 0.1 al 2 % del total disponible en el suelo y se encuentra en regiones de intercambio entre la capa de arcilla y capas de suelo orgánico. Este potasio se denomina intercambiable, porque puede remplazarse por otros cationes como hidrogeno, calcio y magnesio, siendo posible que la planta lo absorba o que drene entre el suelo (Conti, Marta Elvira. Facultad de Agronomía - Universidad de Buenos Aires, 2005).

Las arcillas son cristalinas y están formadas por capas en arreglos laminares de varios elementos químicos, fundamentalmente el oxígeno, el silicio y el aluminio en forma de capas de tetraedros de silicio y de octaedros de aluminio. Los iones potasio ( $K^+$ ) se presentan en

---

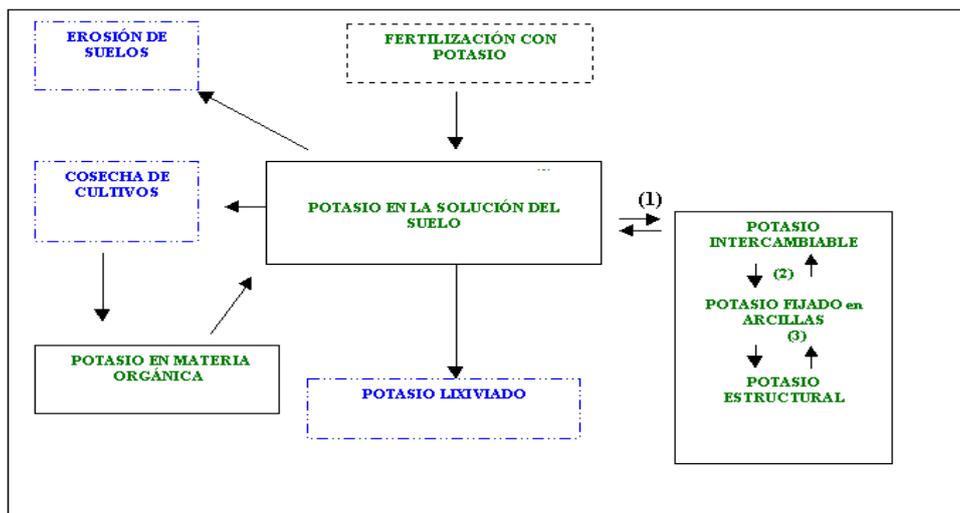
<sup>12</sup> La pedogénesis o evolución de suelo (formación) es el proceso por el cual se crea suelo.

distintas posiciones en las arcillas y pueden ser mantenidos electrostáticamente por las cargas eléctricas negativas situadas en la superficie o en los bordes de las láminas. La fuerza con que los iones K son mantenidos, varía con el tipo de arcilla y la posición del ion en la misma. (Conti, Marta Elvira. Facultad de Agronomía - Universidad de Buenos Aires, 2005)

Investigaciones confirmaron que el sistema agrícola no tiene otra vía de ingreso natural para el balance de potasio que la reposición primaria proveniente de la liberación de los minerales primarios y secundarios, siendo los minerales arcillosos la fuente principal de potasio en el suelo (Sardi y Debreczeni, 1992; Buhman, 1993).

**Modelo de cajas:**

La siguiente figura, muestra las formas más importantes de potasio en el suelo, y los procesos que las relacionan.



**Referencias**



- (1) = Adsorción - Desorción
- (2) = Fijación - Liberación
- (3) = Cristalización - Meteorización

Adaptado de Goulding (1987)

Figura VIII-23 Formas más importantes de potasio en el suelo y procesos relacionados. (Conti, Marta Elvira. Facultad de Agronomía - Universidad de Buenos Aires)

## C. Introducción al Algoritmo de Redes Neuronales de Kohonen (RNK)

Esta arquitectura de red, nombrada así por su creador Tuevo Kohonen, varía considerablemente del modelo más usado, que es el de *feed forward back propagation (FFBP)*; la red neuronal de Kohonen, no solo difiere en cómo es entrenada, sino en cómo recuerda los patrones; a su vez, estas redes neuronales, no usan funciones de activación, capas ocultas, ni predisposición de pesos. (Kohonen, 1984)

La mayor diferencia entre las RNK y la FFBP, es que la red de Kohonen, se entrena de manera no supervisada; esto significa que a la RNK se le presentan datos, pero la salida correcta para los datos, no es especificada; al usar la RNK, estos datos pueden clasificarse en grupos.

En términos generales, al presentarse un patrón de entrada (input) a la red neuronal de Kohonen (RNK), solo una, de las neuronas de salida (output) es seleccionada como ganadora. Esto se denomina un aprendizaje no supervisado competitivo, ya que no existe ninguna salida objetivo hacia la cual la red neuronal deba tender, a su vez, al competir las neuronas por activarse, queda solo una como neurona vencedora y el resto son forzadas a sus valores de respuesta mínimos. El objetivo de este aprendizaje es categorizar los datos que se introducen en la red. Se clasifican valores similares en la misma categoría y, por tanto, deben activar la misma neurona de salida (Heaton, 2005).

Es importante comprender las limitaciones de las RNK, donde al igual que las redes neuronales con dos capas, presentan mejores resultados al usarse en problemas que pueden ser descompuestos linealmente (Heaton, 2005). De otro lado, las redes neuronales de Kohonen, proporcionan ventajas como facilidad de construcción y agilidad en el proceso de entrenamiento.

### 1. Estructura de la RNK

A diferencia de las redes FFBP, las redes neuronales de Kohonen, solo contienen dos capas de neuronas, una de entrada y otra de salida, no tienen capas ocultas. A continuación, se mostrarán los conceptos de entrada y salida de datos para las redes neuronales de Kohonen. (Kohonen, 1984)

### 2. Entrada de datos

La entrada de datos a la RNK, está dada por las neuronas de entrada; estas neuronas tienen asignados números de punto flotante, que corresponden al patrón de entrada de la red. La red neuronal de Kohonen, requiere que estas entradas estén normalizadas, usualmente entre 0 y 1

(0,1) o en el rango entre -1 y 1 (-1,1). Este patrón de entradas a la red, ocasionará que las neuronas de salida reaccionen. (Heaton, 2005)

### 3. Salida de datos

La salida de una red neuronal de Kohonen, varía de la salida de una red tipo FFBP, en que solo una de las neuronas de salida producirá un valor, adicionalmente, este valor puede ser Verdadero o Falso; al presentarse un patrón a una RNK, solo una neurona es escogida como neurona de salida, motivo por el cual, usualmente la salida de la RNK es el índice de la neurona que se activó (ej. La neurona ganadora es la número 3). (Heaton, 2005)

La estructura de una RNK típica se enseña a continuación, en la Figura VII-1.

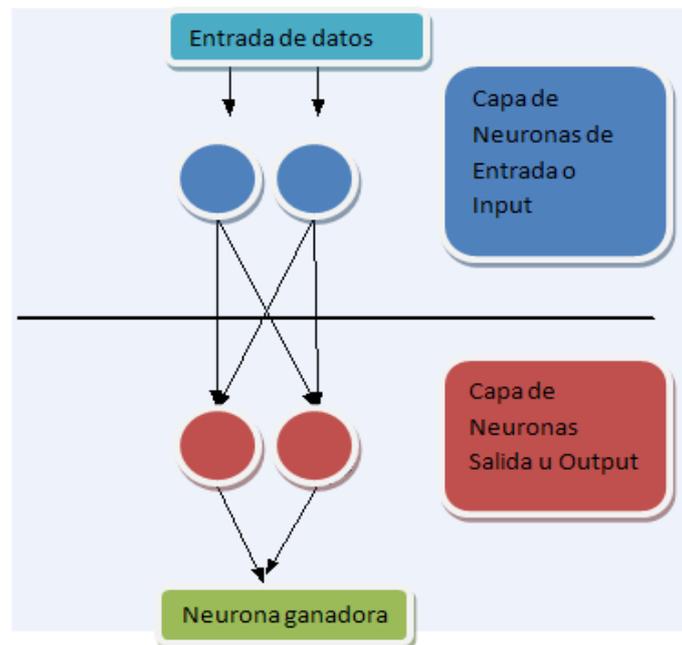


Figura VIII-24 Estructura típica de una Red Neuronal de Kohonen

### 4. Procesamiento de la información

Para examinar como es el procesamiento de información, se debe comprender cuál es el proceso de cálculo que lleva la red neuronal. Estos pasos incluyen normalización del vector de entradas, cálculo de la neurona de salida, mapeo de números en formato bipolar (opcional), selección de la neurona ganadora, entrenamiento de la red, determinación de la razón de aprendizaje y ajuste de pesos.

### 5. Normalización la Entrada

Las redes neuronales de Kohonen, requieren que sus entradas sean normalizadas, siendo este requerimiento uno de los mayores limitantes de las RNK. El rango de las variables de entrada,

debe ser [-1,1], y cada una de las variables de entrada debe poder usar este rango libremente; se ha encontrado que si una o varias neuronas de entrada usan solo el rango entre [0,1], se perjudica el desempeño de la red neuronal. (Heaton, 2005)

Usando la metodología propuesta por Heaton (2005), para normalizar la entrada, primero se debe calcular la magnitud del vector de los datos de entrada; esto se realiza sumando los cuadrados del vector de entrada (ej.  $0.5^2+0.75^2 = 0.8125$ ); con el valor de esta magnitud, es posible determinar el factor de normalización; siendo el factor de normalización igual al recíproco de la raíz cuadrada de la magnitud. (ej.  $1/(0.8125)^{0.5} = 1.1094$ ) El proceso de normalización, será usado al calcular la capa de salida.

## 6. Cálculo de la salida de cada neurona

Usando como ejemplo la Figura anterior, el algoritmo de Kohonen establece que se debe tener en consideración, el vector de entradas y los pesos de conexión entre la neurona 1, de la capa de entradas y los pesos entre esta neurona y cada una de las neuronas de la capa de salida. Una medida usual para relacionar las entradas y los pesos, es calcular la distancia entre los mismos; generalmente se utiliza el concepto de Distancia Euclidiana para este fin. (Buckland, 2005)

Sean  $P = (x_1, y_1, z_1)$  y  $Q = (x_2, y_2, z_2)$ , dos puntos en el espacio. La distancia PQ entre P y Q está dada por:

$$\overline{PQ} = \sqrt{(X_1 - X_2)^2 + (Y_1 - Y_2)^2 + (Z_1 - Z_2)^2}$$

Donde el cálculo de este resultado, debe repetirse para cada neurona de salida y posteriormente normalizarse, una vez calculada y normalizada la salida, Heaton (2005) propone que se mapea en un número bipolar<sup>13</sup>

## 7. Mapeo de números en formato Bipolar

Un número bipolar, es una forma de representación alterna a los números binarios; en el sistema bipolar, el valor CERO binario, se representa como un MENOS UNO, y el UNO se representa como UNO. A causa que las entradas de la red neuronal están normalizadas en este rango, se debe realizar una normalización similar a la salida de las neuronas. Para realizar este mapeo, se debe adicionar UNO y dividir el valor por la mitad, siendo este resultado el valor final de la salida para una neurona. (Heaton, 2005)

---

<sup>13</sup> Representación numérica similar al sistema binario, pero reemplaza los ceros con -1

## 8. Selección de la neurona ganadora

La neurona ganadora es aquella cuya distancia euclidiana entre las entradas y los pesos es la menor. Como se pudo observar, los pesos entre las neuronas de entrada y las de salida, determinan el valor de la salida; a continuación se procederá a explicar el proceso en el que esos pesos son ajustados para producir salidas más adecuadas para una tarea dada; este proceso es el entrenamiento de la red. (Hecht-Nielsen, 1990)

## 9. Entrenamiento de la red

En términos generales, el entrenamiento de las redes neuronales de Kohonen, es competitivo, ya que para cada juego de parámetros de entrenamiento, solo una neurona resulta ganadora; a esta neurona le será reajustado su peso, de manera que reaccione más fuertemente a las entradas la próxima vez; como diferentes neuronas ganan para diferentes patrones, su habilidad para reconocer un patrón específico será aumentada. (Kohonen, 1995)

El proceso de entrenamiento de una RNK, involucra pasar por diferentes épocas (iteraciones), hasta que el error (o distancia entre las entradas y los pesos) de la RNK sea menor a un valor aceptable dado. Este proceso, incluye calcular el error de la red, ajustar los pesos en cada época y establecer cuando no es necesario continuar el entrenamiento.

La RNK es entrenada mediante la repetición de épocas hasta que uno de dos eventos pasa, si el error calculado es aceptable se termina el proceso de entrenamiento, si por el contrario, el error permanece alto respecto al valor aceptable, existen dos opciones, que la razón de cambio de este error sea muy pequeña, ocasionando que este ciclo individual sea abortado y no se realicen épocas adicionales; de ocurrir esto, los pesos se inicializan con valores aleatorios y se iniciara un nuevo ciclo. Este entrenamiento continuara el ciclo previo y realizará de nuevo la verificación de si los pesos producen errores aceptables. (Heaton, 2005)

Como se puede inferir, el proceso iterativo para reducir los errores es clave en el entrenamiento de la red.

## 10. Razón de aprendizaje

La razón de aprendizaje es un valor numérico usado por el algoritmo de aprendizaje, puede ser constante o variable en el proceso, teniendo en cuenta que siempre debe ser un número positivo menor que 1; usualmente la razón de aprendizaje es un número entre 0.4 y 0.5 y se representa con la letra griega alpha ( $\alpha$ ). Generalmente valores más altos de *alpha*, causan que el proceso de aprendizaje sea más rápido, sin embargo pueden causar que la red nunca

converja; esto a causa que las oscilaciones en los vectores de pesos, pueden ser tan grandes que impidan a los patrones de clasificación manifestarse. (Heaton, 2005)

Una variante a la técnica anterior, es iniciar *alpha* con valores altos e ir decreciendo su valor conforme avanza el entrenamiento, permitiendo un entrenamiento inicial más rápido de la RNK e ir depurando el proceso a medida que avanza el mismo. Independientemente de si la razón de aprendizaje se toma como variable o constante, esta razón se usa como parte integral del algoritmo que calcula los pesos de las neuronas. (Heaton, 2005)

## 11. Ajuste de los pesos

La memoria de las RNK, es almacenada dentro de las conexiones ponderadas entre las capas de entrada y de salida; estos pesos son ajustados en cada época o iteración, que busca que la red neuronal presente una respuesta más favorable la próxima vez que el mismo juego de datos de entrenamiento se le presente; estas iteraciones continúan al ingresar nuevos datos a la red y ser los pesos reajustados. Eventualmente, el reajuste de los pesos disminuirá hasta que no sea importante continuar con este juego de pesos, cuando esto ocurre la matriz de pesos se reinicia con valores aleatorios y se crea un nuevo ciclo. (Heaton, 2005)

La matriz de pesos definitiva que será usada, corresponderá a la mejor matriz de pesos determinada en cada uno de los ciclos.

El método original para calcular los cambios en los pesos, o método aditivo, fue propuesto por T. Kohonen (1984) y usa la siguiente ecuación:

$$w^{t+1} = \frac{w^t + \alpha x}{\|w^t + \alpha x\|}$$

Donde la variable  $x$  es el vector de entrenamiento que fue presentado a la red, la variable  $w^t$  es el peso de la neurona ganadora, y la variable  $w^{t+1}$  es el nuevo peso. Las líneas verticales dobles representan la magnitud del vector. (Kohonen, 1984)

Aunque el método aditivo usualmente trabaja bien con las RNK, existen ocasiones en las que el método es excesivamente inestable y no converge, siendo necesario utilizar un método alternativo, como el método sustractivo. (Heaton, 2005)

El método sustractivo usa las siguientes ecuaciones, para transformar los pesos de la red:

$$e = x - w^t$$

$$w^{t+1} = w^t + \alpha e$$

### a) *Calculo del Error*

El propósito de las RNK, es clasificar datos de entrada en varios juegos, luego el error para las redes neuronales de Kohonen, debe ser capaz de medir que tan bien se están clasificando los datos de entrada. Esta característica permite que diversos cálculos de errores hayan sido propuestos para las RNK, sin ser ninguno oficial.

Es importante resaltar que al ser el entrenamiento no supervisado, se debe replantear la formula de calcular el error, siendo aceptable la cotejar el valor esperado del entrenamiento y el valor actual; esto debido a que no existen salidas anticipadas que permitan la comparación entre lo observado y lo simulado.

En términos generales, el error es mínimo cuando la distancia euclidiana entre las entradas y los pesos es cero.

A continuación, se observara un ejemplo de cálculo y aplicación de las RNK, los mapas auto-organizados de Kohonen o SOMs.

## D. **Mapas Auto – Organizados de Kohonen**

Ahora que ya se conocen las características de las redes neuronales de Kohonen (RNK), se procederá a explicar una de sus más conocidas aplicaciones: La creación de mapas auto-organizados, o SOMs por sus siglas en inglés.

Así como las RNK, los SOMs fueron inventados por Teuvo Kohonen, profesor emérito de la Universidad Tecnológica de Helsinki y miembro de la Academia de Finlandia; estos mapas auto-organizados, proveen una forma de representar datos multidimensionales en espacios dimensionales menores, usualmente en una o dos dimensiones; esta proceso es básicamente una técnica de compresión de datos, conocida como **Cuantización Vectorial**; adicionalmente, esta técnica crea una red que almacena información, de manera tal que las relaciones topológicas entre los elementos del conjunto se mantienen. (Kohonen, 1995)

Un ejemplo común de las funcionalidades de los SOMs, es la realización de mapas de colores, basados en sus tres componentes dimensionales (rojo, azul y verde), para obtener un mapa en dos dimensiones; el siguiente ejemplo fue realizado con el programa Self Organizing Map Demo; en este se entrena un SOM para reconocer los colores mostrados en la Figura VIII - 25:

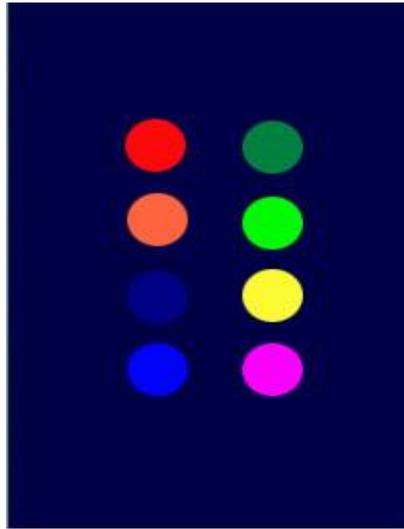


Figura VIII-25 Patrones de Color usados en el ejemplo de Mapas Auto Organizados (Germano, 1999)

Estos colores, se han representado en la RNK, como vectores tridimensionales, y la red ha aprendido a representarlos en un espacio de dos dimensiones; cada uno de los nodos tiene tres pesos, uno por cada elemento del vector de entrada (RGB); para efectos de visualización, cada nodo es representado por una celda rectangular. La Figura VIII – 26, muestra un estado intermedio de organización en el mapa del ejemplo.



Figura VIII-26 Estado intermedio de organización. SOMs. Usando Self Organizing Map Demo. Iteración 41

A continuación, en la Figura VIII-27, se observa que además de realizar clústeres de colores en diferentes regiones del mapa, regiones con propiedades similares se encuentran adyacentes, es así como el clúster amarillo esta cerca al clúster naranja y este al rojo; los clústeres azules están cerca del clúster verde y del violeta. Como se expuso anteriormente, uno de los aspectos más interesantes de las RNK, es que aprenden a clasificar datos sin ser supervisadas.



Figura VIII-27 Clasificación final de colores del SOM, usando SOM demo, luego de mil iteraciones

### 1. Topologías neuronales

Tanto en la literatura sobre el tema, como en los programas de computador analizados, se mencionan tres topologías neuronales, para la capa de salida; estas topologías son:

1. Topología en forma de grilla rectangular.
2. Topología en forma hexagonal
3. Topología de nodos aleatorios.

A continuación se mostraran ejemplos de estas topologías, creados con el Neural Network Toolbox de Matlab 7.1, para Windows (The MathWorks Inc., 2008).

- **Topología en forma de grilla rectangular.**

Esta topología, en términos generales para efectos de programación, inicia un conjunto de neuronas en una grilla rectangular de dimensiones  $N \times M$ , siendo  $N$  el numero de filas y  $M$  el numero de columnas en el arreglo matricial que corresponde a la capa de salida del SOM. Sin embargo al utilizar el Toolbox de Matlab, la notación cambia respecto a la usual y el primer término corresponde al número de columnas deseadas para la grilla, y el segundo a las filas.

El comando **gridtop(m,n)** es el encargado de crear la grilla de neuronas, y el comando **plotsom(Var)** dibuja la grilla de neuronas creada , sin importar la topología usada.

A continuación, en las figuras VIII-28 y VIII-29, se ilustraran dos grillas rectangulares, creadas con el comando **gridtop**. La primera de ellas, muestra una grilla con cinco (5) filas de cuatro (4) elementos (Var = gridtop (4, 5)); la segunda grilla, creada con el comando (Var = gridtop(5,4)), crea una malla de neuronas, con cuatro (4) filas de cinco (5) nodos. Se aprecia adicionalmente, en la primera de estas imágenes, el Command Window de Matlab, y las líneas de comandos que deben escribirse para obtener dicha salida; a su vez, se muestra en pantalla la posición de los diferentes nodos de la grilla Var, en términos de sus coordenadas (X, Y). Matlab utiliza en este caso, una salida de veinte (20) columnas de dos elementos.

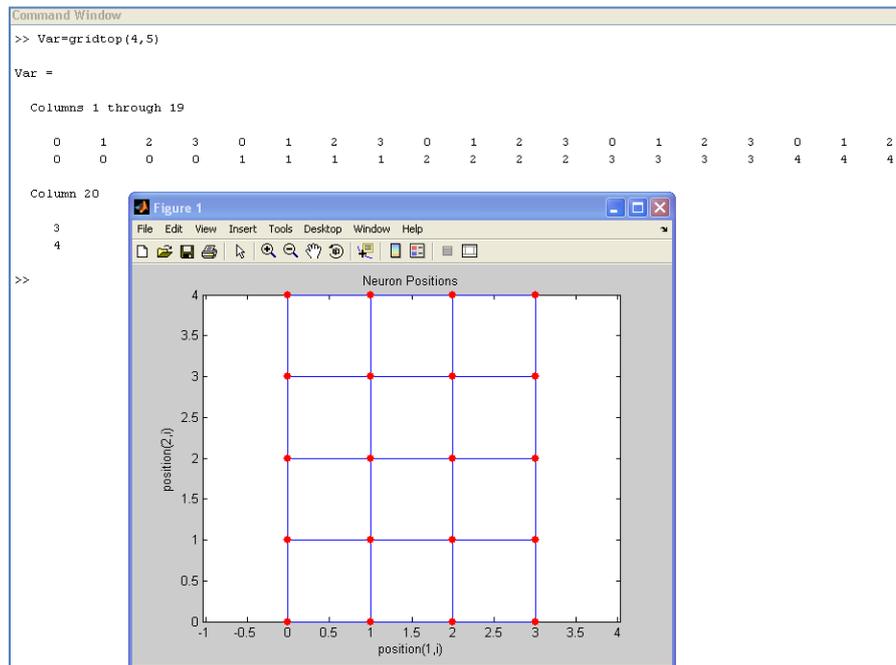


Figura VIII-28 Topología rectangular. Ejemplo usando la función gridtop de Matlab 7.1 Var= gridtop (4,5 )

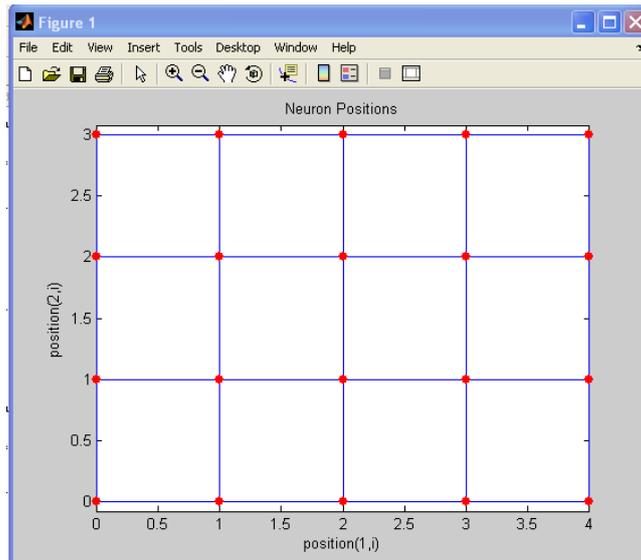


Figura VIII-29 rectangular. Usando la función GRIDTOP de Matlab 7.1. (Var= gridtop (5,4) )

- **Topología en forma hexagonal**

La función **hextop**, del Toolbox de Matlab 7.1, es la encargada de crear una grilla con topología hexagonal. Usando el mismo procedimiento que para el comando **gridtop**, luego se procede a ver la grilla con el comando **plotsom (Var)**.

A continuación, en la Figura VIII-30, se mostrará un grafico que enseña una grilla hexagonal, (Var= hextop(4,5) ), con cuatro (4) nodos en dirección X, por cada una de las cinco (5) filas ubicadas en Y.

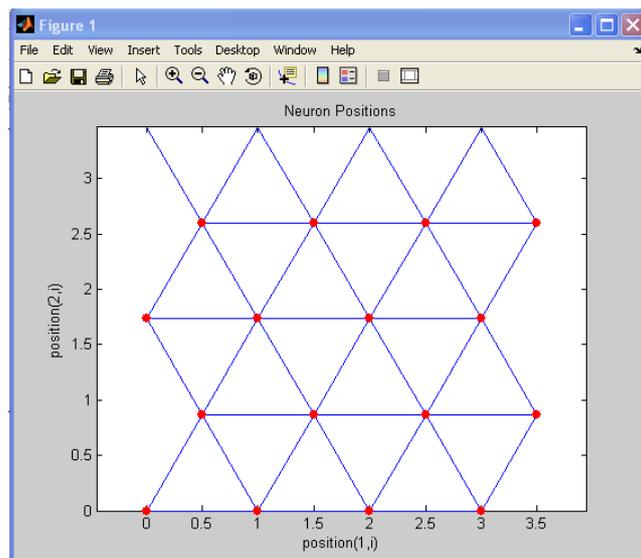


Figura VIII-30 Topología Hexagonal. Función HEXTOP usando Matlab 7.1 (var=hextop(4,5))

- **Topología de nodos aleatorios.**

La topología de nodos aleatorios se crea usando la función **randtop**, y como su nombre lo indica, distribuye aleatoriamente los nodos en la grilla; esto lo realiza asignándole coordenadas aleatorias a los pares de puntos (x,y) del arreglo. Su sintaxis es: `Var=randtop (m,n)`.

La Figura VIII-31, muestra una grilla de veinte nodos generada con el comando **randtop**.

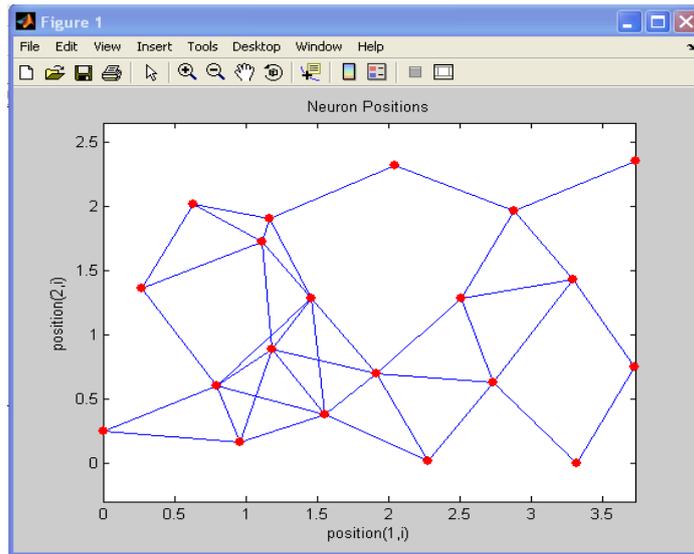


Figura VIII-31 Topología Aleatoria. Función RANDTOP usando Matlab 7.1 (`var=randtop(4,5)`)

## 2. Arquitectura de la red

En el ejemplo analizado, se contempla una arquitectura de red en dos (2) dimensiones; donde cada nodo de la red (neurona de salida) está conectado a la capa de entrada. En la siguiente Figura (VIII-32), se presenta una RNK de dimensiones 4 x 4, conectada a una capa de entrada que representa un vector tridimensional.

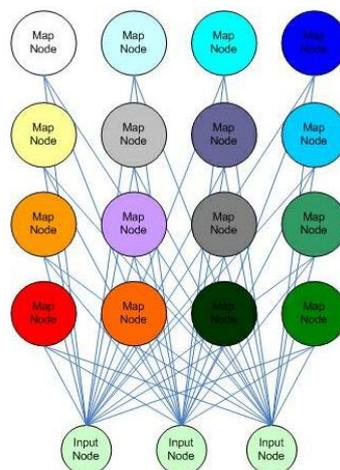


Figura VIII-32 Arquitectura típica de una RNK. (<http://www.generation5.org/content/2004/aiSomPic.asp>)

Cada nodo tiene una posición topológica y contiene un vector de pesos ( $W_n$ ) de la misma dimensión que el vector de entradas ( $V_n$ ). En el ejemplo tratado en el presente capítulo, cada color en el mapa está asociado con un conjunto de pesos (RGB), similar a la representación en la Figura VIII-33.

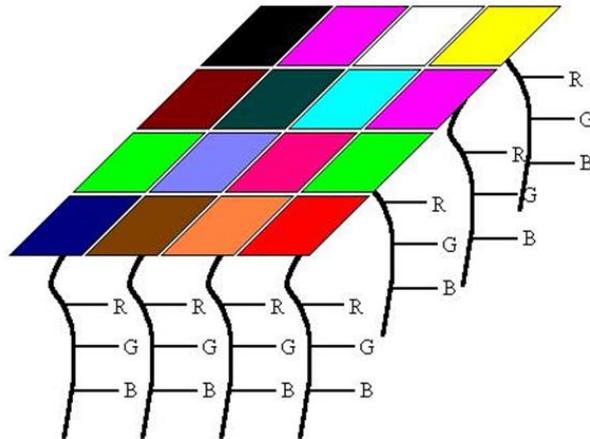


Figura VIII-33 Asociación de colores con vectores RGB. (<http://davis.wpi.edu/~matt/courses/soms/>)

Para el ejemplo dado, se tiene una malla de 40 x 40 (Figura VIII-34), donde cada nodo de la grilla posee tres (3) pesos, uno por cada elemento del vector de entradas (RGB).

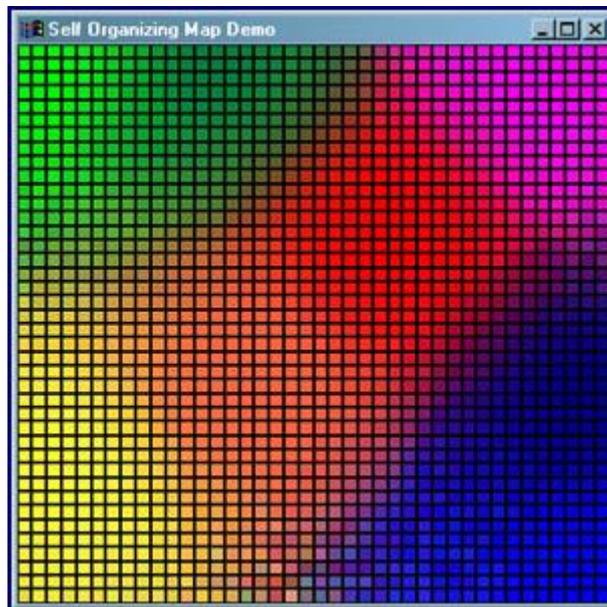


Figura VIII-34 . Grilla de 40 x 40. Ejemplo usando SOM Demo (Buckland, 2005)

### 3. Proceso de entrenamiento

El entrenamiento ocurre en varios pasos y numerosas iteraciones, de la siguiente manera (Buckland, 2005):

1. Los pesos de cada nodo son inicializados.
2. Un vector se escoge del juego de datos de entrenamiento y es presentado a la malla.
3. Cada nodo se examina para calcular que juego de pesos es más similar al vector de entrada. El nodo ganador se determina como la Mejor Unidad de Coincidencia.
4. El radio de la Mejor Unidad de Coincidencia es calculado; este valor inicia con un valor elevado, generalmente el radio de la malla, pero disminuye en cada paso de tiempo. Todos los nodos que se encuentren dentro de este radio, se consideran que pertenecen a la vecindad de la Mejor Unidad de Coincidencia.
5. Para cada uno de los nodos encontrados en el vecindario de la Mejor Unidad de Coincidencia, se ajustan los pesos con el fin de hacerlos “mas similares” al vector de entrada. Entre más cerca se encuentre un nodo a la Mejor Unidad de Coincidencia, sus pesos serán alterados de mayor manera.
6. Repetir el paso 2 por N iteraciones.

#### a) Inicialización de los pesos

Como norma general, los pesos de los nodos se deben inicializar antes del proceso de entrenamiento. A continuación en la Figura VIII-35, se observan tres (3) recuadros, cada uno de ellos representa una forma de inicializar los pesos; generalmente se inicializan con valores aleatorios entre cero y uno (primer recuadro de izquierda a derecha), con valores máximos de las variables (en este ejemplo R, G o B) en las esquinas (imagen central) y valores de las variables (colores R, G, B) equidistantes (tercer recuadro).



Figura VIII-35 Inicialización de pesos. Ejemplo. (<http://davis.wpi.edu/~matt/courses/soms/>)

#### b) Cálculo de la Mejor Unidad de Coincidencia (MUC)

Un método para determinar la MUC, es iterar a través de los nodos y calcular la distancia euclidiana entre el vector de pesos de cada nodo y el vector de entradas actual; donde el nodo

cuyo vector de pesos este más cercano al vector de entradas será considerado la MUC. Según la definición de distancia euclidiana expuesta anteriormente, se puede generalizar de la siguiente forma, donde V es el vector de entrada actual y W es el vector de pesos de un nodo dado (Buckland, 2005):

$$Dist = \sqrt{\sum_{i=0}^{i=n} (V - W)^2}$$

Como ejemplo de cálculo, se consideraran tres colores, el verde, el rojo y el verde claro; al escoger el verde como muestra, se puede determinar que el verde claro está más cerca al verde que al rojo, usando la definición de distancia euclidiana, así:

Sean los códigos RGB de estos colores

Tabla 19. Códigos RGB de los colores. Ejemplo

Códigos RGB de los colores. Ejemplo	
Verde	(0, 6, 0)
Verde Claro	(3, 6, 3)
Rojo	(6, 0, 0)

$$\overline{VcV} = \sqrt{(3 - 0)^2 + (6 - 6)^2 + (3 - 0)^2} = 4.24$$

$$\overline{RV} = \sqrt{(6 - 0)^2 + (0 - 6)^2 + (0 - 0)^2} = 8.49$$

Este proceso se repite en todo el mapa y el peso con la menor distancia es escogido como MUC.

### *c) Determinación de Mejor Unidad de Coincidencia (MUC), en una vecindad*

El proceso iterativo es similar al realizado por las RNK, sin embargo, después de determinar la neurona ganadora, el siguiente paso es calcular cuales otros nodos están cerca de la MUC; este proceso se realiza ajustando sus pesos en el siguiente paso del algoritmo, al determinar cuál es radio de la vecindad y si cada nodo esta dentro de esa distancia radial o no.

Como se observa a continuación, la vecindad está centrada en la MUC (amarillo), y tiene radio R, ilustrado por la flecha (verde). Una característica interesante del algoritmo de Kohonen, es que el área del vecindario, disminuye con el tiempo, al reducir el radio, esto se puede lograr utilizando la función de decaimiento exponencial. En donde sigma ( $\sigma_0$ ) denota la longitud de la grilla en el tiempo  $t_0$ , y lambda ( $\lambda$ ), representa una variable dependiente de sigma y del numero de iteraciones propuestas al correr el algoritmo (Buckland, 2005).

$$\sigma(t) = \sigma_0 \exp\left(-\frac{t}{\lambda}\right) \quad t = 1, 2, 3, \dots$$

En la Figura VIII-36 se clarifica este concepto y muestra como la vecindad decrece con el tiempo, bajo el precepto que el vecindario permanece centrado en la MUC, en la práctica la MUC se mueve según el vector de entradas.

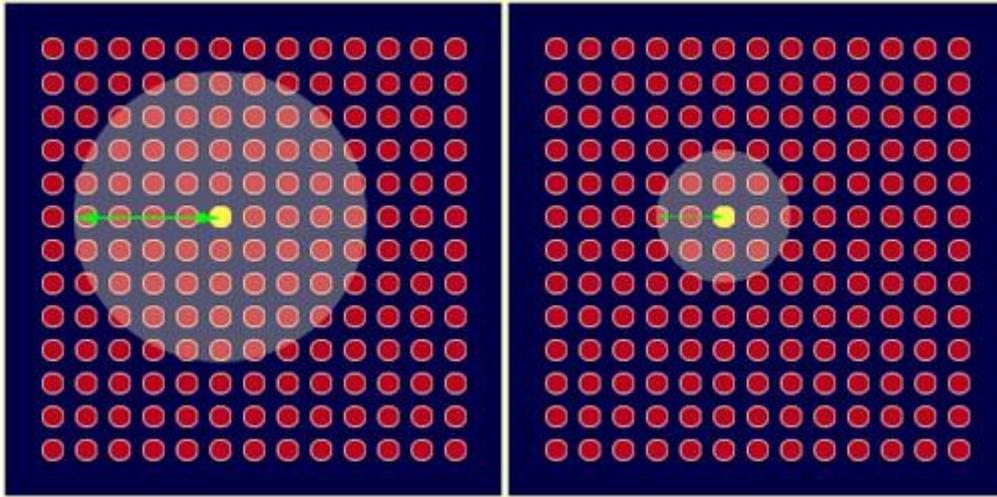


Figura VIII-36 Radios variables de la RNK. (Buckland, 2005)

#### d) Ajuste de los pesos

Cada nodo dentro de la vecindad de la MUC, tiene su vector de pesos ajustado según la siguiente ecuación:

$$W(t+1) = W(t) + \Theta(t) \alpha(t)(V(t) - W(t))$$

Donde  $t$  representa el paso de tiempo y  $\alpha$  es la razón de aprendizaje, constante o decreciente con el tiempo, según lo expuesto en la teoría del algoritmo RNK. Siendo lo usual que esta decrezca con el tiempo. El valor  $\Theta$ , representa la magnitud de la influencia que tiene la distancia entre cada nodo y la MUC, en el proceso de aprendizaje.

Al tomar en cuenta la distancia entre el nodo y la MUC, se obtiene que el proceso de aprendizaje no debe tener gran efecto en los límites de la vecindad de la MUC y debe ser mayor en la MUC; esta analogía permite utilizar el decaimiento gaussiano como ajuste a la ecuación anterior (Buckland, 2005).

$$\Theta(t) = \exp\left(-\frac{dist^2}{2\sigma^2(t)}\right) \quad t = 1, 2, 3, \dots$$

Donde *dist.* es la distancia del nodo a la MUC, y  $\sigma$  es la longitud del vecindario, según lo calculado en el punto anterior. Es importante notar que el valor de  $\Theta$ , puede utilizar otras funciones diferentes a la gaussiana, como la ecuación del sombrero mexicano, la distancia entre cuadrados concéntricos y la distancia entre hexágonos concéntricos (Figura VIII-37), entre otras.

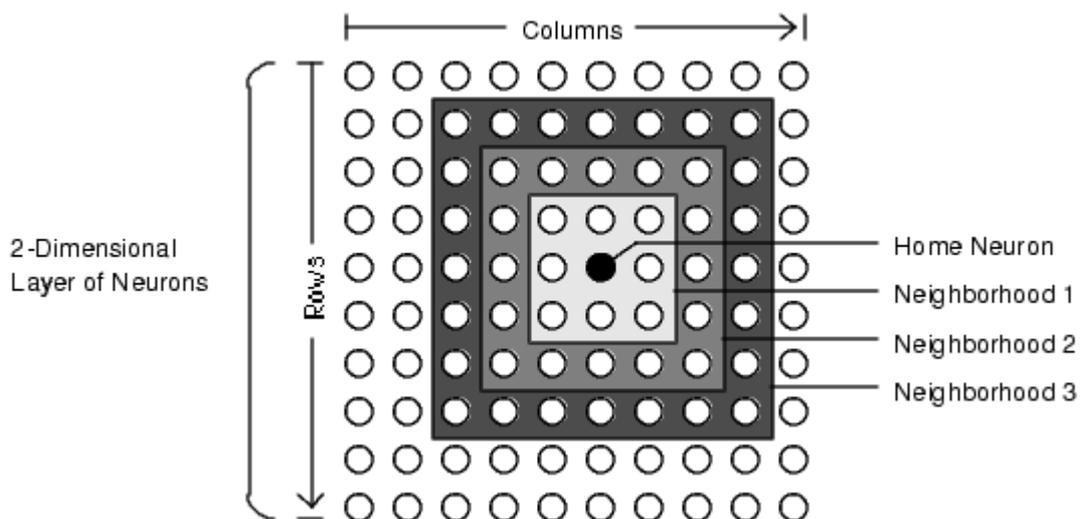


Figura VIII-37 Distancia entre cuadrados concéntricos. Función `boxdist` Matlab 7.1. (The MathWorks Inc., 2008)

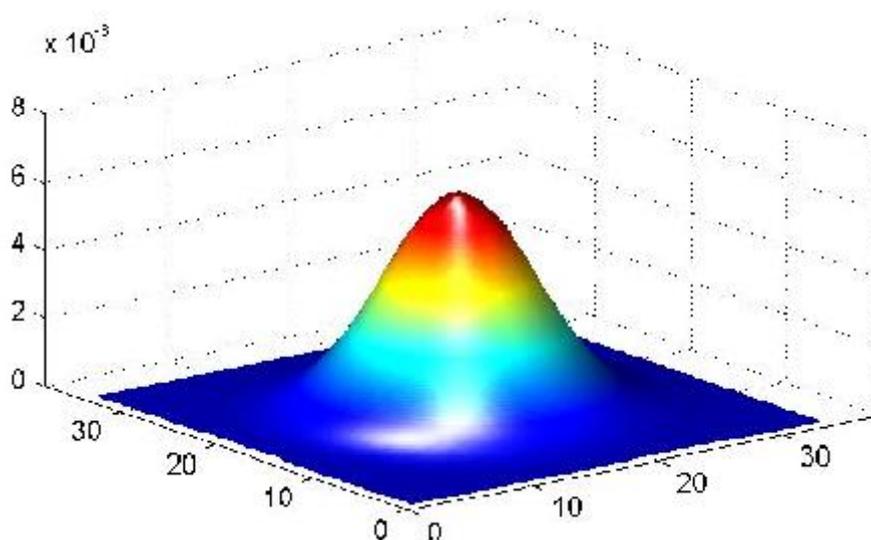


Figura VIII-38 Ajuste de pesos, según el decaimiento gaussiano, valores en colores cálidos implican un mayor ajuste de pesos en la vecindad MUC.

La ecuación de cálculo de pesos, implica que el peso en el instante  $t+1$ , es igual al peso en el instante  $t$ , mas una fracción de la diferencia entre el vector de entradas y peso en el instante  $t$ .

Como se expuso anteriormente, este paso corresponde a la quinta etapa del proceso de entrenamiento, y debe ser repetido por  $N$  número de iteraciones.

### E. Diagrama de Voronoi

Un diagrama de Voronoi, llamado así en honor al matemático ruso Georgy Fedoseevich Voronoi, también se conoce como teselación de Voronoi o Dirichlet, es un tipo de descomposición de un espacio métrico, determinada por las distancias a un juego discreto de objetos en el espacio, como lo puede ser un conjunto de puntos. Su aplicación más simple, es la representación en un plano de un conjunto de puntos  $S$ , para el que el diagrama de Voronoi, es la partición del plano que asocia una región  $V(p)$  con cada punto  $p$ , perteneciente a  $S$ , de forma tal que todos los puntos en  $V(p)$ , están más cerca de  $p$ , que de cualquier otro punto de  $S$ . (Voronoi, 1907)

Los diagramas de Voronoi, son usados en hidrología, geofísica y meteorología, para analizar datos distribuidos espacialmente como precipitaciones, en este caso reciben el nombre de Polígonos de Thiessen; en física de materia condensada, los diagramas se usan con el nombre de Celdas de Wiegner-Seitz.

### F. Teselaciones de Delaunay

En las teselaciones de Delaunay, la circunferencia circunscrita en cada triángulo de la red no contiene ningún vértice de otro triángulo, en donde los centros de los círculos circunscritos, son los vértices de los diagramas de Voronoi (Figura VIII-39). Estas triangulaciones, para los datos de entrada, se realizan utilizando esferas circunscritas, no círculos; y son útiles para representar zonas en un espacio tridimensional donde, no se encuentran más puntos del conjunto de entrada de datos. (Wikipedia, 2008)

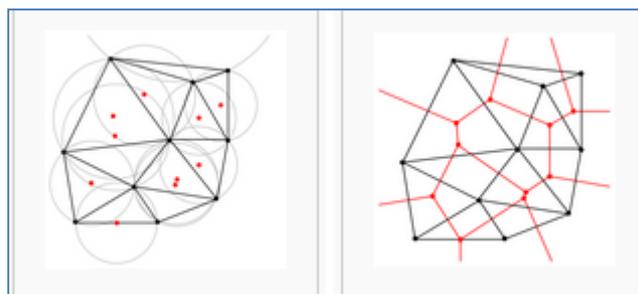


Figura VIII-39. Teselación de Delaunay. (Wikipedia 2006)

## **IX. Delimitación del problema**

El presente capítulo tiene como objetivo delimitar el alcance de este trabajo, el tipo de vigilancia tecnológica a emplear, los ciclos biogeoquímicos a estudiar y las herramientas de vigilancia tecnológica a ser consideradas en el estudio de los mismos.

En particular, se desea analizar las tecnologías aplicadas al estudio de los ciclos biogeoquímicos, con propósitos académicos e investigativos y dar una herramienta que facilite la toma de decisiones, respecto a qué tecnologías se usan a nivel mundial y cuáles se pueden adoptar en nuestro país; este enfoque implica partir de la base que el estudio no se realiza para desarrollar un producto, diferenciarse de la competencia u obtener un beneficio económico.

Teniendo en cuenta estos antecedentes, se procederá a seleccionar el tipo de vigilancia apropiado y la(s) herramienta(s) para lograrla.

### **A. Tipo de Vigilancia Tecnológica a Emplear**

Conforme al esquema desarrollado por el EAFIT ([ver Figura VI-3](#)) (EAFIT, 2008), la actividad de vigilancia se divide en cuatro grandes ramas independientes: 1) la vigilancia tecnológica, 2) la vigilancia competitiva, 3) la vigilancia comercial y 4) la vigilancia del entorno. Un aparte de la Figura VI 3 se muestra a continuación, para facilitar la orientación del lector.

Este estudio se centrará en la vigilancia tecnológica, la forma de realizarla, su objeto de estudio, el propósito de la vigilancia y las herramientas para ejecutarla. Al basarse en las subdivisiones de este tipo de vigilancia, es más sencillo determinar ¿qué ruta seguir?, y ¿cual vigilancia se ajusta mejor a los objetivos de este estudio?

Como se manifestó anteriormente, la siguiente Figura suministra información muy valiosa, ya que gracias a las características de cada subdivisión, se puede llegar al tipo de vigilancia tecnológica que mejor se ajusta a este estudio; en particular, se observa que la vigilancia tecnológica se divide en Vigilancia Científica y Vigilancia Estratégica, teniendo cada nueva subdivisión características únicas, que identifican a partir de qué tipo de información se pretende realizar la actividad de vigilancia.

De la Figura IX-1, se obtiene que la Vigilancia Estratégica se subdivide en tres ramas, la primera de ellas, utiliza relaciones económico-financieras y/o de trabajo entre las empresas del sector, para realizar su objetivo; las otras dos ramas, tienen como eje central a la competencia, la primera analiza su capacidad tecnológica y la segunda realiza seguimiento a la trayectoria de

los artículos de la competencia; de otra parte, la Vigilancia Científica tiene tres ramas independientes, la primera rama usa Ingeniería Inversa de los productos de la competencia; la segunda rama, realiza análisis de patentes mediante tratamientos estadísticos y la tercera rama realiza seguimiento de publicaciones científicas y técnicas.

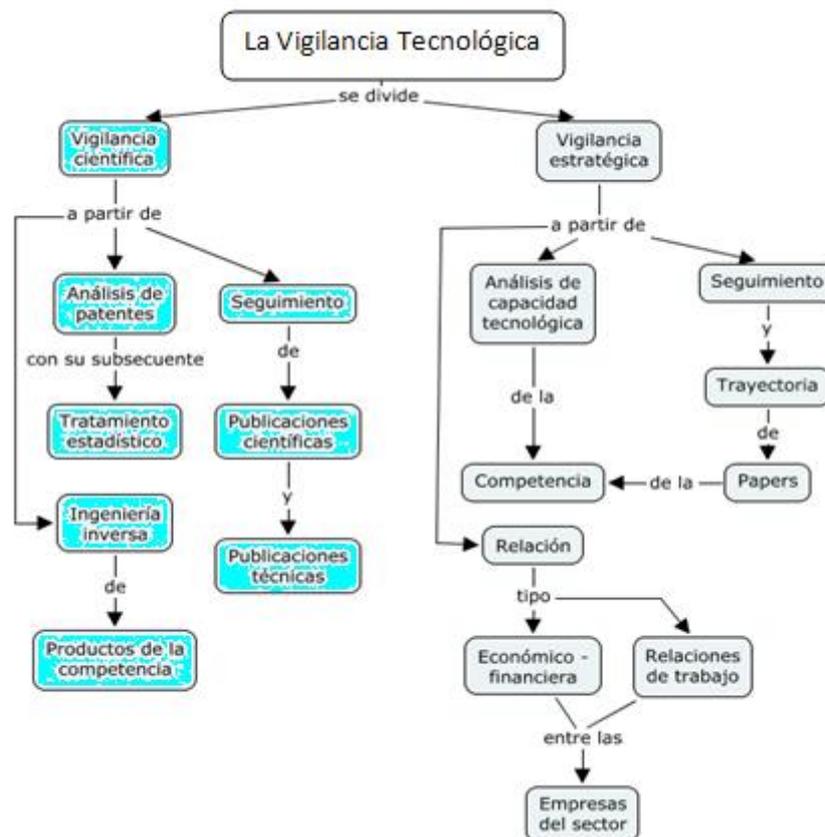


Figura IX-1 Aparte de la Figura VI-3. Tipos de Vigilancia Tecnológica. (EAFIT, 2008)

Según lo expresado en la definición del caso de estudio y la figura anterior, la ruta que se debe seguir para el desarrollo de este documento, está más clara, ya que la sub-rama de Vigilancia estratégica, se centra en el estudio de la competencia y en relaciones entre las empresas del sector a estudiar.

Como se mencionó anteriormente, se parte del principio que este estudio tiene alcances académico - científicos y que no se tiene competencia alguna; adicionalmente, el objeto es obtener información sobre las tecnologías asociadas al estudio de los ciclos biogeoquímicos e hidrológico, no una relación entre las empresas involucradas en el desarrollo de estas herramientas; motivo por el cual se ha descartado la realización de vigilancia estratégica, reduciendo el tipo de vigilancia tecnológica a emplear a vigilancia científica.

Análogamente al proceso de eliminación anterior, se observa que la Figura IX-1 y la definición del caso de estudio, suministran la información necesaria para escoger la(s) rama(s) de la vigilancia científica que cumple(n) los requisitos, para el desarrollo de este trabajo. De esta manera, se descarta inmediatamente realizar vigilancia en base a ingeniería inversa de los productos de la competencia; motivo por el cual sólo quedarían disponibles dos ramas del árbol de tipos de vigilancia tecnológica: la vigilancia científica que realiza análisis de patentes y la vigilancia científica que hace seguimiento de publicaciones científicas y técnicas. (Ver Figura IX-2)

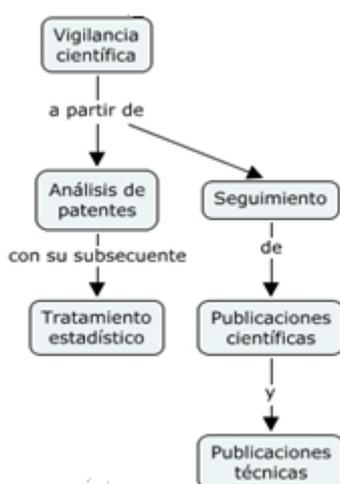


Figura IX-2 Ramas de la Vigilancia Tecnológica Científica viables para este estudio. Aparte de la Figura IV-4. (EAFIT, 2008)

## B. Herramientas de Vigilancia Tecnológica a Incorporar

Dados los tipos de Vigilancia Tecnológica Científica escogidos para el caso de estudio, se espera trabajar con volúmenes considerables de datos alfanuméricos, ya que tanto la actividad de consulta de patentes como el seguimiento de publicaciones científicas y técnicas, implica estar en contacto con bases de datos especializadas, que por lo general crecen día a día, conforme se patentan nuevos desarrollos y se publican nuevos trabajos científicos y técnicos.

Según lo expuesto en el Marco Conceptual, las herramientas de Vigilancia Tecnológica se diferencian según su función, ya sean orientadas a la recopilación y al procesamiento de la información, o al análisis de datos.

Respecto al grupo de análisis de datos, se trabajó con datos provenientes del estudio de patentes y los campos tecnológicos en los que se encuentran, según la clasificación de la USPTO. Esta clasificación se tomó como referencia principalmente porque en la base de datos de la USPTO se encontró la información mundial para cada uno de los campos tecnológicos. De

otra parte, de usarse el estándar IPC-8 de la WIPO (Oficina Internacional de Propiedad Intelectual) se hubiera requerido realizar concordancias entre estos dos estándares, al contar con la información en formato USPTO. Estas concordancias no son equivalencias y se deben tratar como clasificaciones sugeridas entre dos sistemas que pueden ser similares. (USPTO, 2008)

Del primer grupo de herramientas, se trabajará con Redes Neuronales Artificiales, dadas las capacidades de Auto Organización encontradas por T. Kohonen en 1989, especialmente para la creación de Mapas Auto Organizados o SOM, por sus siglas en inglés (Kohonen, 1995), según se mostró en el Marco Teórico del presente documento.

El uso de las RNK y la creación de SOMs que visualicen clústeres de información obtenida en grandes bases de datos, integra dos de las herramientas de Vigilancia Tecnológica expuestas en el Marco Conceptual: Las Redes Neuronales y las Bases de Datos. Siendo las Redes Neuronales de Kohonen, ideales para este trabajo por los siguientes motivos:

1. No requieren etapa de validación del entrenamiento.
2. Poseen capacidades de Auto Organización.
3. Los SOMs representan espacios multidimensionales en dos o tres dimensiones.
4. Se han empleado exitosamente en la clasificación de grandes volúmenes de información, como mapas genéticos, indicadores económicos, colecciones de imágenes y reconocimiento de patrones de voz.
5. El desarrollo realizado con las RNK es replicable a otras áreas del conocimiento.
6. Los SOMs son escalables y pueden actualizarse conforme aumenta el número de registros de entrada.
7. Se pueden crear SOMs con diferente número de clústeres de datos.
8. Tanto las patentes como los documentos científicos y técnicos crecen en número día a día, necesitando de una herramienta que pueda manejar grandes volúmenes de información, para crear las clasificaciones que faciliten su estudio.

## **X. Implementación de la Herramienta para la Toma de Decisiones**

Este capítulo trata sobre la herramienta elegida para tomar decisiones, relacionadas a los ciclos biogeoquímicos y el ciclo hidrológico. Se ha elegido utilizar el algoritmo de Kohonen y su aplicación a la creación de mapas auto-organizados, como herramienta que facilita la visualización y la agrupación de ítems de bases de datos.

### **A. Características de la herramienta desarrollada para la creación de SOMs.**

Una vez conocidas las características del algoritmo de Kohonen, el funcionamiento de las redes RNK y el proceso para la creación de los SOMs, se expondrán a continuación las características principales de la herramienta desarrollada para la creación de Mapas Auto Organizados de Kohonen.

Inicialmente se mostrará la funcionalidad de la herramienta, con el ejemplo tratado anteriormente, sobre la auto organización de un conjunto de puntos en formato RGB, creado aleatoriamente; para luego proceder a aplicar la herramienta en la clasificación de las publicaciones científicas y técnicas, relacionadas con los ciclos biogeoquímicos y el ciclo hidrológico; de esta manera, se implementaría a la actividad de Vigilancia Tecnológica en su componente Científico.

### **B. Lenguaje de programación usado**

Para el desarrollo de la herramienta, se escogió el programa Matlab 7.1 (The MathWorks Inc., 2008), por su facilidad para generar distintos tipos de gráficas (diagramas de dispersión, histogramas, gráficas en 2D y 3D, entre otras) y trabajar datos importados de diversas fuentes, como Hojas de Cálculo, Procesadores de Texto y Bases de Datos, con sus archivos separados por comas (CSV), tabulaciones o espacios.

### **C. Entrada de datos**

Se desarrolló un programa con extensión. M de Matlab, que se encarga de todo el proceso de creación de los Mapas Auto Organizadas, desde la entrada de datos, hasta la salida final en pantalla del SOM, así como de la determinación de algunas características estadísticas del vector de entradas y de la medida de precisión del mapa, según diferentes análisis que se mostrarán posteriormente.

La entrada de datos del programa, es compatible con la usada en el programa SOM PAK, un software de distribución libre, en lenguaje C, creado por el equipo de programadores SOM de la Universidad de Helsinki, donde Kohonen se desempeña como profesor.

El programa creado permite importar los datos desde una hoja de Microsoft Excel, siempre y cuando el formato del archivo tenga la extensión .XLS; a los usuarios de MS Excel 2007, se les sugiere guardar el archivo de datos en ese formato, o utilizar el formato de archivos del paquete SOM PAK, como se mencionó anteriormente, con extensión .DATA, para almacenar el vector de entradas.

El formato .DATA, para aquellos no familiarizados con el mismo tiene la siguiente estructura: en la primer fila del documento se encuentra el numero de dimensiones (**D**) del vector de entrada, en la segunda fila aparecen los nombres de cada una de las dimensiones del vector de entrada, precedida por los caracteres #n; de la fila tres (3) a la fila m, se encuentran los valores de cada conjunto de puntos D dimensional.

Para el ejemplo desarrollado, las primeras cinco filas tendrían la siguiente estructura en el formato .DATA del SOM PAK. Dado un vector de entradas con tres (3) dimensiones (D), una equivalente al valor R (rojo), otra por el valor G (verde) y la ultima con el valor B (azul).

**Tabla 20. Características de un archivo .DATA para un vector de entrada dado (ejemplo)**

Características de un archivo .DATA para un vector de entrada dado (ejemplo)	
<b>Numero de dimensiones (D)</b>	3
<b>Nombres de las dimensiones</b>	#n R G B
<b>Valores punto i (D dimensional)</b>	255 0 0
<b>Valores punto i+1 (D dimensional)</b>	0 255 0
<b>Valores punto M (D dimensional)</b>	0 0 255

El vector de entrada de datos, como se mencionó anteriormente, contiene un conjunto de *m* tripletas de puntos (R,G,B); para este caso de estudio en particular, se crearon aleatoriamente cien (100) tripletas de puntos, usando la función ALEATORIO.ENTRE de MS Excel 2007; con esta función, se puede generar un numero aleatorio, cuyo valor está en un rango comprendido entre los dos argumentos de entrada. La sintaxis de la función es la siguiente:

$$= \text{ALEATORIO.ENTRE}(\text{inferior}, \text{superior})$$

Como es sabido, los valores RGB están comprendidos entre 0 y 255; motivo por el cual se utilizó como valor inferior el 0 y como valor superior 255, al ingresar los argumentos de la función anteriormente mencionada.

#### **D. Pre-tratamiento de la información**

Una vez ingresado el vector de entrada de datos, el programa debe realizar un pre-tratamiento de la información contenida en el mismo. Como se mostró al inicio de este capítulo, las redes neuronales requieren que los datos de entrada sean normalizados, generalmente este proceso se realiza entre 0 y 1, o entre -1 y 1; para este fin, se puede normalizar linealmente, haciendo que la varianza sea 1, o aplicándole el logaritmo natural a los valores.

Como se explicó anteriormente, el método más usado para determinar el error y la distancia entre el vector de entrada y el vector de pesos, es el cálculo de la distancia euclidiana; motivo por el cual, la escala de las variables es muy importante para determinar como el mapa va a ser. Si el rango de una variable dada es mucho mayor que el de las otras, esa variable probablemente dominara la organización en el mapa.

El programa realiza la normalización vectorial, escalando los valores del conjunto de entrada, entre 0 y 1 y complementariamente realiza otra normalización entre -1 y 1; para el primer cálculo, se divide cada elemento del vector de entradas, entre el valor máximo encontrado en las  $m$  tripletas de puntos, para esa dimensión. La segunda normalización, se realiza multiplicando cada elemento del vector de entradas por dos (2) y restándole uno (1) al resultado; así para valores de CERO el resultado sería -1, y para valores de 1 el valor de salida normalizado se conservaría en 1.

#### **E. Representación grafica de la información de entrada**

Al tener los datos de entrada normalizados, se procede a mostrar en pantalla la representación espacial del conjunto de  $m$  puntos, y Diagrama de Voronoi de cada una de sus vistas auxiliares; en el ejemplo tridimensional, se muestra la vista frontal, la superior y la lateral, para el ejemplo de estudio, se tiene una figura con cuatro gráficos incluidos; en la esquina superior izquierda, se encuentra un diagrama de dispersión de los puntos RGB del vector de entrada; en la esquina superior derecha se encuentra un Diagrama de Voronoi, de los puntos en el plano Rojo-Verde; en la esquina inferior izquierda, se encuentra un diagrama análogo, con los puntos del plano Rojo-Azul y finalmente, en la esquina inferior derecha se muestra un diagrama de Voronoi con los puntos del plano Verde-Azul, según lo mostrado en la Figura X-1.

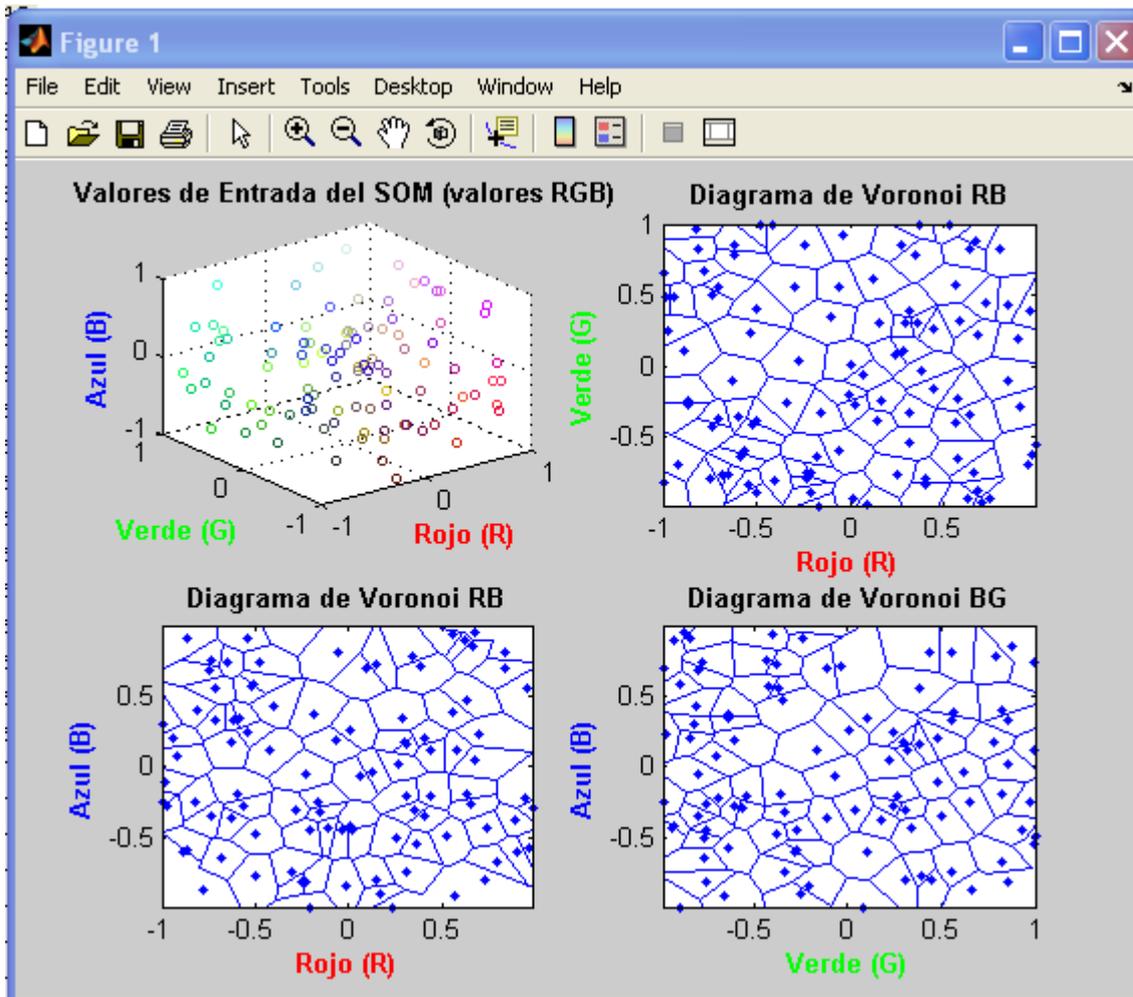


Figura X-1 . Visualización de los datos de entrada. Programa de creación de SOM

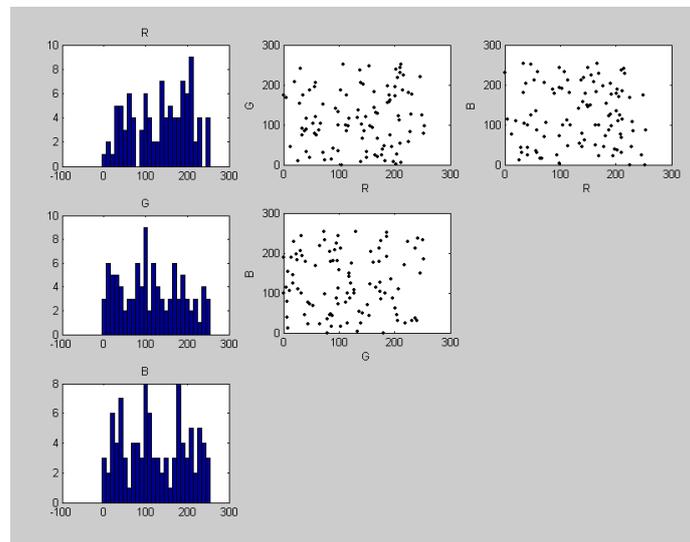
Se escogió, representar los planos auxiliares como Diagramas de Voronoi, por las aplicaciones complementarias que estos pueden ofrecer, respecto a un diagrama común de dispersión o nube de puntos; entre las aplicaciones sobresale su funcionalidad para facilitar la búsqueda de vecinos o elementos más cercanos, aplicación que ha sido aprovechada ampliamente por la industria, ante interrogantes como ubicar la antena más cercana de una red de comunicación móvil y identificar los establecimientos más cercanos a un punto dado, entre otros; a su vez, en el análisis de bases de datos se utiliza para encontrar el elemento más similar a un objeto determinado.

Posteriormente se presentan histogramas de ocurrencias y diagramas de dispersión, de las tres variables contenidas en la matriz de vectores de entrada, como información gráfica complementaria. (Ver Figura X-2)

Complementariamente a la información de los gráficos, el programa muestra en pantalla una tabla con características estadísticas de cada una de las variables de entrada normalizadas entre 0 y 1; en esta tabla se indica el nombre de la variable, el valor mínimo, medio y máximo de la misma, así como la desviación estándar y el número y porcentaje tanto de los datos faltantes como de los únicos. A continuación, se enseña la tabla obtenida para el ejemplo desarrollado en el capítulo (tabla 21):

**Tabla 21. Tabla de características estadísticas. (Ejemplo)**

Nombre	Min	Media	Max	StDev	Faltantes	Unicos
Variable1	0	0.488902	0.992157	0.293828	0 (0%)	88 (88%)
Variable2	0.00784314	0.46498	1	0.310623	0 (0%)	77 (77%)
Variable3	0.00784314	0.494588	0.988235	0.279162	0 (0%)	83 (83%)



**Figura X-2 Graficas auxiliares. Columna 1: Histogramas de Ocurrencias, Columnas 2 y 3: Diagramas de dispersión de las variables.**

## F. Inicialización y Entrenamiento del SOM

Luego de terminar la etapa de pre-tratamiento y de la representación grafica de los datos de entrada, se procede a determinar la topología de la red y a inicializar los pesos de la capa de salida. La topología de la capa de neuronas de salida, se escogió como hexagonal en este ejemplo y la inicialización de los pesos se realizó con dos métodos diferentes, para probar la bondad del entrenamiento del mapa. Primero se utilizó la inicialización lineal (imagen 2, Figura X-3), en donde el azul y el rojo se ubicaron en esquinas opuestas y el verde se localizó en la diagonal principal, luego el azul y el rojo inician un gradiente lineal descendiente hacia el

verde; el segundo método, consiste en la inicialización aleatoria (imagen 3, X-3), proceso que se explicó en la primera parte del capítulo.

Para el entrenamiento, se probó la funcionalidad del algoritmo Batch (imagen 4, Figura X-3) y del algoritmo Secuencial (imagen 5, X-3), y se identificó la última MUC en cada caso ; según la documentación encontrada en el Neural Networks Toolbox de Matlab, el algoritmo Batch es más rápido que el Secuencial, y es el usado por defecto en este Toolbox. (The MathWorks Inc., 2008).

En el algoritmo Batch, los valores del vecindario son tenidos en consideración, a diferencia del Secuencial en el que se realiza el entrenamiento punto a punto. Es catalogado como un algoritmo determinista e iterativo, de naturaleza cuasi Newtoniana (gradiente de segundo orden), en el que se minimiza la distorsión extendida (Ritter, et al., 1992). Se recomienda al lector con interés en profundizar sobre la naturaleza de este algoritmo, leer el trabajo de Fort, Letremy y Cottrell (Advantages and drawbacks of the Batch Kohonen algorithm, 2002).

## **G. Determinación de la calidad del Mapa Auto Organizado**

Luego de haber obtenido los mapas entrenados (ver Figura X-3), es necesario determinar la calidad de los mismos; a pesar de estar basados en el principio de auto organización de Kohonen y de realizarse de modo no supervisado, los Mapas Auto Organizados o SOMs, deben cumplir con dos (2) propiedades principalmente, así como cualquier mapa :

- 1- Exactitud en el representación de los datos
- 2- Exactitud en la representación topológica del juego de datos

Teniendo en cuenta las anteriores consideraciones, se desarrolla una medida para cada una de ellas, de forma que se obtenga un indicador de la calidad del mapa, relacionado con cada una de las propiedades así:

Para 1) Se adapta el concepto de Error Medio de Cuantización (EMC), esta medida determina el error medio entre los vectores de datos y sus Mejores Unidades de Coincidencia o MUCs, en el mapa.

Para 2) La literatura propone diferentes medidas y no existe aún un consenso al respecto. Para efectos del programa realizado, se adopta la Medida del Error Topográfico (MET) como indicador, donde el MET es el porcentaje de vectores de datos, para los cuales sus MUC, no son adyacentes a sus Segundas Mejores Unidades de Coincidencia (SMUC).

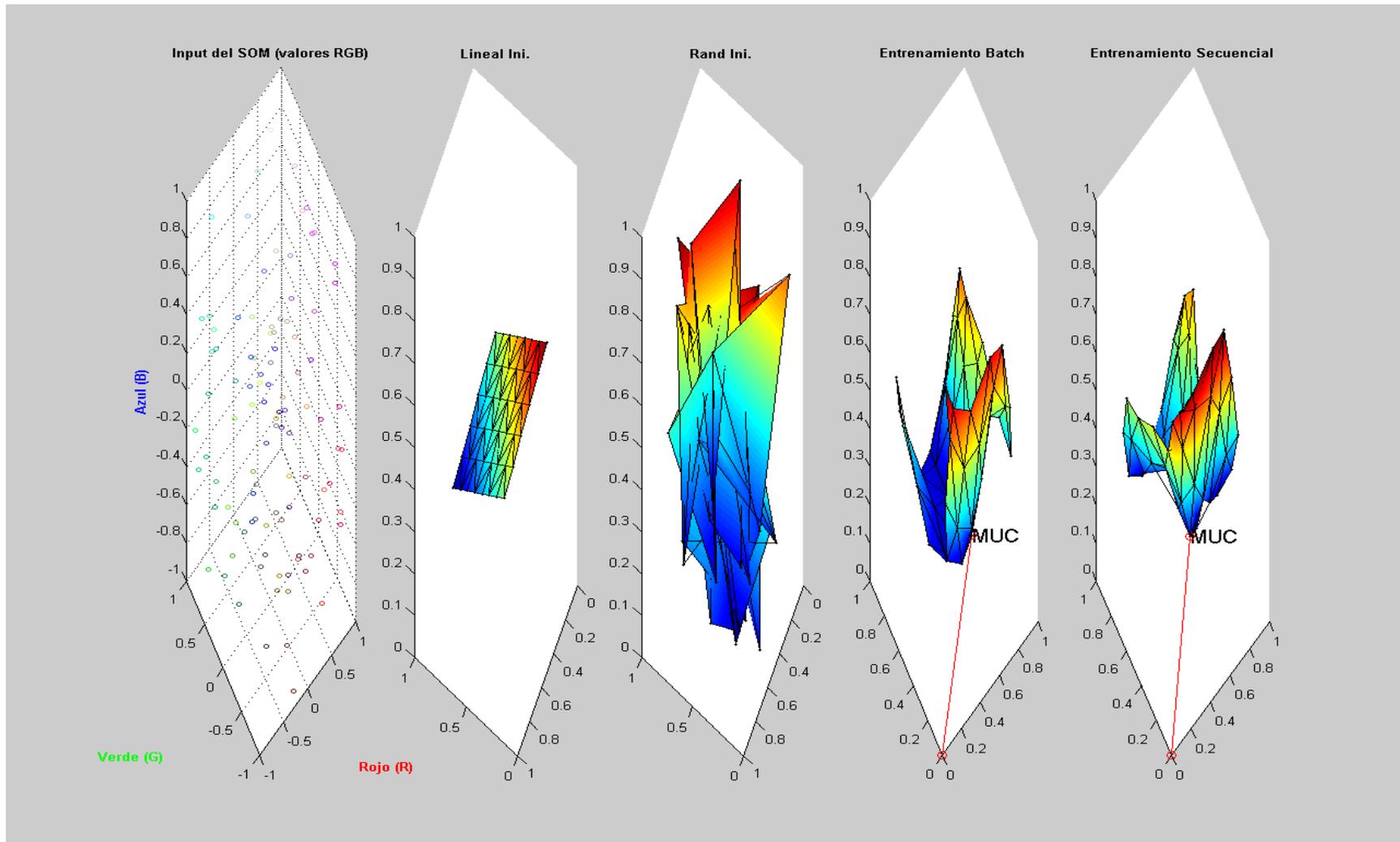


Figura X-3 Mapas de inicialización y entrenamiento. De izquierda a derecha: 1) Valores RGB de entrada al SOM, 2) Mapa de Inicialización lineal, 3) Mapa de Inicialización aleatoria, 4) Mapa de entrenamiento con el algoritmo Batch y 5) Mapa de entrenamiento secuencial

Para el ejercicio tratado en este capítulo, estos indicadores tuvieron los siguientes valores:

Tabla 22. Indicadores de Calidad del Mapa Auto Organizado. (Datos ejemplo)

Tabla de Indicadores de Calidad del Mapa Auto Organizado (SOM)		
Mapa	Error Medio de Cuantización (EMC)	Medida del Error Topográfico (MET)
Mapa de inicialización lineal	0.2737	0
Mapa de entrenamiento usando <i>Batch</i>	0.1725	0.03
Mapa de entrenamiento usando Secuencial	0.205	0.06

Como se observa en la tabla anterior, al desdoblarse el Mapa Auto Organizado de inicialización a R3, se redujo el EMC, pero sufrió la capacidad de representar la topología, MET. Este resultado se espera siempre que se compare con el mapa de inicialización lineal, ya que en este último, siempre estarán adyacentes la MUC y la SMUC, sin embargo, es posible mejorar esta medida, al utilizar un radio de vecindario mayor en el entrenamiento. Adicionalmente, es posible determinar que para este caso de estudio, el algoritmo de entrenamiento Batch, efectivamente presenta menores valores de EMC y de MET, que el algoritmo de entrenamiento Secuencial.

## H. Visualización de Celdas

Esta característica, desarrollada por el “SOM toolbox team” de la Universidad de Helsinki, permite mostrar la grilla de datos del Mapa Auto Organizado, como si esta perteneciera a un espacio bidimensional. La visualización implica graficar una malla regular de celdas, cuyos elementos tienen asociados valores; como es de esperarse, este proceso es más claro, al representar mapas en 1 o 2 dimensiones.

La principal herramienta de visualización de celdas, es conocida como *Unified Distance Matrix (U-Matrix)*, o Matriz - U; como su nombre lo indica, representa una matriz en la que se visualiza la distancia entre unidades vecinas del SOM y sirve para estructura de los clústeres en el mapa; donde valores altos en la Matriz - U indican un borde de clúster, y áreas uniformes de valores pequeños indican la presencia de clústeres por sí mismos.

Adicionalmente se muestran los planos de los componentes (uno por cada variable de entrada), en estos planos se indican los valores de cada una de las unidades o celdas del plano.

Para el ejemplo desarrollado en este capítulo, se obtuvo la salida de datos mostrada en la Figura X-4

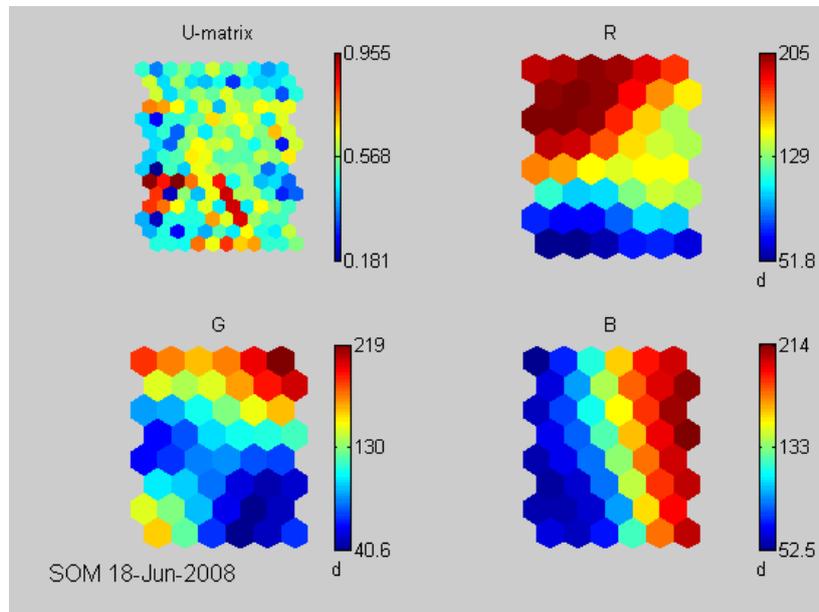


Figura X-4 Representación de Celdas. En la esquina superior izquierda Matriz - U, las otras tres graficas corresponden a los planos de componentes R, G y B. Ejemplo

De la Figura anterior (X-4), se obtiene una auto organización de clústeres, tanto para el componente Rojo, como para el Verde y el Azul, se aprecia como los mayores valores de Rojo se ubican en la esquina superior izquierda, los mayores valores de Azul se encuentran en la margen derecha del mapa y como los mayores valores de verde se encuentran en la esquina superior derecha y se presenta un pequeño clúster en la esquina inferior izquierda. Una vez realizadas estas consideraciones, es más sencillo analizar la Matriz - U, correspondiente a la grafica de la parte superior izquierda de la Figura X-4; en ella se aprecian limites de clústeres, caracterizados por los colores cálidos (rojo, naranja y amarillo) y clústeres de datos, en las zonas con colores fríos (verde, cian y azul); observando la distribución de celdas, se pueden observar clústeres en la parte superior, en la esquina inferior izquierda y un gran clúster en forma de V, en la parte derecha de la matriz, siguiendo la forma aproximada de la zona de verdes en el plano de componente R (Rojo), zona que corresponde con los mayores valores de Azul en el plano B.

En la Figura X-5, se muestra la MUC, ubicada en cada uno de los planos de componentes y en la Matriz - U; como la MUC es única para cada SOM, corresponde a una misma celda en cada plano de componentes y a una región equivalente de esta celda, en la Matriz de Distancias Unificadas.

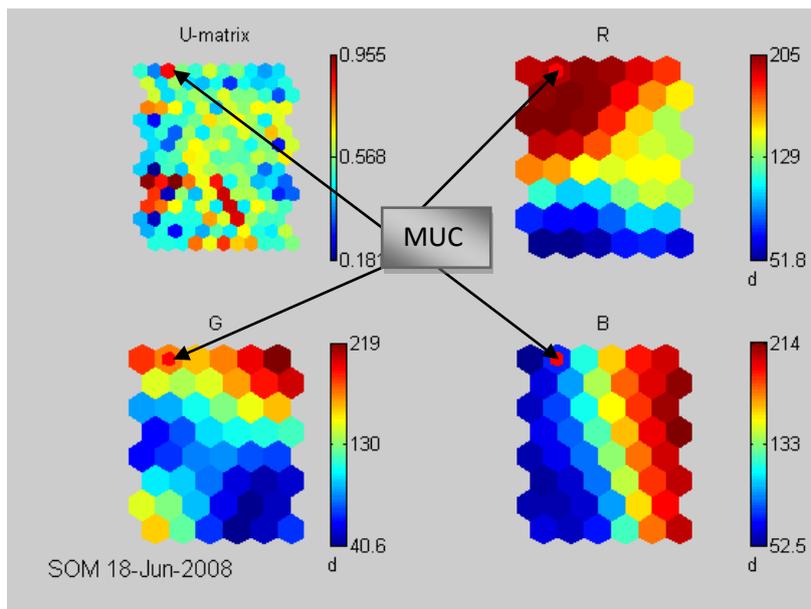


Figura X-5 Identificación de la MUC en los planos de componentes y en la Matriz - U

## XI. Vigilancia Científica sobre los Ciclos Biogeoquímicos y el Ciclo Hidrológico

En los capítulos anteriores, se identificó el tipo de Vigilancia Tecnológica que se realizará sobre los ciclos Biogeoquímicos y el ciclo Hidrológico; así como las características técnicas y conceptuales de la herramienta escogida para la toma de decisiones. Específicamente y para contextualizar al lector, se determinó que este estudio usará la Vigilancia Tecnológica Científica, y se concentrará en dos (2) de sus ramas: el análisis de patentes y el seguimiento de publicaciones científicas.

Respecto a la herramienta desarrollada para facilitar la toma de decisiones, se escogió utilizar el algoritmo de Kohonen y su aplicación para la creación de Mapas Auto Organizados (SOMs), herramienta que es ampliamente tratada en el Marco Teórico.

Como se expresó anteriormente, este capítulo tratará sobre la Vigilancia Tecnológica Científica, realizada sobre los ciclos Biogeoquímicos y el ciclo Hidrológico. Con este fin, el capítulo se dividirá en dos grandes secciones, la primera de ellas, dedicada al análisis de patentes relacionadas a las tecnologías o desarrollos, en áreas del conocimiento vinculadas al estudio de los ciclos mencionados anteriormente; la segunda sección, está dedicada a la rama de la vigilancia tecnológica científica, que se encarga del seguimiento de publicaciones científicas y técnicas; en esta sección, se implementará la herramienta informática

desarrollada en el capítulo anterior, en el estudio de los ciclos biogeoquímicos y el ciclo hidrológico.

### **A. Seguimiento de Publicaciones Científicas y Técnicas**

En esta sección del capítulo, se mostrará la funcionalidad de la herramienta desarrollada, y su capacidad para generar Mapas Auto Organizados (SOMs); concretamente, se realizaron SOMs de las publicaciones científicas y técnicas, que tratan temas relacionados a los ciclos biogeoquímicos y del ciclo hidrológico. Con el fin de obtener estas publicaciones, se utilizó el Catalogo en Línea de la Biblioteca General de la Pontificia Universidad Javeriana, y su opción de búsqueda en Bases de Datos Especializadas.

Del menú de Bases de Datos, se hizo una selección de las que pertenecen conjuntamente a las áreas de Ingeniería y Ciencias Naturales, entre las que se encontraban: Science Direct, Springer Link, ProQuest y Science Magazine. De esta selección, se escogió trabajar con la base de datos ProQuest, que hace parte del *Cambridge Information Group*, por presentar cerca de 10 veces el número de registros respecto a Science Direct y Science Magazine y por tener un formato para exportar los resultados de búsqueda más fácil de tratar utilizando Matlab 7.1. Esta base de datos provee acceso y navegación a más de 125.000 millones de páginas digitales relacionadas con múltiples campos entre los que se encuentran artes, literatura, ciencias sociales, ciencia, tecnología y medicina. (ProQuest, 2008).

Una vez dentro de la Base de Datos de ProQuest, se buscó la información de publicaciones científicas y técnicas en estas cinco (5) categorías:

- *Disertations & Theses*
- *Agriculture Journals*
- *Biology Journals*
- *Research Library*
- *Science Journals*

Los criterios de búsqueda iniciales fueron GEOLOGY, CHEMISTRY, BIOLOGY AND HYDROLOGY; sin embargo fue necesario refinar la búsqueda, ya que estas áreas comprenden temas mucho más diversos que los ciclos biogeoquímicos; ante este escenario, se ingresó como criterio GEOCHEMISTRY AND BIOGEOCHEMISTRY, búsqueda que arrojó 9865 resultados que cumplían con estas dos palabras claves; esta búsqueda sin embargo, no contemplaba al ciclo hidrológico, motivo por el cual se refinó aún más el criterio, y se buscó por registros que pertenecieran a las categorías BIOGEOCHEMISTRY AND HYDROLOGY, arrojando 219 resultados de búsqueda.

Los resultados obtenidos para los criterios de búsqueda mencionados anteriormente, en las bases de datos seleccionadas se observan en la siguiente tabla:

Tabla 23 Registros obtenidos para el criterio BIOGEOCHEMISTRY AND HYDROLOGY, en las diferentes bases de datos consultadas

Base de datos	Número de registros
Science Direct	26
Science Magazine	33
ProQuest	219
Springer Link	1419

De otra parte, surge la necesidad de categorizar los elementos obtenidos usando el conector lógico AND, estos datos pertenecen a las dos categorías mencionadas anteriormente, y representan un conjunto de datos heterogéneo, donde cada registro puede contener palabras claves que lo relacionen con una o más categorías de datos, es así como la palabra BIOGEOCHEMISTRY, implica que el registro que la contiene, pertenece a la categoría de Biología (BIO), Geología (GEO) y Química (CHEM); si a manera de ejemplo, el registro también contiene la palabra HYDROLOGY, se podría clasificar en cualquiera de las tres categorías anteriores y en la categoría de Hidrología (HYDRO). Como se observa, este sería un escenario de clasificación mucho más preciso, que el obtenido usando el conector OR.

**Este escenario llevó a la creación de cuatro grupos de categorías, o dimensiones vectoriales, en términos del algoritmo de entrada (ECO, GEO, CHEM e HYDRO).** De esta manera, dado un registro de entrada de datos, éste estaría identificado por unas coordenadas vectoriales que lo diferencian de los demás registros. Para este caso de estudio, cada registro de entrada tiene un valor en cada una de las cuatro dimensiones vectoriales creadas.

Respecto a las dimensiones vectoriales, retomando el ejemplo tratado en el capítulo VIII, sobre los vectores de colores, se tenía que para ese caso, cada registro de entrada de datos tenía un valor tridimensional, al poseer un valor numérico para el componente Rojo (R), otro valor para su componente Verde (G) y un último valor que correspondía al Azul (B). Análogamente, la categorización propuesta en este capítulo, incluye que cada registro tenga un valor BGQH, donde la primera dimensión es el valor del componente Ecológico - Biológico o ECO, el segundo valor corresponde al componente Geológico o GEO, el tercer valor corresponde al componente Químico o CHEM, y el cuarto valor corresponde a la cuarta dimensión, la dimensión Hidrología o HYDRO.

Este hecho, llevó a determinar una métrica que fuera útil para catalogar el valor de cada ítem en la entrada de datos, presentándose varios retos en el proceso; primero que todo, los datos de entrada, a diferencia del ejemplo RGB del capítulo VIII, son oraciones compuestas de letras, y números, no sólo números; el algoritmo categoriza conjuntos numéricos y puede usar máximo una etiqueta alfanumérica de identificación por registro, a su vez, las oraciones no tienen igual número de caracteres ni de palabras. Siendo necesario realizar un pretratamiento de la información para poder utilizarla en la generación de SOMs.

## 1. Pre tratamiento

Como se mencionó anteriormente, con el fin de solucionar algunos de los retos planteados, se tuvo que realizar un pretratamiento de la información obtenida. Para esto se debe estar familiarizado con la forma en que los registros son presentados por ProQuest.

Los registros de la Base de Datos de ProQuest tienen la siguiente estructura: número de registro en el primer renglón, título en el segundo, el tercer renglón inicia con la palabra "by", que determina el autor, tipo de documento, universidad, año y número de páginas, luego sigue un ";" y un código alfanumérico. A continuación, se muestra como ejemplo el registro 110 de la búsqueda.

110.

[Hydrologic controls of nitrogen cycling in an Adirondack watershed](#)

by McHale, Michael Robert, Ph.D., State University of New York College of Environmental Science and Forestry, 1999, 230 pages; AAT 9959010

Inicialmente de cada uno de los 219 registros, se extrajo sólo el segundo renglón, correspondiente al título; este campo tiene longitud variable y puede contener letras, caracteres especiales y números.

Primero se intentó leer estos datos desde MSEXCEL, usando la función XLSREAD() empleada para leer los valores RGB, del ejemplo tratado en el capítulo VII; sin embargo, esta función sólo lee datos numéricos y fue necesario buscar maneras alternas de leer datos de texto en Matlab. Para quienes no están familiarizados con este programa, las funcionalidades principales del mismo son numéricas y presenta grandes facilidades para el tratamiento y generación de gráficos; no se caracteriza propiamente por el tratamiento de datos de texto.

Para sortear este inconveniente, fue necesario crear un procedimiento que leyera e importara los datos de texto, en un formato soportado por Matlab; la primera opción fue importar los datos en formato CHAR, donde cada letra de un renglón de entrada corresponde a una posición en un vector de dimensiones [1 x n], donde n representa el número de caracteres y

de columnas en el vector. Sin embargo, computacionalmente no es un proceso óptimo, ya que al realizarlo, se están creando  $m$  (219) vectores  $n$ -dimensionales, y como cada fila en el conjunto de datos de entrada, contiene un número ( $n$ ) diferente de caracteres, estos vectores no se pueden trabajar de manera matricial ya que sus dimensiones no concuerdan, dado que no tienen el mismo número de columnas.

Este hecho y la necesidad de crear un procedimiento automático, que pudiera manejar gran cantidad de registros de entrada, llevó a la creación de una función que leyera todo el conjunto de datos de entrada, fila a fila, y contara cuantos caracteres tenía cada una de ellas, luego buscará el valor más grande de caracteres en una fila, y creará un arreglo de caracteres de tamaño [Num. Filas x  $n$ ], donde Num. Filas corresponde al número de registros (219) y  $n$ , al número máximo de caracteres; este arreglo, al ser dimensionalmente uniforme, sí puede ser llenado con los registros del archivo de texto en cada fila.

Una vez creado el arreglo de caracteres con los datos de entrada, se procede a utilizar la función CELLSTR() de Matlab, para convertir un arreglo de caracteres en un arreglo de celdas; el siguiente ejemplo creado desde la línea de comandos de Matlab, ayudará a clarificar este concepto:

```
>> c='arreglo de caracteres'
c =
arreglo de caracteres
>> a=cellstr(c)
a =
    'arreglo de caracteres'
>> whos
      Name      Size      Bytes  Class
-----
a           1x1         102  cell array
c           1x21          42  char array
```

Figura XI-1. Ejemplo arreglo de caracteres en Matlab

En el ejemplo,  $c$  es un arreglo de caracteres, y  $a$  es un arreglo de celdas; como se observa al utilizar el comando WHOS,  $c$  tiene dimensiones [1 x21] y  $a$  tiene dimensiones [1 x 1], aunque su almacenamiento consume 2.4 veces más bytes que  $c$ .

Esta conversión, permite trabajar con arreglos de celdas de tamaño [Num. Filas x 1], y cada fila representa un renglón del archivo de texto de entrada. A su vez, el tener los datos en formato CELL ARRAY, permite usar la función REGEXP(), que busca dentro de un arreglo un conjunto de

caracteres dado, y devuelve un arreglo de celdas, con el numero de ocurrencias del conjunto buscado dentro del arreglo o una fila del mismo.

Usando el ejemplo anterior, se buscó el conjunto de caracteres 'ar' dentro del arreglo de celdas **a**, usando la función REGEXP(), como resultado se obtuvo un arreglo de celdas **p**, de dimensiones [1x2], donde el número 2, representa el numero de ocurrencias del conjunto dentro de **a**.

```
a =  
    'arreglo de caracteres'  
  
>> p=regexp(a, 'ar')  
  
p =  
    [1x2 double]
```

Figura XI-2. Ejemplo función REGEXP en Matlab

Al poder realizar este tipo de búsquedas dentro de un arreglo de caracteres, se procedió a buscar la existencia de los conjuntos de caracteres "HYDRO", "ECO", "GEO" y "CHEM", dentro del arreglo de datos de entrada; esta experiencia fue exitosa parcialmente, ya que luego de realizar una comprobación visual entre el conjunto y los datos de salida de la función, se detectó que sólo encontró algunas ocurrencias y que otras fueron pasadas por alto, así mismo, un gran número de registros de entrada no obtuvo ocurrencias de ninguno de los conjuntos de caracteres estudiados.

Ante este evento, se buscó la causa de estos errores de detección, y se encontró que la función REGEXP, es sensible a las mayúsculas; motivo por el cual se utilizó la función LOWER(), para convertir todos los caracteres de los registros del arreglo de entrada, a minúsculas. Esta solución, llevó a superar el inconveniente de los errores de detección.

Respecto al error de clasificación, se comprobó que efectivamente algunos registros no contenían ninguno de los conjuntos de caracteres usados. Esto llevó a la creación de familias de palabras o prefijos que estuvieran relacionadas, con el fin de hacer más flexible el criterio de búsqueda.

En este orden de ideas, se crearon cuatro grandes familias (categorías), cada una relacionada con uno de los conjuntos de búsqueda, usados anteriormente ("HIDRO", "ECO", "GEO" y "CHEM"), según se muestra en la siguiente tabla:

Tabla 24. Categorías y Elementos usados para la clasificación de los textos.

Categoría	Elementos	Categoría	Elementos	Categoría	Elementos	Categoría	Elementos
ECO		GEO		CHEM		HYDRO	
	veget		soil		metal		water
	organic		land		solute		river
	ecolog		mountain		isotop		lake
	microb		mine		ion		stream
	nutrient		field		acid		aqu
	bacter		lito		nitr		acuatic
	forest		topo		phosph		snow
	plant		physic		carbo		drain
	agric				chlor		catchment
					seleni		wet
					manganes		flow
					mercur		evap
					iron		precip
					potas		clima
					sodi		atmos
					sulfur		basin
					oxi		

El procedimiento para crear las familias consistió en determinar grandes grupos o categorías de palabras que estuvieran relacionadas con los parámetros de búsqueda, es así como al utilizar la información de las diferentes bases de datos, como por ejemplo Springer Link. Se obtuvo que para los 1491 resultados obtenidos, 491 estaba catalogados bajo el calificativo ECOLOGY, sugiriendo la necesidad de incluir una familia de datos relacionada a la ecología; de otra parte, 437 datos se caracterizaron en GEOSCIENCES, 1104 bajo EARTH & ENVIRONMENTAL SCIENCE, 180 en GEOCHEMISTRY y 173 en BIOCHEMISTRY. Esta clasificación *a priori*, realizada por Springer Link sirvió de guía para crear las familias de palabras que posteriormente serían representadas por clústeres de datos al crear el mapa auto organizado, utilizando las redes neuronales de Kohonen.

Una vez implementado en el programa, el algoritmo de búsqueda para cada uno de los elementos, se procedió a realizar una prueba de escritorio para verificar las bondades de esta nueva categorización; obteniéndose una mejora sustancial respecto a la aproximación anterior. Sin embargo, la matriz de salida, sólo presentaba arreglos vacíos donde no había ocurrencias y un arreglo de celdas [n x m], donde sí se presentaban. Este hecho, llevó a

considerar la creación de una aplicación que funcionara en base a valores booleanos, pero aún así, no constituiría la métrica necesaria para poder catalogar los datos de entrada.

Ante este evento, se propuso un sistema de clasificación no binario, que no sólo detectara la presencia o no de determinado elemento de una categoría en los datos de entrada, sino que asignara un valor positivo de pertenencia a una familia de categorías dada. De esta manera, se logró asignar a todas las filas del arreglo de entrada una magnitud vectorial en cada una de las dimensiones planteadas.

La métrica planteada determina el grado de pertenencia a una clase, según las siguientes consideraciones básicas:

1. Existen cuatro familias de palabras, una por cada dimensión vectorial.
2. Las familias o categorías están integradas por conjuntos de caracteres que generalmente se encuentran en palabras relacionadas con cada familia.
3. Dentro de una familia, cada conjunto de caracteres tiene un número de caracteres determinado [N].
4. Cada conjunto de caracteres puede o no encontrarse en un campo de la entrada de datos.
5. Un conjunto de caracteres de dimensión N tiene N subconjuntos de caracteres que lo forman.
6. Cada subconjunto de caracteres tiene un valor dimensional correspondiente al número de caracteres dentro del mismo.
7. El máximo valor dimensional para un conjunto de caracteres dado es la suma de los valores dimensionales de los subconjuntos que lo forman.
8. El valor dimensional de cualquier palabra respecto a una familia, se encuentra sumando los valores dimensionales de todos los conjuntos de caracteres, pertenecientes a dicha familia.

Para este caso de estudio, dado que los resultados de la búsqueda en ProQuest, fueron principalmente en inglés, con algunas ocurrencias en francés y alemán, se optó por no usar la raíz en español de las palabras. (Ej. HYDRO, no HIDRO)

A manera de ejemplo, se estudiará el valor dimensional de la partícula "*nutrient*", esta partícula tiene 8 caracteres (N), y posee 8 (N) subconjuntos de caracteres: n, nu, nut, nutr, nutri, nutrie, nutrien, y nutrient.

Conforme a las consideraciones 6 y 7, cada subconjunto tiene un valor dimensional, y el valor dimensional de “nutrient” está dado por la suma de los subconjuntos que lo forman; así según la tabla 25, el valor dimensional de “nutrient” es 36, y los valores de sus subconjuntos, aumentan desde 1 hasta 8.

**Tabla 25. Valor dimensional para la partícula NUTRIENT**

Elemento	Nutrient									
Número de caracteres	8									
Valor por combinación	1 (n)	2 (nu)	3 (nut)	4 (nutr)	5 (nutri)	6 (nutrie)	7 (nutrien)	8 (nutrient)	TOTAL	36

Es importante notar la importancia de la creación de familias de palabras relacionadas ya que flexibiliza la categorización realizada por el algoritmo, al no tratar simplemente la correspondencia booleana entre cada palabra del título y el nombre o palabra clave del clúster de datos. Estas familias de datos son modificables y se espera en siguientes ejercicios incluir la mayor cantidad posible de descriptores o palabras clave relacionadas a cada una de ellas, para que el proceso de categorización y de creación de los clústeres pueda hacerse con documentos en diferentes idiomas.

De manera análoga se realizó una tabla (26), que muestra los valores dimensionales de todos los miembros de cada una de las categorías, o familias planteadas, y una que muestra los máximos valores posibles para cada una de las familias (Tabla 27).

**Tabla 26. Valores dimensionales máximos si se encierran los caracteres de la entrada de datos en cualquier posición de una palabra**

Valor dimensional máximo si se encuentran los caracteres de la entrada de datos en cualquier posición de una palabra													
Categorías	miembros	Numero letras	Letra 1	Letra 2	Letra 3	Letra 4	Letra 5	Letra 6	Letra 7	Letra 8	Letra 9	TOTAL	Valor por Letra encontrada
<b>ECO</b>		3	1	2	3							6	0.1667
	veget	5	1	2	3	4	5					15	0.0667
	organic	7	1	2	3	4	5	6	7			28	0.0357
	ecolog	6	1	2	3	4	5	6				21	0.0476
	microb	6	1	2	3	4	5	6				21	0.0476
	nutrient	8	1	2	3	4	5	6	7	8		36	0.0278
	bacter	6	1	2	3	4	5	6				21	0.0476
	forest	6	1	2	3	4	5	6				21	0.0476
	plant	5	1	2	3	4	5					15	0.0667
	agric	5	1	2	3	4	5					15	0.0667
<b>GEO</b>		3	1	2	3							6	0.1667
	soil	4	1	2	3	4						10	0.1000
	land	4	1	2	3	4						10	0.1000
	mountain	8	1	2	3	4	5	6	7	8		36	0.0278
	mine	4	1	2	3	4						10	0.1000
	field	5	1	2	3	4	5					15	0.0667
	lito	4	1	2	3	4						10	0.1000
	topo	4	1	2	3	4						10	0.1000
	physic	6	1	2	3	4	5	6				21	0.0476
<b>CHEM</b>		4	1	2	3	4						10	0.1000
	metal	5	1	2	3	4	5					15	0.0667
	solute	6	1	2	3	4	5	6				21	0.0476
	isotop	6	1	2	3	4	5	6				21	0.0476
	ion	3	1	2	3							6	0.1667
	acid	4	1	2	3	4						10	0.1000
	nitr	4	1	2	3	4						10	0.1000
	phosph	6	1	2	3	4	5	6				21	0.0476
	carbo	5	1	2	3	4	5					15	0.0667
	chlor	5	1	2	3	4	5					15	0.0667
	seleni	6	1	2	3	4	5	6				21	0.0476
	manganes	8	1	2	3	4	5	6	7	8		36	0.0278
	mercur	6	1	2	3	4	5	6				21	0.0476
	iron	4	1	2	3	4						10	0.1000
	potas	5	1	2	3	4	5					15	0.0667
	sodi	4	1	2	3	4						10	0.1000
	sulfur	6	1	2	3	4	5	6				21	0.0476

	oxi	3	1	2	3							6	0.1667
<b>HYDRO</b>		4	1	2	3	4						10	0.1000
	water	5	1	2	3	4	5					15	0.0667
	river	5	1	2	3	4	5					15	0.0667
	lake	4	1	2	3	4						10	0.1000
	stream	6	1	2	3	4	5	6				21	0.0476
	aqu	3	1	2	3							6	0.1667
	acuatic	7	1	2	3	4	5	6	7			28	0.0357
	snow	4	1	2	3	4						10	0.1000
	drain	5	1	2	3	4	5					15	0.0667
	catchment	9	1	2	3	4	5	6	7	8	9	45	0.0222
	wet	3	1	2	3							6	0.1667
	flow	4	1	2	3	4						10	0.1000
	evap	4	1	2	3	4						10	0.1000
	precip	6	1	2	3	4	5	6				21	0.0476
	clima	5	1	2	3	4	5					15	0.0667
	atmos	5	1	2	3	4	5					15	0.0667
	basin	5	1	2	3	4	5					15	0.0667

Tabla 27 Valores dimensionales máximos posibles

Dimensión	Valor Máximo posible
<b>ECO</b>	199
<b>GEO</b>	128
<b>CHEM</b>	284
<b>HIDRO</b>	267

De la tabla 26, se aprecia que la palabra con el mayor número de caracteres (9), es “catchment” que pertenece a la familia HYDRO. Este hecho, determinó el máximo número de columnas a usar, para calcular los valores dimensionales de cada palabra.

Inicialmente se realizó el cálculo dimensional para cada familia de elementos, según las consideraciones básicas, expuestas anteriormente. De esta forma, se procedió a crear una matriz de dimensiones [Num.Filas x 9], para cada uno de los elementos de cada familia, donde en cada columna se calcula el valor dimensional para un subconjunto de caracteres, al verificar la ocurrencia de cada subconjunto de caracteres (cada columna), en cada una de las filas de los datos de entrada. Luego se procedió a crear una matriz auxiliar de UNOS, de las mismas dimensiones, y se utilizó el comando IEMPTY para determinar que celdas del arreglo, no contenían datos. Posteriormente, se restó de la matriz de unos, la matriz IEMPTY, así se creó una matriz ISNOTEMPTY; esta última matriz, muestra en cada una de sus posiciones, valores de 1 si hay algún elemento o de 0 si está vacía; de esta manera se obtuvo una matriz de dimensiones [Num.Filas x 9], que muestra para cada fila (título de la publicación), si tiene o determinado subconjunto de caracteres.

Sin embargo, en la métrica propuesta, cada columna tiene un valor dimensional, correspondiente al número de la misma (1 a 9), esto implicó que la matriz ISNOTEMPTY fuera modificada, de manera que las ocurrencias de cada columna fueran multiplicadas por el índice de la misma, de esta manera si se encuentra un valor de 1 en la columna 4, por ejemplo, este 1 se convierte en 4; según la métrica propuesta esto implicaría que se encontró en alguna parte del título analizado un subconjunto idéntico a la columna 4. Finalmente, se suman todos los valores dimensionales de cada columna (subconjunto), y se asigna este valor, como valor dimensional total de la fila analizada, para ese elemento determinado de una familia.

Este proceso se repite para cada uno de los elementos de una familia dada, y se consolidan los valores dimensionales en columnas, donde el número de columnas está dado por la cantidad de elementos que contenga una familia determinada. Para el caso de la familia ECO, se obtienen 10 columnas, el número de filas de esta matriz consolidada es igual a Num.Filas, al igual que en las anteriores matrices.

Para los primeros cuatro registros de entrada analizados en la familia ECO, se obtiene la matriz de salida (AUXBIO), que se muestra en la Figura XI.3

```

tab =

A laboratory study on the development of a biological pollution control system for contaminated soils
A model of carbon and nitrogen dynamics for temperate grasslands and forests
A unified conceptual model of coupled reactive transport involving organic and inorganic species in groundwater using
Alcohol-enhanced, passive bioreactors for treatment of acid mine drainage

auxbio =

  6   3   3   1   3   1   1   6   1   1
  1   0   3   1   6   1   1  21   1   1
  1   3  28   3   1   1   1   1   3   1
  6   3   3   1   3   1   1   6   1   3

```

**Figura XI-3. Matriz AUXBIO, con los valores dimensionales para cada una de las palabras de la familia ECO, de los primeros cuatro registros de entrada de datos.**

Como se ve en la Figura XI-3, la matriz AUXBIO cuenta con 10 columnas, una por cada elemento de la familia ECO; si se observa con atención la columna número 8, correspondiente al elemento "forest", se aprecia un valor dimensional de 21 en la segunda fila; al leer el título del segundo campo de entrada, se encuentra el elemento "forest" en la última palabra del mismo (forests).

Una vez obtenida la matriz auxiliar (en este caso AUXBIO), se tienen los valores para determinar qué tanto pertenece cierto elemento (título) a cada dimensión; se propone utilizar cualquiera de los siguientes dos métodos:

1. Utilizar el valor máximo de cada fila, como valor dimensional.
2. Utilizar la suma de los valores de la fila, como valor dimensional.

Para este trabajo, se escogió la segunda opción, ya que así se valora si una frase determinada tiene dos (2) o más palabras que pertenecen a la familia de datos analizada.

Este proceso para determinar los valores dimensionales, se repite para cada dimensión vectorial (Figura XI- 4), y se consolidan los datos obtenidos en una matriz de pesos, que tiene dimensiones [Num.filas x Num. Familias]; de esta manera a cada elemento de la matriz de entrada, le corresponde un valor dimensional por cada número de familias analizadas (Figura XI- 5).

A laboratory study on the development of a biological pollution control system for contaminated soils  
 A model of carbon and nitrogen dynamics for temperate grasslands and forests  
 A unified conceptual model of coupled reactive transport involving organic and inorganic species in groundwater  
 Alcohol-enhanced, passive bioreactors for treatment of acid mine drainage

auxbio =

6	3	3	1	3	1	1	6	1	1
1	0	3	1	6	1	1	21	1	1
1	3	28	3	1	1	1	1	3	1
6	3	3	1	3	1	1	6	1	3

auxhydro =

1	0	1	3	3	1	1	1	1	3	0	1	3	1	1	3
0	0	1	3	3	1	1	1	1	3	0	1	1	1	1	3
1	15	3	1	1	1	3	1	1	1	1	1	1	3	1	3
1	0	1	1	1	1	3	1	15	1	0	1	1	1	1	6

auxchem =

1	3	3	1	6	1	1	1	3	1	1	1	3	1	3	3	1	1
1	1	1	1	1	1	10	1	15	1	1	1	1	1	1	1	1	1
3	1	1	3	1	3	3	1	1	3	3	1	1	1	3	1	1	1
1	3	1	1	3	10	1	1	1	1	1	1	3	1	1	1	1	1

auxgeo =

1	10	3	1	6	1	1	3
3	1	10	3	3	1	1	1
1	1	1	3	1	6	3	1
3	1	1	1	10	1	1	3

Figura XI-4. Matrices auxiliares con valores dimensionales para cada elemento de las familias ECO, HYDRO, CHEM y GEO. Primeros 4 elementos analizados

Al utilizar el método 2 para determinar los pesos, se obtiene una matriz como la que se observa en la Figura XI. 5. En esta Figura, las columnas 1 a 4, corresponden a las dimensiones ECO, GEO, CHEM e HIDRO, respectivamente. La matriz de pesos, representa los valores de entrada de datos, para iniciar el proceso de elaboración del SOM, usando el algoritmo de Kohonen.

pesos =			
26	26	35	24
36	23	41	21
43	17	32	38
28	21	33	35

XI-5. Matriz de pesos para los cuatro primeros elementos analizados, cada columna representa una dimensión vectorial.

Se mencionó en el capítulo “Características de la herramienta desarrollada para la creación de SOMs”, se utilizó el estándar de entrada de datos del paquete SOM PAK, creado por Universidad de Helsinki en 1992 (ver tabla 17); este formato, es fácilmente obtenible si se copia la matriz de pesos desde Matlab a Excel, y luego se utiliza la función ESPACIOS(), que

quita todos los espacios del texto excepto los espacios individuales entre palabras; posteriormente, estos datos se copian y pegan como valores, en la tercera fila de la primera columna, de una nueva hoja de cálculo; donde la primera fila se llena con el número de dimensiones y la segunda con los caracteres #n, seguidos de los nombres de las mismas, separados por un espacio.

Esto concluye la etapa de pretratamiento de la información.

## 2. Proceso de entrada de datos

Como se mostró en el capítulo VIII, una vez ingresada la matriz de datos de entrada, es necesario un proceso de normalización, para que estos valores se encuentren entre cero y uno, o menos uno y uno. En este caso de estudio, se optó por normalizar entre cero y uno.

Los datos de entrada, están dados por cuatro (4) dimensiones; en la Figura XI-6, se aprecian cuatro graficas generadas con esta información así: en la parte superior izquierda, se presenta un diagrama de dispersión tridimensional, con ejes ECO, GEO y CHEM, la barra de color, determina la magnitud de los puntos en la dimensión HIDRO (cuarta dimensión).

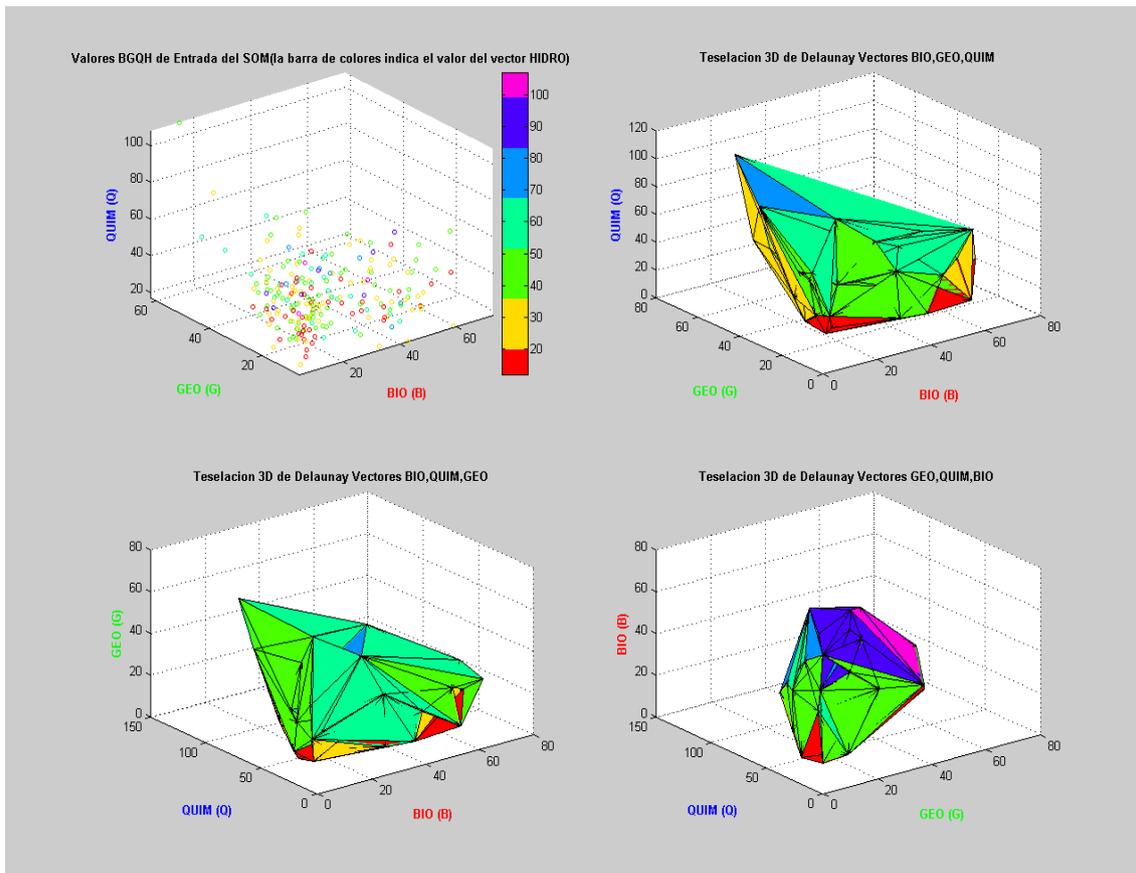


Figura XI-6 Graficas relacionadas a la entrada de datos. Se muestra un diagrama de dispersión tridimensional y tres Tesselaciones tridimensionales de Delaunay.

Las otras tres gráficas, corresponden a teselaciones tridimensionales de Delaunay, un concepto relacionado con los diagramas de Voronoi, expuestos en el Marco Teorico.

Posteriormente, y como se hizo en el ejemplo de agrupación de los vectores RGB, se procedió a realizar histogramas de ocurrencias y diagramas de dispersión de cada una de las variables.

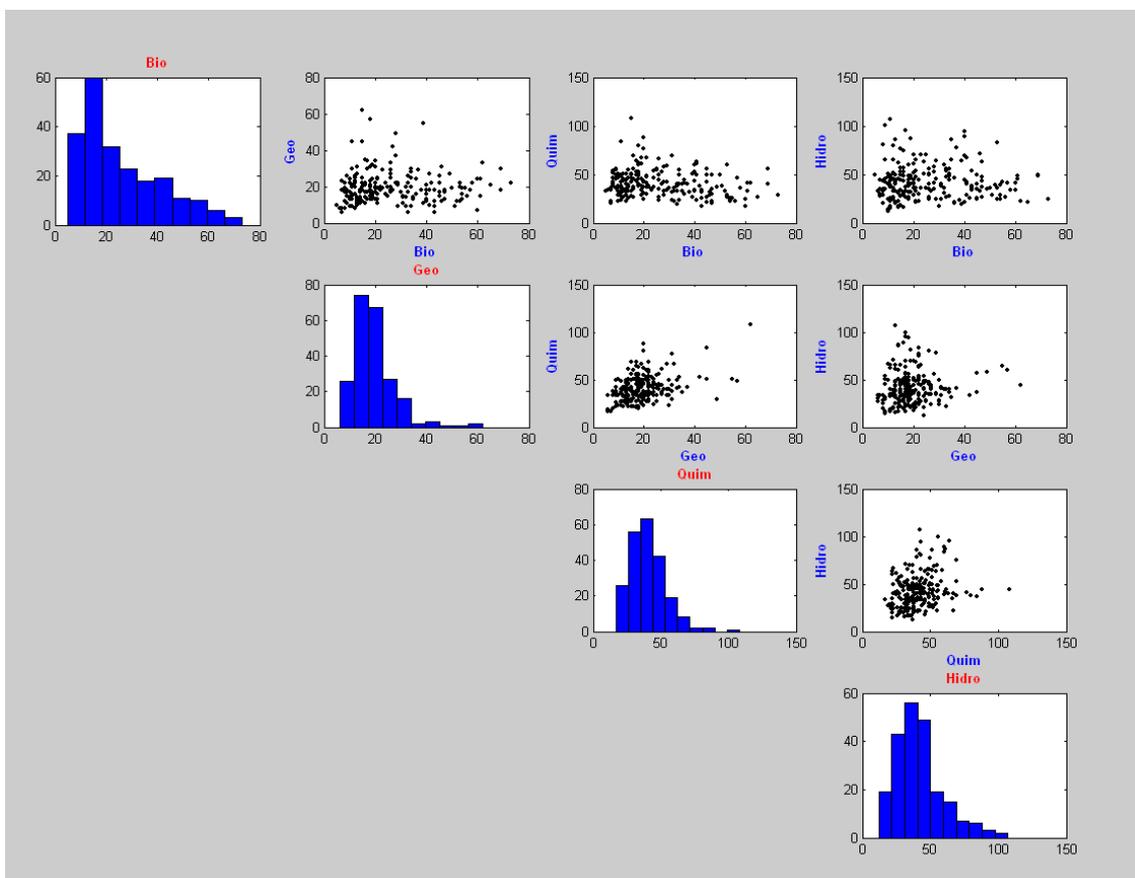


Figura XI-7 Histogramas y Diagramas de Dispersión de las cuatro dimensiones utilizadas.

La Figura anterior, más la información estadística obtenida del programa (Figura XI-8), facilitan la interpretación de los datos de entrada, en donde las variables 3 y 4 (CHEM e HIDRO), presentan los mayores valores mínimos, medios y promedio, de las cuatro variables analizadas; respecto a los valores mínimos, es interesante notar que se esperaba que estos valores fueran cercanos al número de elementos de cada familia, como ocurrió efectivamente (5/10 para ECO, 6/9 para GEO, 17/18 para CHEM y 12/17 para HIDRO), estos valores mínimos implican que sólo se encontró la primera letra de cada uno de los integrantes de cada familia, en el mejor de los casos, siendo el vector analizado perteneciente a otra familia sin lugar a dudas.

Acercas de los valores medios, era de esperar que fueran mayores en las familias 3 y 4, ya que éstas tienen un mayor número de palabras o elementos, esto no implica necesariamente que

en los títulos de publicaciones utilizados como datos de entrada, se presente una predominancia de obras que incluyen aspectos químicos e hidrológicos; a su vez, los valores máximos de clasificación obtenidos, indican que aunque las variables 3 y 4 presentan los mayores valores máximos, si se obtienen los valores máximos relativos al valor máximo posible por familia (Tabla 26), es el grupo GEO, el que presenta mayores valores (48%), las otras tres dimensiones tienen valores máximos relativos similares (36%,38% y 40%); esto implica que la mejor clasificación de un vector de entrada, se obtuvo en la familia GEO; en las otras tres familias, se esperaba que los valores máximos relativos no fueran muy altos, dado que la posibilidad de encontrar un título con tantas palabras es considerablemente más baja.

Nombre	Min	Media	Max	StDev	Faltantes	Unicos
Variable1	5	25.8539	73	15.5848	0 (0%)	58 (26%)
Variable2	6	19.4977	62	8.50546	0 (0%)	37 (17%)
Variable3	17	40.8721	108	13.5061	0 (0%)	54 (25%)
Variable4	12	41.9178	107	17.6368	0 (0%)	68 (31%)

Figura XI-8 Información estadística de los datos de entrada

Una vez realizadas estas gráficas auxiliares, se trata la información según lo expuesto en el capítulo VIII; se procede a normalizarla, se crean matrices de pesos iniciales del SOM y se continúa con el proceso de entrenamiento. La Figura XI-9, muestra el diagrama de dispersión de las variables de entrada, según las consideraciones expuestas anteriormente, y dos diagramas de inicialización de pesos para el SOM, el del centro representa la inicialización lineal y el de la derecha la inicialización aleatoria.

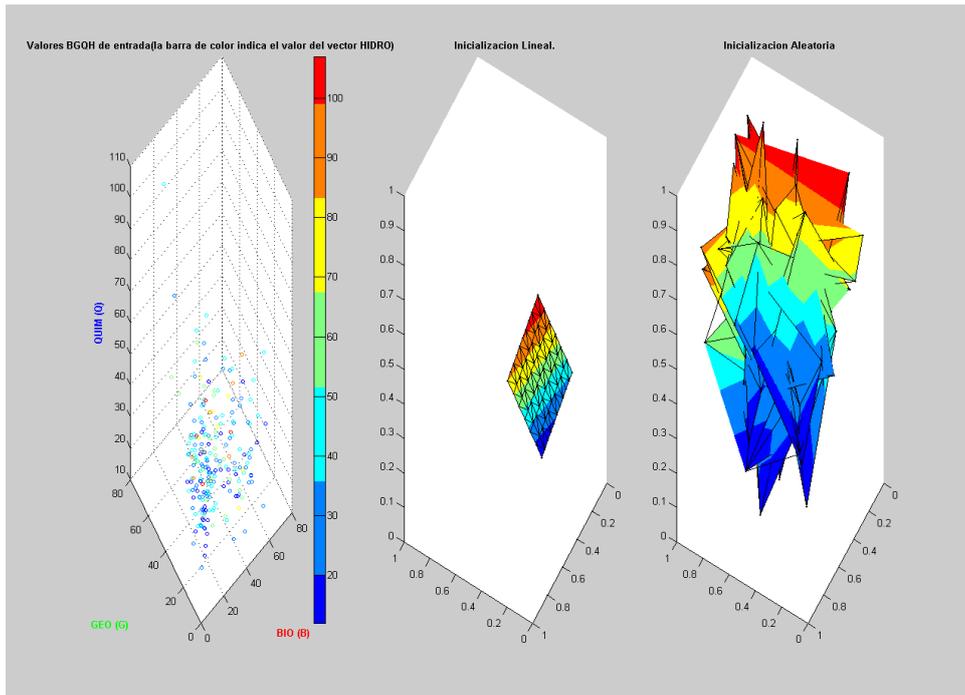


Figura XI-9 Inicialización de pesos para el SOM, usando procesos lineales (centro) y aleatorios (derecha)

Posteriormente, se procedió a entrenar los SOMs usando el algoritmo BATCH y el SECUENCIAL; a continuación se presentan los mapas obtenidos para la decima corrida de entrenamiento, y la ubicación de la MUC para su última iteración:

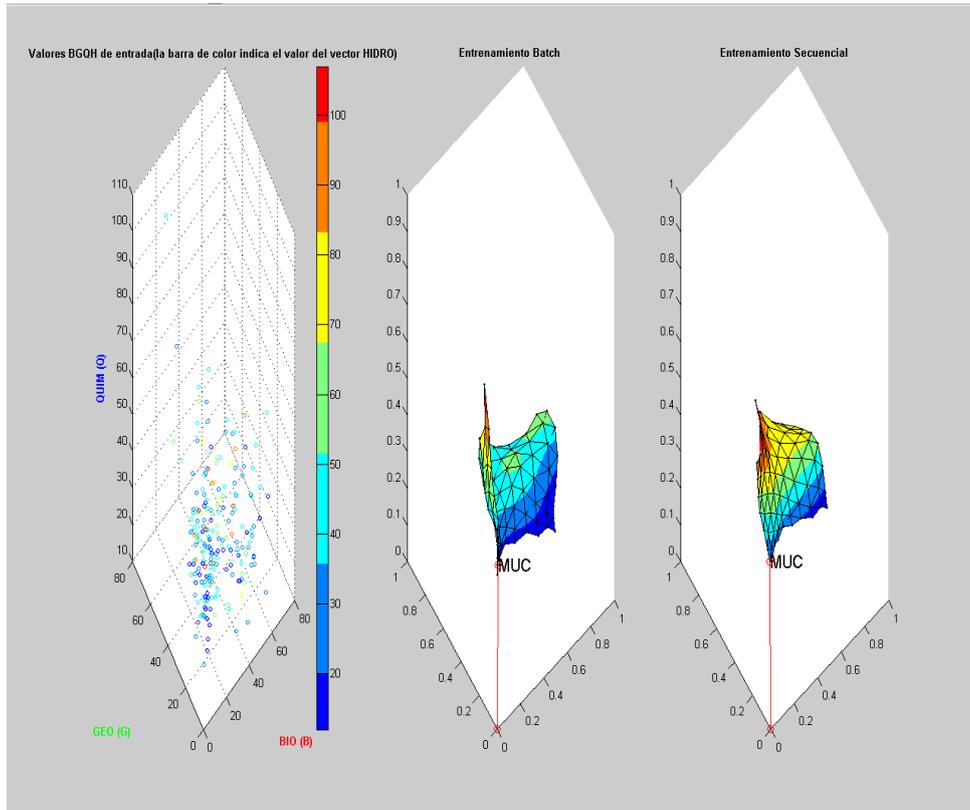


Figura XI-10. Mapas de entrenamiento utilizando el algoritmo BATCH (centro) y el SECUENCIAL (derecha). Valores de entrada (izquierda)

A continuación, se procede a realizar los SOMs de Kohonen para cada una de las variables y la grafica de Matriz - U, para representar en un espacio de dos dimensiones, los datos de entrada cuatro dimensionales. En la Figura XI-11 se muestran cuatro gráficas de clústeres, uno para cada familia y un diagrama de Pie; en la Figura XI-12, se presenta la Matriz - U, que se analizará posteriormente.

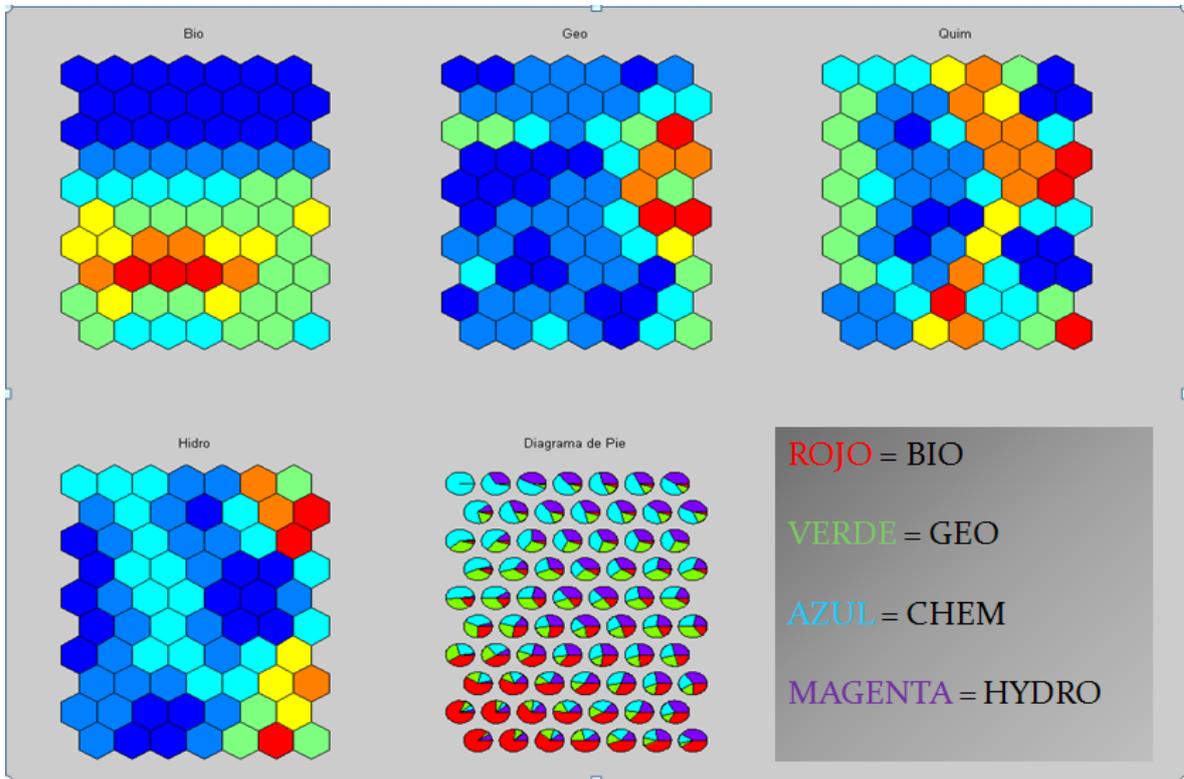


Figura XI-11 SOMs y Diagramas de Ocurrencias para el caso de estudio.

Como se mencionó anteriormente, la Figura XI-11, muestra el diagrama de Pie, para los valores relativos de las magnitudes vectoriales usadas (ECO, GEO, CHEM e HIDRO), según el siguiente código de colores:

1. ROJO: Corresponde al valor de la dimensión o familia ECO
2. VERDE: Corresponde al valor de la dimensión o familia GEO
3. AZUL: Corresponde al valor de la dimensión o familia CHEM
4. MAGENTA: Corresponde al valor de la dimensión o familia HIDRO

## B. Análisis de patentes

Como se mencionó anteriormente el estudio de las patentes relacionadas a las clases tecnológicas asociadas al estudio de los ciclos biogeoquímicos e hidrológico se encuentra en el Tomo 2.

## XII. Resultados y Discusiones

Como se puede observar en la Figura IX-12 (Matriz – U generada), las mayores distancias entre celdas continuas se representan con colores cálidos, correspondiendo las zonas con estos valores a los bordes de los clústeres, y los colores fríos corresponden a la ubicación de los clústeres en sí, ya que las distancias euclidianas entre las celdas es baja; al comparar celda a celda la Matriz-U con cualquiera de los diagramas de valores vectoriales relativos, se obtiene que:

- El clúster correspondiente a las publicaciones cuyo valor dimensional predominante se encuentra en la familia CHEM inicia en la zona superior izquierda del mapa y limita hacia el sur, con el clúster ECO, en las celdas amarillas ubicadas en la parte inferior de la matriz.
- El clúster ECO, correspondiente a las publicaciones cuyo valor dimensional predominante se encuentra relacionado a altos valores de la familia ECO, se encuentra abajo del clúster CHEM, y está ubicado en la parte inferior izquierda de la matriz, limitando a la derecha con celdas en color amarillo que marcan el inicio del clúster HIDRO.
- El clúster HIDRO, se encuentra en la esquina inferior derecha, al lado del clúster ECO, y se extiende al norte hasta un conjunto de celdas naranjas que rodean el clúster GEO, también se encuentra en la esquina superior derecha, como extensión del clúster inferior.
- El clúster GEO, está ubicado dentro de un conjunto de celdas naranjas, ubicadas en las últimas columnas, al oriente del mapa, está rodeado por el clúster HIDRO, por el clúster CHEM al occidente y una zona con valores altos de la familia ECO al sur occidente.

Tabla 28 Familia predominante en cada celda del mapa auto organizado

FAMILIA PREDOMINANTE EN CADA CELDA DEL MAPA AUTO ORGANIZADO									
		Columna No.							
		1	2	3	4	5	6	7	8
Fila No	1	CHEM	CHEM	CHEM	CHEM	CHEM	CHEM	CHEM	HYDRO
	2	CHEM	CHEM	HYDRO	HYDRO	CHEM	CHEM	HYDRO	HYDRO
	3	CHEM	CHEM	CHEM	HYDRO	HYDRO	HYDRO	HYDRO	HYDRO
	4	CHEM	CHEM	HYDRO	HYDRO	HYDRO	HYDRO	HYDRO	HYDRO
	5	GEO	GEO	HYDRO	HYDRO	HYDRO	HYDRO	GEO	GEO
	6	ECO	ECO	ECO	ECO	HYDRO	HYDRO	GEO	GEO
	7	ECO	ECO	ECO	ECO	ECO	ECO	ECO	GEO
	8	ECO	ECO	ECO	ECO	ECO	ECO	HYDRO	HYDRO
	9	ECO	ECO	ECO	ECO	ECO	ECO	ECO	HYDRO

La anterior tabla muestra la familia predominante en cada una de las 72 celdas del mapa auto organizado generado. Utilizando la misma convención de colores de la figura XI -11 para facilitar la visualización. Como se observa los clústeres están ubicados en las zonas correspondientes a la familia predominante. Las variaciones entre la figura XI-12 y la tabla anterior, se deben a celdas con valores dimensionales similares entre 2 o más familias, específicamente el caso del pequeño clúster GEO (verde) que se aprecia en la fila 5 columnas 1 y 2 de la tabla, y en los límites entre el clúster CHEM e HYDRO en la parte superior de la tabla.

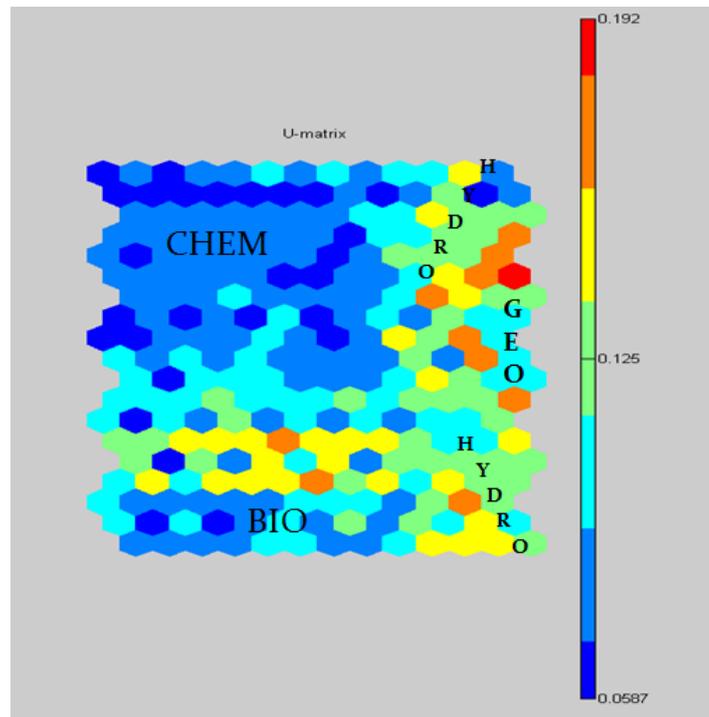


Figura XII-1 Matriz - U para el caso de estudio.

En la Matriz - U generada, los colores cálidos representan los bordes de los clústeres, y los fríos representan los clústeres, según las gráficas de la Figura XI-11, los objetos que obtuvieron los mayores valores en la familia ECO, se encuentran en la parte baja (fila 1, columna 2. Figura XI – 11), los que obtuvieron los mayores valores de la familia GEO, se encuentran en la zona media de las últimas columnas de celdas (fila 1, columna 3 Figura XI – 11), el grupo CHEM, obtuvo los mayores valores en la esquina superior derecha (fila 2, columna 1. Figura XI – 11), y el grupo HIDRO, en las esquinas superior derecha e inferior derecha. Estos comportamientos facilitan la interpretación de la Matriz - U, en donde se presentan bordes de clústeres, representados por amarillo en la parte superior de la antepenúltima fila, zona en la que se presentaron los mayores valores para la familia ECO; un clúster se encuentra en la parte superior derecha, delimitado por 2 celdas amarillas y un conjunto de celdas naranjas, en esta zona, se encuentran elementos con valores muy altos de las dimensiones HIDRO y CHEM; la familia GEO se encuentra en una zona demarcada por el conjunto de celdas naranjas que delimitó la clasificación anterior, y cuatro celdas naranjas, que delimitan un clúster de celdas cian y verdes; finalmente otro clúster de celdas verdes se observa en la esquina inferior derecha cerca de la zona donde se presentan valores altos de la familia HIDRO. La gran zona azul de la esquina superior izquierda, limita con las cuatro zonas anteriores y sus valores se hacen más pequeños, a medida que se alejan de estos límites; esto significa, que dentro de este clúster se encuentran los registros de documentos que tratan al menos dos familias, y cuyos valores dimensionales son similares.

En la Figura XI-13, los tamaños de los hexágonos blancos, determinan cuantas veces una celda dada fue la MUC. Según se explicó en el capítulo VIII, la determinación de la MUC, es fundamental para realizar el proceso de auto organización, la presencia de varias MUC es indicador de una serie de valores de entrada heterogénea, y su dispersión espacialmente en la Matriz - U, indicador de un buen proceso de auto organización en el mapa final.

**Tabla 29 Número de ocurrencias por celda del SOM generado**

		No. de ocurrencias por celda del SOM generado.							
		Columna No.							
		1	2	3	4	5	6	7	8
Fila No	1	11	3	3	6	2	3	6	2
	2	7	3	2	4	3	3	1	1
	3	0	4	7	3	2	1	4	2
	4	3	2	4	3	2	3	7	1
	5	2	2	1	4	6	2	3	2
	6	6	3	1	3	5	7	0	3

7	3	0	1	4	1	1	5	1
8	1	0	5	1	3	0	2	1
9	10	5	0	5	1	2	5	4

La tabla 29 muestra utilizando el mismo código de colores usado previamente, el número de títulos de la base de datos, clasificado en cada celda del SOM generado. El total de ocurrencias es 219 al igual que el número de registros clasificado. Las celdas con valores de CERO, representan celdas que simplemente no tuvieron una MUC, para el vector de entradas dado.

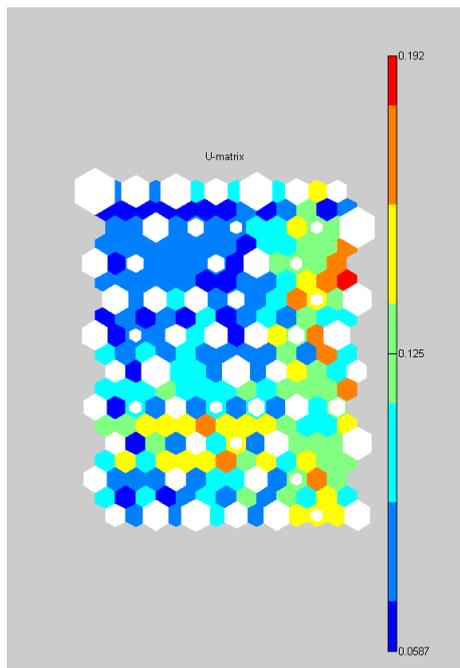


Figura XII-2 Histograma de ocurrencias en la Matriz - U. Grandes hexágonos blancos representan mayores ocurrencias como MUC para una celda.

La Figura IX-14, muestra la matriz de distancia tridimensional, como complemento a la información suministrada por la Matriz - U, en esta Figura, se pueden apreciar claramente cuatro (4) picos de valores altos, cada uno representando un dimensión, y zonas con colores fríos, como el azul, en el que la variabilidad entre los registros de entrada es mínima, como lo es su distancia euclidiana.

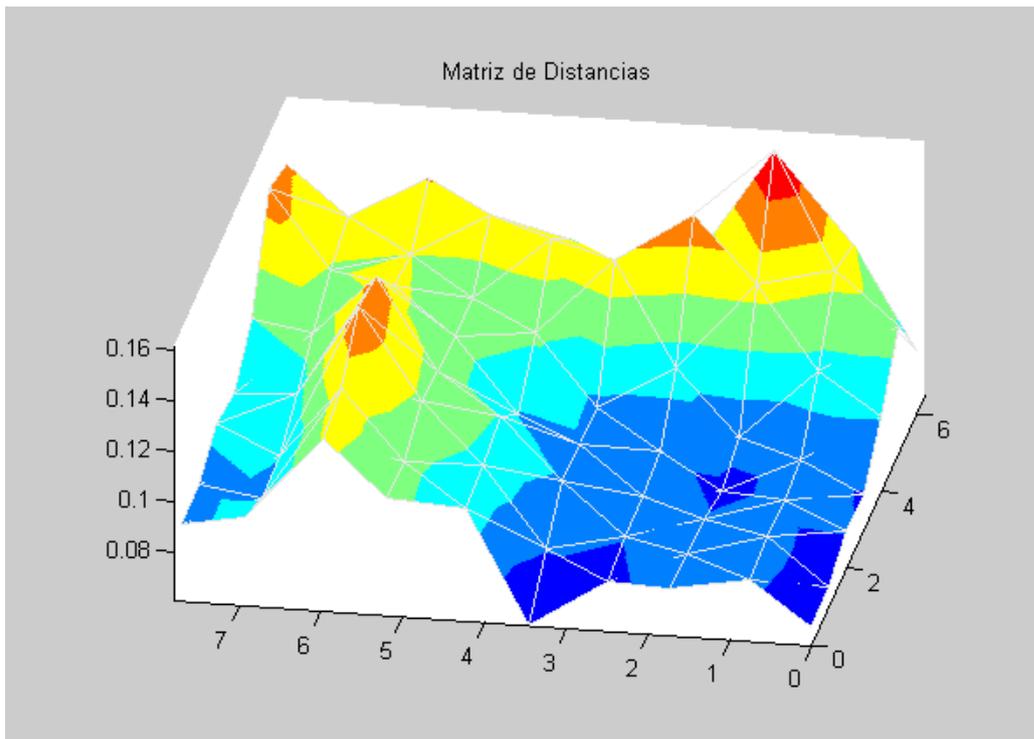


Figura XII-3. Matriz de distancias vectoriales para los vectores organizados por el SOM

La Tabla 30 muestra el porcentaje de ocurrencias en cada una de las 72 celdas del SOM generado. Se aprecian valores superiores al 4% en la esquina superior izquierda e inferior izquierda, correspondientes a las celdas con 11 y 10 ocurrencias respectivamente. En el caso de la celda del clúster CHEM, corresponde a los títulos cuyos valores dimensionales en cada una de las cuatro familias es bajo, sugiriendo para futuros desarrollos ampliar el número de palabras asociadas a cada familia, para mejorar así la clasificación. Por el contrario la celda del clúster ECO-BIO, representa 10 títulos que fueron catalogados principalmente en esta categoría.

Tabla 30 Porcentaje de ocurrencias por celda del SOM generado

Porcentaje de ocurrencias por celda del SOM generado									
		Columna No.							
		1	2	3	4	5	6	7	8
Fila No	1	5.023	1.370	1.370	2.740	0.913	1.370	2.740	0.913
	2	3.196	1.370	0.913	1.826	1.370	1.370	0.457	0.457
	3	0.000	1.826	3.196	1.370	0.913	0.457	1.826	0.913
	4	1.370	0.913	1.826	1.370	0.913	1.370	3.196	0.457
	5	0.913	0.913	0.457	1.826	2.740	0.913	1.370	0.913
	6	2.740	1.370	0.457	1.370	2.283	3.196	0.000	1.370
	7	1.370	0.000	0.457	1.826	0.457	0.457	2.283	0.457
	8	0.457	0.000	2.283	0.457	1.370	0.000	0.913	0.457
	9	4.566	2.283	0.000	2.283	0.457	0.913	2.283	1.826

Finalmente, la tabla 31 muestra la distribución de celdas del SOM entre las cuatro familias o clústeres creados. Se observa un predominio de las familias ECO e HYDRO con cerca de un tercio de las celdas asignadas para cada uno, seguido por CHEM y un pequeño clúster de la familia GEO compuesto por 7 celdas (9.7 %).

**Tabla 31 Distribución de celdas dentro de la clasificación obtenida por el SOM**

<b>FAMILIA</b>	<b>Número de Celdas</b>	<b>Porcentaje</b>
<b>ECO</b>	24	33.33
<b>GEO</b>	7	9.72
<b>CHEM</b>	16	22.22
<b>HYDRO</b>	25	34.72
<b>TOTAL</b>	72	100

La tabla 32 muestra el porcentaje de títulos catalogados en cada familia o clúster, usando el SOM generado, como se observa los clústeres ECO, CHEM e HYDRO aportan más del 30 % de los títulos clasificados cada uno y el clúster GEO sólo clasificó el 5.94 % de los mismos. Esto indica una tendencia similar en cuanto a la temática de las obras científicas y técnicas publicadas, con casi iguales porcentajes de categorización.

De otra parte las tablas anteriores muestran que la clasificación de obras en estos 3 clústeres es efectiva y por lo tanto valida la importancia de cada una de estas familias; sin embargo, al encontrar valores inferiores al 10 % en la distribución total de celdas del SOM y en el porcentaje de títulos clasificados en la familia GEO, surge el interrogante de la relevancia de esta categoría como clúster principal. Se sugiere para siguientes desarrollos utilizar GEO como subcategorías de otra familia, probablemente ECO por la similitud en sus temáticas.

**Tabla 32 Porcentaje de títulos catalogados en cada familia usando el SOM**

<b>FAMILIA</b>	<b>PORCENTAJE</b>
<b>ECO</b>	30.14
<b>GEO</b>	5.94
<b>CHEM</b>	30.14
<b>HYDRO</b>	33.79

### **XIII. Conclusiones y comentarios finales**

En este capítulo se desarrollan las conclusiones y comentarios finales del presente trabajo investigativo. Esta investigación busca ser un primer aporte a nivel nacional, en la utilización de

mapas auto organizados de Kohonen, para la visualización de clústeres de información no numérica, específicamente la relacionada con títulos de publicaciones científicas indexadas en bases de datos especializadas.

De otra parte, esta investigación relaciona tres áreas aparentemente disimiles, como la Vigilancia Tecnológica, la Biogeoquímica y la Inteligencia Artificial, para desarrollar un producto único, que suministra a los encargados de tomar decisiones relacionadas a la investigación y desarrollo de productos afines a los ciclos biogeoquímicos e hidrológico, información agrupada según cuatro áreas de interés (Ecología-Biología, Geología, Química o Hidrología), de las últimas publicaciones científicas disponibles en bases de datos especializadas. Esta información se clasifica y agrupa utilizando el algoritmo de las redes neuronales de Kohonen, y posteriormente se visualiza usando mapas auto organizados.

De esta forma, el presente trabajo se presta para realizar tanto conclusiones generales de la investigación, como conclusiones acerca de la Vigilancia Tecnológica, el algoritmo de Kohonen y la relación de estos elementos con los ciclos biogeoquímicos, desde el trabajo investigativo.

### **A. Conclusiones generales de la investigación**

Respecto a las conclusiones generales, esta investigación suministra una primera aproximación, desde la ingeniería y las ciencias de la tierra, al estudio de los ciclos biogeoquímicos e hidrológico, así como indica la importancia de estudiar y comprender el sistema terrestre.

Adicionalmente, el presente trabajo aporta elementos relacionados con la importancia de realizar actividades de Vigilancia Tecnológica, e incorporar esta herramienta en el proceso de toma de decisiones.

Sobre la clasificación de grandes volúmenes de datos, el presente trabajo realizó un aporte nacional, al implementar el algoritmo de Kohonen, para agrupar información no numérica y su posterior visualización usando mapas auto organizados.

Un aporte adicional es la implementación de un sistema de clasificación dimensional para determinar el grado de pertenencia de una oración (título) a una determinada clase o familia. Generalmente esta clasificación es booleana y solo determina existencia o no existencia de una palabra.

## **B. Conclusiones y comentarios finales desde el trabajo investigativo respecto a la Vigilancia Tecnológica.**

Desde el trabajo investigativo realizado, se concluye que en relación a la Vigilancia Tecnológica, se optó por trabajar la Vigilancia Tecnológica Científica, sub rama que involucra el estudio de patentes y de publicaciones científicas y técnicas. En el caso de estudio, estas patentes y publicaciones están relacionadas con los ciclos biogeoquímicos y el ciclo hidrológico. (Ver tomo 2 para mayor información sobre patentes asociadas)

El estudio de patentes se realizó en base a información de los últimos cinco años, relacionada con los campos tecnológicos asociados al estudio de los ciclos biogeoquímicos e hidrológico. Esta información se obtuvo de la Oficina de Patentes y Marcas de Estados Unidos, e incluye catorce campos tecnológicos y el número de patentes realizadas anualmente por cada país, en cada campo. De este estudio se desprenden las siguientes conclusiones para el período 2003-2007:

- Los países con mayor número de patentes realizadas en los campos tecnológicos relacionados con el tema de estudio son Estados Unidos, Japón y Alemania, respectivamente.
- Zonalmente, los líderes en desarrollo de patentes en los campos tecnológicos estudiados son:
  - Estados Unidos (América del Norte)
  - Japón (Asia)
  - Alemania (Europa)
  - Israel (Lejano y Medio Oriente)
  - Suráfrica (África)
  - Australia (Oceanía)
  - México (Centro América)
  - Brasil (América del Sur)
- El número de registros de patentes en Norteamérica, es tres veces más grande que el valor registrado para Europa y superior en un tercio al número de patentes desarrolladas en Asia.
- La actividad de patentes en Centro y Sur América (incluyendo México), es levemente superior a la registrada en África, y una decima parte de la registrada en el Lejano y Medio Oriente, para las clases tecnológicas analizadas.

- Del total de treinta patentes registradas en Centro y Sur América, para las clases tecnológicas analizadas, sólo UNA patente fue desarrollada en Colombia y pertenece a la clase 374 (Pruebas y mediciones termales).
- La clase tecnológica analizada con mayor número de patentes es la 345 con 11461 registros, y la de menor número es la 260 con 7 registros.

### **C. Conclusiones y comentarios finales desde el trabajo investigativo respecto al algoritmo de Kohonen.**

Se utilizó el algoritmo de Kohonen y su aplicación de la creación de mapas auto-organizados, como herramienta que facilita la visualización y la agrupación de ítems de bases de datos. Es importante notar que el algoritmo de Kohonen, no requiere que las salidas de la red neuronal se validen con un conjunto de valores de salida dado; es precisamente esta una de sus mayores ventajas, junto con la facilidad de implementación del algoritmo y evitarle al creador de la red la realización de clasificaciones “manuales”, para validar la información. Siendo ideal para trabajar con grandes volúmenes de datos que requieran ser clasificados.

Respecto a los mapas auto organizados, su mayor virtud es representar en un espacio de dos dimensiones, un conjunto de vectores n-dimensional. Es así como en el ejemplo del capítulo VIII, se representó en una matriz bidimensional un conjunto de valores tridimensional, correspondiente a los vectores RGB del espectro de colores.

### **D. Conclusiones respecto a la aplicación del algoritmo de Kohonen en la Vigilancia Tecnológica.**

Se encontró muy útil la capacidad de crear mapas auto organizados con el algoritmo de Kohonen, ya que uno de los principales retos de la Vigilancia Tecnológica, es lidiar con volúmenes de información cada vez mayores, como número de publicaciones y número de patentes. En este orden de ideas, resulta ser una solución adecuada para visualizar y organizar información obtenida de grandes bases de datos.

De otra parte, los mapas auto organizados ofrecen una manera alterna de visualizar la información, dado que es posible asignarle color, intensidad, brillo u otras características a los clústeres de información formados luego del proceso de entrenamiento de la red; estas mismas características lo han llevado a ser usado en la clasificación de moléculas químicas o de células con cáncer. En el caso de su aplicación a la Vigilancia Tecnológica, puede clasificar cualquier vector de datos n-dimensional, independientemente de si sus valores son numéricos o alfanuméricos. Esta funcionalidad abre un gran espectro de posibilidades para clasificar,

tecnologías, patentes, autores, empresas, inventores o incluso países, según determinadas características de interés para la Vigilancia Tecnológica que se quiera llevar a cabo.

### **E. Conclusiones respecto a la aplicación del algoritmo de Kohonen en la Vigilancia Tecnológica de los ciclos biogeoquímicos e hidrológico.**

Los SOMs, de Kohonen fueron generados como herramienta de visualización de la información contenida en las bases de datos especializadas, sobre publicaciones realizadas en las áreas de biogeoquímica e hidrología. Esta herramienta probó ser útil y versátil a la hora de representar e identificar los diferentes clústeres de información creados; es así como el mapa auto organizado creó un clúster relacionado con cada una de las dimensiones existentes en el vector de entrada creado. De esta manera en la visualización de la Matriz – U de salida, se identifican los clústeres ECO, HIDRO, CHEM y GEO.

Como se mencionó en el capítulo IX, en la Matriz - U generada, los colores cálidos representan los bordes de los clústeres, y los fríos representan los clústeres *per se*, según las graficas de la Figura IX-12, los objetos que obtuvieron los mayores valores en la familia ECO, se encuentran en la parte baja, los que obtuvieron los mayores valores de la familia GEO, se encuentran en la zona media de las ultimas columnas de celdas, el grupo CHEM, obtuvo los mayores valores en la esquina superior derecha, y el grupo HIDRO, en las esquinas superior derecha e inferior derecha.

Estos comportamientos, sumados a los gráficos de Barras y Pie creados para cada una de las celdas en la grilla del mapa de salida, determinan que la ubicación de los clústeres es la siguiente:

1. ECO: ubicado en la parte inferior izquierda del SOM, en las zonas donde las graficas de barras y pie, presentan grandes valores relativos de rojo, respecto a los demás componentes.
2. GEO: este clúster se identifica con el color verde en los diagramas de pie y de barras, y corresponde a la zona intermedia de las últimas columnas del SOM.
3. CHEM: se identifica con el color azul y se ubica en la esquina superior izquierda del SOM, su área limita al sur con el clúster ECO y al oriente con los clústeres GEO e HIDRO.

4. HIDRO: este clúster se identifica con el color morado en los diagramas de barras y de pie, está ubicado en la esquina superior derecha, sobre el clúster GEO, y en la esquina inferior derecha.

Lo anterior muestra la funcionalidad para agrupar la serie de datos inicial y desarrollar clústeres de información alrededor de cada una de las dimensiones del vector de entrada. De esta manera ahora corresponde una zona en el SOM a cada una de las publicaciones estudiadas, y estas están relacionadas entre sí, mediante valores de pertenencia a cada uno de los clústeres; este nivel de pertenencia como se explicó en el capítulo VIII, es proporcional al inverso de la distancia euclidiana entre el centro del clúster y el vector dado.

#### XIV. Recomendaciones generales

- Se recomienda a manera de proyecto complementario, crear una aplicación que funcione en línea y permita tener acceso a los datos de las publicaciones científicas, mediante una interfaz que relacione el mapa auto organizado de Kohonen obtenido y sus celdas, con los registros de entrada de datos. (ver figura XIV-1 para obtener un ejemplo de aplicación)

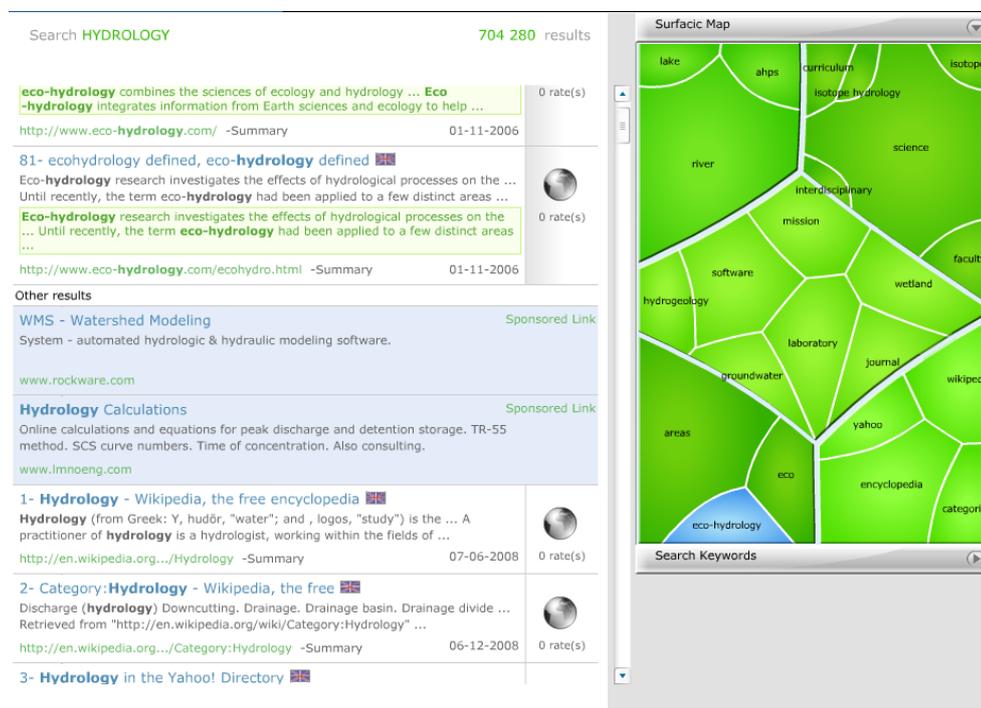


Figura XIV-1. Interfaz gráfica del programa Kvisu. En el ejemplo mostrado se buscó HYDROLOGY se encontraron 704.280 resultados y se creó un mapa temático en la parte derecha de la figura. (KVISU, 2007)

- Por los resultados mostrados, se recomienda extender el uso del algoritmo de Kohonen a otros campos que requieran la clasificación de grandes volúmenes de información.
- Se recomienda utilizar la herramienta para catalogar información multidimensional. Ya se probó su utilidad y eficiencia para realizar clústeres en 3 y 4 dimensiones.
- Se propone realizar futuros estudios para probar la herramienta en la creación de 5 o más clasificaciones, así como desarrollar un método para determinar las familias de los clústeres automáticamente. Se sugiere implementar búsquedas que cuenten todas las palabras en los registros y luego determinar las de mayor ocurrencia entre los sustantivos.
- Se espera en siguientes ejercicios incluir la mayor cantidad posible de descriptores o palabras clave relacionadas a cada una de las familias de datos, para que el proceso de categorización y de creación de los clústeres pueda hacerse con documentos en diferentes idiomas.

## XV. Bibliografía

*Advantages and drawbacks of the Batch Kohonen algorithm.* **Fort, Jean-Claude, Letremy, Patrick y Cottrell, Marie. 2002.** 2002.

**Asimov, Isaac. 1980.** *Breve historia de la química.* Madrid : Alianza, 1980.

**Bensaude-Vincent, B. y Stengers, I. 1997.** *Historia de la química.* Madrid : Addison-Wesley, 1997.

**Buckland, Mat. 2005.** Neural Network Tutorial in Plain English. *AI Junkie.* [En línea] 20 de Abril de 2005. [Citado el: 13 de Mayo de 2008.] <http://www.ai-junkie.com/ann/som/som1.html>.

**Cabezas, Raul, y otros. 2007.** *Agenda Propectiva de Investigacion y Desarrollo Tecnologico de la Cadena Lactea Colombiana.* 2007.

**California State University Monterey Bay. 2006.** Life and Biogeochemical Cycles. [En línea] 11 de Noviembre de 2006. [Citado el: 23 de Abril de 2008.] <http://essp.csumb.edu/esse/climate/climatebiogeo.html>.

**Combs, Richard et al. 2006.** *Competitive Intelligence Handbook.* Chicago, IL : s.n., 2006.

**Conti, Marta Elvira. Facultad de Agronomía - Universidad de Buenos Aires. 2005.** El sitio agricola.com. *Dinámica de la Liberación y Fijación de Potasio en el Suelo.* [En línea] 2005. [Citado el: 28 de Mayo de 2008.] <http://www.elsitioagricola.com/articulos/conti/Liberacion%20y%20Fijacion%20Potasio.asp>.

**Corsi, Pietro. 2008.** Les Oeuvres de Lavoisier. [En línea] Universidad de Paris, 2008. [Citado el: 08 de 07 de 2008.] <http://histsciences.univ-paris1.fr>.

**EAFIT. 2008.** Atlas EAFIT. [En línea] 01 de Marzo de 2008. [Citado el: 13 de Mayo de 2008.] [http://200.12.187.213:8001/servlet/SBReadResourceServlet?rid=1175636096906\\_326331909\\_865&partName=htmltext](http://200.12.187.213:8001/servlet/SBReadResourceServlet?rid=1175636096906_326331909_865&partName=htmltext).

**Escuela Tecnica Superior de Ingenieros de Telecomunicaciones, Universidad Politecnica de Madrid. 2007.** Analisis de Patentes. *Como procedimiento de apoyo al diagnostico tecnologico .* [En línea] Grupo de Gestion de la Tecnologia, Septiembre de 2007. [Citado el: 23 de Abril de 2008.] <http://www.getec.etsit.upm.es/docencia/gtecnologia/diagnostico/analisispatentes.htm>.

**Fundacion COTEC. 1999.** *Vigilancia Tecnologica, Documentos COTEC sobre oportunidades tecnologicas.* Valencia, España : Graficas Arias Montano, 1999.

**Garcia, Mary Ruth . 2005.** *Curso de Ecologia y Medio Ambiente - Universidad Nacional de Colombia.* Bogota : s.n., 2005.

**Generation 5. 2008.** Self Organizing Map AI for Pictures. [En línea] 14 de Junio de 2008. [Citado el: 15 de Junio de 2008.] <http://www.generation5.org/content/2004/aiSomPic.asp>.

**Germano, Tom. 1999.** Self Organizing Maps. [En línea] 23 de Marzo de 1999. [Citado el: 12 de Junio de 2008.] <http://davis.wpi.edu/~matt/courses/soms/#Scale%20Neighbors>.

**Hassoun, Mohamad H. 1995.** *Fundamentals of Artificial neural Networks.* Cambridge, Massachusetts : MIT Press, 1995.

**Heaton, Jeff. 2005.** *Introduction to Neural Networks with JAVA.* St. Louis : Heaton Research, 2005.

**Hecht-Nielsen. 1990.** *Neurocomputing.* Massachusetts : Addison-Wesley, 1990.

**Instituto de Microbiologia Bioquimica, Universidad de Salamanca. 2005.** *Ciclos Biogeoquimicos.* Salamanca, España : s.n., 11 de Marzo de 2005.

**IPCC. 2007.** *An Assessment of the Intergovernmental Panel on Climate Change.* Valencia, España : Cambridge University Press, 2007.

**Kohonen, T. 1984.** *Self-Organization and Associative Memory.* Berlin : Springer-Verlag, 1984.

—. **1995.** *Self-Organizing Maps.* Berlin : Springer-Verlag, 1995.

**KVISU. 2007.** KVISU. [En línea] 2007. <http://www.kvisu.com/>.

**Leicester, H.M. 1967.** *Panorama histórico de la química.* Madrid : Alhambra, 1967.

**Maspons, Ramon y Escorsa, Pere. 2007.** *De la Vigilancia Tecnologica a la Inteligencia Competitiva.* 2007.

*Plan Tecnológico de Navarra.* **Agencia Navarra de Innovacion. 2007.** 2007, Innovamos, págs. 1-38.

**ProQuest. 2008.** ProQuest. [En línea] 2008. [Citado el: 20 de 05 de 2008.] <http://www.proquest.com/division/aboutus/>.

**Ritter, H, Martinetz, T y Shulten, K. 1992.** *Neural Computation and Self-Organizing Maps. an Introduction.* Reading : Addison-Wesley, 1992.

**Sanchez, Ivon. 2008.** La Capacidad Tecnologica. *Slideshare.* San Salvador : s.n., 2008.

**Scientific Committee on Problems of the Environment (SCOPE) Series . 2003.** *Interactions of the Major Biogeochemical Cycles: Global Change And Human Impacts.* s.l. : Island Press, 2003.

*Seminario de Capacitación sobre Patentes, en el marco del proyecto "Redes de Investigación y Propiedad Intelectual"* . **SECOPI y SIC. 2007.** Bogota : s.n., 2007.

*Simulating effects of land use changes on carbon fluxes: past contributions to atmospheric CO2 increases and future commitments due to losses of terrestrial sink capacity.* **K. M. STRASSMANN, F. JOOS, G. FISCHER. 2008.** 2008, Tellus B. International Meteorological Institute of Stockholm.

—. **STRASSMANN, K. M., JOOS, F. y FISCHER, G. 2008.** s.l. : International Meteorological Institute of Stockholm, 2008, Tellus B.

**Suárez Menéndez, Juan José. 2008.** *HITOS DE LA HISTORIA DE LA QUÍMICA.* 2008.

**The MathWorks Inc. 2008.** *Neural Network Toolbox™ User's Guide.* [aut. libro] Howard Demuth, Mark Beale y Martin Hagan. Natick : MathWork, 2008.

**Universidad de California, Irvine. 2007.** *Earth Systems Graduate Course.* Irvine : s.n., 2007.

**University Corporation for Atmospheric Research (UCAR). 2007.** *Windows to the Universe.* [En línea] 7 de Mayo de 2007. [Citado el: 13 de Abril de 2008.] <http://www.windows.ucar.edu/tour/link=/earth/Life/biogeochem.html>.

**University of Colorado, Geological Sciences. 2007.** *Introduction to Biogeochemical Cycles.* [En línea] 2007. [Citado el: 18 de Mayo de 2008.] [www.colorado.edu/GeolSci/courses/GEOL1070/chap04/chapter4.html](http://www.colorado.edu/GeolSci/courses/GEOL1070/chap04/chapter4.html).

**USPTO. 2008.** *USPC-to-IPC Reverse Concordance.* [En línea] 2008. [Citado el: 08 de 07 de 2008.] [http://www.uspto.gov/go/classification/international/ipc/ipc8/ipc\\_concordance/ipcse1.htm](http://www.uspto.gov/go/classification/international/ipc/ipc8/ipc_concordance/ipcse1.htm).

**Van der Reis, Jan. 2007.** *Ayudando a las PYMES de habla Hispana.* [En línea] 29 de 11 de 2007. [http://www.vanderreis.com/blog/inteligencia\\_competitiva/](http://www.vanderreis.com/blog/inteligencia_competitiva/).

**Villee, Claude A. 1990.** *Biología*. Mexico : McGraw-Hill Interamericana de Mexico S.A de C.V., 1990.

**Voronoi, Georgy Fedoseevich. 1907.** *Nouvelles applications des paramètres continus à la théorie des formes quadratiques*. s.l. : Journal für die Reine und Angewandte Mathematik, 1907.

**Wikipedia. 2008.** Ingeniería Inversa. *Teselaciones de Delaunay*. [En línea] 12 de Mayo de 2008. [Citado el: 13 de Junio de 2008.] [http://es.wikipedia.org/wiki/Ingenier%C3%ADa\\_inversa](http://es.wikipedia.org/wiki/Ingenier%C3%ADa_inversa).