

UNIVERSIDAD DE CANTABRIA

ESCUELA TÉCNICA SUPERIOR DE INGENIEROS
DE CAMINOS, CANALES Y PUERTOS

TESIS DOCTORAL

Desarrollo de una metodología para el estudio de la
morfología de playas basado en mapas auto-organizativos de
imágenes digitales.

Presentada por: **OMAR QUETZALCÓATL GUTIÉRREZ GUTIÉRREZ**

Dirigida por: **ERNESTO MAURICIO GONZÁLEZ RODRÍGUEZ
RAÚL MEDINA SANTAMARÍA**

Santander, julio 2011

Planteamiento de una metodología para clasificar una colección de IMDP según las características morfológicas

5.1. Introducción

En la sección anterior se realizaron los experimentos necesarios para asegurar que la clasificación obtenida al entrenar un mapa con las IMDP sea útil en la clasificación de las mismas en función de las características morfológicas que se observan. Los experimentos consistieron en entrenar los SOMs usando diversos parámetros, obtenidos a partir de las IMDP, y métricas, que fueron implementadas a la técnica. Se encontró que la clasificación obtenida usando los valores de luminosidad de las IMDP es la que más información de las características morfológicas de la playa proporciona. En el caso de la métrica, los mejores resultados se obtuvieron usando la correlación.

Los resultados encontrados en el capítulo anterior permiten plantear la aplicación de la técnica para clasificar colecciones de IMDP, para lo cual es necesario establecer

critérios y desarrollar una metodología. Esta metodología debe resumir y especificar los tratamientos realizados previamente a la colección para solventar los efectos externos que afectan a las IMDP, establecer criterios para el entrenamiento de un mapa, la aplicación de un modelo morfológico al mapa entrenado y la formación de una serie de evolución. En este capítulo se revisan estos temas y se propone una metodología para ser aplicada a una colección cualquiera.

5.2. Desarrollo metodológico

Una metodología que permita analizar una colección de IMDP y generar una serie morfológica se debe contar con tres elementos principales (Figura 5.1): el primero consiste en la preparación de la colección de IMDP para ser analizada; la segunda parte es el entrenamiento de la colección en un mapa y finalmente en la tercera parte se genera la serie temporal según un modelo o característica morfológica.

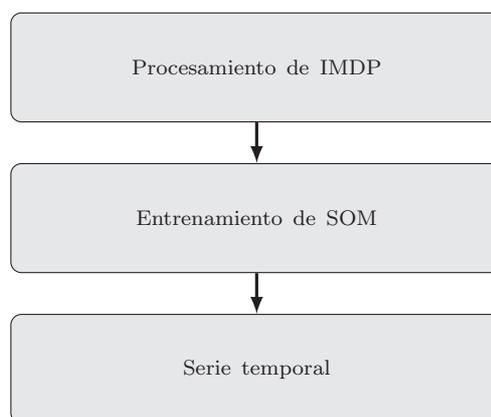


Figura 5.1. Se indican los tres elementos a considerar en el planteamiento de una metodología de clasificación de IMDP por medio de SOMs.

Cada uno de estos elementos está formado por procesos independientes que a continuación son analizados.

5.2.1. Procesamiento de IMDP

Cualquier colección de IMDP georeferenciada en la que el número de imágenes sea elevado, puede ser clasificada por medio de los SOMs. A partir de la clasificación obtenida se pueden generar grupos en función de las características de las imágenes. Como se mostrará más adelante estos grupos pueden estar relacionados, por citar algunos ejemplos, con la posición de la línea de costa, con la presencia o forma de barras, o con un modelo morfodinámico. A partir de estos grupos se puede analizar la evolución de la característica observada.

La colección de imágenes antes de ser entrenada debe de ser procesada. Este procesamiento incluye la reducción de las IMDP a la zona de interés con el propósito de, primero reducir el número de dimensiones del problema y segundo eliminar información no necesaria para la clasificación. A su vez algunos procesos externos como la variación del nivel del mar por la marea, la meteorología y el horario en que fue obtenida cada IMDP tienen efectos sobre las colecciones y las clasificaciones, por lo que deben de ser eliminados o al menos reducidos al máximo.

La reducción de dimensiones del problema consiste en descartar de las IMDP las secciones que no tienen interés en el análisis, *i. e.* zonas que siempre son agua o tierra. También se recomienda girar las IMDP de tal forma que la playa sea paralela a la misma. Los detalles de estos procedimientos se presentan en el apartado 4.2.1

El efecto de la marea en las IMDP se traduce en cambios en la posición de la línea de costa, la forma más sencilla de descartar este efecto es utilizar en la clasificación sólo IMDP obtenidas a un mismo nivel de marea. Los detalles de este procedimiento se muestran en el apartado 4.2.2. Por otro lado, las variaciones meteorológicas y horarias producen cambios en las tonalidades, con lo que se tienen IMDP oscuras obtenidas a últimas horas del día, brillantes y con un exceso de luz cuando fueron obtenidas cerca del mediodía, muy pálidas o azules en los días de lluvia, o rojizas en amaneceres, atardeceres. Estos efectos se eliminan al transformar las IMDP de color a valores de

luminosidad o tonos de grises (apartado 4.4) y al aplicar la ecualización del histograma. Con esta técnica los valores de luminosidad se redistribuyen uniformemente en todo el espectro del visible, con lo que IMDP oscuras o claras son corregidas a valores intermedios; los detalles de esta técnica se describen en 4.4.1.1. Una vez realizada la ecualización en la colección persisten dos categorías de imágenes, en la primera la playa es oscura y el agua clara, y en la segunda ocurre lo contrario. Es necesario extraer de la colección al máximo posible alguna de estas categorías. Esto se realiza seleccionando dos píxeles típicos de cada zona y obteniendo la distribución de valores de cada píxel. Mediante un análisis de las distribuciones se puede descartar una de las categorías. Los detalles de este procedimiento se describen en la sección 4.4.1.2.

Finalmente después de una inspección rápida de algunas IMDP un área reducida representativa de la playa debe ser seleccionada para utilizarla en los entrenamientos. Esta área debe de contener una muestra de la variabilidad que se desea clasificar en las IMDP. Los datos de entrenamiento se construyen siguiendo las indicaciones descritas en el anexo B y usando la fecha de cada IMDP como etiqueta.

5.2.2. Entrenamiento de los SOMs

Una vez se tienen preparados los datos se puede entrenar SOMs usando la correlación como métrica. En el entrenamiento además de la métrica a usar es de importancia determinar el tamaño del mapa más adecuado para el entrenamiento. Posteriormente al entrenamiento se deben determinar los prototipos de cada unidad y analizar el mapa obtenido. A continuación se describe las condiciones para seleccionar un tamaño de mapa adecuado y como deben de ser determinados los prototipos de las unidades.

5.2.2.1. Selección del tamaño del mapa

Una vez que esta preparada la colección de IMDP, determinar el tamaño del mapa a usar es lo más importante para la clasificación. Entrenar un mapa con una colección de IMDP puede ser comparado con la acción de clasificar una serie de documentos en un archivador, donde los documentos son las IMDP, el archivador es el mapa y el criterio para clasificar los documentos, digamos *palabras clave*, es el equivalente a la métrica de los SOMs (ver figura 3.2).

El resultado, la eficiencia de la clasificación, dependerá primero de la naturaleza (cantidad y variedad) de los documentos, de los criterios utilizados y finalmente del número de separadores en el archivador. Por tanto, es claro que ante la necesidad de clasificar un conjunto de datos dados, bajo una métrica fija, la calidad de la clasificación depende solamente del número de separadores del archivador, así como, que no existe un número correcto o único de unidades para realizar la clasificación.

En el caso de las colecciones de IMDP, previamente se determinó que los parámetros a entrenar y la métrica más adecuados son los valores Y de las IMDP y la correlación respectivamente. Por lo tanto, al igual que en el ejemplo mencionado previamente, la calidad de la clasificación depende solamente de el mapa elegido. Desafortunadamente no existe una formulación para determinar las dimensiones óptimas del mapa, por lo que solo se puede establecer límites. El límite superior de número de unidades K_{max} , está dado por el valor mínimo obtenido del número de IMDP a entrenar y el máximo de unidades posibles en función del PC a utilizar. *Vesanto et al.* (2000) proporciona una estimación de la memoria utilizada por un PC en un entrenamiento en función del número de casos o IMDP n , el número de variables o dimensiones por caso r y el número de unidades K

$$M_t = 8 \times (5Kr + 4nr + 3K^2), \quad (5.1)$$

de la que se obtiene, suponiendo una memoria máxima M_{max} a utilizar, un número

máximo de unidades K_{max} ,

$$K_{max} = -\frac{5}{6}r + \sqrt{\frac{25r^2 - 48nr + 1.5M_{max}}{36}}, \quad (5.2)$$

la figura 5.2 muestra los contornos de K_{max} obtenidos para valores de típicos de n y d en la colecciones de IMDP, considerando una $M_{max} = 1GB$. Los valores son en general altos comparados con los valores típicos de n , por lo tanto, podemos considerar que $K_{max} = n$.

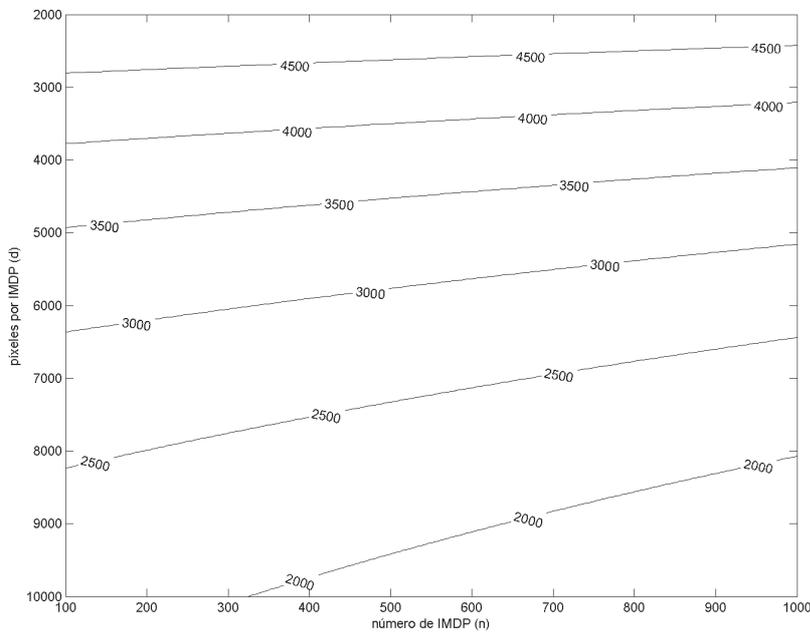


Figura 5.2. Se indican los valores posibles de K_{max} como función del numero de IMDP, n y el numero de píxeles por IMDP, r y considerando una $M_{max} = 1 GB$.

Por otro lado, el número mínimo de unidades K_{min} en el mapa coincide con el numero de categorías en que se desea clasificar las IMDP, *e.g.*, el modelo de *Wright y Short* (1984), contiene 6 categorías, una para cada estado morfodinámico, por lo tanto $K_{min} = 6$.

De esta forma para entrenar una colección de IMDP es necesario utilizar un numero de unidades tal que, $K_{min} \leq K \leq K_{max}$. Estos límites no acotan lo suficiente el

problema y ya que no existe una fórmula para determinar un número ideal, es necesario probar con distintos valores de K hasta encontrar una clasificación adecuada. Tomando de nuevo el ejemplo del archivador y los documentos, es necesario probar con distintos archivadores hasta encontrar aquel que proporcione la clasificación más adecuada. Usando mapas con número alto/bajo de unidades mejor/peor es la clasificación pero el mapa es de menor/mayor funcionalidad. Por tanto es necesario buscar un equilibrio entre la funcionalidad y la clasificación que se desea obtener. Para que un mapa sea manipulable, debe de reducir el trabajo de clasificación mínimo un 50 %, por lo tanto se fija el límite superior de unidades en $K_{max} = 0.50n$, en el otro extremo, la mapa mínimo debe de tener las unidades suficientes para permitir organizar toda la colección, por lo que se considera que debe de tener tantos prototipos como al menos el 10 % de IMDP. Así se considera apropiado utilizar un mapa con $0.10n \leq K \leq 0.50n$ unidades. Utilizar $K > 0.50n$ implica un esfuerzo muy alto en asignar una categoría a cada prototipo (equivalente a asignar categorías a la colección completa) y usando $K < 0.10n$ se corre el riesgo de que no existan suficientes unidades en el mapa para separar las IMDP según las características deseadas. La figura 5.3 indica el intervalo recomendado de número de unidades para una colección de IMDP. Se recomienda realizar entrenamientos usando distintos mapas hasta encontrar la clasificación más adecuada. Otro criterio a tomar en cuenta es que mientras mayor es el mapa, la probabilidad de encontrar unidades vacías o con una sola IMDP aumenta. Este efecto debe de ser evitado.

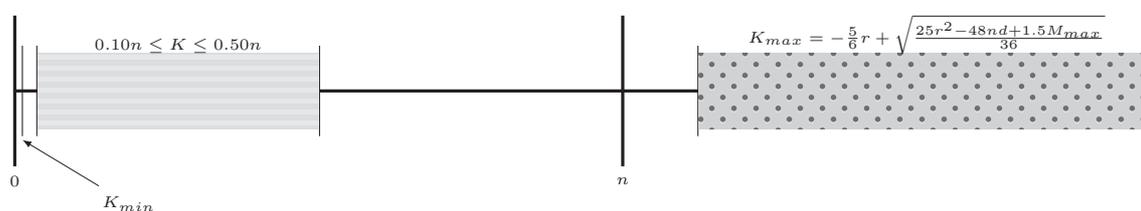


Figura 5.3. Escala de tamaños de mapas.

5.2.2.2. Determinación de prototipos

Al final de cada entrenamiento se obtiene un mapa de prototipos, donde cada prototipo representa las características de las IMDP contenidas en cada unidad. A cada paso del entrenamiento los prototipos son re-calculados, puesto que las IMDP contenidas en cada unidad del mapa varían. Al finalizar el entrenamiento, cada prototipo define las características de cada unidad del mapa. Estos prototipos no forman parte de la colección de IMDP, sino que son el resultado de las operaciones realizadas entre las IMDP de cada unidad. Por lo tanto para construir el SOM ya entrenado es necesario determinar la IMDP más *cercana* a cada prototipo. Esta corresponde a las IMDP con correlación más alta con el prototipo de cada unidad.

$$IMDP_{prot_k} = \max_k(\text{CO}(IMDP_k, prot_k)) \quad (5.3)$$

5.2.2.3. Evaluación de los entrenamientos

La evaluación de los mapas entrenados, generalmente se lleva a cabo usando la distancia media (q_e) y el error topográfico (T_e). Sin embargo, estas medidas como se explica en la sección 4.4.2, no son adecuadas para determinar la calidad de la clasificación de las IMDP en un SOM entrenado. En la misma sección se utilizaron medidas de errores basadas en la comprobación de las IMDP contenidas en cada unidad de los mapas entrenados. Estas medidas no son prácticas puesto que requieren de la comprobación manual de cada unidad. Debido a esto, la evaluación del mapa entrenado, en esta etapa del procedimiento se realiza analizando los prototipos, su distribución en el mapa y considerando los siguientes criterios:

1. Los prototipos del mapa deben mostrar una organización *i. e.* mientras más parecidos sean dos prototipos más cercanos deben de aparecer en el mapa

2. El número de IMDP por unidad: Sí el número es demasiado alto puede haber mezcla de categorías en algunas unidades indicando un mapa muy pequeño, por el contrario si el numero es bajo o hay algunas unidades vacías, indica que el mapa es grande.

El error total obtenido en la clasificación se determina una vez reconstruida la serie temporal.

5.2.3. Reconstrucción de una serie de evolución

El resultado final de la aplicación de la metodología es la obtención de una serie temporal de evolución de alguna característica de las IMDP. Esta serie se construye a partir de una medición realizada o la aplicación de un modelo a los prototipos. La información temporal se recupera de las fechas de las IMDP almacenadas en las etiquetas. Una vez que la serie es reconstruida es necesario evaluar la calidad de la misma. Si la calidad no es suficiente será necesario repetir el procedimiento usando un tamaño de mapa distinto.

5.2.3.1. Clasificación y formación de supergrupos

Cualquier modelo o medida de alguna característica morfológica detectable en las IMDP es válida para clasificar los prototipos del mapa *e.g.*, el modelo de *Wright y Short* (1984), la posición de la línea de costa, la forma o posición de las barras. En un mapa *bien* entrenado se observan regiones en las que las unidades presentan la misma categoría. A estas regiones se les denomina supergrupos. Lo ideal es encontrar un solo supergrupo por categoría, sin embargo dependiendo de la naturaleza de las categorías, es posible encontrar más de uno por cada categoría *e.g.* Si la característica de un supergrupo es la presencia de una barra, y esta cambia de forma a través del tiempo, es posible que IMDP con barra se encuentren ubicadas en distintas zonas del

mapa. En algunas ocasiones dependiendo del tipo de medición realizada y del tamaño del mapa no es conveniente formar supergrupos.

5.2.3.2. Construcción de una serie de evolución

La construcción de la serie temporal se realiza a partir de los valores asignados a los prototipos. Utilizando estos valores y la fecha de cada IMDP se reconstruye una serie de evolución.

5.2.3.3. Evaluación de la serie

La serie obtenida representa la variación de las IMDP en el tiempo según la clasificación asignada a cada supergrupo. La calidad de la serie esta afectada por dos tipos de errores: uno sistemático debido a las IMDP mal clasificadas por la técnica y otro que es función del tamaño de mapa elegido y la formación de los supergrupos. Una vez construida la serie es fácil identificar los errores sistemáticos y corregirlos, sin embargo detectar los errores aleatorios debidos al tamaño del mapa es más complicado y requiere de un procedimiento recursivo.

1. Error sistemático:

La técnica clasifica las IMDP en función de los valores de luminancia, que después del tratamiento realizado a la colección relacionada con las características morfológicas de la playa. Sin embargo existe un error asociado del orden del 5% (ver tabla 4.3). Este error, una vez reconstruida la serie, es fácil de identificar puesto que aparece como variaciones de categoría por periodos muy cortos en situaciones de estabilidad. Para ejemplificar esto suponemos una colección de IMDP entrenada en un SOM en el que se forman 5 supergrupos llamados *A*, *B*, *C*, *D* y *E*. Cada supergrupo esta formado por un numero distinto de unidades del mapa. Usando las fechas de las IMDP y las categorías de los supergrupos se construye una serie temporal, tal como se ha explicado a lo largo de esta metodología. La figura 5.4

muestra un segmento de esta serie hipotética, en la que se observa un periodo de estabilidad de IMDP clasificadas en la categoría B , sin embargo, hay una IMDP que indica la categoría D , esta representa una IMDP mal clasificada puesto que no es posible que la playa experimentara un cambio tan rápido. Estos errores

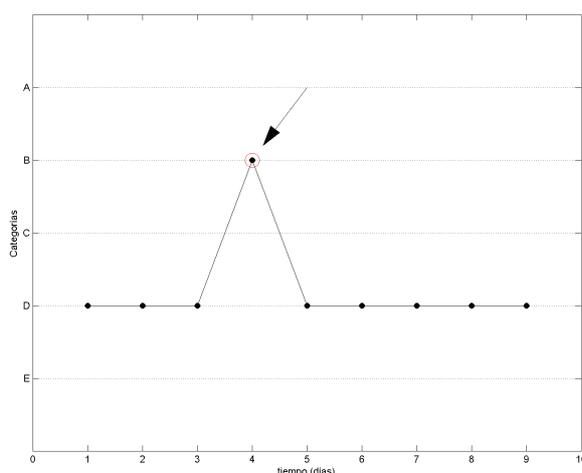


Figura 5.4. Ejemplo de un error sistemático típico de la técnica. En la serie se observa un valor atípico producto de una IMDP mal clasificada

pueden ser cuantificados y eliminados aplicando un filtro de media o moda móvil a la serie. La longitud del filtro, L_f , esta determinada por la escala temporal del proceso que se desea observar o medir en la serie, puesto que toda variabilidad de escala menor que el filtro es eliminada también. La proporción de elementos corregidos a cada aplicación del filtrado puede determinarse como

$$E_s(t) = \frac{(S(t) \neq S_e(t))_n}{L_f}, \quad (5.4)$$

donde $S(t)$ y $S_e(t)$ son las serie antes y después del filtrado. La proporción media de elementos corregidos es

$$E_m = \frac{\sum E_s(t)}{n} \times 100 = \frac{\sum (S(t) \neq S_e(t))}{n} \times 100, \quad \text{si } n \gg L_f \quad (5.5)$$

Para demostrar la capacidad del filtro de eliminar este tipo de errores se creó artificialmente, como ejemplo, una serie de evolución de 4500 elementos (días) a partir de un hipotético mapa entrenado en el que se determinaron 5 supergrupos, llamados *A*, *B*, *C*, *D* y *E*. Un segmento de 500 elementos de la serie se muestra en la figura 5.5_a). Esta serie fue modificada, de tal forma que la categoría del 10 % (450 elementos) de los valores fue alterada (Figura 5.5_b). Un error del 10 % es elevado en comparación con el error sistemático de la técnica, sin embargo es un ejemplo adecuado para demostrar el efecto del filtrado. La serie graficada en 5.5_c) es el resultado de aplicar a la serie un filtro de moda móvil de $L_f = 5$ días. En la gráfica 5.5_d) se muestran la variación temporal de E_s , esta variación tiene importancia para determinar si estos errores están asociados a alguna categoría del mapa. La proporción total de datos corregidos en la serie es del 8 %, lo que nos indica que usando un filtro $L_f = 5$ se corrigen alrededor del 80 % de los datos que la técnica clasifica mal.

2. Error aleatorio:

Al eliminar los errores sistemáticos, se descartan aquellos casos en los que, en una sección de la serie, la categoría de una o varias IMDP no coincide con el resto a su alrededor. Sin embargo, las secciones de longitud mayor que el filtro en las que la categoría permanece constante no son alteradas. De aquí surge la pregunta: ¿Son todos estos segmentos de la serie reales? El número de segmentos con distintas categorías depende del tamaño del mapa y de los supergrupos, en caso de ser formados, formados. Por tanto, al variarlos se obtienen series distintas. Este error que depende totalmente de las elecciones realizadas se considera de carácter aleatorio.

- Dimensiones del mapa:

El número de prototipos disponibles para clasificar la colección de IMDP depende de las dimensiones del mapa. Si el número de prototipos es elevado, la diversidad de los mismos también lo será y por tanto la similitud entre las IMDP de cada unidad aumenta. Por el contrario si el número de prototipos

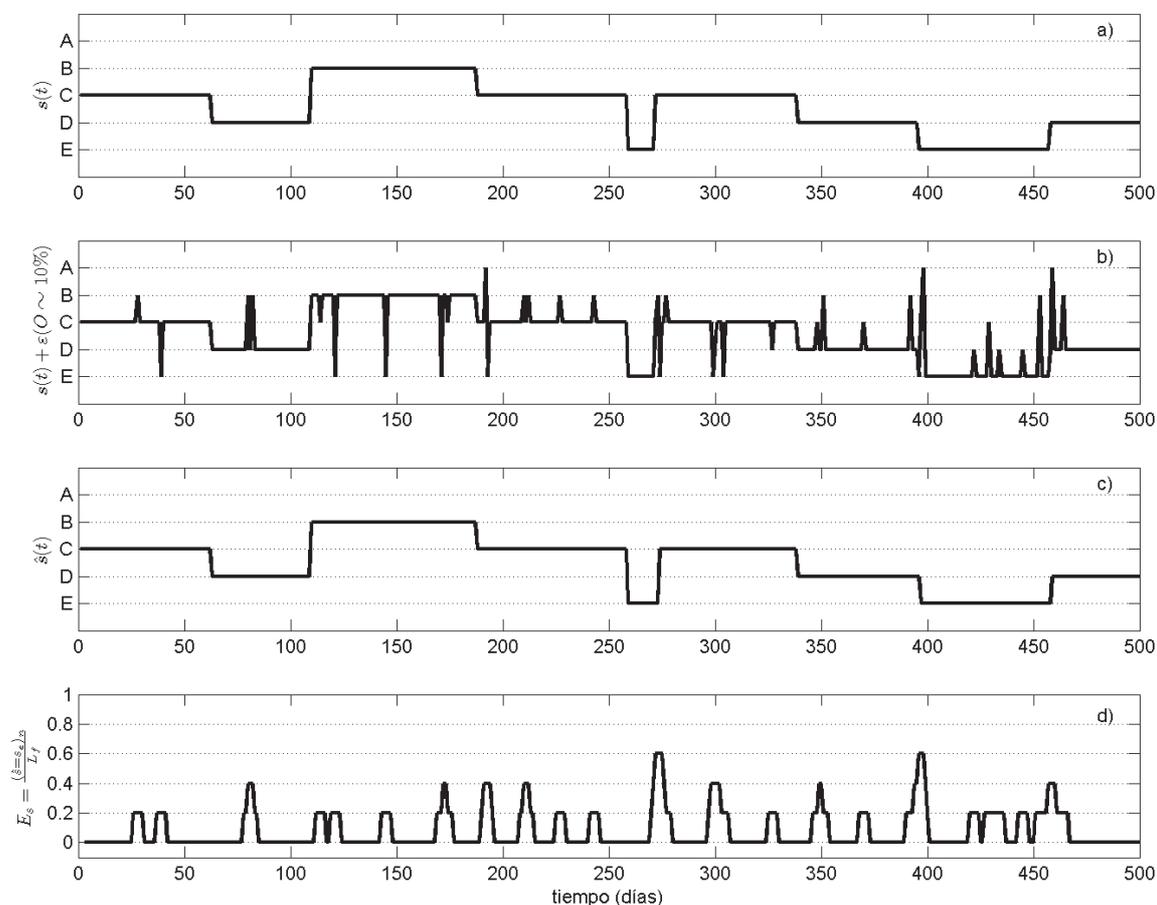


Figura 5.5. Ejemplo de aplicación de un filtro modal a una serie para corregir el error sistemático. a) una serie artificial que muestra la evolución de una variable entre 5 estados; en b) la serie original fue alterada en un 10% de los elementos, simulando el error sistemático obtenido en la clasificación de las IMDP en un SOM; c) muestra la serie obtenida después de aplicar un filtro modal de 5 días y en d) se muestra el residuo entre ambas series.

disminuye, la similitud de IMDP en cada unidad disminuye, puesto que la técnica esta forzada a clasificar todas las IMDP. Por tanto la calidad de la serie obtenida depende del tamaño de mapa utilizado.

- Formación de supergrupos:

En el caso de que se decida utilizar solamente unas pocas categorías para clasificar el mapa *i.e.* un modelo discreto, o cuando el numero de prototipos es elevado se pueden formar supergrupos de características similares. Si a

cada supergrupo se le asigna un valor representativo, la serie contara con discontinuidades, por tanto la decisión de incluir una unidad del mapa en un supergrupo o no afecta a la calidad de la serie obtenida.

El resultado son cambios en la serie en función de la conformación de los supergrupos y del tamaño del mapa, creando saltos o ciclos de evolución distintos. La figura 5.6 muestra, como ejemplo tres series discretas obtenidas usando distinto numero de prototipos, por lo que presentan categorías distintas para un mismo tiempo.

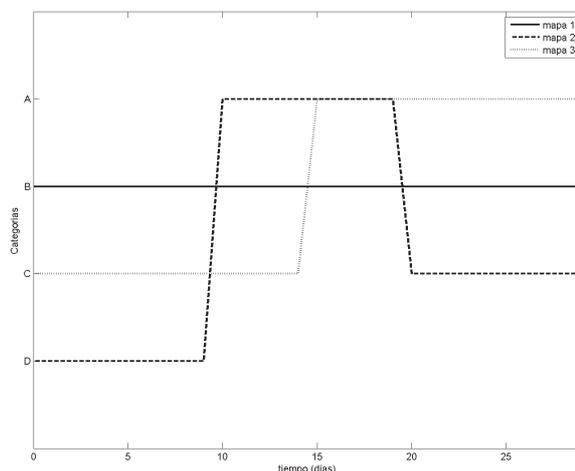


Figura 5.6. Ejemplo de un error aleatorio. Se muestran tres series obtenidas de entrenar mapas de distintos tamaños y con distintos supergrupos, por lo que la evolución que muestra cada serie es distinta.

Ya que cada problema requiere de un tamaño de mapa y de una configuración de supergrupos distintos, es necesario determinar la serie de evolución usando mapas de distintos tamaños. La recomendación general en el uso de SOMs es realizar varios entrenamientos, iniciando con un mapa grande y reduciendo las dimensiones a cada nuevo entrenamiento. De esta forma se obtiene el menor mapa que proporciona la clasificación óptima. Sin embargo esto no es aplicable a las colecciones de IMDP, puesto que un mapa de dimensiones grandes siempre tendra una mejor clasificación que uno de dimensiones menores, y una vez clasi-

ficado este, es innecesario realizar nuevos entrenamientos. Aunque el esfuerzo de clasificación puede ser elevado debido al número de prototipos.

Por el contrario aquí lo que se busca es clasificar lo mejor posible la colección y el menor número de prototipos, por lo que el procedimiento se debe realizar en sentido opuesto. Iniciar con un mapa pequeño, que es fácil de clasificar y formar supergrupos, e ir aumentando el tamaño del mapa. A cada mapa se obtiene una serie y se compara con la anterior, cada nuevo entrenamiento la serie será más compleja. Conforme el mapa aumenta la serie se estabilizará y habrá pocos cambios. El mapa correspondiente a la serie en la que no hay cambios se puede considerar como el óptimo y por lo tanto la evolución descrita por la serie como real.

5.3. Metodología final

Finalmente, una vez analizados los tres componentes principales de la clasificación de la colección de IMDP, se establecen los pasos de la metodología. Los pasos de esta se resumen a continuación:

■ Preparación de la colección de IMDP

1. Seleccionar el área de interés:

Reducir las IMDP al área de playa que interesa clasificar, de esta forma se busca eliminar las características de las IMDP ajenas al problema (ver apartado 4.2.1). También es recomendable girar las IMDP hasta obtener una playa paralela a la imagen.

2. Eliminar los efectos de la marea:

El efecto de la marea se refleja en cambios en la posición de la línea de costa y por supuesto afectan la clasificación. La forma de eliminar este efecto es considerar una IMDP por ciclo de marea a un nivel constante. Los niveles

entre media marea y bajamar son los más adecuados puesto que permiten observar mejor las variaciones morfológicas de la playa. Los detalles de este procedimiento se muestran en el apartado 4.2.2

3. Eliminar efectos meteorológicos y horarios:

La eliminación de los efectos meteorológicos y horarios se realiza transformando las IMDP de color a Luminosidad (ecuación 4.1) y ecualizando el histograma de las IMDP. Los detalles de este procedimiento se muestran en la sección 4.4.1.1.

4. Descartar categoría minoritaria:

Al ecualizar las IMDP se forman dos categorías *i. e.* una con arena clara y agua oscura y otra de forma opuesta. Es necesario eliminar una de estas categorías. Ver la sección 4.4.1.2

5. Seleccionar área de análisis:

Finalmente se requiere elegir una área reducida donde el fenómeno o característica a clasificar se presente en las IMDP. Se recomienda un tamaño del área de aproximadamente 5000 píxeles.

6. Preparar datos de entrada:

Se construye un archivo de entrada, donde cada IMDP corresponde a un set de parámetros a entrenar y la fecha debe ser usada como etiqueta de cada IMDP.

■ **Entrenamiento del SOM**

7. Seleccionar dimensiones del mapa:

el tamaño del mapa depende de cada caso a clasificar por lo que se debe experimentar con diferentes dimensiones. Los tamaños recomendados están limitados entre un 10 % y 50 % del número de IMDP de la colección. Ver el apartado 5.2.2.1

8. Entrenar el mapa:

El entrenamiento se realiza utilizando la correlación como métrica. El uso de esta métrica proporciona la mejor relación en calidad de la clasificación y tiempo requerido (ver 4.5.1).

9. Determinar los prototipos del mapa:

Una vez entrenado el mapa es necesario determinar las IMDP prototipos o más cercanas a los prototipos de cada unidad del mapa, para representar la clasificación. Los detalles en apartado 5.2.2.2.

10. Evaluar mapa:

En este paso es posible analizar el mapa entrenado y determinar subjetivamente si la organización de las IMDP es adecuada, si no es así se debe volver a 7 elegir un nuevo tamaño de mapa y repetir los pasos. Los criterios básicos de la evaluación se describen en el apartado 5.2.2.3.

■ Construcción de la serie temporal

11. Aplicar modelo:

Una vez determinados los prototipos, se asigna una categoría a cada unidad del mapa. La categoría asignada debe de estar relacionada con un modelo o una característica morfológica de las IMDP.

12. Formar supergrupos:

Basándose en las categorías asignadas a cada unidad se forman supergrupos y a cada IMDP contenida en las unidades de cada supergrupo se le asigna un valor relacionado con la categoría.

13. Reconstruir serie temporal:

Utilizando la fecha de cada IMDP y los valores de las categorías asignadas se reconstruye una serie de evolución.

14. Evaluar la serie obtenida:

Finalmente la serie se evalúa para asegurarse de obtener un resultado fiable. En el uso de la técnica existen dos tipos de errores que determinan la calidad de la serie: el primero, sistemático del orden del 5 %, puede ser eliminado aplicando un filtro de moda móvil a la serie obtenida. La longitud del filtro esta relacionada con la duración de los procesos que se quieren observar. El segundo es un error aleatorio, pues depende del tamaño del mapa elegido y de la configuración de los supergrupos. Debido a que se desconoce la serie real, la única forma de contrastar la serie obtenida es probando con mapas de diversos tamaños hasta encontrar el más adecuado, por lo que es necesario repetir la metodología a partir del paso 7. Más detalles de como realizar la evaluación de la serie se encuentran en el apartado 5.2.3.3.

La figura 5.7 muestra de forma esquemática los pasos a seguir en la clasificación de la colección e IMDP.

5.4. Conclusiones

En este capítulo se desarrollo una metodología que permite clasificar cualquier colección de IMDP, para lo cual fue necesario establecer los criterios y pasos requeridos para asegurar una clasificación adecuada. La metodología se encuentra formada por tres partes, en la primera las IMDP son procesadas para eliminar fenómenos externos e internos que afectan la clasificación; en el segundo un SOM es entrenado y en el tercero se reconstruye la serie de evolución y evalúa el resultado.

El procesamiento de las IMDP incluye la reducción y rotación del área de interés, la eliminación de los efectos producidos por la marea, meteorología y cambios horarios, y finalmente la creación de los datos de entrada para el entrenamiento.

La segunda parte de la metodología consiste precisamente en el entrenamiento del SOM utilizando las IMDP previamente procesadas. El entrenamiento se realiza utilizan-

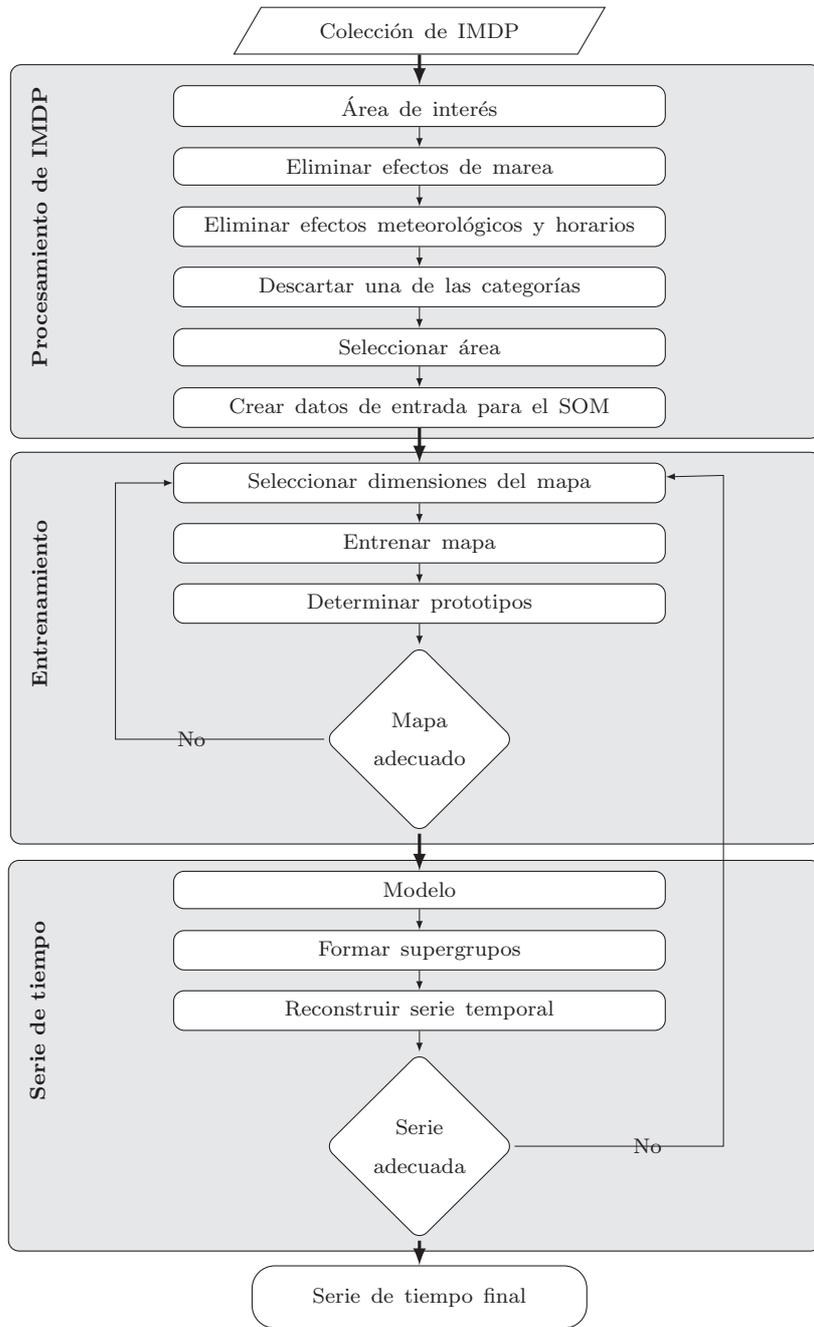


Figura 5.7. representación esquemática de la metodología.

do la correlación como métrica y posteriormente se determinan las IMDP prototipos de cada unidad del mapa.

Finalmente, en la tercera parte, se asigna una categoría a cada unidad del mapa y forman supergrupos, con lo que se reconstruye una serie temporal. El procedimiento

CAPÍTULO 5. PLANTEAMIENTO DE UNA METODOLOGÍA PARA CLASIFICAR UNA COLECCIÓN DE IMDP SEGÚN LAS CARACTERÍSTICAS MORFOLÓGICAS

tiene dos fuentes de error, uno sistemático que es eliminado utilizando un filtro de moda móvil, y un segundo aleatorio que depende de las dimensiones de mapa elegidas en el entrenamiento y los supergrupos. Para asegurar el menor error aleatorio es necesario repetir el procedimiento para asegurar que la serie obtenida es convergente.