

UNIVERSIDAD NACIONAL AUTÓNOMA DE
MÉXICO

FACULTAD DE ESTUDIOS SUPERIORES

ACATLÁN

INTERPOLACIÓN OBJETIVA ESPACIO-
TEMPORAL DE DATOS SATELITALES DE
NIVEL DEL MAR UTILIZANDO FUNCIONES
EMPÍRICAS ORTOGONALES COMPLEJAS

T E S I S
PARA OBTENER EL GRADO DE:
LICENCIADO EN ACTUARÍA
P R E S E N T A:
ARTURO MÉNDEZ SÁNCHEZ

Director de tesis:
Dr. Jorge Zavala Hidalgo



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Quiero dedicar este trabajo a mis padres por su guía, paciencia y,
más que nada, su sacrificio.

Gracias.

Agradecimientos

Quiero agradecer al Dr. Jorge Zavala Hidalgo, por su guía, paciencia y ayuda en la realización de este trabajo, a la M en C. Rosario Romero Centeno, por sus opiniones y correcciones en la escritura de este trabajo, al Actuario Harvey Spencer Sánchez Restrepo, por su amistad y motivación en la búsqueda de este proyecto.

Agradezco a mis maestros Víctor Manuel Ulloa Arellano, Luis Alejandro Tavera Pérez y Juan Carlos Cárdenas Martínez, por sus valiosas sugerencias para la realización de este trabajo y por su amistad, ayuda, guía y confianza en mí.

Agradezco a la Facultad de Estudios Superiores Acatlán, por haberme formado como el profesionalista que ahora soy, y al Centro de Ciencias de la Atmósfera por permitirme realizar mi trabajo en sus instalaciones.

Este trabajo fue realizado con el apoyo del proyecto CONACYT SEP-2003-C02-45634, Circulación oceánica del Golfo de México. Primera etapa: Estudio de los flujos de escala sinóptica a lo largo y a través de la plataforma continental mexicana del Golfo de México mediante modelación numérica y análisis de observación satelital.

| | |
|--|-----------|
| INTRODUCCIÓN..... | 2 |
| CAPÍTULO 1 FUNCIONES EMPÍRICAS ORTOGONALES..... | 4 |
| 1.1 DEFINICIÓN DE LAS FUNCIONES EMPÍRICAS ORTOGONALES..... | 4 |
| 1.2 INTERPRETACIÓN ESTADÍSTICA Y FÍSICA DE LAS FUNCIONES EMPÍRICAS ORTOGONALES..... | 11 |
| CAPÍTULO 2 FUNCIONES EMPÍRICAS ORTOGONALES COMPLEJAS | 14 |
| 2.1 FUNCIONES EMPÍRICAS ORTOGONALES COMPLEJAS | 14 |
| 2.2 LA TRANSFORMADA DE HILBERT | 18 |
| 2.3 DESCOMPOSICIÓN EN VALORES SINGULARES COMPLEJOS | 24 |
| CAPÍTULO 3 LOS DATOS SATELITALES DE ALTIMETRÍA Y SU IMPORTANCIA EN LA OCEANOGRAFÍA..... | 29 |
| 3.1 ANTECEDENTES DE LA ALTIMETRÍA SATELITAL | 29 |
| 3.2 LOS PROYECTOS TOPEX/POSEIDON Y JASON-1 | 31 |
| 3.3 LA VISIÓN DE LOS MARES EN LA ERA DE LA ALTIMETRÍA | 36 |
| CAPÍTULO 4 METODOLOGÍA..... | 39 |
| 4.1 PROCESAMIENTO DE LOS DATOS DE ALTIMETRÍA | 39 |
| 4.2 OBTENCIÓN DE LOS DATOS A LO LARGO DE LAS TRAYECTORIAS DEL SATÉLITE | 41 |
| CAPÍTULO 5 INTERPOLACIÓN ESPACIAL Y/O TEMPORAL USANDO FUNCIONES EMPÍRICAS ORTOGONALES COMPLEJAS | 46 |
| 5.1 DESCRIPCIÓN DE LA TÉCNICA DE INTERPOLACIÓN | 46 |
| 5.2 GENERACIÓN DE MALLAS ORTOGONALES UTILIZANDO FEOCS | 49 |
| <i>Experimento 1</i> | 49 |
| <i>Experimento 2</i> | 54 |
| <i>Experimento 3</i> | 57 |
| <i>Experimento 4</i> | 60 |
| CAPÍTULO 6 RESULTADOS..... | 63 |
| CONCLUSIONES..... | 86 |
| REFERENCIAS | 87 |

Resumen

En este trabajo se propone un método alternativo de interpolación de datos satelitales de nivel del mar utilizando funciones empíricas ortogonales complejas. A partir de series de tiempo reales se aplica la transformada de Hilbert para formar series de tiempo complejas. Los datos se organizan en matrices representando distintos puntos del espacio en los renglones y distintos tiempos de muestreo en las columnas. Posteriormente se realiza una descomposición en valores y vectores propios de la matriz, se obtienen los componentes principales mediante la proyección de los datos originales al multiplicarlos por la matriz de vectores propios. Los elementos complejos de la matriz de componentes principales son interpolados a una malla de mayor densidad en el espacio, interpolando por separado la amplitud y fase de cada componente, para posteriormente sumarlos y recuperar la parte real. Se evalúa el método mediante varios ejemplos sintéticos y con datos de altimetría satelital. Se describe también la altimetría satelital, su historia, técnicas y usos.

Introducción

El lanzamiento de los satélites científicos ha proporcionado a los investigadores acceso a enormes volúmenes de datos, presentando así la oportunidad para el desarrollo de nuevas técnicas de procesamiento y análisis de datos satelitales.

Uno de los instrumentos disponibles para estudios oceanográficos son los satélites de altimetría, los cuales usan un radar de microondas para determinar la elevación instantánea de la superficie del océano. Por ejemplo, el satélite Topex/Poseidon (T/P) que proporciona datos de altimetría, tiene una órbita que se repite aproximadamente cada 10 días, con 6.2 kilómetros entre una observación y otra, y la distancia entre pasos o trayectorias contiguas es de 315 kilómetros. Sin embargo, a pesar de la amplia cobertura del satélite T/P, quedan grandes huecos entre las trayectorias cuya extensión es mayor comparada con los fenómenos de mesoescala¹ (Jacobs et al., 2001). Un ejemplo de estos fenómenos son los remolinos oceánicos, los cuales tienen asociadas variaciones del nivel del mar y cuyos diámetros son de aproximadamente 100 kilómetros. La Corriente del Lazo en el Golfo de México genera remolinos anticiclónicos (que giran en el sentido de las manecillas del reloj) con una altura del nivel del mar de alrededor de 50 centímetros. Cuando se observan remolinos con datos del satélite T/P, aparecen dificultades para detectarlos cuando están localizados entre los pasos del satélite. La interpolación espacial tradicional de datos de satélite da como resultado una discontinuidad de los remolinos en mapas espaciales de altura de la superficie del océano. Como consecuencia, en muchos casos, es necesario utilizar un método de interpolación para formar parte del procesamiento de datos en la creación del conjunto de valores que se analizará (Peng et al., 2004).

Una técnica para analizar la variabilidad de series de tiempo es la de los componentes principales. En el campo oceanográfico, el análisis de funciones empíricas ortogonales (FEOs) retoma muchas de las herramientas de los componentes principales como: la formación de FEOs a través de la matriz de dispersión y la reconstrucción de un campo menos variable al original preservando la mayor variabilidad del conjunto de datos y, en ciencias sociales se le utiliza como análisis factorial, en el cual se busca

¹ La mesoescala es la extensión de fenómenos marinos o atmosféricos que va desde decenas hasta cientos de kilómetros.

encontrar subgrupos con características similares dentro del conjunto general de datos. Las FEOs proveen una descripción compacta de la variabilidad temporal y espacial de las series de datos en términos de funciones ortogonales o modos estadísticos (Emery and Thomson, 1995). Así mismo, la técnica ha sido muy usada para explorar las relaciones espaciales y temporales dentro de un conjunto de datos. Sin embargo, el análisis de FEOs detecta solamente ondas estacionarias y no ondas en movimiento (Horel, 1984). Para estudiar fenómenos que se propagan en forma de ondas se sugiere utilizar una matriz de covarianzas con retrasos o el análisis de funciones empíricas ortogonales complejas.

Las construcciones matemáticas utilizadas en el análisis de FEOs permiten extender las aplicaciones a campos complejos mediante la formación de matrices imaginarias con la transformada de Hilbert y, posteriormente, obteniendo los eigenvectores de la matriz de dispersión. La generalización del método se conoce como análisis de funciones empíricas ortogonales complejas (FEOCs), el cual ayuda a la detección eficiente de patrones o fenómenos que se propagan.

En este trabajo se presenta una alternativa al método de interpolación tradicional de datos basada en el análisis de FEOCs. La técnica de interpolación, que hace uso de las FEOCs, recobra los rasgos de movimiento extrayendo información de un arreglo bidimensional de datos de p puntos en el espacio y n puntos en el tiempo. Los modos estadísticos o eigenvectores asociados con la mayor parte de la varianza de la matriz de datos son escogidos entre todos los modos ya que explican de manera eficiente la variabilidad de los datos. La amplitud y la fase de las FEOCs se colocan en una malla con diferente densidad, ya sea en tiempo o en espacio, y al recuperar la parte real se obtiene el campo interpolado.

Capítulo 1 Funciones Empíricas Ortogonales

En este capítulo se introducen conceptos y fundamentos de las aplicaciones utilizadas, analizando las interpretaciones y aplicaciones físicas, estadísticas y matemáticas de las funciones empíricas ortogonales (FEOs). En muchas ramas de la investigación los componentes principales y las FEOs son herramientas de análisis de datos de gran utilidad.

Los meteorólogos y oceanógrafos han usado las FEOs para estudiar la variabilidad espacial y/o temporal de los campos de variables físicas, diferenciando entre señal y ruido en los datos y encontrando relaciones entre los parámetros que son de bastante interés para ciertos estudios.

La ventaja del análisis con FEOs es el proveer una descripción compacta de la variabilidad espacial y/o temporal de series de tiempo en términos de funciones ortogonales, o “modos” estadísticos. Frecuentemente la mayor parte de la varianza está asociada a las primeras funciones ortogonales, cuyo significado puede vincularse a posibles procesos conocidos.

1.1 Definición de las Funciones Empíricas Ortogonales

En investigaciones estadísticas es común tener un gran número de variables de estudio, las cuales pueden estar correlacionadas entre sí. Esta correlación entre las variables se representa por medio de una matriz de varianzas-covarianzas \mathbf{R} , y a las funciones empíricas ortogonales se les define ordinariamente como los eigenvectores de la matriz \mathbf{R} . La construcción de la matriz de componentes principales se hace proyectando la información original sobre las FEOs para así estudiar un conjunto nuevo de variables incorrelacionadas entre si, o en otras palabras se utilizan los eigenvectores provenientes de los modos que expliquen el mayor porcentaje de varianza para así generar nuevas variables que sean linealmente independientes. El análisis mediante FEOs sirve para estudiar campos en que la información se presenta en arreglos bidimensionales, una dimensión representa

los puntos de estudio, la otra dimensión representa tiempos distintos en que se realizan (Hartmann, 2003). Las FEOs nacen de un problema estadístico, el cual busca la obtención de datos de estudio que no estén correlacionados entre sí, formando combinaciones lineales de la información original, de tal manera que se represente en forma compacta la variabilidad de los datos iniciales con un número menor de FEOs.

Los puntos y sus n mediciones son representados en la matriz \mathbf{X} donde los renglones representan los puntos del campo de interés y las columnas diferentes tiempos estudiados y, así la matriz toma la forma $\mathbf{X} = \{x_{ij} : 1 \leq i \leq p, 1 \leq j \leq n\}$, o en forma más explícita

$$\mathbf{X} = \begin{pmatrix} x_{11} & \dots & x_{1n} \\ \dots & \dots & \dots \\ x_{p1} & \dots & x_{pn} \end{pmatrix} = [x_1 \dots x_p]^T .$$

El procedimiento común para la obtención de las FEOs resta el promedio temporal a cada serie de tiempo y usualmente se divide entre su desviación estándar.

En el análisis de campos geofísicos, frecuentemente se esta interesado en las variaciones sobre escalas de tiempo desconocidas, esto es, si en los datos se conocen ciclos anuales, su detección es usualmente irrelevante a los resultados. En casos como este, sería deseable remover las variaciones conocidas antes de aplicar el análisis con FEOs. Esto puede ser realizado calculando el ciclo anual climatológico y restándolo del campo de estudio. Así quedan solo las desviaciones del ciclo anual, o las anomalías de interés para la realización del estudio. Puede generalizarse el proceso anterior removiendo otras variaciones conocidas en los datos los cuales pueden ser asociados a otras escalas de tiempo, las cuales se deseen excluir del análisis por alguna razón particular. En todo caso la elección queda en manos del investigador y el tipo de estudio a realizar.

El promedio temporal de la i -ésima variable se estima con la siguiente fórmula:

$$\mu_i = \frac{\sum_{j=1}^n x_{ij}}{n} ,$$

el promedio μ_i es restado a cada serie de tiempo, formando elementos y_{ij} de la forma

$$y_{ij} = x_{ij} - \frac{\sum_{j=1}^n x_{ij}}{n} = x_{ij} - \mu_i \quad ,$$

y el dividir los y_{ij} entre su desviación estándar depende del tipo de investigación realizada. La estructura de $\mathbf{Y} = \{y_{ij}\}$ está organizada de la siguiente forma: los renglones representan las anomalías de las posiciones de estudio y las columnas simbolizan el número de observaciones o pasos en el tiempo,

$$\begin{array}{c} \text{Observaciones} \rightarrow \dots \\ \mathbf{Y} = \begin{pmatrix} y_{11} & \dots & y_{1n} \\ \dots & \dots & \dots \\ y_{p1} & \dots & y_{pn} \end{pmatrix} \dots \downarrow \text{Variables} \quad . \end{array}$$

La matriz de dispersión, \mathbf{R} , se obtiene mediante:

$$\mathbf{R} = \mathbf{Y} \mathbf{Y}^T / (n-1) \quad ,$$

$$\mathbf{R} = (r)_{ij} = \sum_{l=1}^n \frac{y_{il} y_{jl}}{n-1} \quad , \text{ donde}$$

$$\mathbf{R} = \begin{pmatrix} \sum_{i=1}^n \frac{y_{1i} y_{1i}}{n-1} & \dots & \sum_{i=1}^n \frac{y_{1i} y_{pi}}{n-1} \\ \dots & \dots & \dots \\ \sum_{i=1}^n \frac{y_{pi} y_{1i}}{n-1} & \dots & \sum_{i=1}^n \frac{y_{pi} y_{pi}}{n-1} \end{pmatrix} = \begin{pmatrix} r_{11} & \dots & r_{1p} \\ \dots & \dots & \dots \\ r_{p1} & \dots & r_{pp} \end{pmatrix} \quad ,$$

obsérvese que $r_{ij} = r_{ji}$, lo cual implica que la matriz \mathbf{R} es simétrica y cuadrada. Utilizando resultados de la teoría del álgebra lineal, la matriz \mathbf{R} tiene asociado un conjunto de p vectores propios ortonormales y p valores propios, los cuales son utilizados para la formación de las FEOs. La matriz \mathbf{R} tiene tamaño $p \times p$, pero usualmente solo sus primeros k valores propios son distintos de cero, donde $k \leq \min(p, n)$. Por lo tanto, la dimensión de \mathbf{R} es k . Solamente k FEOs pueden ser determinadas debido a que cada FEO está asociada a un

porcentaje de varianza de los datos, pero si la varianza asociada a esa FEO es nula entonces no se toma en cuenta.

Una alternativa interesante para el cálculo de las FEOs se da obteniendo el siguiente producto matricial:

$$\mathbf{S} = \mathbf{Y}^T \mathbf{Y} / (n-1) \quad \text{ó}$$

$$\mathbf{S} = (s)_{ij} = \sum_{k=1}^p \frac{y_{ki} y_{jk}}{n-1} \quad ,$$

cuya diagonal estima la varianza total de las posiciones de estudio, dado que:

$$\text{diag}(\mathbf{R}) = \sum_{k=1}^p r_{kk} = \sum_{k=1}^p \sum_{l=1}^n \frac{y_{kl}^2}{n-1} \quad \text{intercambiando las sumas}$$

$$\sum_{l=1}^n \sum_{k=1}^p \frac{y_{kl}^2}{n-1} = \sum_{l=1}^n s_{ll} = \text{diag}(\mathbf{S}) \quad ,$$

de esta forma se da una alternativa mas para el cálculo de la dispersión de un conjunto de datos. Aunque la interpretación de \mathbf{S} no es del todo obvia tiene una propiedad útil para el cálculo de las FEOs, debido a que presenta una alternativa para el caso cuando $p > n$ en donde se tendrá un mayor trabajo computacional para el cálculo de los vectores propios de \mathbf{R} al de \mathbf{S} .

\mathbf{R} representa la matriz de varianzas-covarianzas de las variables, teniendo en la diagonal la varianza de cada variable. La matriz \mathbf{S} , estima la variabilidad de las series de tiempo con respecto a cada campo de interés en los renglones. A primera vista las matrices parecen ser distintas pero ambas matrices ayudan a estimar la varianza total de los datos resultado de la suma de la diagonal de ambas matrices.

Para encontrar los patrones en los datos se procede a descomponer la matriz \mathbf{R} en valores y vectores propios resolviendo la siguiente ecuación:

$$\mathbf{R}\mathbf{E} = \mathbf{E}\mathbf{\Lambda} \quad ,$$

donde \mathbf{R} , \mathbf{E} y $\mathbf{\Lambda}$ son matrices de tamaño $p \times p$, $\mathbf{\Lambda}$ es una matriz que contiene en la diagonal los valores propios de \mathbf{R} en orden decreciente con $\lambda_i > \lambda_j$ para $i > j$. La matriz $\mathbf{\Lambda}$ es de la forma:

$$\mathbf{\Lambda} = (\delta_{ij}\lambda_{ij}) = \begin{pmatrix} \lambda_1 & \dots & 0 \\ \dots & \dots & \dots \\ 0 & \dots & \lambda_p \end{pmatrix}, \delta_{ij} \text{ la delta de kroneker.}$$

Como la matriz \mathbf{R} es real y simétrica eso implica que los valores propios son reales y mayores e iguales a cero. Por otro lado \mathbf{E} es una matriz que contiene en sus columnas los vectores propios de \mathbf{S} ; los vectores \mathbf{e}_i . La matriz \mathbf{E} es de la forma:

$$\mathbf{E} = \begin{pmatrix} \mathbf{e}_{11} & \dots & \mathbf{e}_{p1} \\ \dots & \dots & \dots \\ \mathbf{e}_{1p} & \dots & \mathbf{e}_{pp} \end{pmatrix} = [\mathbf{e}_1 \dots \mathbf{e}_p] \quad .$$

La propiedad $\mathbf{E}\mathbf{E}^T = \mathbf{I}_p$, donde \mathbf{I}_p es la matriz identidad de tamaño $p \times p$, significa que los vectores propios son ortogonales con respecto al tiempo o a las observaciones muestreadas, en términos de columnas \mathbf{e}_i es de tamaño $p \times 1$ $\{1 \leq i, j \leq p\}$; $\mathbf{e}_i^T \mathbf{e}_j = \delta_{i,j}$, es decir $\delta_{i,j} = 1$ si $i = j$ y 0 en otro caso. Cada vector propio está asociado a una función empírica ortogonal, recibe el nombre de empírica debido a que proviene de los datos y ortogonal porque está incorrelacionada con respecto al tiempo o a las variables según sea \mathbf{R} ó \mathbf{S} .

El método de análisis de FEOs o componentes principales esta muy documentado por varios autores, lo que hace la diferencia son las distintas aplicaciones en las que es utilizado. Por ejemplo, en la mayoría de las aplicaciones estadísticas de los componentes principales puede encontrarse el cribado de datos (localización de datos outliers), agrupación de elementos, análisis discriminante y aplicaciones en la regresión lineal, por mencionar algunos ejemplos.

Dado el problema a resolver para obtener los valores propios; $\mathbf{R}\mathbf{E} = \mathbf{E}\mathbf{\Lambda}$ y la propiedad de la matriz ortogonal \mathbf{E} ; $\mathbf{E}\mathbf{E}^T = \mathbf{E}^T\mathbf{E} = \mathbf{I}_p$, se tiene que, $\mathbf{R} = \mathbf{E}\mathbf{\Lambda}\mathbf{E}^T$. Esta identidad es valida dado que $\mathbf{E}^T = \mathbf{E}^{-1}$, por lo que la ecuación anterior puede escribirse como $\mathbf{R} = \mathbf{E}\mathbf{\Lambda}\mathbf{E}^{-1}$, y en forma escalar como:

$$r_{ij} = \sum_{l=1}^p \mathbf{e}_{il} \mathbf{e}_{jl} \lambda_l \quad .$$

Los elementos de la diagonal de \mathbf{R} estiman la varianza de cada variable es decir

$$\sum_{k=1}^p r_{kk} = \sum_{k=1}^p \sum_{l=1}^p e_{kl}^2 \lambda_l = \sum_{l=1}^p \sum_{k=1}^p \lambda_l e_{kl}^2 =$$

$$\sum_{l=1}^p \lambda_l \sum_{k=1}^p e_{lk}^2 = \sum_{l=1}^p \lambda_l \quad \text{dado que } \sum_{l=1}^p e_{lk}^2 = 1 .$$

$\sum_{k=1}^p r_{kk} = \sum_{l=1}^p \lambda_l$ señala que el total de la variabilidad también puede ser estimada con la

suma de los elementos de la diagonal de la matriz \mathbf{L} , que contiene los valores propios de \mathbf{R} .

Comúnmente la mayor parte de la varianza se concentra en pocos λ s. El porcentaje de varianza asociado a cada λ se estima mediante el cálculo:

$$\frac{\lambda_i}{\lambda_1 + \lambda_2 + \dots + \lambda_p} = i * 100 = \text{porcentaje de varianza asociado al valor propio } \lambda_i.$$

Para obtener toda la varianza se tiene:

$$\frac{\lambda_1 + \lambda_2 + \dots + \lambda_p}{\lambda_1 + \lambda_2 + \dots + \lambda_p} = i * 100 = 100\% \text{ de la varianza explicada.}$$

Realizando la proyección de las FEOs sobre las perturbaciones \mathbf{Y} , se indaga en el comportamiento de las variables al paso del tiempo, mismas que ahora se proyectan en un espacio coordinado en el cual no hay correlación entre ellas, es decir en términos estadísticos son linealmente independientes. Matricialmente se forma una matriz \mathbf{A} que comúnmente se le llama matriz de componentes principales,

$$\mathbf{A} = \mathbf{E}^T \mathbf{Y}, \mathbf{A} = \{a_{ij}: 1 \leq i \leq p, 1 \leq j \leq n\} \quad ,$$

$$a_{ij} = \sum_{l=1}^p e_{li} y_{lj} = e_{1i} y_{1j} + \dots + e_{pi} y_{pj} \quad .$$

La representación matricial de la matriz \mathbf{Y} en términos de componentes principales es la siguiente:

$$\mathbf{Y} = \mathbf{I}_n \mathbf{Y} = \mathbf{E} \mathbf{E}^T \mathbf{Y} = \mathbf{E} (\mathbf{E}^T \mathbf{Y}) \text{ y dado que } \mathbf{A} = \mathbf{E}^T \mathbf{Y}$$

$$\mathbf{Y} = \mathbf{E} \mathbf{A} \quad ,$$

o en notación escalar

$$y_{ij} = \sum_{l=1}^p e_{il} a_{lj} \quad .$$

Y cuando el porcentaje de varianza deseado es obtenido con $h < k$ eigenvalores se define una matriz \mathbf{E}' en cuyas columnas estarán los eigenvectores utilizados para representar el porcentaje de varianza deseado. El tamaño de \mathbf{A}' es de $h \times p$

$$a_{ij} = \sum_{l=1}^h e_{li} y_{lj} \quad , \quad \mathbf{A}' = \mathbf{E}'^T \mathbf{Y} \quad \text{donde} \quad \mathbf{E}' = [\mathbf{e}_1 \dots \mathbf{e}_h] \quad .$$

El objetivo principal de la descomposición en FEOs es la reconstrucción de un campo aproximado, compacto y menos variable que el original. Para compactar los datos y obtener la mayor parte de su varianza, se realiza la multiplicación de la matriz \mathbf{E}' por la matriz de componentes \mathbf{A}' que es la expresión de los datos originales en un nuevo sistema coordenado, a este análisis algunas veces se le llama rotación de FEOs. La matriz $\hat{\mathbf{Y}}$ y sus elementos \hat{y}_{ij} se calculan de acuerdo con:

$$\hat{y}_{ij} = \sum_{l=1}^h e_{li} a_{lj} \quad , \quad \hat{\mathbf{Y}} = \mathbf{E}' \mathbf{A}' \quad .$$

La matriz $\hat{\mathbf{Y}}$ retiene la mayoría de la varianza del campo \mathbf{Y} . El porcentaje de varianza que se desea preservar o, en términos prácticos, el número $h < p$ a seleccionar, que es el número de valores propios deseados para el tamaño de \mathbf{E} ; es elegido por cada investigador de acuerdo con el problema particular que se está estudiando.

Y en el caso de utilizar la matriz \mathbf{S} , se tiene la alternativa de resolver el problema matricial

$$\mathbf{S} \mathbf{U} = \mathbf{\Gamma} \mathbf{U} \quad ,$$

$$\mathbf{\Gamma} = \{\gamma_{ij} \delta_{ij}\}$$

donde la matriz \mathbf{U} contiene los vectores propios de la matriz \mathbf{S} y $\mathbf{\Gamma}$ los valores propios asociados a \mathbf{S} , donde \mathbf{S} , \mathbf{U} y $\mathbf{\Gamma}$ son de tamaño $n \times n$. Posteriormente para la obtención de \mathbf{E} , se necesita realizar primeramente el siguiente producto matricial:

$$\mathbf{D} = \mathbf{Y} \mathbf{U} \quad , \text{ donde } \mathbf{D} \text{ tiene tamaño } p \times n$$

y la relación que hay entre \mathbf{D} y \mathbf{E} es la siguiente:

$$\mathbf{E} = [\mathbf{e}_1 \ \mathbf{e}_2 \ \dots \ \mathbf{e}_n] \quad \text{y}$$

$$\mathbf{D} = [\mathbf{d}_1 \ \mathbf{d}_2 \ \dots \ \mathbf{d}_n]$$

$$\mathbf{e}_k = l_k \mathbf{d}_k \quad , \quad 1 \leq k \leq p.$$

Donde l_k es el k -ésimo valor singular (ver sección 2.3) de la matriz \mathbf{Y} . Con esta aproximación se recuperan solo p vectores propios de \mathbf{E} , pero son los que se asocian a una varianza positiva de la matriz de dispersión \mathbf{R} . Es decir, la totalidad las FEOs asociadas a la matriz \mathbf{Y} son recuperadas con ayuda de la matriz \mathbf{S} .

1.2 Interpretación estadística y física de las funciones empíricas ortogonales.

Para la interpretación física y estadística de las FEOs se puede recurrir a su significado matemático o al problema que les dió origen. En ambos casos se tiene que las FEOs representan la varianza tan eficientemente como sea posible, siendo ortogonales. Algunas veces esas construcciones matemáticas dejan fuera ciertas estructuras científicamente interesantes, pero eso no ocurre frecuentemente.

Basado en la eficiencia de la descomposición estadística, un conjunto de pocos FEOs generalmente es suficiente para describir la variabilidad dentro de un gran conjunto de datos. Frecuentemente es aconsejable usar FEOs como filtro para eliminar escalas de variabilidad indeseables. En muchos casos, un número pequeño de FEOs es usado para reconstruir una buena aproximación del conjunto original de datos. Haciendo eso se pueden filtrar escalas de variabilidad que no son coherentes en la entrada del campo de datos y,

además, son menos importantes en su contribución a la varianza del mismo. Debe observarse que el análisis con FEOs puede producir estructuras parecidas a ondas, pero eso no se debe interpretar como que los datos contienen estructuras de onda coherentes. A continuación se hacen algunas sugerencias para identificar si las FEOs producidas en un análisis reflejan estructuras en los datos:

- La varianza de los datos explicada por las FEOs es mayor de la que se esperaría si los datos no tuvieran estructura.
- Se tiene alguna razón a priori para esperar la estructura que se está encontrando. Por ejemplo, se explica la estructura encontrada en términos de alguna teoría.
- El qué tan robustas son las estructuras ante la selección del dominio. En otras palabras, si se cambia el dominio del análisis, la estructura sufre un cambio significativo o no. Por ejemplo si la estructura está definida en un espacio geográfico, y si se cambia el tamaño de la región, entonces cambia la estructura significativamente o no.
- Qué tan robustas son las estructuras en el muestreo usado, es decir, si se divide el muestreo dentro de una selección aleatoria reducida y se hace el análisis sobre cada mitad, se consigue consistentemente o no la misma estructura (Hartmann D. L., 2003).

En la interpretación del significado de las FEOs, se necesita tener en mente que mientras las FEOs ofrecen una comprensión de la variabilidad estadística del campo de datos, ello no necesariamente corresponde a modos dinámicos reales. Esto se da en el caso de un proceso físico individual que se propaga a través de varios modos. En otros casos, más de un proceso físico puede contribuir a la varianza contenida en una FEO. Los modos estadísticos derivados del procedimiento de FEOs deben ser considerados en un buen sentido para aceptar mecanismos físicos, no asociar algún proceso físico mediante los modos obtenidos. Es probable que la fuerte variabilidad asociada con los modos dominantes se atribuya a procesos físicos reconocibles. Una pista de los mecanismos físicos asociados con el modo de una FEO algunas veces puede ser encontrada en un componente principal de \mathbf{A} . Algo puede ser conocido acerca de la variabilidad temporal de un proceso, que puede ser reflejado en el componente principal. Esto puede sugerir una relación que no

es claramente aparente observando en la estructura espacial de la FEO solamente. Los modelos de FEOs son construidos para representar de manera óptima la varianza en un campo y no conexiones físicas o correlaciones máximas. Cuando no se incluyen modos de eigenvalores pequeños, se puede enfocar la atención en aquellos vectores propios donde los valores propios tienen un significado estadísticamente amplio, los cuales, regularmente, son los primeros valores propios. Desafortunadamente, un significado estadístico no garantiza una interpretación física. Desde un punto de vista estadístico los primeros vectores propios en la expansión no son más que combinaciones lineales de los y_{ij} que representan la estructura de la matriz de covarianza más eficientemente. Para establecer el significado físico de esos vectores propios es necesario mostrar, además, que las estructuras que representan son consistentes con leyes físicas (esto no solamente se aplica a las FEOs sino también a toda herramienta estadística usada de manera diagnóstica para definir la estructura de un fenómeno físico). Para este fin, es altamente deseable que investigaciones de ese tipo sean conducidas en conjunción con estudios de modelación dinámica y/o gráfica (Wallace J. M. 1972).

Capítulo 2 Funciones Empíricas Ortogonales Complejas

Los componentes principales han sido ampliamente usados para explorar la relación espacial y temporal de grandes cantidades de información. La ventaja de los componentes principales es su habilidad para expresar la variabilidad del conjunto de datos con un número pequeño de modos o FEOs. Sin embargo, el análisis convencional de componentes principales detecta solamente oscilaciones estacionarias, no oscilaciones en movimiento. Para el estudio de fenómenos ondulatorios en movimiento se puede usar el análisis de Funciones Empíricas Ortogonales Complejas (FEOCs).

En la sección 2.1 de este capítulo se documenta la generalización del tratamiento de matrices sobre campos reales a campos complejos, presentando la teoría para aplicar análisis de FEOs sobre dominios imaginarios o reales. En la sección 2.2 se define la transformada de Hilbert, se estudian algunas de sus propiedades con ayuda de la teoría de series de Fourier. La propiedad de la transformada de Hilbert más utilizada en este trabajo es aquella que transforma una serie real en una serie compleja, agregando información a la serie de tiempo y obteniendo los parámetros de la amplitud y fase en las series de datos resultantes.

2.1 Funciones empíricas ortogonales complejas

Los fenómenos físicos que pueden representarse con variables complejas son los que pueden depender de dos parámetros, como el movimiento de cuerpos (vientos, corrientes, etc.) o en el análisis de espectros de frecuencia de series de Fourier; por mencionar algunos ejemplos. Cuando se está trabajando con vientos, la parte real puede ser interpretada como el movimiento zonal y la imaginaria como el movimiento meridional.

En principio se darán definiciones generales acerca de operaciones utilizadas en matrices de entradas complejas. La estructura de las matrices será representada por \mathbf{X} , una matriz de p renglones y n columnas (\mathbf{X} de dimensiones pxn), con elementos $\{x_{ij} = a_{ij} + i b_{ij}, i = \sqrt{-1}, 1 \leq i \leq p, 1 \leq j \leq n\}$, donde a_{ij} es la parte real de x_{ij} ($a_{ij} = \Re(x_{ij})$) y b_{ij} es la parte imaginaria de x_{ij} ($b_{ij} = \Im(x_{ij})$). El conjugado complejo de x_{ij} se denota por x_{ij}^* y está definido como:

$$x_{ij}^* = a_{ij} - i b_{ij}$$

La norma $\|x_{ij}\|^2$ y el módulo $\|x_{ij}\|$ de x_{ij} se definen como:

$$\|x_{ij}\|^2 = x_{ij} x_{ij}^* = a_{ij}^2 + b_{ij}^2$$

$$\|x_{ij}\| = (x_{ij} x_{ij}^*)^{1/2}$$

La operación transpuesta de \mathbf{X} , denotada por \mathbf{X}^T , se define como:

$$\mathbf{X}^T = \{x_{ji}\}.$$

Otra operación matricial necesaria es el conjugado de la matriz $\mathbf{X} = \{x_{ij}\}$, definida como $\mathbf{X}^* = \{x_{ij}^*\}$. Algunas propiedades de estas operaciones son las siguientes:

$$(\mathbf{X}^*)^* = \mathbf{X}$$

$$(\mathbf{X}^T)^T = \mathbf{X}$$

Si \mathbf{X} es de pxn y \mathbf{Y} de nxm

$$(\mathbf{XY})^T = \mathbf{Y}^T \mathbf{X}^T$$

$$(\mathbf{XY})^* = \mathbf{X}^* \mathbf{Y}^*$$

La identidad $(\mathbf{X}^*)^T = (\mathbf{X}^T)^*$ se denota por \mathbf{X}^H , es decir $\mathbf{X}^{*T} = \mathbf{X}^{T*} = \mathbf{X}^H$, y se conoce como el hermitiano conjugado de la matriz \mathbf{X} . Una matriz cuadrada \mathbf{S} de tamaño nxn es hermitiana simétrica si $\mathbf{S}^H = \mathbf{S}$, lo cual es una analogía con una matriz simétrica \mathbf{R} de entradas reales que cumple con la propiedad $\mathbf{R}^T = \mathbf{R}$. Como en el caso real, la matriz \mathbf{X} de tamaño pxn representa las p variables en un muestreo de tamaño n , donde el muestreo puede consistir de observaciones de un fenómeno de interés en n distintos tiempos, y de esta manera obtener p series de tiempo de tamaño n .

Posteriormente se forma una matriz $\mathbf{Y} = \{y_{ij}\}$, que contiene las perturbaciones obtenidas al sustraer la media muestral a cada variable:

$$y_{ij} = x_{ij} - \frac{\sum_{j=1}^n x_{ij}}{n} = x_{ij} - \mu_i$$

donde μ_i contiene tanto el promedio real como el imaginario debido a que las partes reales se suman entre ellas y lo mismo las partes imaginarias. Las matrices de dispersión **R** y **S** se obtienen respectivamente de:

$$\mathbf{R} = \mathbf{Y}\mathbf{Y}^H/(n-1)$$

$$\mathbf{S} = \mathbf{Y}^H\mathbf{Y}/(n-1)$$

La carga computacional es menor al calcular los eigenvectores de una matriz de 100×100 que de una matriz de 1000×1000 . Si la matriz de datos tiene tamaño 1000×100 es decir 1000 posiciones de estudio y 100 muestreos puede ser utilizada la matriz **S** para calcular la varianza de los datos. Este tipo de matrices se obtienen en estudios con datos satelitales de altimetría donde el satélite toma muestras de todo el planeta, en periodos de 10 días, en donde 100 muestreos corresponden a 1000 días de investigación, poco más de tres años.

Ambas matrices sirven para estimar la varianza de los datos dado que tienen las propiedades de sus análogas reales y, debido a que generalmente solo se utilizan los primeros modos para representar la varianza total de los datos, puede utilizarse **R** o **S** según sea $p < n$ ó $n < p$, **R** es de $p \times p$, y **S** es de $n \times n$.

R y **S** tienen elementos complejos fuera de su diagonal, son hermitianas y su descomposición en valores propios contiene elementos complejos en la matriz de eigenvectores y eigenvalores reales mayores o iguales a cero.

El problema a resolver en el caso de **S**, es el siguiente:

$$\mathbf{S}\mathbf{E} = \mathbf{E}\mathbf{\Lambda}$$

De donde se obtiene como resultado una matriz $\mathbf{\Lambda}$ de $n \times n$ con los eigenvalores de **S** en su diagonal ordenados de manera descendente y donde sólo algunos de ellos son

distintos de cero. La matriz \mathbf{E} de $n \times n$ resulta ser unitaria (es decir el equivalente a ortogonal en el campo real), \mathbf{E} contiene los eigenvectores complejos \mathbf{e}_i , asociados a un porcentaje del campo de varianza de estudio, proporcional al eigenvalor λ_i . Un cambio contrastante con respecto al análisis de FEOs sobre campos reales es que los \mathbf{e}_i tienen elementos complejos e_{ij} . Por la identidad de Euler los e_{ij} pueden ser representados en términos de su amplitud r_{ij} y su fase θ_{ij} .

\mathbf{E} con elementos $e_{ij} = r_{ij} \exp(i\theta_{ij})$

$$r_{ij} = \left((\Re(e_{ij}))^2 + (\Im(e_{ij}))^2 \right)^{1/2}$$

$$\theta_{ij} = \arctan \left(\frac{\Im(e_{ij})}{\Re(e_{ij})} \right)$$

En consecuencia, la matriz \mathbf{A} de componentes principales, resultado de la proyección de \mathbf{E} sobre la matriz de perturbaciones \mathbf{Y} , tiene elementos complejos.

$$\mathbf{A} = \mathbf{E}^T \mathbf{Y}$$

$$a_{ij} = p_{ij} \exp(i\phi_{ij})$$

$$p_{ij} = \left((\Re(a_{ij}))^2 + (\Im(a_{ij}))^2 \right)^{1/2}$$

$$\phi_{ij} = \arctan \left(\frac{\Im(a_{ij})}{\Re(a_{ij})} \right)$$

El objetivo en las FEOs es la comprensión del campo de estudio en una versión menos ruidosa y variable. La matriz $\hat{\mathbf{Y}}$ puede ser construida proyectando esos modos provenientes de los eigenvalores que explican un mayor porcentaje de varianza y sus correspondientes vectores propios que forman una matriz \mathbf{E}' . Tomando los primeros h términos

$$a'_{ij} = \sum_{l=1}^h \mathbf{e}_l y_{li}, \quad \mathbf{A}' = \mathbf{E}'^H \mathbf{Y} \quad \text{donde} \quad \mathbf{E}' = [\mathbf{e}_1 \dots \mathbf{e}_h]$$

$$\hat{y}_{ij} = \sum_{l=1}^h e_{il} a'_{lj}, \quad \hat{\mathbf{Y}} = \mathbf{E}' \mathbf{A}' \quad .$$

Debido a las propiedades análogas de \mathbf{E} y \mathbf{E}' .

$$\mathbf{E} \mathbf{E}^H = \mathbf{E}^H \mathbf{E} = \mathbf{I}_n \quad .$$

Y cuando $h = n$ $\mathbf{E}' = \mathbf{E}$, pero en caso contrario hay dos casos

$$\mathbf{E}' \mathbf{E}'^H \approx \mathbf{I}_n \quad .$$

Se tiene una aproximación a \mathbf{I}_n que casi se garantiza la igualdad cuando se utilizan los modos que explican al menos el 90% de la varianza explicada de los datos. Análogamente para $h < n$

$$\mathbf{E}'^H \mathbf{E}' = \mathbf{I}_h \quad .$$

2.2 La transformada de Hilbert

Se mostrará que esta transformada es un caso particular de una convolución de la Transformada de Fourier (TF) denotada por “ \mathbf{F} ” de 2 funciones. Como se mencionó anteriormente la Transformada de Hilbert (TH) denotada por “ \mathbf{H} ” ayuda a la formación de series de tiempo complejas a partir de la serie de tiempo original. La transformada de Hilbert es la siguiente:

$$\mathbf{H}(f(t)) = H(x) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{f(t)}{t-x} dt = f(t) * \left(\frac{-1}{\pi x} \right)$$

La convolución de 2 funciones está definida como:

$$f(x) * g(x) = \int_{-\infty}^{\infty} f(t) g(x-t) dt$$

Y la transformada inversa de Hilbert será

$$f(t) = \mathbf{H}^{-1}(H(x)) = -\frac{1}{\pi} \int_{-\infty}^{\infty} \frac{H(x)}{x-t} dx = -\mathbf{H}(H(x))$$

(Lokenath, 1995) es decir la inversa de la TH es “menos TH”. Las propiedades de la TH se presentan sin prueba a continuación:

- 1) $\mathbf{H}(f(t+a)) = H(x+a)$, donde a es una constante

$$2) \mathbf{H}(f(at)) = H(ax)$$

$$3) \mathbf{H}(f(-at)) = -H(-ax)$$

$$4) \mathbf{H}(f'(t)) = \frac{d}{dx} H(x)$$

$$5) \mathbf{H}(tf(t)) = xH(x) + \frac{1}{\pi} \int_{-\infty}^{\infty} f(t) dt$$

$$6) \mathbf{F}(\mathbf{H}(f(t))) = (i \operatorname{sgn}(k)) \mathbf{F}(f(t)), i = \sqrt{-1}$$

$$7) \|\mathbf{H}(f(t))\| = \|f(t)\|, \text{ donde la norma de } f, \|f(t)\| = \left[\int_{-\infty}^{\infty} f(x) \overline{f(x)} dx \right]^{1/2} = \left[\int_{-\infty}^{\infty} |f(x)|^2 dx \right]^{1/2}$$

$$8) \mathbf{H}(f(t)+g(t)) = H(x) + G(x)$$

La transformada de Fourier definida por

$$\mathbf{F}(f(x)) = F(k) = \int_{-\infty}^{\infty} f(x) e^{-ikx} dx$$

La función signo de k se define como

$$\operatorname{sgn}(k) = 1, \text{ si } k > 0 \text{ ó } -1, \text{ si } k < 0$$

Las TH del seno y el coseno son de interés para el trabajo y por eso serán desarrolladas:

$$\mathbf{H}(\operatorname{sen}(wt)) = H(x) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\operatorname{sen}(wt)}{t-x} dt$$

$$= \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\operatorname{sen}(w(t-x) + wx)}{t-x} dt$$

$$= \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\operatorname{sen}(w(t-x)) \cos(wx) + \cos(w(t-x)) \operatorname{sen}(wx)}{t-x} dt$$

aplicando identidades trigonométricas se obtiene

$$\frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\operatorname{sen}(w(t-x)) \cos(wx)}{t-x} dt + \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\cos(w(t-x)) \operatorname{sen}(wx)}{t-x} dt$$

$$= \frac{\cos(wx)}{\pi} \int_{-\infty}^{\infty} \frac{\operatorname{sen}(w(t-x))}{t-x} dt + \frac{\operatorname{sen}(wx)}{\pi} \int_{-\infty}^{\infty} \frac{\cos(w(t-x))}{t-x} dt$$

haciendo $t-x = T$

$$= \frac{\cos(wx)}{\pi} \int_{-\infty}^{\infty} \frac{\text{sen}(wT)}{T} dT + \frac{\text{sen}(wx)}{\pi} \int_{-\infty}^{\infty} \frac{\cos(wT)}{T} dT$$

La segunda integral es impar y en un dominio simétrico de integración respecto a cero la integral vale cero. La primera integral es par y resulta 2 veces el valor de la integral evaluada de cero a infinito

$$\begin{aligned} & \frac{\cos(wx)}{\pi} \int_{-\infty}^{\infty} \frac{\text{sen}(wT)}{T} dT + \frac{\text{sen}(wx)}{\pi} \int_{-\infty}^{\infty} \frac{\cos(wT)}{T} dT \\ &= \frac{\cos(wx)}{\pi} \int_{-\infty}^{\infty} \frac{\text{sen}(wT)}{T} dT = \frac{\cos(wx)}{\pi} 2 \int_0^{\infty} \frac{\text{sen}(wT)}{T} dT . \end{aligned}$$

Utilizando siguiente resultado de cálculo integral:

$$\int_0^{\infty} \frac{\text{sen}(wT)}{T} dT = \frac{\pi}{2} ,$$

se concluye que

$$\frac{\cos(wx)}{\pi} 2 \frac{\pi}{2} = \cos(wx) .$$

Por lo tanto $\mathbf{H}(\text{sen}(wt)) = \cos(wx)$.

Similarmente se puede demostrar que $\mathbf{H}(\cos(wt)) = -\text{sen}(wx)$.

Hay una relación interesante entre la TH y la TF. La TF puede ser expresada como la suma de una función real par y una imaginaria pura impar (Hsu, 1973). Ahora si se obtiene la TF de una función par o impar, el resultado será una función real o una función imaginaria pura respectivamente. Para una función causal¹ la parte real de la TF será la

¹ Si $f(t)$ es causal implica $f(t) = 0$ si $t < 0$

transformada de Hilbert de la parte imaginaria, e inversamente la transformada de Hilbert de la parte imaginaria será la parte real. Esto se muestra a continuación.

La TF de $f(t)$

$$F(f(t)) = F(k) = \int_{-\infty}^{\infty} f(t)e^{-ikt} dt$$

puede ser escrita como:

$$\begin{aligned} F(f(t)) = F(k) &= \int_{-\infty}^{\infty} f(t)\cos(kt)dt + i(-1) \int_{-\infty}^{\infty} f(t)\text{sen}(kt)dt \\ &= R(k) + iI(k) \end{aligned}$$

donde

$$R(k) = \int_{-\infty}^{\infty} f(t)\cos(kt)dt \quad ,$$

$$I(k) = (-1) \int_{-\infty}^{\infty} f(t)\text{sen}(kt)dt \quad .$$

De las integrales anteriores se puede ver que si $f(t)$ es una función par $I(k)$ será cero. Por otro lado si $f(t)$ es impar $R(k)$ será cero. Dado que toda función $f(t)$ puede ser expresada como la suma de una función par y una función impar.

$$f(t) = \frac{\{f(t) + f(-t)\}}{2} + \frac{\{f(t) - f(-t)\}}{2}$$

$$= f_p(t) + f_i(t) \quad , \text{ donde}$$

$$f_p(t) = \frac{\{f(t) + f(-t)\}}{2} \quad ,$$

$$f_i(t) = \frac{\{f(t) - f(-t)\}}{2} \quad ,$$

$$F(f_p(t)) = R(k) \text{ y}$$

$$F(f_i(t)) = iI(k) .$$

Dado que $f(t)$ es causal se tiene que para $t < 0$

$$\begin{aligned} f_p(t) &= \frac{1}{2}f(-t) \\ f_i(t) &= -\frac{1}{2}f(-t) \quad , y \\ f_p(t) &= -f_i(t) \end{aligned}$$

y para $t > 0$

$$\begin{aligned} f_p(t) &= \frac{1}{2}f(t) \quad y \\ f_i(t) &= \frac{1}{2}f(t) \end{aligned}$$

es decir

$$f_p(t) = f_i(t) .$$

En otras palabras

$$\begin{aligned} f_p(t) &= f_i(t) \operatorname{sgn}(t) \quad \text{o análogamente} \\ f_i(t) &= f_p(t) \operatorname{sgn}(t) . \end{aligned}$$

aplicando la TF a ambos lados de la igualdad y tomando en cuenta el teorema de la convolución de la TF que enuncia

$$F(h(t)g(t)) = \frac{1}{2\pi} H(k)*G(k)$$

donde

$$\begin{aligned} F(h(t)) &= H(k) \\ F(g(t)) &= G(k) \end{aligned}$$

Y el transformado de signum de t es

$$F(\operatorname{sgn}(t)) = -\frac{2i}{k}$$

se tiene que

$$F(f_p(t)) = F(f_i(t) \operatorname{sgn}(t))$$

$$R(k) = \frac{1}{2\pi} F (f_i(t)) * F (-\frac{2i}{k})$$

$$R(k) = \frac{1}{2\pi} \int_{-\infty}^{\infty} -2i \frac{iI(x)}{k-x} dx$$

$$R(k) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{I(x)}{k-x} dx$$

Similarmente para $f_i(t)$

$$F (f_i(t)) = F (f_p(t) \operatorname{sgn}(t))$$

$$iI(k) = \frac{1}{2\pi} F (f_p(t)) * F (-\frac{2i}{k})$$

$$iI(k) = \frac{1}{2\pi} \int_{-\infty}^{\infty} -2i \frac{R(x)}{k-x} dx$$

$$I(k) = -\frac{1}{\pi} \int_{-\infty}^{\infty} \frac{R(x)}{k-x} dx$$

Resumiendo resultados se tienen las siguientes relaciones

$$R(k) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{I(x)}{k-x} dx$$

$$I(k) = -\frac{1}{\pi} \int_{-\infty}^{\infty} \frac{R(x)}{k-x} dx,$$

dada una función de dominio y rango real $f(t)$. Puede ser asociada una función de rango complejo cuya parte real es $f(t)$ y la parte imaginaria es $\mathbf{H}(f(t))$, la TH de $f(t)$

$$f(t) + i H(x) \quad .$$

Este tipo de problemas se encuentran en análisis de señales y en óptica cuando la variable independiente es el tiempo, la función compleja asociada es conocida como señal analítica y su TH es llamada función de cuadratura (Lokenath, 1995).

Un sistema se define como la transformación de una función de entrada en una función de salida. Si la parte real de la función de un sistema $J(k)$ causal está dada es posible encontrar la parte imaginaria con la TH. Si $J(k)$ tiene la forma:

$$J(k) = R(k) + iI(k)$$

Ahora de la teoría de Fourier se sabe que toda serie de tiempo $s(t)$ puede ser aproximada con cierto error mediante una serie infinita de Fourier de la forma

$$s(t) = \sum_{n=0}^{\infty} a_n \cos(nt) + b_n \text{sen}(nt)$$

y su TH es

$$\hat{s}(t) = \sum_{n=0}^{\infty} a_n (-1) \text{sen}(xt) + b_n \cos(xt) \quad ,$$

formando así una serie compleja

$$S(t) = s(t) + i \hat{s}(t)$$

donde la parte real de $S(t)$ es igual a $s(t)$

$$\Re\{S(t)\} = s(t)$$

y la parte compleja

$$\Im\{S(t)\} = \hat{s}(t) \quad .$$

Esta es la característica principal que será utilizada para la formación de series de tiempo complejas y, posteriormente, de matrices reales a matrices complejas para así obtener información adicional.

2.3 Descomposición en valores singulares complejos

Para realizar la descomposición de una serie de tiempo en modos ortogonales se utiliza la descomposición en valores singulares complejos (DVSC) en la que no es necesario calcular la matriz de covarianzas aunque hay una relación directa de los eigenvectores de la matriz de covarianzas y los valores singulares de la DVSC de la matriz de datos. Esto lo hace un método más sencillo y eficiente. Haciendo uso de DVSC se

pueden obtener no solo los eigenvalores y eigenvectores sino también funciones de fase y amplitud temporal y espacial todo esto en un solo proceso (Susanto et al., 1998).

Retomando la matriz de datos \mathbf{Y} de tamaño pxn , la cual representa los datos de p variables y n muestreos a los que se les sustrajo su media, donde los muestreos pueden ser interpretados como pasos de tiempo. Antes de aplicar el método de DVSC a la matriz de datos \mathbf{Y} , esta es transformada en una matriz compleja \mathbf{Z} , haciendo uso de la TH, donde la parte real de \mathbf{Z} es \mathbf{Y} , esto es

$$\mathbf{D} = H(\mathbf{Y}) \quad ,$$

$$\Re(\mathbf{Z}) = \mathbf{Y} \quad \text{y}$$

$$\mathbf{Z} = \mathbf{Y} + i \mathbf{D} \quad .$$

La TH aplicada a la matriz \mathbf{Y} toma el i -ésimo renglón como una función y obtiene su TH en el i -ésimo renglón de \mathbf{D} . El método de DVSC aplicado a \mathbf{Y} resulta:

$$\mathbf{Z} = \mathbf{U} \mathbf{\Gamma} \mathbf{V}^H$$

Siendo \mathbf{V}^H el transpuesto conjugado de \mathbf{V} . \mathbf{U} , de tamaño pxp , tiene en sus columnas los eigenvectores correspondientes al producto de las matrices $\mathbf{Z}\mathbf{Z}^H$ y las columnas de \mathbf{V} son los eigenvectores de $\mathbf{Z}^H\mathbf{Z}$, \mathbf{V} de tamaño nxn . Los r valores singulares en la diagonal de $\mathbf{\Gamma}$ son las raíces cuadradas de los eigenvalores distintos de cero de $\mathbf{Z}\mathbf{Z}^H$ y $\mathbf{Z}^H\mathbf{Z}$, además r representa el rango de \mathbf{Z} . Para matrices complejas $\mathbf{\Gamma}$ permanece real, \mathbf{U} y \mathbf{V} se vuelven unitarias es decir cumplen las siguientes propiedades.

$$\mathbf{U}\mathbf{U}^H = \mathbf{I}_p$$

$$\mathbf{V}^H\mathbf{V} = \mathbf{I}_n$$

Se puede mostrar fácilmente que las columnas de \mathbf{U} son los eigenvectores de $\mathbf{Z}\mathbf{Z}^H$, y las columnas de \mathbf{V} son los eigenvectores de $\mathbf{Z}^H\mathbf{Z}$ realizando los siguientes desarrollos:

$$\mathbf{Z}\mathbf{Z}^H = (\mathbf{U} \mathbf{\Gamma} \mathbf{V}^H) (\mathbf{U} \mathbf{\Gamma} \mathbf{V}^H)^H = (\mathbf{U} \mathbf{\Gamma} \mathbf{V}^H) (\mathbf{V} \mathbf{\Gamma}^H \mathbf{U}) = \mathbf{U} \mathbf{\Gamma} \mathbf{\Gamma}^T \mathbf{U}^H$$

$$\mathbf{Z}^H\mathbf{Z} = (\mathbf{U} \mathbf{\Gamma} \mathbf{V}^H)^H (\mathbf{U} \mathbf{\Gamma} \mathbf{V}^H) = (\mathbf{V} \mathbf{\Gamma}^H \mathbf{U}^H) (\mathbf{U} \mathbf{\Gamma} \mathbf{V}) = \mathbf{V} \mathbf{\Gamma}^T \mathbf{\Gamma} \mathbf{V}^H$$

Y debido a que $\mathbf{\Gamma}$ contiene elementos reales, el operador hermitiano aplicado a $\mathbf{\Gamma}$ se transforma en su transpuesta. Retomando la descomposición en valores propios de la matriz de dispersión de \mathbf{Z} .

$$\mathbf{R} = \mathbf{Z} \mathbf{Z}^H / (n-1)$$

$$\mathbf{R} = \mathbf{E} \mathbf{\Lambda} \mathbf{E}^H$$

De esta manera se muestra que los eigenvalores de la descomposición en valores propios es el cuadrado de los valores singulares. Dado que $\mathbf{\Gamma}$ es una matriz diagonal con elementos γ_i .

$$\mathbf{R} = \mathbf{E} \mathbf{\Lambda} \mathbf{E}^H$$

$$\mathbf{Z}\mathbf{Z}^H / (n-1) = \mathbf{E} \mathbf{\Lambda} \mathbf{E}^H$$

$$\mathbf{Z}\mathbf{Z}^H = \mathbf{E} (n-1) \mathbf{\Lambda} \mathbf{E}^H$$

$$\mathbf{Z}\mathbf{Z}^H = \mathbf{U} \mathbf{\Gamma} \mathbf{\Gamma}^T \mathbf{U}^H$$

De donde se obtienen las siguientes igualdades

$$\mathbf{E} = \mathbf{U}$$

$$\mathbf{\Lambda} = \mathbf{\Gamma} \mathbf{\Gamma}^T / (n-1)$$

$$\lambda_i = \gamma_i^2 / (n-1), 1 \leq i \leq r$$

Para la obtención de la matriz de componentes principales \mathbf{A} se proyectan los eigenvectores de la matriz \mathbf{U} sobre la matriz \mathbf{Z} .

$$\mathbf{A} = \mathbf{U}^T \mathbf{Z}$$

Una ventaja de esta matriz es que en el análisis tradicional de FEOs la matriz de componentes principales es la parte real de \mathbf{A} , en este tipo de análisis se obtiene esa información además de la de amplitud y fase de las series de tiempo complejas las cuales están dadas por:

$$r_{ij} = \|\mathbf{a}_{ij}\|^{1/2} = (\mathbf{a}_{ij}^* \mathbf{a}_{ij})^{1/2} \text{ y}$$

$$\theta_{ij} = \arctan\left(\frac{\Im(\mathbf{a}_{ij})}{\Re(\mathbf{a}_{ij})}\right).$$

La ventaja de obtener matrices con entradas complejas es que se tiene información de fase y amplitud. La formación de series de tiempo en la matriz de componentes principales agrega información del fenómeno de estudio.

Las medidas de fase y amplitud de \mathbf{U} y \mathbf{A} , constituyen una descripción general de rasgos en el campo de estudio \mathbf{Y} y son determinados de la descomposición realizada. Estas medidas son de ayuda en el estudio exploratorio de los datos. La amplitud r_{ij} da una medida de la variabilidad espacial en la magnitud de la estructura del modo, la fase θ_{ij} describe la variación espacial de la fase asociada con periodicidades en el campo \mathbf{Y} .

Para un modo dado h , la evolución de la fase θ_{ij} revela información acerca de la existencia de cuasi-periodicidades en el campo. De hecho, si θ_{ij} se incrementa monótonamente de 0 a 2π sobre algún intervalo, se puede inferir que existe cierta ciclicidad en los datos.

Ahora si se desea extraer solo aquellos modos donde la varianza es mayor se utilizan solo los eigenvectores correspondientes a esos γ_i de mayor valor. Construir una matriz \mathbf{U}' que utiliza solo $h < p$ eigenvectores en la proyección sobre \mathbf{Y} .

$$\mathbf{A}' = \mathbf{U}'^T \mathbf{Z}$$

Para la reconstrucción de un campo menos variable que el original se obtiene la parte real de $\hat{\mathbf{Z}}$ estimada.

$$\hat{\mathbf{Z}} = \mathbf{U}^* \mathbf{A}'$$

$$\hat{\mathbf{Y}} = \Re(\hat{\mathbf{Z}})$$

El porcentaje de variabilidad representado por $\hat{\mathbf{Y}}$ depende del estudio realizado en específico. Con pocos modos es posible encontrar estructuras en el conjunto de datos que podrían no ser notables a primera vista.

Capítulo 3 Los datos satelitales de altimetría y su importancia en la oceanografía

En este capítulo se describe la altimetría satelital y, en particular, el proyecto Topex/Poseidon (T/P) cuyos datos son utilizados en la prueba del método de interpolación propuesto en este trabajo. También se discute la importancia de tener mejoras sustanciales en la calidad de estos datos y algunas investigaciones y aplicaciones de la altimetría satelital.

3.1 Antecedentes de la altimetría satelital

Los océanos transportan calor y materia, los cuales constantemente son intercambiados con la atmósfera, regulando así el clima de nuestro planeta. La principal fuente de calor de la Tierra es el Sol, pero debido a la esfericidad del planeta y a la inclinación del eje polar, el Sol no calienta la Tierra de manera uniforme lo cual provoca el fenómeno de las estaciones. El calor emitido por el Sol es recibido por la atmósfera, la tierra y el mar; ese calor es redistribuido desde las zonas cercanas al ecuador, que reciben más calor que el resto del planeta, hacia latitudes más altas por un proceso de circulación natural. La circulación en la atmósfera es a través de complejos sistemas de viento y en los océanos por medio de las corrientes oceánicas. El viento mueve la superficie del mar y la diferencia de densidades en el océano crea una circulación entre el fondo y la superficie. Además de estos factores que producen las corrientes y la circulación en el océano, se puede agregar el movimiento de rotación de la tierra y las fuerzas gravitacionales ejercidas por el Sol y la Luna, las cuales generan las mareas y causan variaciones periódicas en el nivel del mar.

Las constantes mejoras tecnológicas, además de los acuerdos internacionales para fortalecer la investigación, han permitido cristalizar proyectos como los de los satélites de investigación y el desarrollo de instrumentos cada vez más precisos para la medición de las variables oceánicas de interés. De todos los instrumentos usados para estudiar los océanos y la atmósfera, los satélites son los únicos con la capacidad de monitorear en forma regular áreas extensas sobre la Tierra.

Los primeros indicios de que la percepción remota tenía algo que ofrecer a la oceanografía vinieron probablemente de tres fuentes: vuelos de avión que transportaban sensores de microondas, fotografías tomadas desde las naves Mercury, Geminis y Apollo, y los primeros satélites meteorológicos que transportaban video y mejores sensores.

En los años setentas, los satélites comenzaron a transmitir datos acerca de la física, química y dinámica de la tierra, mar, atmósfera y biosfera. En esta década los primeros datos de altimetría se utilizaron para el mapeo de la topografía de la superficie del océano. El primer país en colocar un satélite con sensores de altimetría fue Estados Unidos, en las misiones Skylab y Geos3, la Seasat en 1978 y la Geosat en 1985. En los noventas se llevaron a cabo las misiones ERS-1, Topex/Poseidon, ERS-2 y Jason-1 de las cuales aún hay algunas vigentes. Una breve descripción es presentada a continuación.

ERS-1 (ESA Remote-Sensing Satellite 1) fue lanzado en julio de 1991, desconectado en junio de 1996 y retirado en marzo del 2000. La misión estuvo a cargo de la Agencia Espacial Europea (ESA, por sus siglas en inglés), y entre los instrumentos que llevaba el satélite se incluía un altímetro (aparato para medir la altura del nivel del mar). El objetivo del ERS-1 era el tomar medidas de la atmósfera de la tierra y propiedades de la superficie usando radares (técnicas de microondas). El satélite realizó mediciones de viento, velocidad y dirección de la superficie del océano, altura y temperatura de la superficie del mar y vapor de agua de la atmósfera, entre otras, además de imágenes obtenidas por radar de la tierra, hielo y regiones oceánicas.

El satélite Topex/Poseidon (T/P), fue lanzado en Agosto 10 de 1992 con el objetivo de “observar y entender la circulación oceánica”, como parte de un proyecto conjunto entre la *National Aeronautics and Space Administration* (NASA), que es la agencia espacial estadounidense, y el *Centre National d'Etudes Spatiales* (CNES), que es la agencia espacial francesa. El T/P carga dos radares de altimetría y sistemas de determinación de órbita

precisa. En la siguiente sección (3.2) se describen con más detalles las misiones T/P y Jason-1.

El satélite ERS-2 tiene objetivos similares al ERS-1 pero con mejoras en los sistemas para proveer imágenes ópticas, medidas de la concentración de ozono en la troposfera y en la estratosfera y mejoras en la determinación de la órbita y posición del satélite con respecto a puntos de localización en tierra. Su lanzamiento fue en abril 21 de 1995 y aún continúa proporcionando datos de alta calidad.

Estas tres misiones forman parte de programas meteorológicos y oceanográficos internacionales, tales como el World Ocean Circulation Experiment (WOCE) y el Tropical Ocean and Global Atmosphere (TOGA), ambos vinculados con el World Climate Research Programme (WCRP programa de investigación climática mundial).

El Jason-1 fue lanzado en diciembre 7 del 2001, como parte de una misión conjunta de la CNES y la NASA, con el objetivo principal de medir la topografía de la superficie del océano, por lo menos con la misma precisión que el T/P, para continuar estudiando la circulación global de los océanos.

En un futuro cercano se espera poder seguir con el estudio de los datos de satélite y se trabajará por tener continuidad de los mismos. La precisión con la que los sensores satelitales obtienen las medidas supera la esperada inicialmente, lo que asegura la aplicación de modelos basados en datos de entrada cada vez más reales.

3.2 Los proyectos Topex/Poseidon y Jason-1

Como se mencionó en la sección anterior, la misión T/P es conducida conjuntamente por la NASA y el CNES. Los datos del satélite T/P son distribuidos a la comunidad científica internacional vía el Physical Oceanography Distributed Active Archive Center (PODAAC) y el centro AVISO/Altimetry. La NASA y el CNES procesan datos de sus instrumentos (desde datos crudos hasta datos geofísicos) e intercambian datos procesados. Los centros de procesamiento realizan funciones tales como el monitoreo de los instrumentos, pre-procesamiento, procesamiento de datos científicos, verificación de datos y la determinación precisa de la órbita.

Los objetivos de la misión comenzaron a cumplirse desde el lanzamiento del satélite en Agosto 10 de 1992, con un tiempo esperado de vida de 5 años (por ahora es una misión extendida). El satélite T/P realiza mediciones de la altura precisa de la superficie del océano, usando dos sistemas de radar, para estudiar la dinámica de la circulación de los océanos del mundo. El satélite T/P (Figura 3.2.1) establece las bases para el monitoreo oceánico de largo plazo desde el espacio, lo que llevará a mejorar el entendimiento del papel de los océanos en el cambio climático global. Otras aplicaciones incluyen las mareas oceánicas, la forma de la Tierra, la altura de las olas y la velocidad del viento. Con el objeto de ser útil para el estudio de la circulación oceánica, especialmente en la escala de los giros oceánicos y de las cuencas, se realizaron numerosas mejoras en el T/P con respecto a los sistemas de altimetría previos, incluyendo:

- Un diseño específico del satélite, mejoras a los sensores, sistemas de trayectoria del satélite y configuración de órbita.
- Un modelo óptimo de gravedad para la determinación precisa de la órbita.
- Sistemas en tierra dedicados a operaciones de la misión.

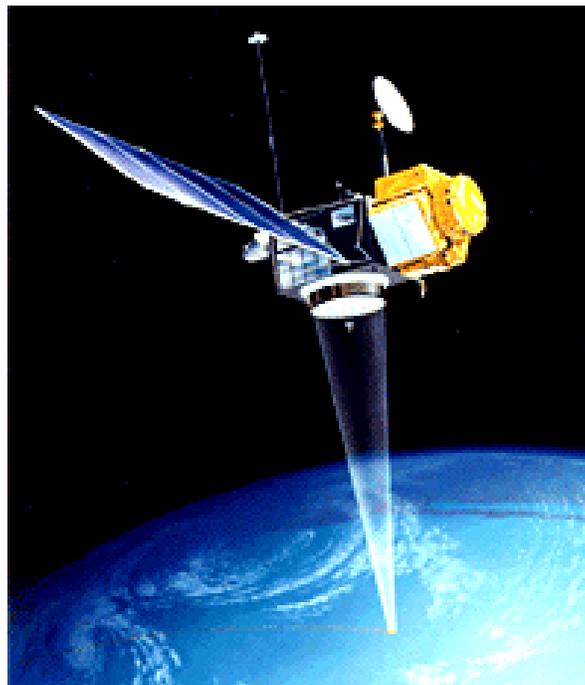


Figura 3.2.1 Satélite Topex Poseidon

Las trayectorias del satélite T/P siguen una órbita casi circular, a 1336 kilómetros sobre la superficie de la Tierra, con una inclinación de 66° con respecto al plano ecuatorial. El satélite cubre el 95% de los océanos libres de hielo cada 10 días.

Cada 10 días (10 días y 22 horas exactamente), después de circular la tierra 127 veces, el satélite pasa por los mismos puntos (con un rango de 1 km), describiendo una malla regular de trayectorias o pasos ascendentes y descendentes sobre la tierra. La órbita sobrevuela dos sitios de calibración: un sitio de la NASA cerca de Punta Conception, California (plataforma Harvest, $239^\circ 19'$ Este, $34^\circ 28'$ Norte, revolución número 22, paso ascendente 43) y un sitio del CNES en Isla Lampedusa, Italia ($12^\circ 57'E$, $35^\circ 52'N$, revolución 111, paso descendente 222).

Hay algunos términos que son utilizados para entender las mediciones de altimetría y se mencionan a continuación:

- Elipsoide de referencia: definición de la forma no esférica de la Tierra con radio ecuatorial de 6378.1363 kilómetros y un coeficiente de achatamiento de $1/298.257$, específico para la misión T/P.

- Altura: se refiere a la distancia de la superficie del océano sobre el elipsoide de referencia.

- Rango: es la distancia desde el centro de masa del satélite a la superficie de la Tierra. Así, la medida del altímetro está referida como “rango” o “rango de altimetría”, no como altura.

- Altitud: se refiere a la distancia del centro de masa del satélite sobre un punto de referencia. El punto de referencia estará usualmente sobre el elipsoide de referencia o el centro de la Tierra.

- Una revolución: es una vuelta a la Tierra por el satélite.

- Un paso del satélite: es media revolución. El número de pasos se incrementa con el tiempo. Los pasos con número impar corresponden a órbitas ascendentes, es decir, de latitud mínima a latitud máxima. Los pasos con número par corresponden a órbitas descendentes, es decir, de latitud máxima a latitud mínima.

- Un ciclo: es el periodo de tiempo que tarda el satélite en sobrevolar una misma posición, es decir, un ciclo está compuesto de 127 revoluciones.

El diseño del T/P le permite cargar seis instrumentos científicos, 4 de la NASA y 2 del CNES, de los cuales cuatro son operacionales y dos son experimentales (Tabla 3.2.1).

| Nombre del sensor | Tipo de sensor | Origen |
|-------------------------|---|----------------------|
| NRA (Topex) | Radar de altimetría | NASA operacional |
| SSALT (POSEIDON) | Radar de altimetría de estado sólido | CNES experimental |
| LRA | Retroreflector láser | NASA operacional |
| TMR | Radiómetro TOPEX de microondas | NASA operacional |
| DORIS | Receptor del sistema de trayectoria Doppler | CNES operacional |
| GPSDR | Receptor de sistema de posicionamiento global | NASA experimental |

Tabla 3.2.1 Instrumentos del satélite

El altímetro NRA opera a dos frecuencias distintas simultáneamente y las mediciones se combinan para obtener la distancia entre el satélite y la superficie del mar. Los altímetros funcionan con sistemas de radar, emiten ondas de radio que son reflejadas por la superficie de agua y regresan al satélite. Un complejo aparato mide el tiempo que le toma a la señal viajar de ida y regreso, proporcionando la medida de altimetría deseada (dado que la señal de propagación es conocida, es decir, la velocidad de la luz es de alrededor de 300,000 km/seg). Después de numerosas correcciones, la precisión instantánea de la distancia medida es de 3 cm.

El altímetro TOPEX es estadounidense y opera 90% del tiempo. El altímetro POSEIDON es Francés, fue instalado experimentalmente y opera el faltante 10% del tiempo (actualmente 10 días en un renglón de 100 días).

El T/P tenía una vida esperada de 5 años, por lo que se lanzó el J-1 para seguir con la recolección de información en la misma trayectoria. Durante el ciclo 365 se realizó una

maniobra de colocación del Jason-1 en el lugar del T/P. Posteriormente se desplazó el T/P quedando un espacio entre ellos de 150 km (tomando como punto de referencia el ecuador).

Los datos del T/P continúan siendo recolectados y analizados ya que el satélite permanece en operación. Para asegurarse de la continuidad del monitoreo del planeta y del clima, CNES y NASA decidieron, el 20 de Diciembre de 1996, extender su cooperación en altimetría satelital trabajando conjuntamente en el satélite Jason-1, el sucesor del T/P. Para la elaboración de este pequeño satélite, que pesa solamente una quinta parte del peso de su predecesor, sólo se gastó una tercera parte con respecto al T/P; se construyó sobre la plataforma Proteus por una colaboración entre CNES y Alcatel Space, fue lanzado el 7 de Diciembre del 2001 y carga instrumentos estadounidenses y franceses igual que su antecesor. Cuesta mucho menos, pero su perfeccionamiento permite obtener la misma calidad de datos que el T/P, con los siguientes objetivos:

- Brindar un servicio operacional que transmita la información en menos de 3 horas después de ser tomada.
- Proporcionar estimaciones de la altura significativa de las olas y de la velocidad del viento sobre el océano.
- Mantener la continuidad en la recolección de datos para el estudio de altimetría satelital.
- Incrementar el entendimiento de la circulación oceánica y su influencia sobre los cambios climáticos.
- Medir el nivel del océano y sus variaciones a escala global.

El satélite Jason-1 carga cinco instrumentos, tres del CNES y dos de la NASA, entre los cuales cuenta con un retroreflector láser (que mide la distancia con referencia a la estación de tierra), un receptor GPS (que mide la distancia con referencia al satélite) y un receptor Doris (que indica la velocidad con la que el satélite se aleja o se acerca hacia las estaciones en tierra). Combinando sus resultados, se consigue la posición exacta del satélite con una precisión de 3 cm. Cuenta también con un radar de altimetría de estado sólido y un radiómetro de microondas.

Los datos crudos son primero grabados a bordo y después transmitidos a tierra. Los datos son enviados a dos centros de procesamiento, uno en Pasadena, California (USA), y el otro en Toulouse (Francia).

3.3 La visión de los mares en la era de la altimetría

Uno de los usos de los datos tomados por los satélites es estimar las corrientes oceánicas, lo cual ayuda a los científicos a entender el entorno y el clima de la Tierra. Entre mayor conocimiento se tenga del océano, la población mundial se verá beneficiada debido a:

- El mejoramiento de las actividades económicas ligadas al océano.
- El suministro de información a personas cuya residencia esté cerca de ríos, lagos u océano.
- Mejorar pronósticos de tiempo en el mar necesarios para la navegación de barcos, buques y demás vehículos marinos.

Los oceanógrafos están interesados en observar, entender y predecir el comportamiento del océano. Debido al alto costo para realizar las observaciones, es casi imposible tener una cobertura amplia en tiempo y espacio de los océanos a partir de muestreos en sitio (buques, boyas, etc.). Para el estudio de las corrientes oceánicas el problema puede ser parcialmente resuelto con la introducción de más satélites, pero aún se encuentran huecos en la información en lugares por donde no pasa la trayectoria del satélite. Dichos huecos son un problema fundamental para muchos estudios geofísicos. Frecuentemente son consecuencia de un muestreo irregular o desigual, o pueden resultar de valores removidos o erróneos (Peng et al., 2004).

Cada región de la tierra tiene sus propias condiciones meteorológicas, las cuales determinan su clima. Los océanos tienen una gran capacidad de almacenamiento de calor, el cual es transferido del mar hacia la atmósfera y ayuda a regular la temperatura de la Tierra. Como ejemplo se puede mencionar la corriente del Golfo, la cual circula por el mar Caribe, la costa este de los Estados Unidos y la costa occidental de Europa, ayudando a regular la temperatura del aire por donde pasa, siendo ésta la razón principal por la cual el oeste de Europa tiene un clima menos extremo que algunas regiones que se encuentran a la misma latitud en el continente americano. Similarmente existen corrientes frías que producen el efecto contrario en otras regiones. El perfeccionamiento en el conocimiento de

la circulación oceánica permite entender el clima, especialmente fenómenos como El Niño, que traen consigo consecuencias climáticas tales como sequías, inundaciones y ciclones o huracanes, entre otros.

El océano es difícil de ser monitoreado, es inmenso, tormentoso, hostil y difícil de acceder, además, las “estructuras” que se desean monitorear son generalmente más pequeñas que las de la atmósfera. Asimismo, las variaciones de algunas variables medidas en el océano son muy pequeñas, como es el caso de variaciones de un centímetro por segundo en las corrientes oceánicas, décimas de grado en la temperatura de la superficie del agua, o unos cuantos milímetros en el promedio del nivel del mar; todo esto requiere que los instrumentos usados deben ser muy precisos.

Para estudiar los océanos tenemos que monitorearlos regularmente, recolectar medidas rápidamente y repetir el proceso de medición con cierta frecuencia, pero con suficiente precisión para que algún fenómeno de corta vida sea observado. La solución moderna consiste en instalar instrumentos de medición sobre satélites artificiales. Las ventajas de los satélites son las siguientes: tienen una vasta área de cobertura, la capacidad de muestrear regiones de difícil acceso, tomar medidas continuamente con un solo instrumento, la facilidad de calibración, entre otras. A esto se le puede agregar la información complementaria de otros instrumentos colocados en el océano, los cuales todavía se siguen usando.

Dependiendo de los instrumentos montados en los satélites, es posible medir:

- la temperatura de la superficie,
- la velocidad y la dirección del viento en la superficie,
- la extensión de áreas cubiertas por hielo,
- la altura de las olas,
- el color del agua (debido a la cantidad de plancton en la superficie).

La altimetría satelital ha hecho posible medir el nivel del océano y registrar sus variaciones en escalas de centímetros. Los satélites como el Jason-1 permiten tener acceso a datos de alta calidad de manera rápida.

El 26 de diciembre del 2004 un tsunami, creado por un terremoto cerca de la isla de Sumatra en Indonesia, devastó las costas cercanas y ocurrió que los satélites Jason-1 y T/P pasaron sobre el área justamente después del terremoto (alrededor de 2 horas después). Así,

las variaciones de la altura de la superficie del océano fueron observadas en mar abierto, mostrando una primera elevación de alrededor de 50 cm en un área de 400 a 500 km y una segunda elevación de 40 cm.

Con las mediciones de altimetría se presenta la oportunidad de tener información de fenómenos no observados con satélites anteriores y, de esta manera, aplicar nuevos modelos que sirvan para ampliar el conocimiento de los fenómenos naturales lo cual ayudará al mejoramiento de las condiciones de vida en el planeta.

Capítulo 4 Metodología

En este capítulo se describen detalladamente los algoritmos y procesos que se siguen para el procesamiento de los datos de altimetría, desde la captura de los datos hasta la obtención del producto final. Se llevó a cabo un proceso previo de control de calidad de la información en donde se eliminaron datos erróneos y se hizo una interpolación lineal para cubrir los huecos debidos a fallas del satélite.

4.1 Procesamiento de los datos de altimetría

El pre-procesamiento y procesamiento de los datos crudos de altimetría obtenidos por los satélites es realizado por la NASA y el CNES para que puedan ser usados por los científicos. Esta información puede ser obtenida de los siguientes sitios Web: para los datos del T/P se ingresa a *ftp://podaac.jpl.nasa.gov/pub/sea_surface_height/topex_poseidon/tp_atg/data/* y para los datos del Jason-1 a *ftp://podaac.jpl.nasa.gov/pub/sea_surface_height/jason/j1_atg/data/*. El manual de lectura de los datos altimetría está ubicado en la dirección *ftp://podaac.jpl.nasa.gov/pub/sea_surface_height/jason/j1_atg/doc/*. Entre los tópicos contenidos en el manual se describen los “Along Track Gridded Sea Surface Height Anomaly Products”, es decir, los productos generados de la anomalía de la altura de la superficie del mar (SSHA, por sus siglas en inglés) en la malla de posiciones a lo largo de los pasos del satélite, derivados de la altura de la superficie del océano obtenida por el T/P (TPSSHA) y de las anomalías de la altura de la superficie del océano obtenidas del Jason-1 (J1SSHA). Los valores procesados son respecto a la media de la altura de la superficie del mar y están colocados en una malla con valores de latitud y longitud fijos para todos los ciclos.

A continuación se retoman dos definiciones del capítulo anterior que ayudan a entender la organización de los datos.

Un paso del satélite es media revolución alrededor de la Tierra, recorriendo de un valor extremo de latitud al otro valor extremo. Un paso ascendente comienza en 66.03° de latitud sur y termina en 66.03° de latitud norte, y un paso descendente lo hace de manera opuesta. Para el Jason-1 y el T/P los pasos con número impar son ascendentes y con número par descendentes.

Un ciclo comprende el recorrido del satélite hasta que vuelve a pasar por el mismo punto sobre la superficie terrestre, contiene 254 pasos y representa una colección de registros correspondientes a un tiempo aproximado de 10 días.

Los datos utilizados son derivados de las mediciones originales de TPSSHA y J1SSHA, tomando el promedio de 10 mediciones para generar el dato deseado, además de otros cálculos en los que se quitan efectos atmosféricos y factores que influyen en los datos.

La información oceanográfica de SSH se encuentra en registros o archivos de capacidad determinada; cada registro corresponde a un ciclo y tiene un tamaño fijo de $255 \times 3127 \times 2 = 1594770$ bytes (aproximadamente 1.6 mega bytes). Cada registro contiene 254 pasos, pero puede presentarse el caso de pasos faltantes los cuales tienen todos los valores iguales a 32767. Puede suceder que haya valores de 32767 en posiciones sobre el mar en donde el satélite tuvo alguna falla o en posiciones correspondientes a tierra.

Para determinar la latitud y la longitud de cada paso del ciclo se usan las trayectorias de referencia ascendente y descendente y los puntos de la órbita del satélite que cruzan por el ecuador. Los archivos de trayectorias se encuentran con los nombres `reftrk.ascending` y `reftrk.descending`, los cuales pueden obtenerse en el mismo sitio de los SSH. Estos archivos contienen arreglos bidimensionales de 122 renglones por dos columnas, la primera columna contiene los valores de latitud y la segunda los de longitud.

Debido al cambio de órbita del satélite T/P por la colocación del Jason-1 hay dos archivos de cruce con el ecuador, uno para antes y otro para después del cambio, cuyos nombres son `eqxl_tp_then_j1.dat` y `eqxl_new_tp.dat` respectivamente, que contienen arreglos de 254 posiciones. La forma de dar las coordenadas de cada paso es la siguiente: primero se revisa si el dato proviene de la primera o segunda trayectoria; para el paso i -ésimo se debe de sumar a toda la columna de longitud el valor del i -ésimo cruce con el ecuador y dejar estos valores en el rango de -180 a 180, que son las longitudes este y oeste

de referencia, dejando los valores de latitud intactos. De esta forma los 254 pasos tienen una localización en tierra única que se repite cada 10 días.

El encabezado de los registros contiene la siguiente información: la agencia que lo produjo, el nombre del satélite del que proviene, el número de ciclo, el día de inicio y fin en que se tomó la medida, etc. Para facilitar su lectura se acompaña de un programa en lenguaje de programación C.

4.2 Obtención de los datos a lo largo de las trayectorias del satélite

Como se mencionó anteriormente, la organización de los registros de la SSH es de un ciclo por archivo y posiciones fijas para cada paso. Debido a que la información se presenta en espacio y el estudio realizado analiza variaciones en tiempo, primeramente se necesita formar arreglos matriciales bidimensionales en donde los renglones representen las posiciones muestreadas del planeta y las columnas la evolución en el tiempo. Se formaron 254 matrices correspondientes a los pasos del satélite, que en realidad serán series de tiempo de 3127 posiciones del planeta. El número de columnas será igual a la cantidad de ciclos de estudio. Tomando en cuenta que para el T/P se deberán tener 2 archivos, uno para antes del cambio de trayectoria en el ciclo 365 (matriz de 3127 posiciones y 365 columnas) y otro para su nueva órbita, mientras que para el Jason-1 sólo se tendrá un tipo de archivo.

A continuación se presenta un resumen de los algoritmos usados para la extracción de los datos y selección de la zona de estudio. Posteriormente se explica se explica el proceso mediante al cual se llenaron los huecos correspondientes a las fallas del satélite. Que como se verá en las gráficas del capítulo 6 de resultados, ocurre frecuentemente.

Primeramente se presenta una explicación del proceso de obtención de los datos de SSH para el satélite T/P antes del cambio de órbita. Para el Jason-1 y para el T/P en la nueva posición, se realizaron procesos similares. En la tabla 4.2.1 está el algoritmo que cambia el formato de los datos para poder manipularlos más fácilmente.

| Entradas | Procesos | Salidas |
|---|--|--|
| Archivos: - tpatgssha22.011 - tpatgssha22.012 ... - tpatgssha22.365 | Obtener solamente los datos de altimetría de los archivos tpatgssha22 provenientes del ciclo 11 al ciclo 365. Generar un nuevo registro por ciclo sin encabezado y cuyo contenido es una matriz con los datos de SSH. El registro creado tendrá nombre <i>ciclostp###</i> y una matriz <i>ciclo###</i> de 254 renglones y 3127 columnas. | Archivos: - <i>ciclostp011</i> - <i>ciclostp012</i> ... - <i>ciclostp365</i> |

Tabla 4.2.1 Algoritmo ciclos_tp_generales: cambia el formato de los datos crudos para se manipulado más fácilmente.

La tabla 4.2.2 presenta el algoritmo para acomodar la información por pasos.

| Entradas | Procesos | Salidas |
|--|--|---|
| Archivos: - <i>ciclostp011</i> - <i>ciclostp012</i> ... - <i>ciclostp365</i> | Tomar la matriz <i>ciclo---</i> de los archivos <i>ciclostp---</i> y acomodarlos de manera tal que las columnas sean distintos ciclos y los renglones sean las posiciones fijas de un paso específico. Generar archivos <i>tp_paso###</i> , cuyo contenido son matrices <i>paso###</i> , de 3127 renglones y 355 columnas. | Archivos: - <i>tp_paso001</i> - <i>tp_paso002</i> ... - <i>tp_paso254</i> |

Tabla 4.2.2 Algoritmo cargapasos: acomoda los datos de SSH por pasos.

Hasta este punto se obtienen los datos de SSH deseados, las matrices de los registros *tp_paso---* tienen 355 columnas debido a que durante los primeros 10 ciclos se calibró el satélite. El siguiente programa (*coord_paso*) toma la referencia de cruce con el ecuador para la creación de las coordenadas de los pasos. Los archivos *reftrk.ascending* y *reftrk.descending* son matrices de 254 renglones y 2 columnas, en la primer columna están los valores de latitud fijos, tanto para un paso ascendente como para uno descendente, los valores en la segunda columna son las referencias respecto al ecuador y cambian con respecto al número de paso. El archivo *eqxl_tp_then_j1.dat* es un arreglo de 254 cifras,

cada una debe de ser sumada a la segunda columna dependiendo del número de paso correspondiente y hacer un pequeño cálculo para la longitud correcta.

| Entradas | Procesos | Salidas |
|--|--|-------------------------|
| Archivos: -eqxl_tp_then_j1.dat -reftrk.ascending -reftrk.descending | Tomar la posición k de eqxl_tp_then_j1.dat y sumarla a la segunda columna de reftrk.ascending y reftrk.descending, según si k es impar o par. Si el valor resultante es mayor a 180 restar 360 y si es menor a -180 sumar 360. Por último, se guardan estos valores en dos matrices, una para los valores en longitud y una para los valores en latitud: lonoldtp y latoldtp, respectivamente en el archivo coord_old. Las columnas representan los pasos. | Archivos: -coord_old |

Tabla 4.2.3 Algoritmo coord_paso: crea un archivo de posiciones de los pasos de la región de estudio.

Los programas ciclos_tp_generales, cargapasos y coord_paso registran la información necesaria para la entrada del modelo de interpolación propuesto en este trabajo, pero la cantidad de información por paso es mucha y como consecuencia la carga computacional es considerable. Dado que el objetivo puede ser bien demostrado con la elección de una subregión de todo el planeta, se eligió una zona de estudio cercana a México, de 180° a 80° oeste y de 8° a 18° norte. En el programa coord_paso_especif se le asigna la SSH a la región de prueba para tener un solo archivo con posición y SSH.

| Entradas | Procesos | Salidas |
|---|--|---|
| Archivos: -coord_old -tp_paso001 -tp_paso002 -... -tp_paso254 Información a priori -latitude de interés -longitud de interés | Seleccionar aquellos pasos que cubren la zona de estudio y asignar su SSH correspondiente. Guardar la información en matrices con latitud, longitud y altimetría específica (lat---, lon--- y ssh) para cada paso. Y esto es guardado en los archivos tpcoord_ssh_paso###. | Archivos: -tpcoord_ssh_paso# -tpcoord_ssh_paso## En este caso el número de archivos depende de los pasos en la región de prueba. |

Tabla 4.2.4 Algoritmo coord_paso_especif: genera archivos que contienen ssh y su posición.

Debido a que el satélite presenta fallas ocasionalmente, es necesario interpolar linealmente la información con el fin de cubrir esos espacios para que no causen incongruencias en los resultados. La corrección aplicada es un promedio de los puntos más cercanos de la matriz en la región elegida. La matriz es un arreglo de posiciones contra tiempo, y para realizar la interpolación se da prioridad a un promedio en espacio (columnas) en vez de tiempo (renglones).

| Entradas | Procesos | Salidas |
|---|--|--|
| Archivos: - tpcoord_ssh _paso# - tpcoord_ssh _paso## | Buscar en los arreglos huecos y sustituirlos por un promedio entre las 2 ó 3 posiciones más cercanas en la columna. Si no hay puntos cercanos en 3 posiciones buscadas, promediar con los datos en el renglón contiguo. Si no hay datos en dicho renglón, promediar con las 4 ó 5 posiciones más cercanas en la columna. Todo esto identificando los valores en tierra primeramente. Por último guardar los registros corregidos en los archivos altim180_80W_8_18N_interpolada# | Archivos: - altim180_80W_8_18N_ Interpolada# - altim180_80W_8_18N_ Interpolada## El número de archivos de salida debe de ser igual al de los de entrada. |

Tabla 4.2.5 Algoritmo proa_fig_interpolada: corrige la información linealmente y genera unos archivos corregidos con posición y ssh.

De esta manera se obtienen los datos de altimetría provenientes del T/P, que posteriormente servirán como entradas del modelo. El producto final son los archivos altim180_80W_8_18N_interpolada. La información contenida en estos archivos se utiliza para graficar las posiciones de estudio a través del tiempo.

Para el Jason-1 se realizó un proceso análogo, solo cabe remarcar que la información de este satélite es la continuación de los primeros ciclos y sitios de estudio del T/P. Después del desplazamiento del T/P y colocación del Jason-1, se tiene un muestreo mucho mayor del planeta dado que se sigue captando la información de los puntos iniciales además de los nuevos lugares de muestreo del satélite T/P, que hasta inicios del 2005 son más de 400 ciclos para el T/P y alrededor de 120 para el Jason-1.

Capítulo 5 Interpolación espacial y/o temporal usando funciones empíricas ortogonales complejas

En este capítulo se explica y desarrolla el método de interpolación que se propone en este trabajo haciendo uso de los conceptos presentados en las secciones anteriores. Cabe aclarar que el método puede ser aplicado a distintos campos de investigación en que se desee tener una mayor densidad de puntos en espacio y/o en tiempo. En la sección 5.2 se presentan los resultados de experimentos en los que se generan distintas mallas que simulan las trayectorias de los satélites y funciones semejantes al comportamiento de los datos. A partir de los resultados obtenidos se muestra la manera en que trabaja el método. En la última sección del capítulo se aplica el método de interpolación en espacio a datos reales de altimetría obtenidos de los satélites T/P y Jason-1.

5.1 Descripción de la técnica de interpolación

El objetivo de este proceso es extraer información estadística de señales que se propagan en un arreglo \mathbf{X} de tamaño p por n , formado por p posiciones en el espacio y n muestreos en el tiempo. Realizando el análisis de FEOCs se obtienen eigenmodos complejos después de aplicar la transformada de Hilbert a la matriz \mathbf{X} y formando la matriz \mathbf{Z} (Ver Cap. 2), es decir:

$$\mathbf{Y} = \mathbf{H}(\mathbf{X}) \quad \text{y}$$

$$\mathbf{Z} = \mathbf{X} + i \mathbf{Y}.$$

En el análisis de FEOCs se calcula la matriz de dispersión \mathbf{R} y se obtienen sus eigenmodos para posteriormente reconstruir los datos mediante la multiplicación de la matriz \mathbf{E}' , compuesta por h eigenvectores complejos en sus columnas correspondientes a los modos que representan mayor varianza

$$\mathbf{R} = \mathbf{Z}\mathbf{Z}^H/(n-1)$$

$$\mathbf{R}\mathbf{E} = \mathbf{E}\mathbf{\Lambda}$$

$$\hat{\mathbf{Z}} = \mathbf{Z}\mathbf{E}'(\mathbf{E}')^H \quad .$$

Haciendo la multiplicación de cada eigenvector de \mathbf{E} por la matriz \mathbf{Z} resulta un vector \mathbf{a}_i de tamaño p .

$$\mathbf{a}_i = \mathbf{Z}\mathbf{e}_i \quad \text{para } 1 \leq i \leq n \quad .$$

A este vector se le conoce como función de espacio y a cada eigenvector \mathbf{e}_i función de tiempo. Si se suman todos los productos de las funciones de espacio y el conjugado transpuesto de las funciones de tiempo se recupera el campo \mathbf{Z} ,

$$\sum_{i=1}^n \mathbf{a}_i \mathbf{e}_i^H = \sum_{i=1}^n \mathbf{Z}\mathbf{e}_i \mathbf{e}_i^H = \mathbf{Z} \sum_{i=1}^n \mathbf{e}_i \mathbf{e}_i^H = \mathbf{Z} \mathbf{I}_n = \mathbf{Z}.$$

Esta descomposición es usada para realizar la interpolación en espacio y/o tiempo la cual se describe a continuación.

Las funciones de espacio y tiempo obtenidas anteriormente tienen entradas complejas y pueden ser expresadas en términos de sus componentes de amplitud y de fase.

$$\mathbf{a}_i = \{\mathbf{a}_{ij}\} = R(ij)\exp(i\theta_{ij}) \quad \text{y}$$

$$\mathbf{e}_i = \{\mathbf{e}_{ij}\} = E(ij)\exp(i\varphi_{ij}).$$

Y la parte real del producto

$$\Re \{ \mathbf{a}_i \mathbf{e}_i^T \} = \Re \left\{ \sum_{j=1}^n R(ij) E(ij) \exp(i[\theta_{ij} + \phi_{ij}]) \right\} = \sum_{j=1}^n E(ij) R(ij) \cos(\theta_{ij} + \phi_{ij}) ,$$

es solo una parte del campo \mathbf{X} reconstruido, la parte imaginaria tiene la forma

$$\Im \{ \mathbf{a}_i \mathbf{e}_i^H \} = \Re \left\{ R(i) E(i) \exp(i[\theta_i + \phi_i]) \right\} = E(i) R(i) \sin(\theta_i + \phi_i) ,$$

donde $E(i)$ es la variabilidad de la amplitud en espacio asociado a cada eigenmodo dado, y $R(i)$ es la función de amplitud espacial. El término $\theta_i + \phi_i$ representa la fase en una posición y tiempo dado.

Tanto la parte real como la imaginaria aportan información del campo \mathbf{X} , si esto es analizado por separado y posteriormente se utiliza para la reconstrucción de un dominio más grande y fino que el original usando la corrección de información de fase y amplitud, se realiza un proceso llamado interpolación en espacio, donde el objetivo buscado es tener datos en posiciones en las cuales no se muestreó.

Ahora se presenta la manera en que la parte real e imaginaria de los productos son utilizadas para interpolar. Cuando se tiene un número complejo de la forma

$$z_1 = a_1 + i b_1 ,$$

puede ser escrito en términos de su amplitud y su fase

$$z_1 = \rho_1 \exp(i\theta_1)$$

en vez de interpolar las partes real e imaginaria, lo que se interpola son la fase y la amplitud, por separado, conservando la forma del movimiento que tenía la onda o función original.

Por ejemplo dados los números complejos z_1 y z_2 de la forma

$$z_1 = a_1 + i b_1 ,$$

$$z_2 = a_2 + i b_2 ,$$

y realizando la interpolación de $\hat{z}_{3/2}$ linealmente se tendría:

$$\hat{z}_{3/2} = (a_1 + a_2)/2 + i(b_1 + b_2)/2$$

mientras que con la interpolación propuesta resultaría:

$$\hat{z}_{3/2} = (\rho_1 + \rho_2)/2 \exp [i (\theta_1 + \theta_2)/2]$$

o sustituyendo por sus equivalentes de z_1 y z_2 resulta.

$$\hat{z}_{3/2} = (\sqrt{a_1^2 + b_1^2} + \sqrt{a_2^2 + b_2^2})/2 \exp [i (\arctan\left(\frac{b_1}{a_1}\right) + \arctan\left(\frac{b_2}{a_2}\right))/2] \quad ,$$

en donde claramente se ve una diferencia entre ambos promedios. Cuando se realiza una interpolación con amplitud y fase se conservan propiedades de onda.

5.2 Generación de mallas ortogonales utilizando FEOCs

Los ejemplos que se presentan a continuación permiten analizar qué tan bien trabaja el método propuesto en situaciones en las que la malla de muestreo tiene distinta estructura de la que se desea utilizar, para lo cual se cambian las características que definen a una malla, por ejemplo: el espaciamiento entre cada punto de la malla ó la frecuencia de mediciones en el mismo punto. Los remolinos oceánicos son claros ejemplos de fenómenos que pueden ser observados con la altimetría con un muestreo espacial pobre por lo que son elegidos para realizar una simulación para examinar el método.

Experimento 1

La malla de la trayectoria del satélite T/P sirve adecuadamente para probar el método propuesto y así observar el proceso de interpolación en los lugares donde no se toman mediciones, con la ventaja de conocer los verdaderos valores en los puntos estimados.

La región seleccionada está al sur de México entre las coordenadas 150° y 120° de longitud oeste y de 8° a 18° latitud norte, la cual se representa en la figura 5.2.1.

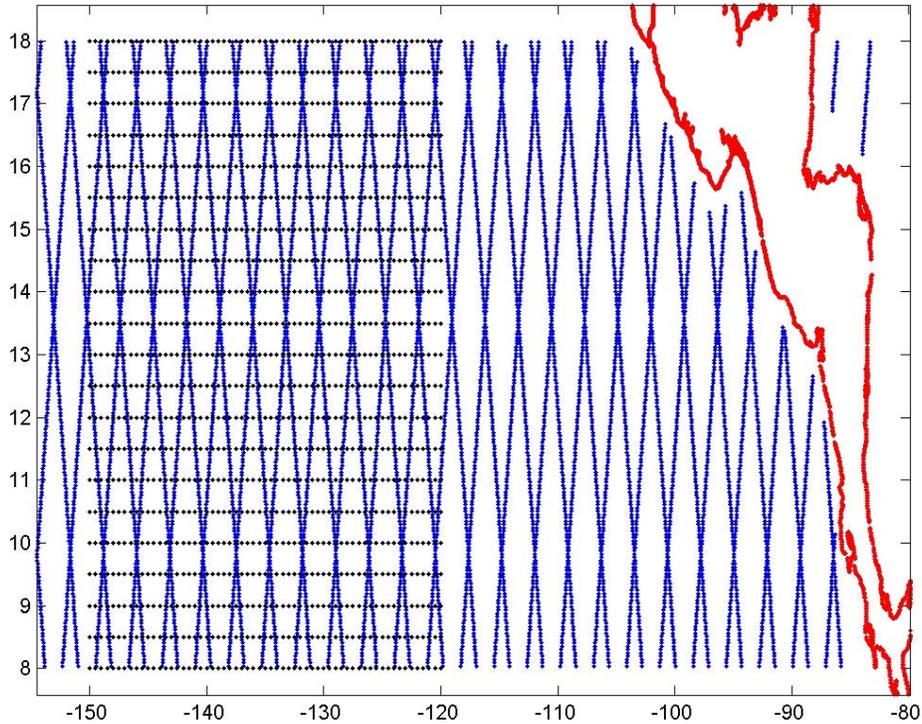


Figura 5.2.1 Malla del satélite Topex/Poseidon (azul), malla de prueba (negro). Con rojo la costa de centroamerica y sur de México.

Una función de forma gaussiana en dos dimensiones es utilizada para simular el remolino oceánico en distintos ciclos de muestreo y tiene la siguiente función analítica:

$$Z = \frac{\exp\left(-\frac{x^2 + y^2}{5}\right)}{5\pi}, \quad \text{el dominio de } x \text{ y } y \text{ es de } -3.5 \leq x, y \leq 3.5.$$

En la figura 5.2.2 se presentan la superficie y curvas de nivel de la función de estudio.

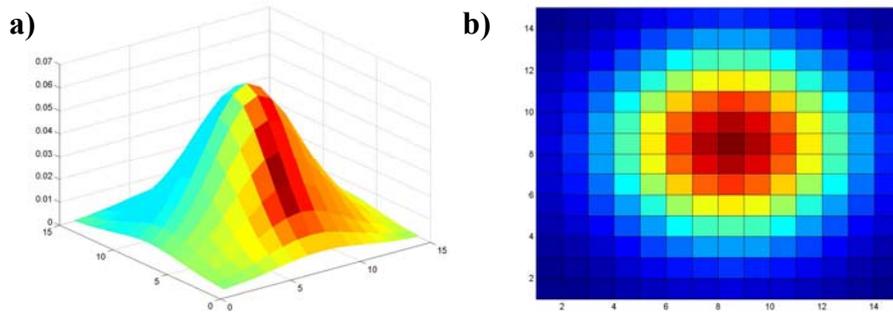


Figura 5.2.2 a) Superficie gaussiana. b) Curvas de nivel de la misma superficie.

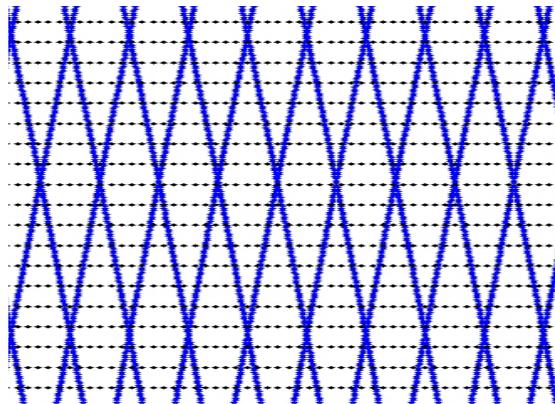


Figura 5.2.3 Malla de valores de la función Z (puntos negros) y trayectorias del satélite T/P (puntos azules) en la región seleccionada.

De esta manera la función Z simula la perturbación asociada al remolino en la malla propuesta, es decir, la figura 5.2.3 presenta dos mallas, una de color negro y una de color azul, la malla negra es el dominio de la función Z en el cual hay una distancia de .5 grados entre cada punto, fijando los valores en el eje vertical y posteriormente haciendo incrementos negativos de .5 grados en el eje horizontal. Se simula el movimiento, de este a oeste, de un remolino oceánico, los valores obtenidos se interpolan a los puntos azules para cada tiempo. Posteriormente, con los valores de las trayectorias se interpola a los puntos negros con dos métodos de interpolación: el lineal y el propuesto con FEOCs. Los resultados obtenidos se presentan en imágenes en las que se observa y, posteriormente evalúa, el comportamiento de la interpolación. Las imágenes producidas constan de 3 recuadros donde el primero presenta la función original y las posiciones de muestreo, el

segundo la reconstrucción que se obtiene con el método lineal, y por último se presenta la reconstrucción de la imagen haciendo uso de FEOCs.

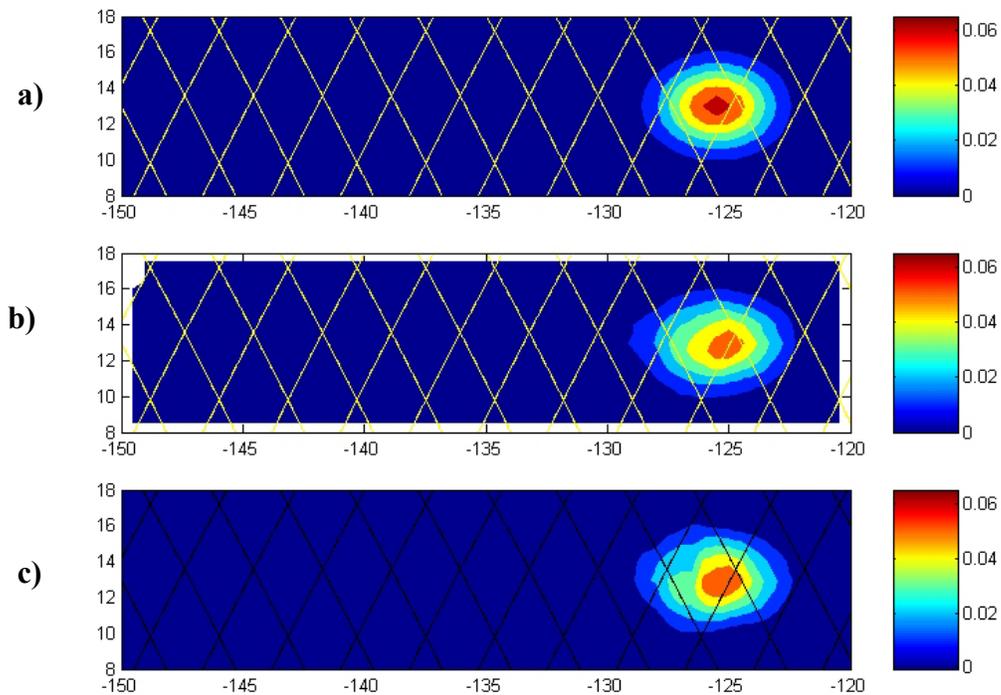


Figura 5.2.4 a) Función gaussiana en dos dimensiones, b) reconstrucción utilizando interpolación lineal, y c) reconstrucción utilizando interpolación con FEOCs. La barra de colores de la derecha representa el rango de valores que la función Z toma respecto del dominio seleccionado.

Con esta información es posible obtener el error cuadrático medio entre los datos originales de las interpolaciones lineal y con FEOCs, mostrando las diferencias en los resultados de ambas interpolaciones.

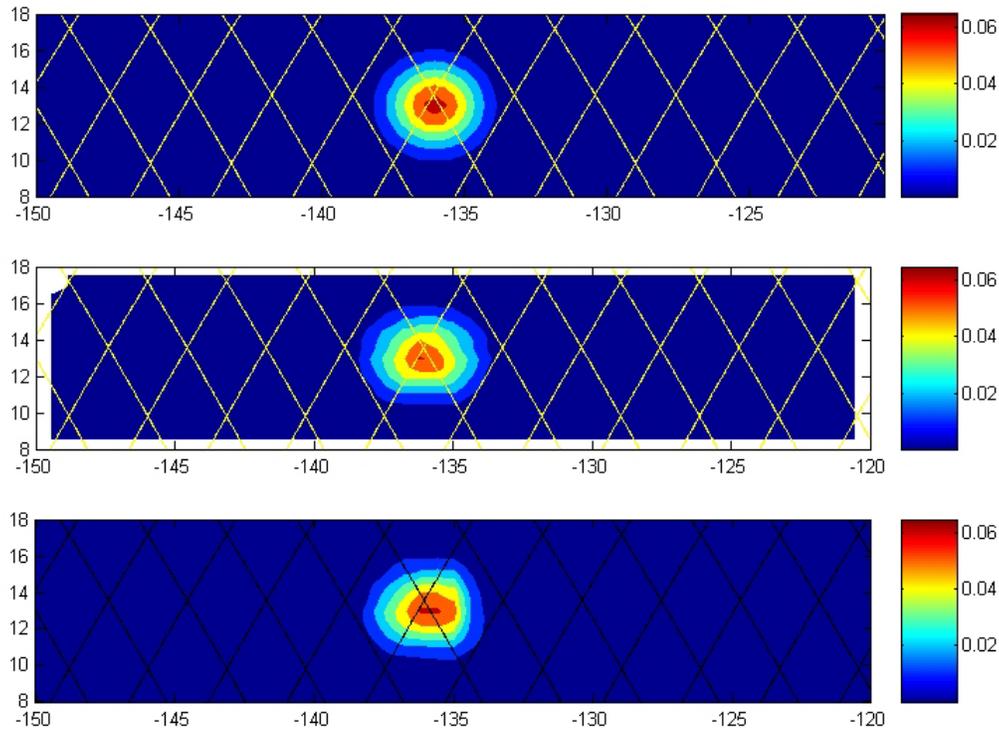


Figura 5.2.5 Diferencias entre la interpolación lineal y con FEOCs. Nótese las diferencias en el centro de ambas interpolaciones, con FEOCs se recupera información real fuera de la malla original, a diferencia de la interpolación lineal.

Después de examinar las figuras anteriores se detectan algunas diferencias, por ejemplo, se aprecia que el tipo de malla altera la interpolación en zonas cercanas a los puntos de muestreo. En la generada usando FEOCs se observa información que el método lineal no recupera (Fig. 5.2.5). Se calculan los errores cuadráticos medios para 40 ciclos o tiempos. Al comparar este parámetro se observa una diferencia entre ambos métodos de .004 siendo menor el obtenido con la interpolación lineal por lo que se concluye que el método tiene deficiencias. En los siguientes capítulos se muestra que el método tiene problemas asociados a la interpolación en fase.

| Interpolación | Lineal | FEOC |
|---------------|--------|-------|
| E. C. M. | 0.014 | 0.017 |

Tabla 5.2.1 Errores cuadráticos medios de las interpolaciones lineal y con FEOCs.

Experimento 2

En la figura 5.2.6 se presentan las mallas del experimento 2, la malla de color azul cielo representa el dominio de la función Z , tiene forma rectangular, con un espaciamiento fijo de $.25^\circ$ entre un punto y otro. La malla de color negro representan las posiciones de estudio para este experimento y tiene un espacio de 2° entre una trayectoria y otra, y los puntos a lo largo de cada trayectoria un espaciamiento de $.25^\circ$. La región simulada para ambas mallas va de 150° a 110° de longitud oeste y 5° a 20° de latitud norte.

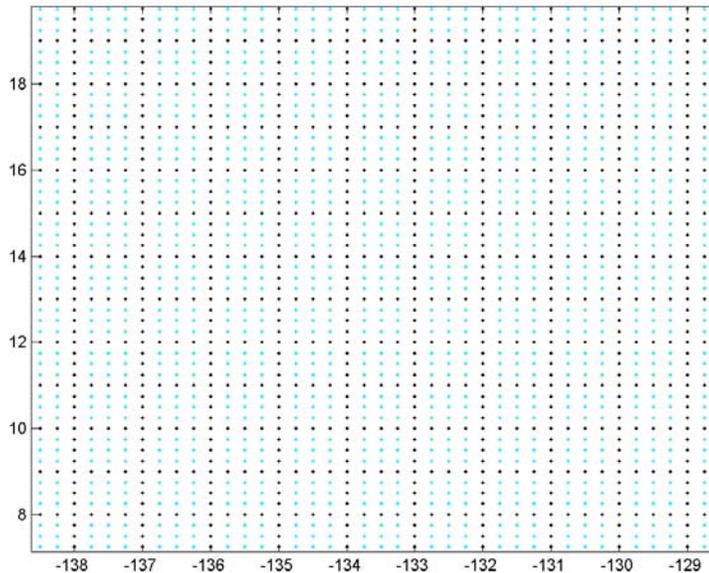


Figura 5.2.6 Malla del experimento 2. Los puntos azules representan el dominio de la función Z . Los puntos negros representan la malla de muestreo.

Se elige esta estructura debido a que es posible tener un muestreo simple siguiendo los puntos de la malla, como lo es a longitud constante, siguiendo una trayectoria ascendente o descendente a la vez. Además se representa la función que simula la perturbación producida por un remolino en forma análoga a la detectada con altimetría satelital.

A la función Z de este experimento se le conoce como sombrero mexicano y tiene una variación en forma oscilatoria y su expresión analítica es la siguiente:

$$Z = \frac{\text{sen}(R)}{R + .5}, \text{ donde } R = \sqrt{x^2 + y^2} \text{ el dominio para } x \text{ e } y \text{ es el siguiente:}$$

$$-46.5 \leq x \leq 27.5 \text{ y } -15 \leq y \leq 15$$

En la figura 5.2.7 se presentan la superficie y curvas de nivel de Z .

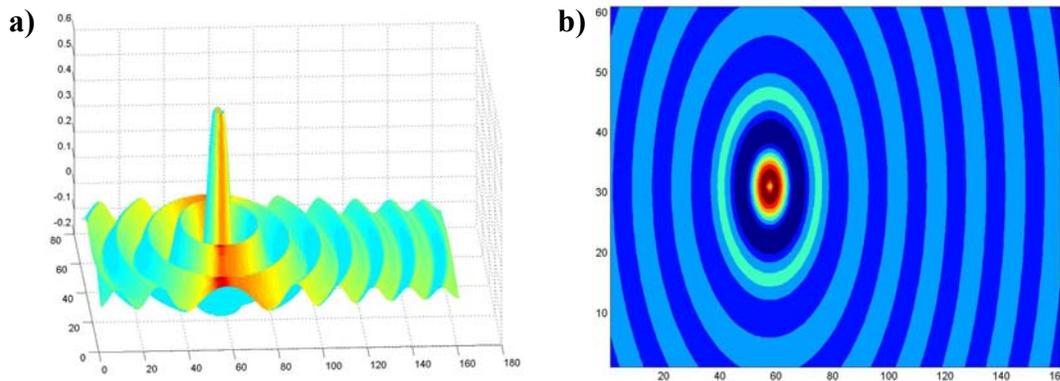


Figura 5.2.7 a) Superficie de la función Z y b) curvas de nivel de la función Z .

El proceso de cálculo de datos para las trayectorias de estudio es análogo al del experimento 1; primero se evalúa la función en una posición inicial y , posteriormente, se va desplegando a lo largo del eje horizontal en incrementos de .25 unidades. Cada nueva posición representa un nuevo ciclo. De esta manera se generan tantos tiempos de estudio simulados como incrementos en la horizontal. Para cada tiempo se muestran los valores a la malla negra generando así los valores de las trayectorias en los puntos deseados, y finalmente se recupera la función Z interpolando linealmente y con FEOCs a los puntos azules en la figura 5.2.6.

Los resultados se presentan en una imagen dividida en 2 recuadros donde el primero es la reconstrucción de la función Z con interpolación lineal y el segundo con FEOCs. La representación visual de los datos es muy similar para ambos métodos y las características de la función, en general, están bien representadas.

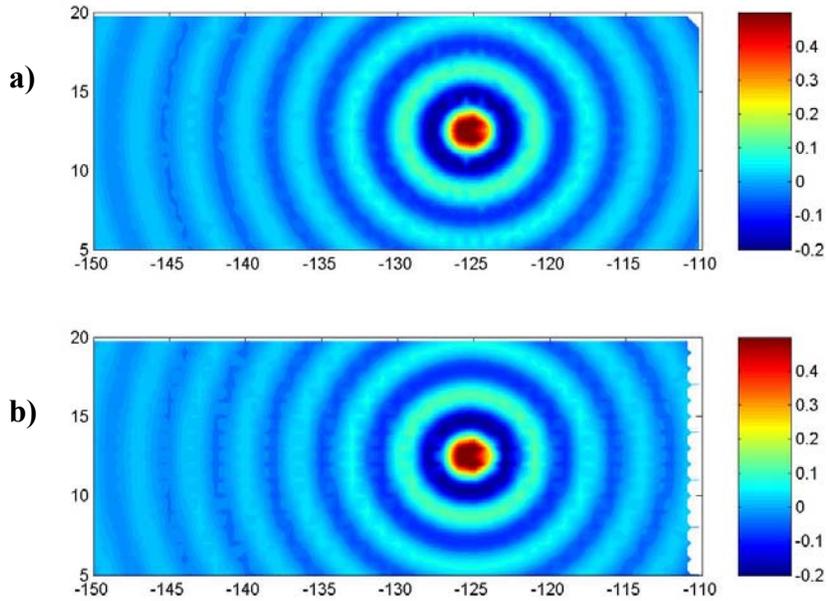


Figura 5.2.8 a) Interpolación lineal y, b) interpolación con FEOCs.

Los errores cuadráticos medios son presentados en la tabla 5.2.2, este parámetro sirve como una medida para cuantificar la diferencia entre los valores reales y los estimados, en este caso el error fue menor para la interpolación con el método propuesto.

| Interpolación | Lineal | CEOF |
|---------------|--------|--------|
| E. C. M. | 0.0072 | 0.0055 |

Tabla 5.2.2- Errores cuadráticos medios para el experimento 2 de interpolación lineal e interpolación con FEOCs

El método con FEOCs tiene un error cuadrático menor. Tiene además la ventaja de realizar la interpolación sin aplicar alisamiento a los datos como lo hace la interpolación lineal, representando correctamente la información en posiciones donde no la había.

Experimento 3

La propuesta del presente trabajo se basa en una interpolación de datos complejos realizando una interpolación en amplitud y fase de las matrices de información. Sin embargo, dependiendo del software de interpolación que se utilice. Al realizar algunas correcciones en el programa de computo estos parámetros pueden tener una mejora, pero ésto solo después de un extenso análisis (comentado en la sección de conclusiones) y un perfeccionamiento en los algoritmos computacionales, ésto fácilmente puede ser implementado en los experimentos 1 y 2.

En el experimento 3 se utilizan los ciclos 369 a 378 del satélite T/P y del 26 al 35 del J-1, los ciclos pertenecen al periodo de septiembre 18 de 2002 a abril 6 de 2003. El experimento tiene lugar en la zona ubicada de 180° a 85° oeste y de 8° a 18° norte. Los veinte ciclos utilizados tienen la característica de que corresponden a un período en que ambos satélites estuvieron en órbita, de esta forma los datos del satélite T/P son usados para estimar los valores de altimetría en las trayectorias del J-1 mediante interpolación lineal y con FEOCs.

La región seleccionada y las trayectorias de ambos satélites se presentan en la figura 5.2.9.

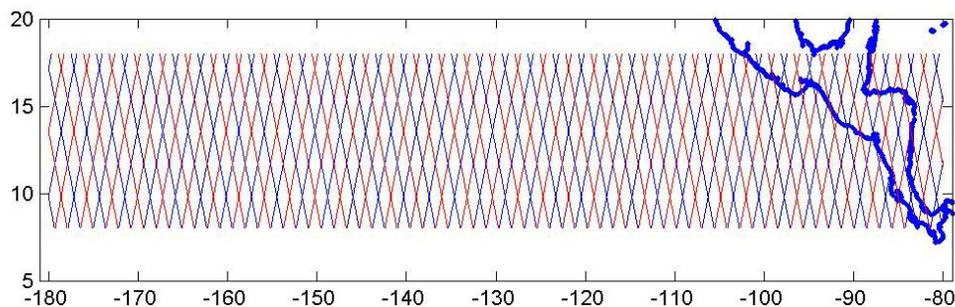


Figura 5.2.9 Línea azul trayectorias del satélite J-1. Línea roja trayectoria del satélite T/P.

Una de las imágenes producidas al interpolar la zona de estudio es presentada en la figura 5.2.10.

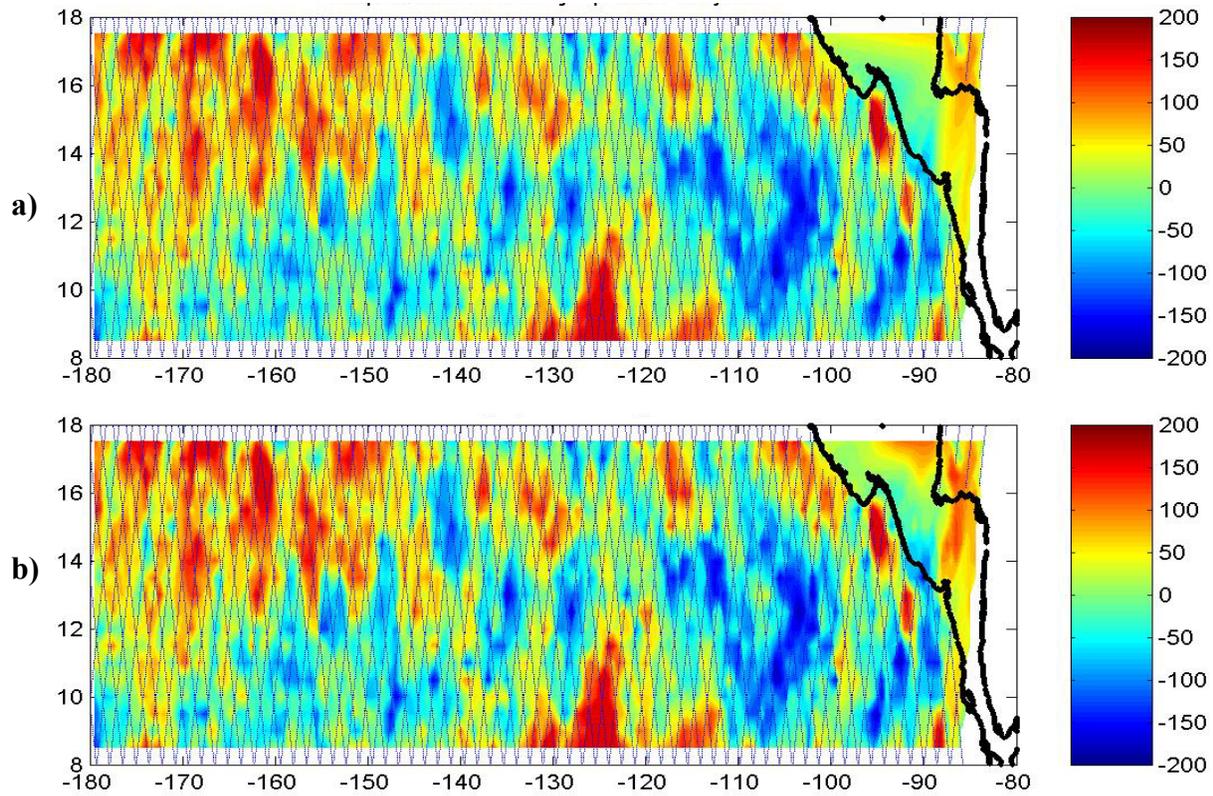


Figura 5.2.10 a) Interpolación lineal utilizando ambos satélites. b) Interpolación utilizando FEOCs.

Los resultados presentados en la tabla 5.2.3 indican que el error cuadrático medio es muy similar con ambos métodos aunque ligeramente mayor con el método propuesto (se consideran 10 ciclos de prueba).

| Interpolación | Lineal | FEOCs |
|---------------|--------|----------|
| E. C. M. | 680.92 | 695.7199 |

Tabla 5.2.3 Errores cuadráticos medios para el experimento 3 de interpolación lineal e interpolación con FEOCs.

Una medida de la energía en un sistema atmosférico es función de la desviación de las líneas de igual presión con respecto a la media espacial. Se puede definir un índice como la suma de los cuadrados de la diferencia de los valores observados en distintos

tiempos con respecto a su media. Esta medida es una varianza y la denotamos por ECD. El cual se presenta en la siguiente fórmula:

Dada la matriz original de datos \mathbf{A} con columnas \mathbf{a}_i con $1 \leq i \leq n$

$$EC(j) = \frac{(a_j - \mu_2(a_j))^T * (a_j - \mu_2(a_j))}{p},$$

donde $\mu_2(a_j) = \frac{\sum_{i=1}^p a_{ij}}{p},$

y al final

$$ECD = \sum_{j=1}^n \frac{EC(j)}{n}$$

Obsérvese que ECD es distinto a la varianza temporal de los datos.

| Interpolación | Lineal | FEOCs |
|---------------|--------|--------|
| ECD | 412.25 | 417.13 |

Tabla 5.2.4 Escala de conservación de energía del sistema para el experimento 3, representando la ECD con interpolación lineal y de FEOCs.

En la tabla 5.2.4 se presenta la tabla de conservación de energía del sistema, de donde puede concluirse que ambos métodos tienen un nivel en los ECM semejantes pero el método propuesto interpola correctamente la información sin aplicar el alisamiento del método lineal a los datos, es decir una interpolación que no elimina las variaciones de interés conservando así las señales de los datos.

Experimento 4

En las pruebas anteriores se usó el método de interpolación espacial con FEOCs a mallas con mayor densidad espacial. Sin embargo las aplicaciones del método propuesto son mayores y a continuación se presenta un ejemplo de cómo puede ser usado para generar un muestreo con mayor frecuencia temporal. La información de este experimento es presión a nivel del mar recolectada cada 6 horas en una región limitada entre 230° a 285° de longitud oeste y 10° a 50° de latitud norte. Utilizando la información de cada 12 horas se busca estimar el valor del campo en el tiempo restante. En este experimento, no se estiman nuevas posiciones sino que se busca información en los mismos puntos, es decir, tomar los valores de cada 12 horas para estimar los datos de cada 6 horas. Para la obtención de la información restante se realiza lo siguiente: a la matriz de eigenvectores se le interpolan nuevas columnas al tiempo deseado, dado que el número de columnas en \mathbf{E} depende directamente de la cantidad de muestreos realizados, al interpolar nuevos eigenvectores, teóricamente se tienen más tiempos de muestreo de la información deseada. Debido a que \mathbf{E} es la matriz de eigenvectores de una matriz compleja \mathbf{Z} , \mathbf{E} tiene elementos complejos y es ahí donde el método de interpolación es usado descomponiendo las entradas complejas en información de amplitud y fase que posteriormente es utilizado en el cálculo de los eigenvectores interpolados.

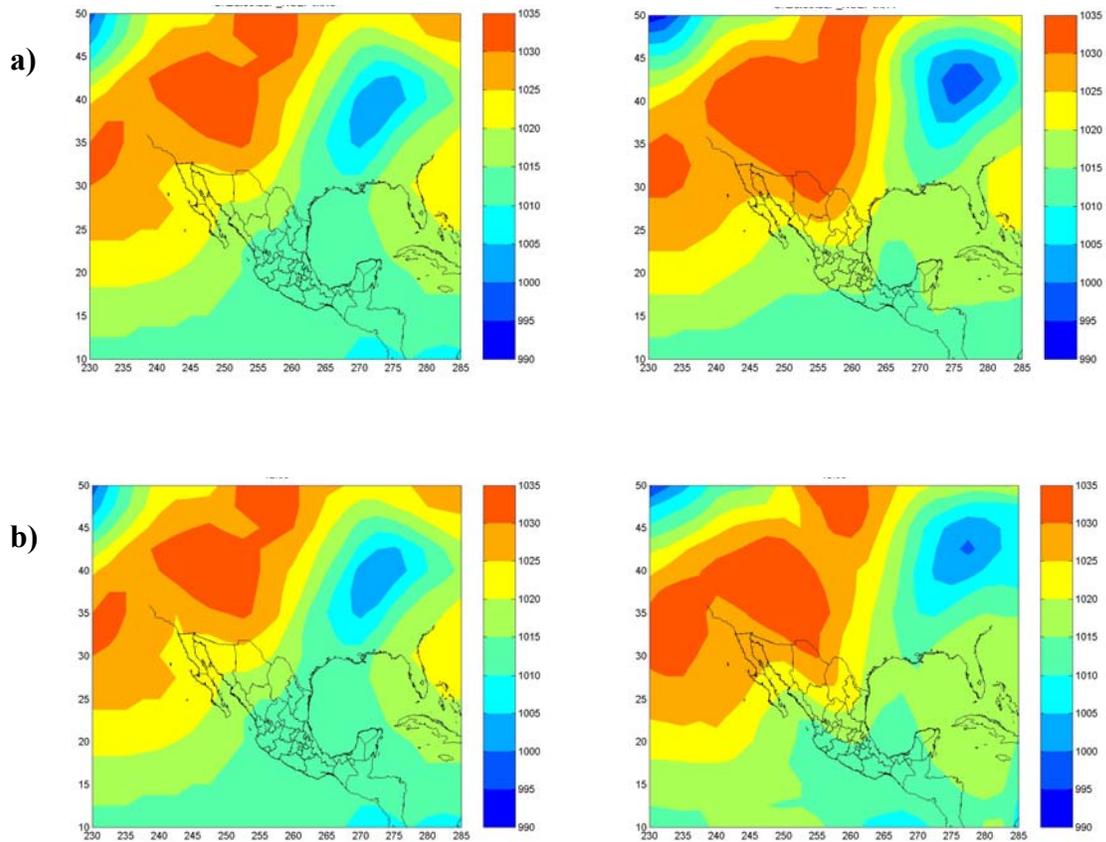


Figura 5.2.11 a) Imágenes provenientes de datos reales de presión a nivel del mar, y b) imágenes generadas interpolando con FEOCs en tiempo para los datos de a).

Puede apreciarse que las anomalías más grandes en general están bien representadas. Los errores cuadráticos medios están presentados en la tabla 5.2.4.

| Interpolación | Lineal | FEOCs |
|---------------|--------|--------|
| E. C. M. | 4.4883 | 4.9288 |

Tabla 5.2.5 Errores cuadráticos medios para el experimento 4 de interpolación lineal y con FEOCs.

Ambos métodos dan porcentaje de error similar con una diferencia de casi el 1% de error con los datos originales, esto a favor del método lineal. La preservación de energía de los datos se presenta en la tabla 5.2.5.

| Interpolación | Original | Lineal | CEOF |
|---------------|----------|--------|--------|
| E. C. D. | 5.2574 | 4.4883 | 4.9288 |

Tabla 5.2.6 Escala de conservación de energía del sistema atmosférico para el experimento 4, presentando la ECD original, con interpolación lineal y la resultante con interpolación FEOCs.

Es decir las variaciones de interés son mejor preservadas en los datos generados con FEOCs que con la interpolación lineal, también en una interpolación en tiempo.

Capítulo 6 Resultados

Los huecos o agujeros en la información son un problema en muchas investigaciones geofísicas. En el caso de los satélites Topex/Poseidon y Jason-1 los huecos en la información son provocados por fallas de los instrumentos que llevan a bordo, lo que genera inconsistencias al realizar la interpolación de datos. El problema se remedia parcialmente implementando subrutinas de control de calidad de la información antes de aplicar directamente la interpolación con FEOCs.

En el capítulo 4 se explica la forma en la que se resuelve el problema, interpolando linealmente la información en aquellos lugares en donde el satélite tuvo fallas (programa prog_fig_interpolada). Como un ejemplo de los huecos en la información, en la figura 6.1 se muestran gráficas de posiciones muestreadas por el satélite Topex/Poseidon. Dado que el satélite pasa por las mismas posiciones éstas deberían corresponder a tierra ó mar, pero cuando hay alguna avería o descompostura de algún altímetro del satélite la información es interpretada como una posición en tierra y por ello hay puntos blancos (que representan posiciones en tierra) donde en realidad hay mar en las gráficas de la figura 6.1. Ejemplos de las imágenes una vez que se han corregido los datos se presentan en la figura 6.2.

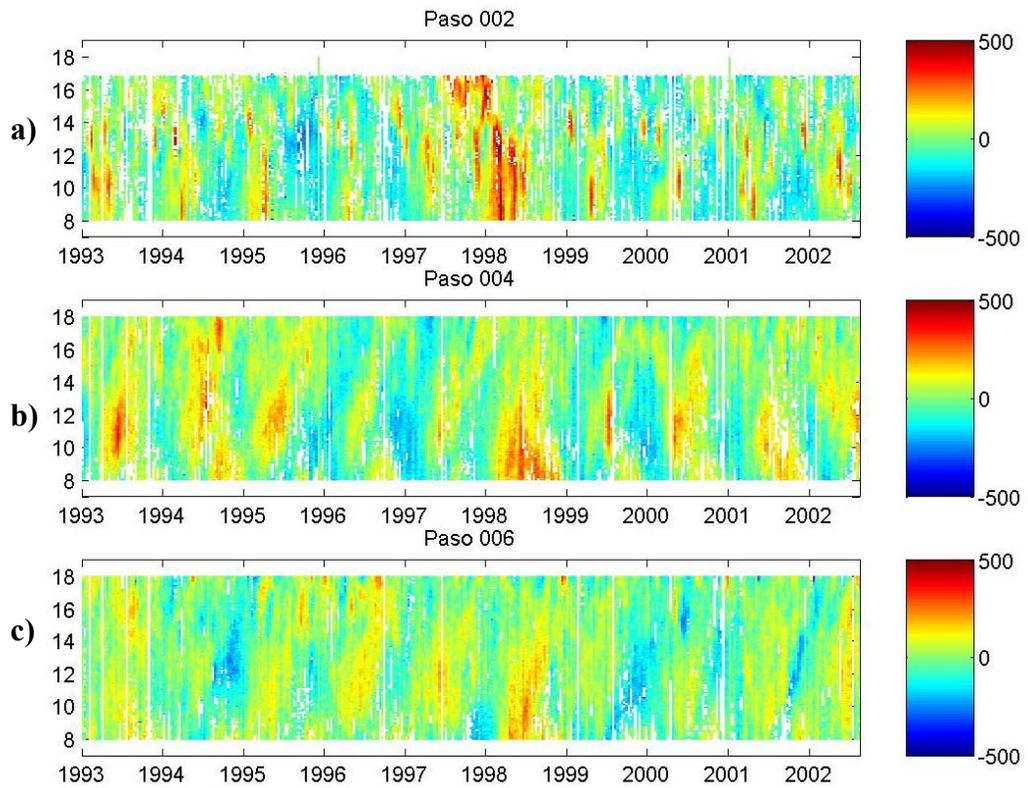


Figura 6.1 Series de tiempo de los datos originales de SSH para: a) Paso 2, b) Paso 4 y c) Paso 6. El eje horizontal representa el número del ciclo de muestreo y el vertical la posición en latitud. La barra de colores representa los valores de la anomalía en milímetros. Los puntos de color blanco representan posiciones en tierra o fallas del instrumento.

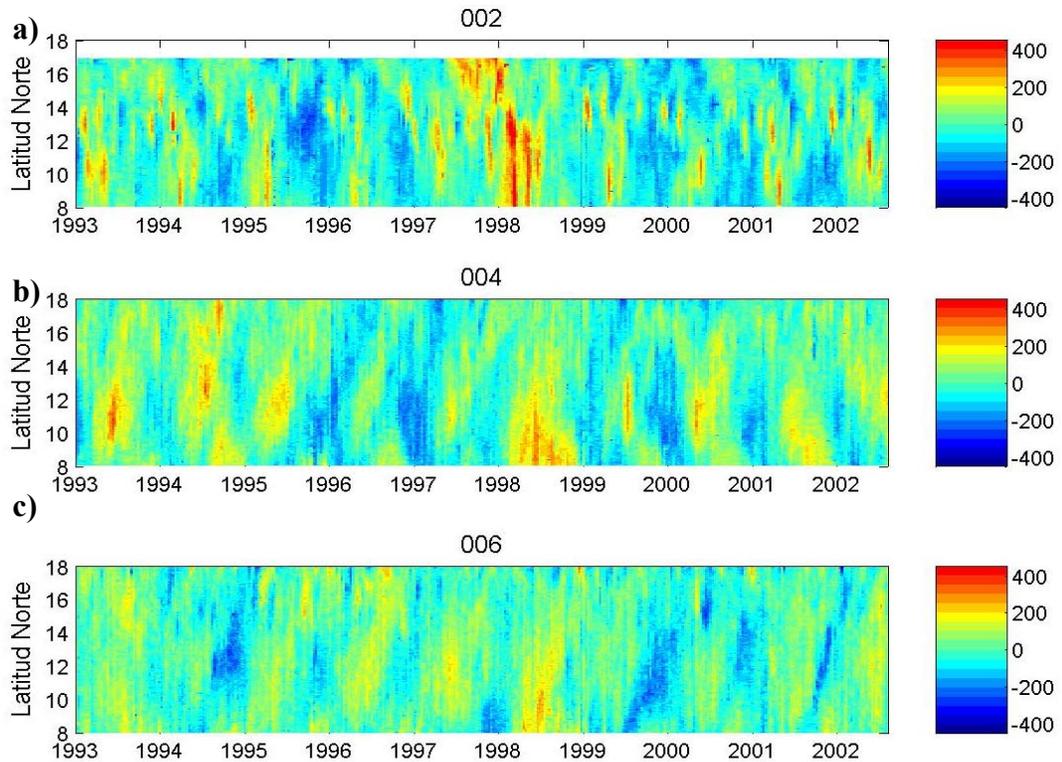


Figura 6.2 Series de tiempo de datos corregidas para: a) Paso 2, b) Paso 4 y c) Paso 6. La escala horizontal representa la fecha del ciclo de muestreo. La barra de colores representa los valores de la anomalía en milímetros.

Las gráficas de la figura 6.3 ilustran el número de fallas o averías del satélite T/P para los pasos 2, 4 y 6. En cada paso el satélite muestrea 3127 posiciones en tierra, tomadas cada 6 kilómetros, y se estudiaron 355 lecturas de cada posición correspondientes a 355 ciclos. El eje horizontal de las gráficas representa el número de posición del paso, y el eje vertical representa el número de datos no útiles del satélite. En el caso óptimo el eje vertical sólo tendría 2 valores; 0 y 355, es decir, 100% y 0% de datos nulos, que representarían posiciones en mar o tierra, respectivamente.

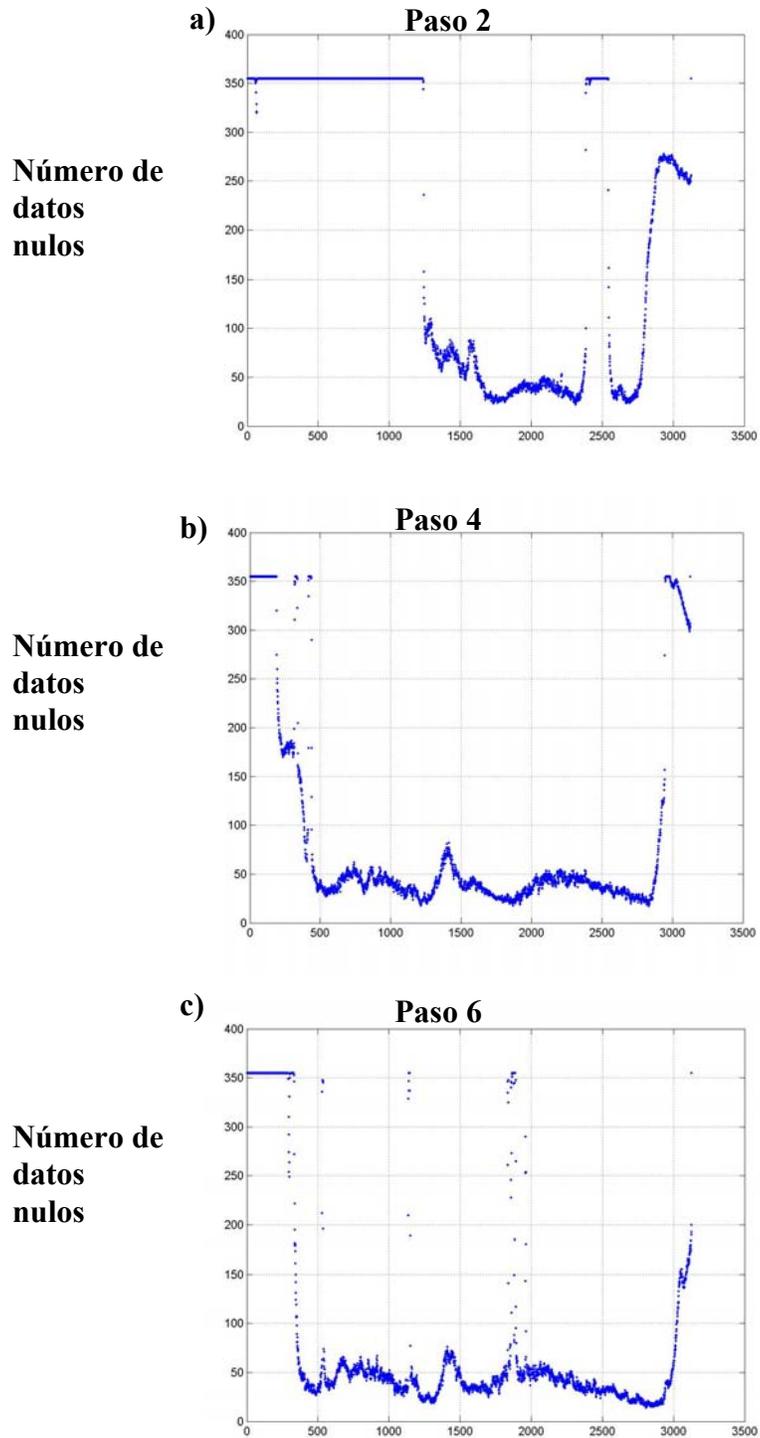


Figura 6.3 Gráficas correspondientes al número de datos no útiles o nulos del satélite T/P para: a) Paso 2, b) Paso 4 y c) Paso 6. El número 355 representa el 100% de datos nulos que corresponden a una posición en tierra, y el 0 representa 0% de datos nulos que correspondería a una posición en mar.

El objetivo del trabajo no es la corrección de los datos de altimetría satelital, pero dado que la información generada puede ser utilizada para otros análisis e investigaciones, se organizó una base de datos de la región comprendida entre 180° y 80° de longitud oeste y entre 8° y 18° de latitud norte para el período de 1993 a 2005. En la figura 6.4 se muestran los mapas de posición de los pasos ascendentes y descendentes del satélite T/P en esta región. Las imágenes generadas pueden servir para realizar estudios sobre remolinos ciclónicos, características de mesoescala, análisis estadísticos de remolinos oceánicos en la zona, tamaño, tiempo de vida, ocurrencia en años afectados por el fenómeno de El Niño o estudios sobre el cambio climático.

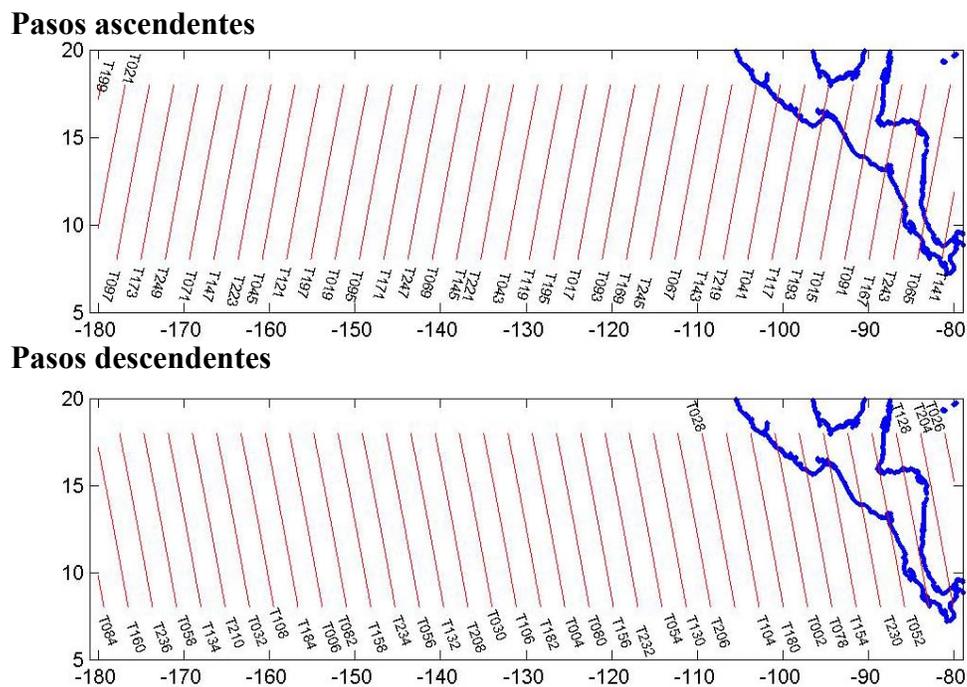


Figura 6.4 Mapas de posición de los pasos ascendentes (gráfica superior) y descendentes (gráfica inferior) del satélite T/P en la región de estudio. Los pasos ascendentes tienen números impares y los descendentes números pares.

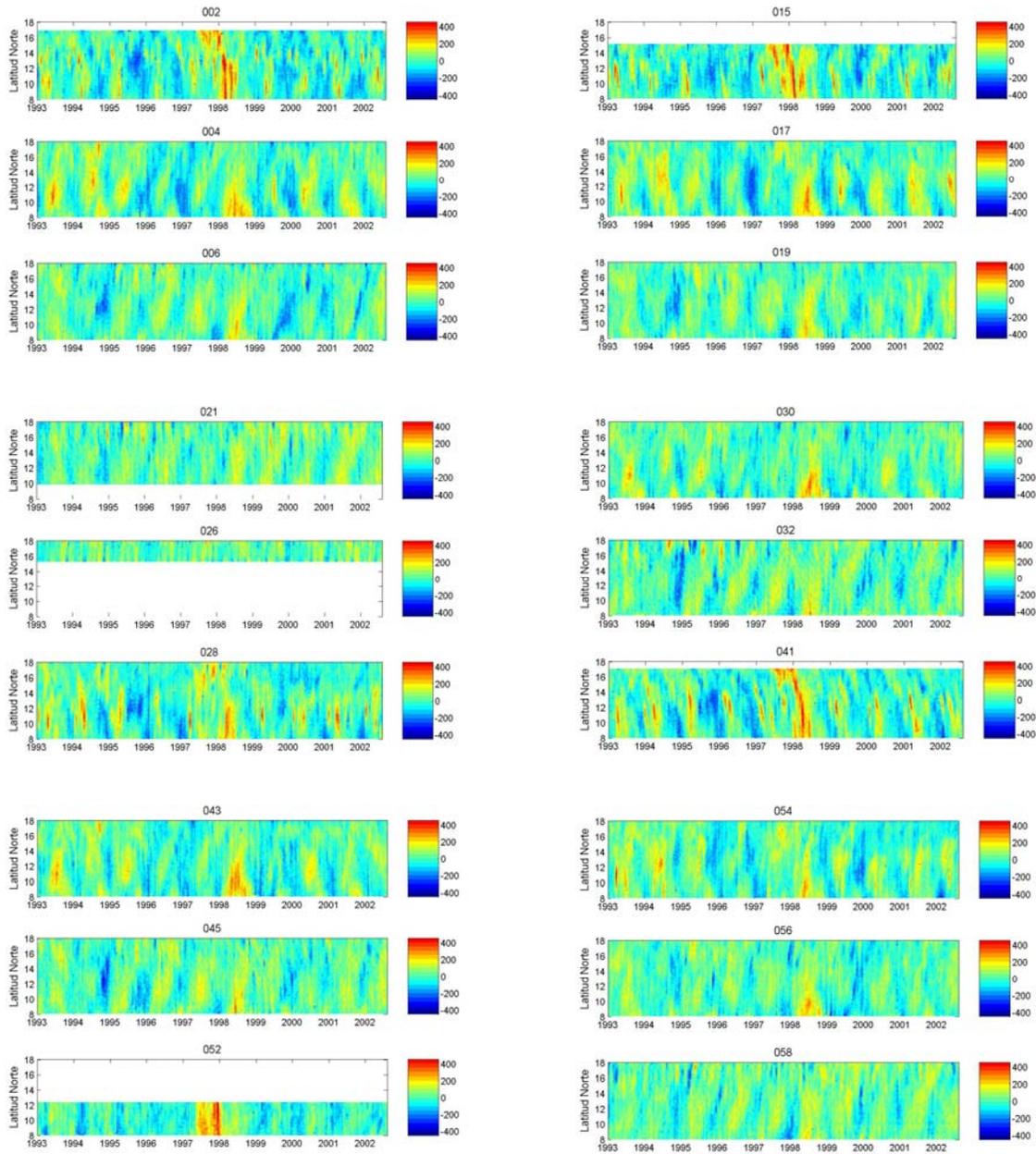


Figura 6.5 Series de tiempo de los datos de altimetría (mm) del satélite T/P, para los ciclos 10 a 365, correspondientes a los pasos mostrados en la figura 6.4 dentro de la región de estudio. El número en la parte superior de cada gráfica indica el paso correspondiente del satélite.

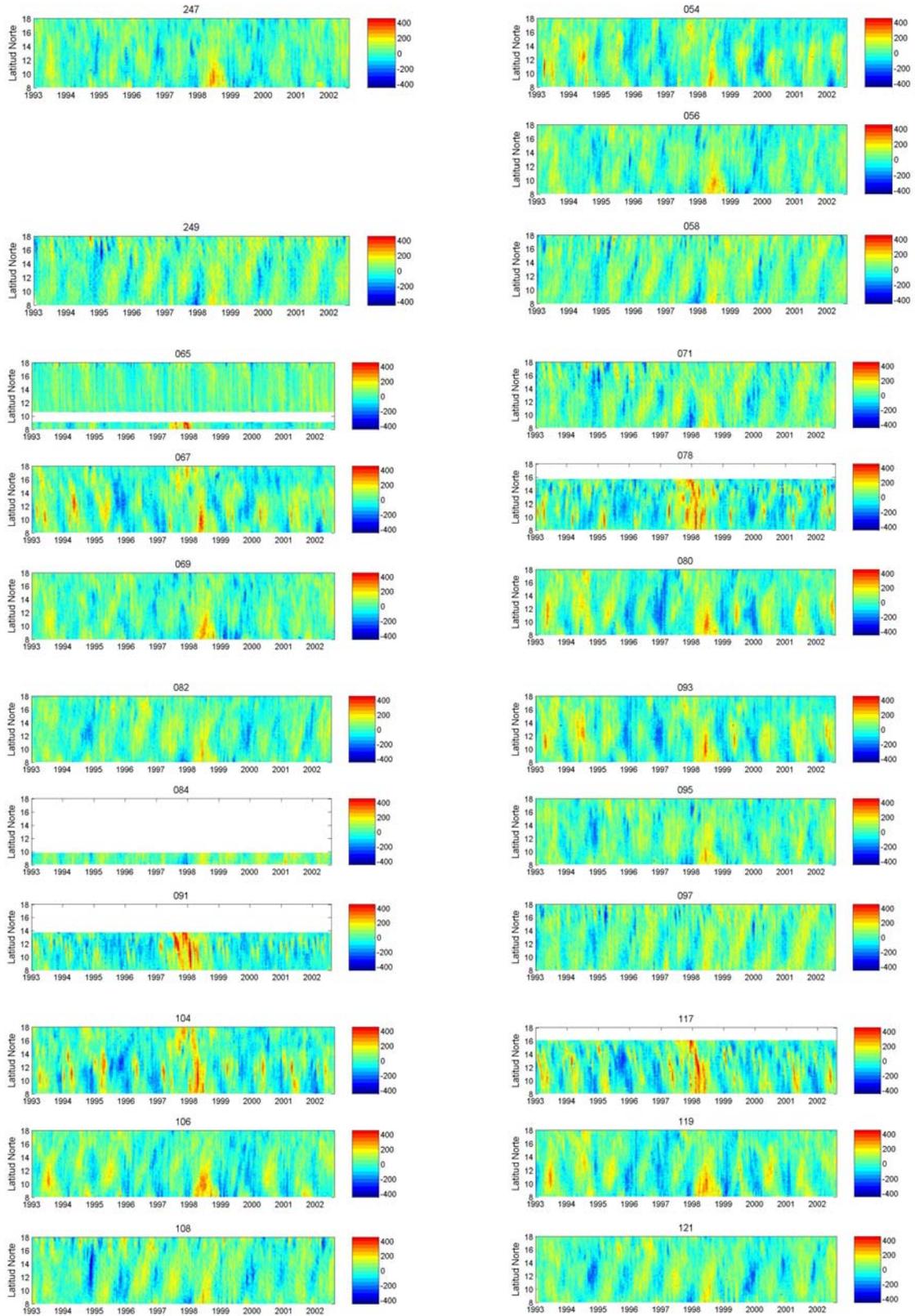


Figura 6.5 (Continuación)

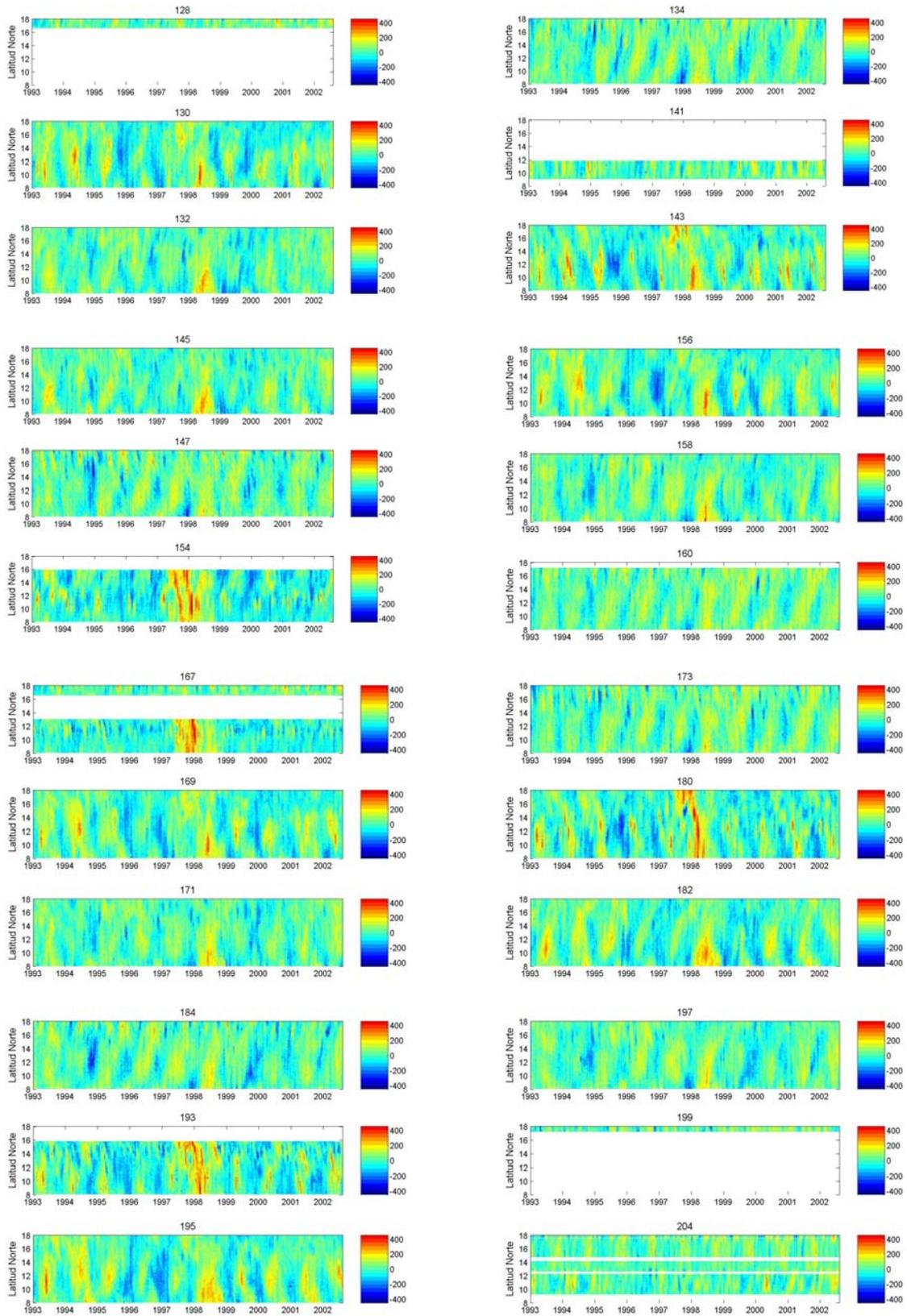


Figura 6.5 (Continuación)

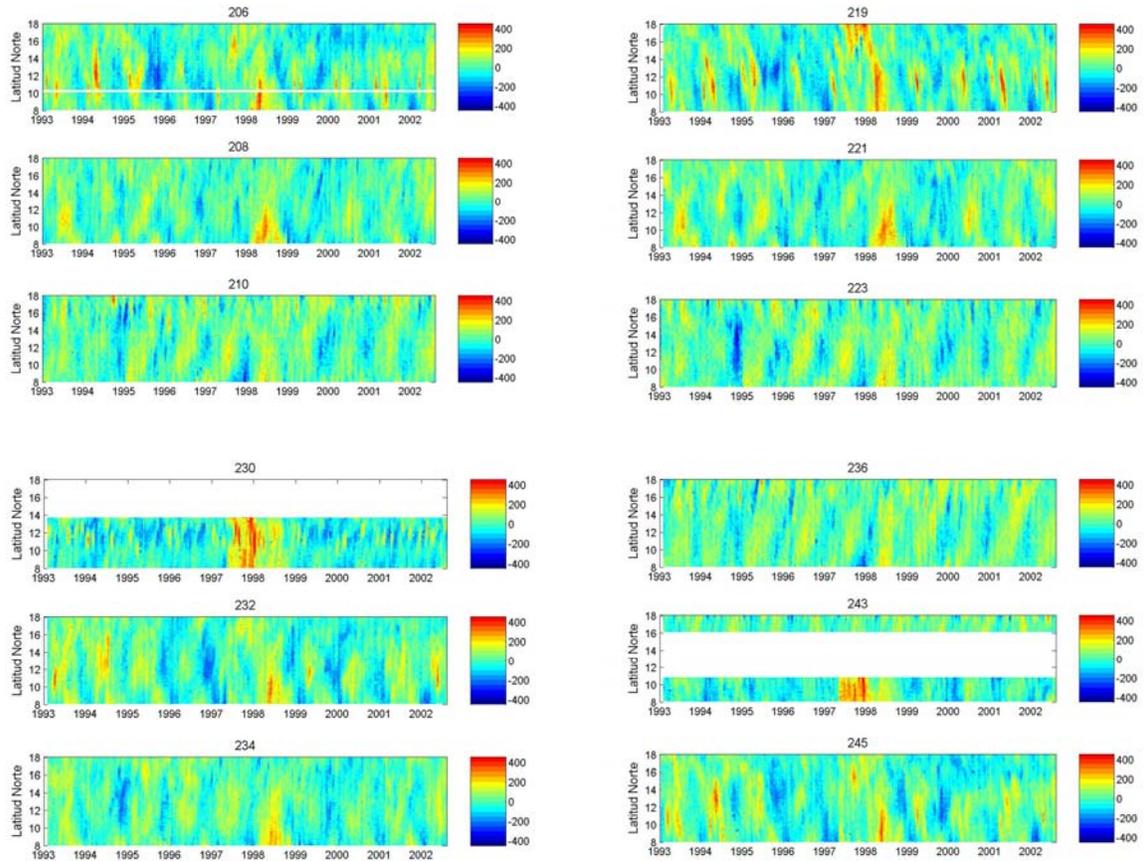


Figura 6.5 (Continuación)

Antes de aplicar el método de interpolación utilizando FEOCs a los datos de satélite, éste se probó en experimentos diseñados para observar su funcionamiento. En el capítulo 5, se presentó el experimento 1 en donde la altimetría es simulada con una perturbación generada mediante la función gaussiana en dos dimensiones y se usó la malla del satélite T/P limitada a la zona de estudio.

Cuando se busca la obtención de un muestreo más denso que el original utilizando FEOCs, se debe de tomar en cuenta el número de modos estadísticos en la construcción del muestreo deseado, recordando que la cantidad de modos está asociada a la variabilidad de los datos iniciales. Debido a esto, primero se emplearon todos los modos en la reconstrucción de la función gaussiana. La figura 6.6 presenta los resultados del experimento, donde se compara la función gaussiana original y las reconstrucciones obtenidas con el método de interpolación lineal y con el uso de FEOCs. Se observa que

ninguno de los métodos recupera la función original, sin embargo en algunas gráficas la reconstrucción con FEOCs presenta datos que la interpolación lineal no produce.

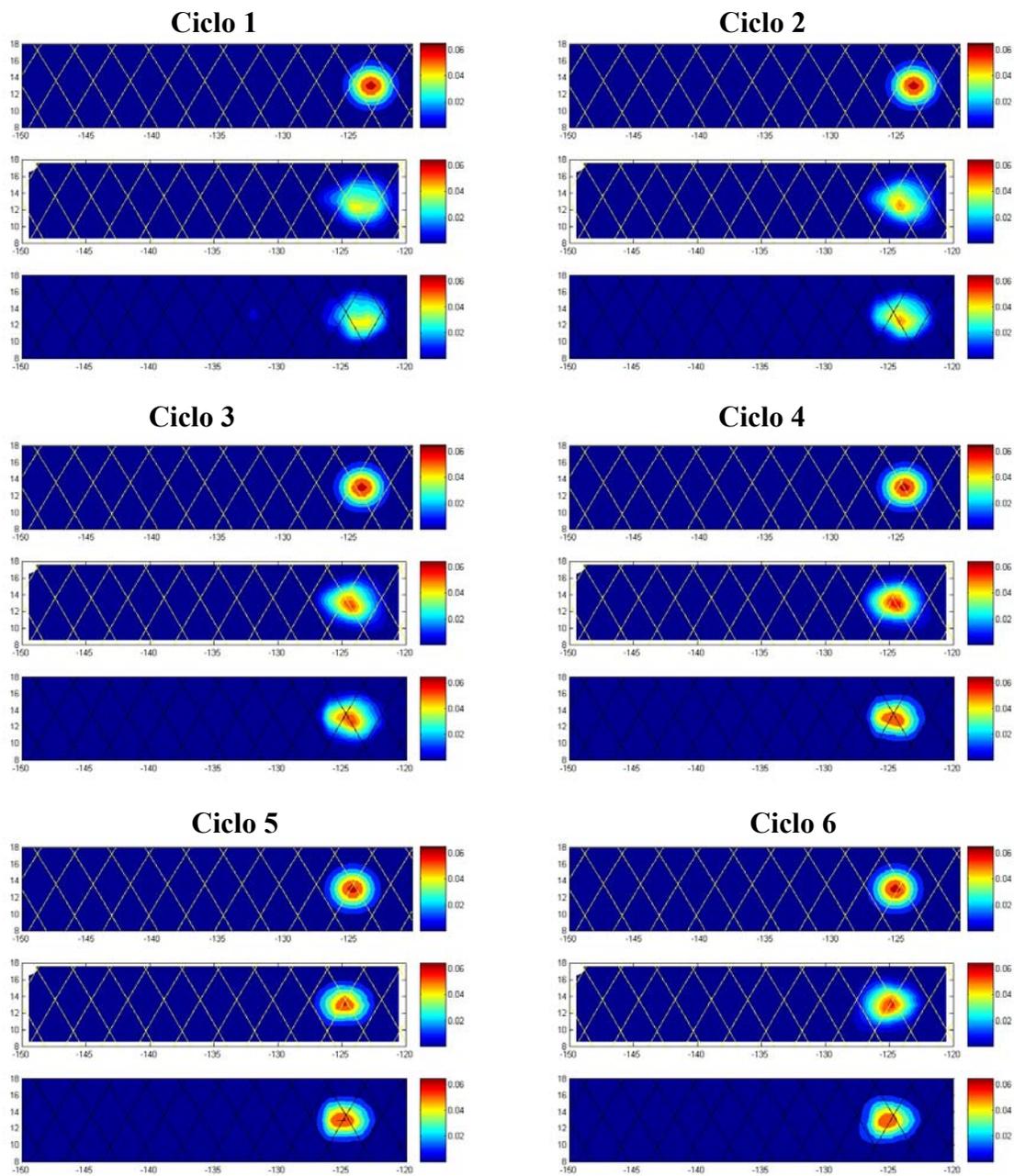


Figura 6.6 Simulación de datos de altimetría generados mediante una función gaussiana (gráfica superior) comparada con las reconstrucciones obtenidas mediante interpolación lineal (gráfica intermedia) y haciendo uso de las FEOCs (gráfica inferior), para los ciclos 1 a 14.

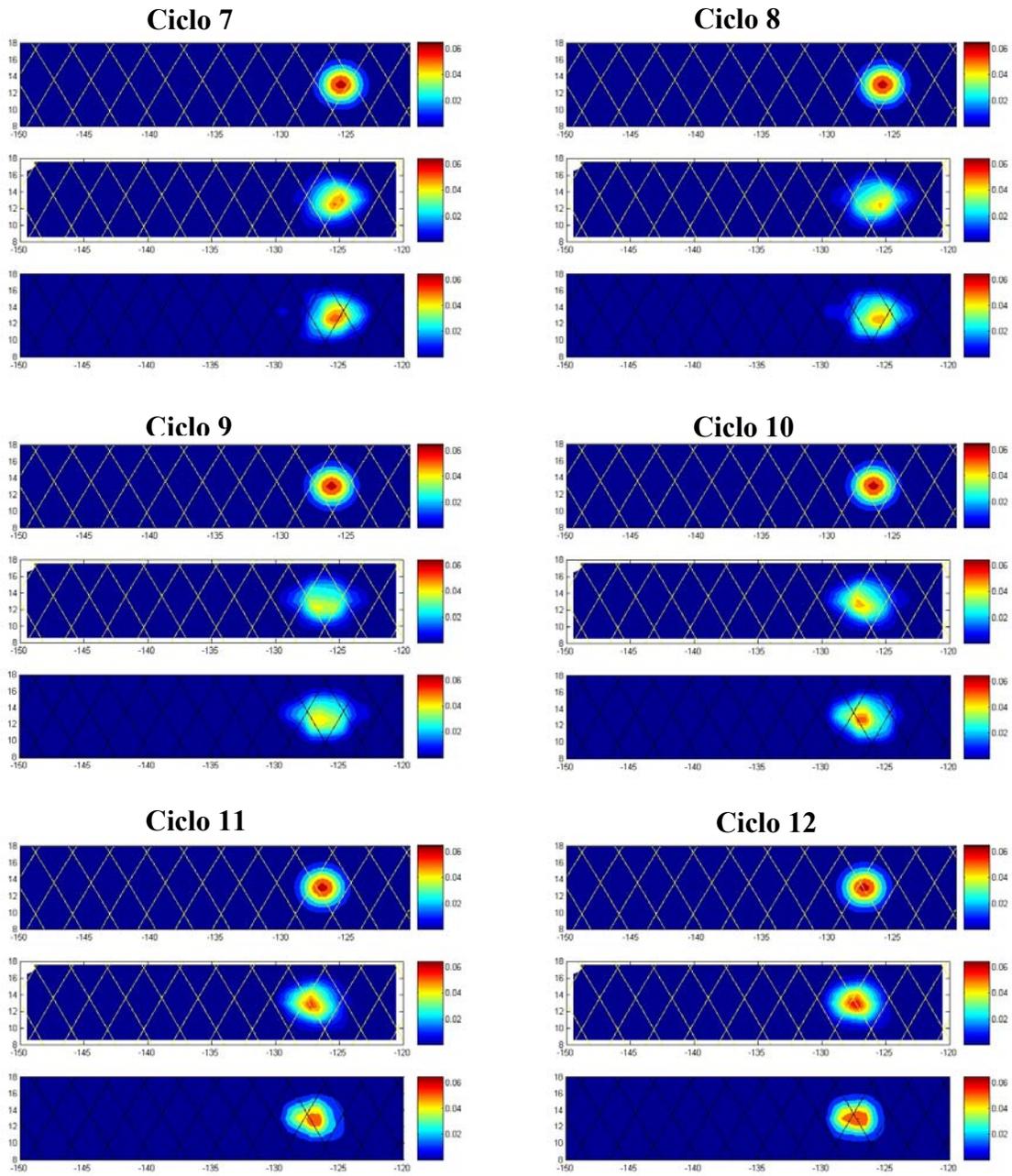


Figura 6.6 (Continuación).

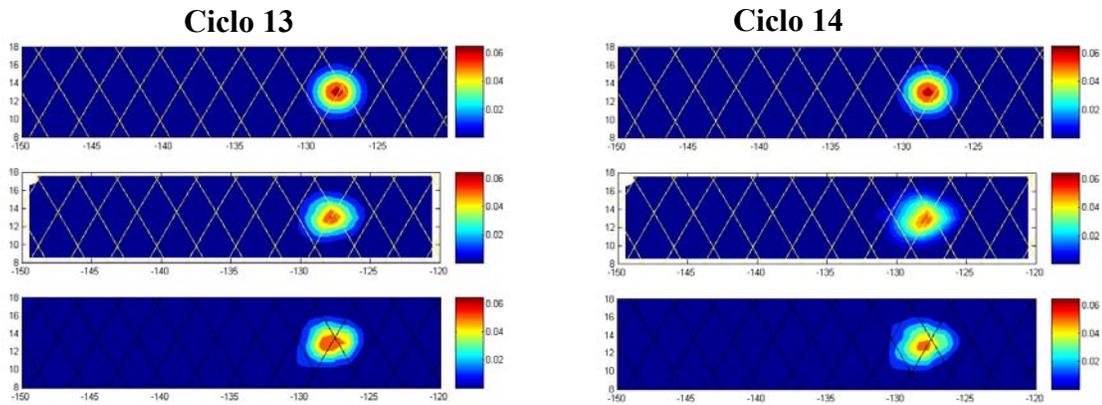


Figura 6.6 (Continuación).

Debido a que la interpolación con FEOCs genera matrices de entradas complejas y posteriormente interpola los parámetros de amplitud y fase, se examinaron la amplitud y la fase al momento de generar las primeras interpolaciones. La amplitud no presentó mayores problemas pero al examinar la fase se observaron saltos irregulares en donde los datos no deberían tenerlos, por lo que se optó en un principio por analizar la interpolación de la fase modo por modo. Los resultados fueron muy interesantes debido a que, conforme se empleaban más modos estadísticos en la reconstrucción, la fase presentaba más saltos, de manera que los primeros modos generaban información correcta pero después de cierto número la información era degenerada. Estos análisis se muestran en las figuras 6.6 (con todos los modos incluidos), 6.7 (incluyendo un sólo modo) y 6.8 (incluyendo 5 modos). En las figuras 6.7 y 6.8 el recuadro superior es una gráfica de fase en donde la dirección de las flechas indica el ángulo y el recuadro inferior es la función reconstruida, para los ciclos que se indican en la parte superior.

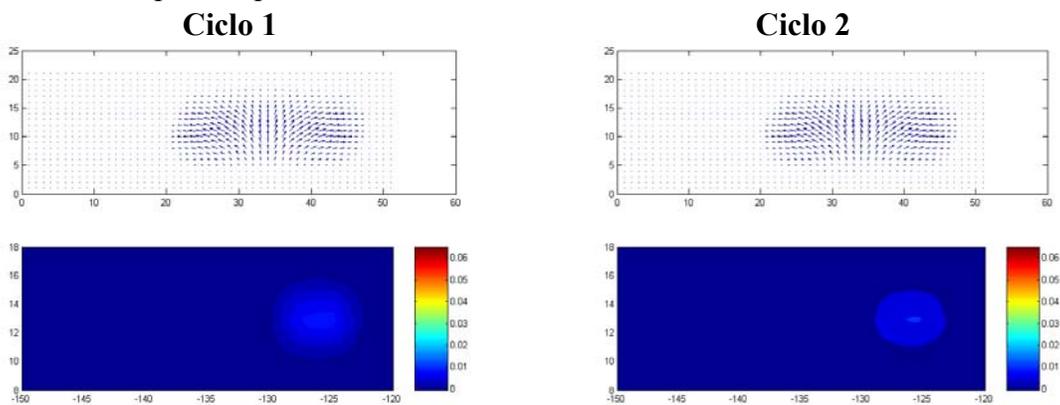


Figura 6.7 Gráficas de fase (recuadro superior) y reconstrucción de la función gaussiana en dos dimensiones (recuadro inferior) incluyendo un solo modo, para los ciclos indicados en la parte superior.

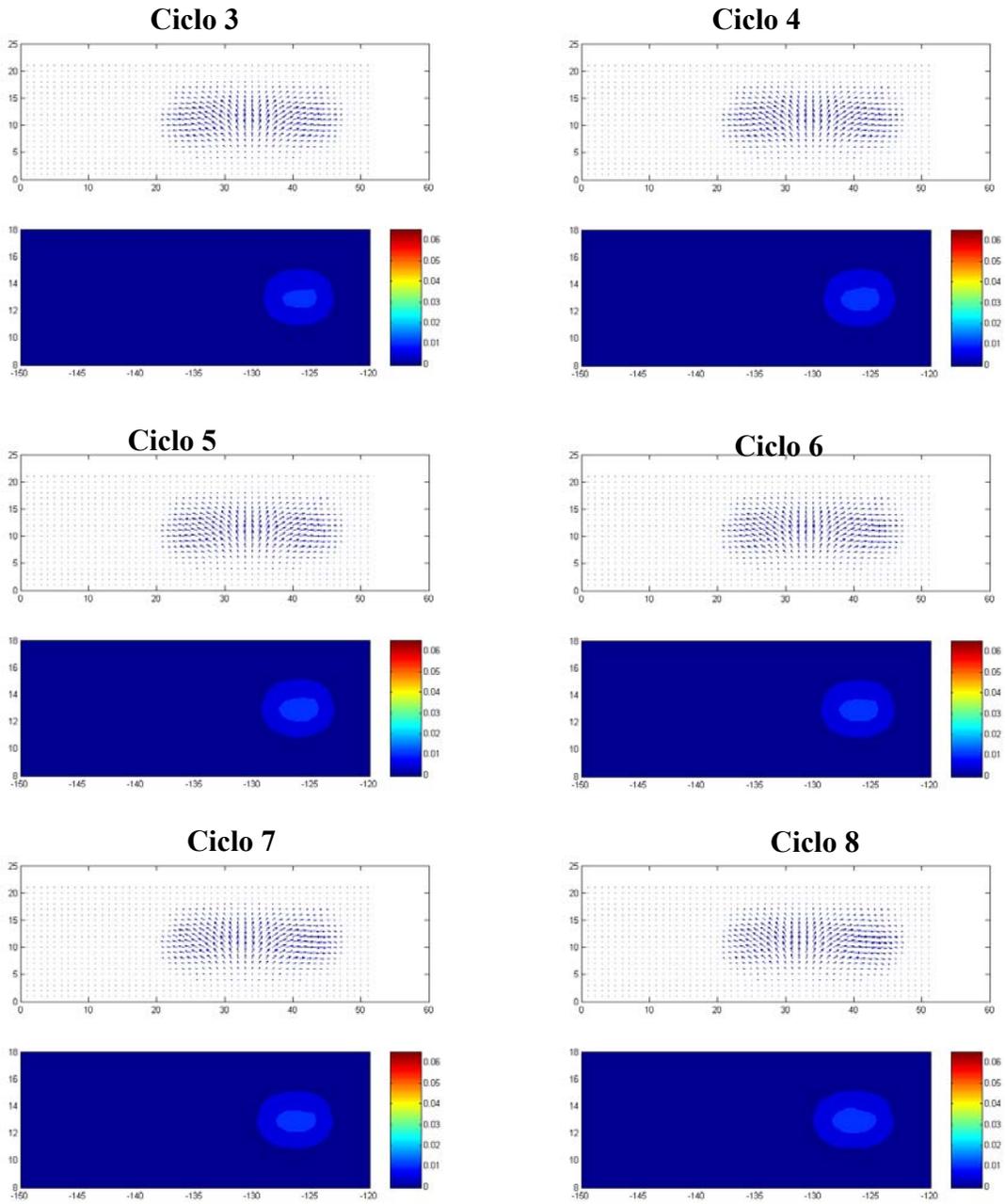


Figura 6.7 (Continuación).

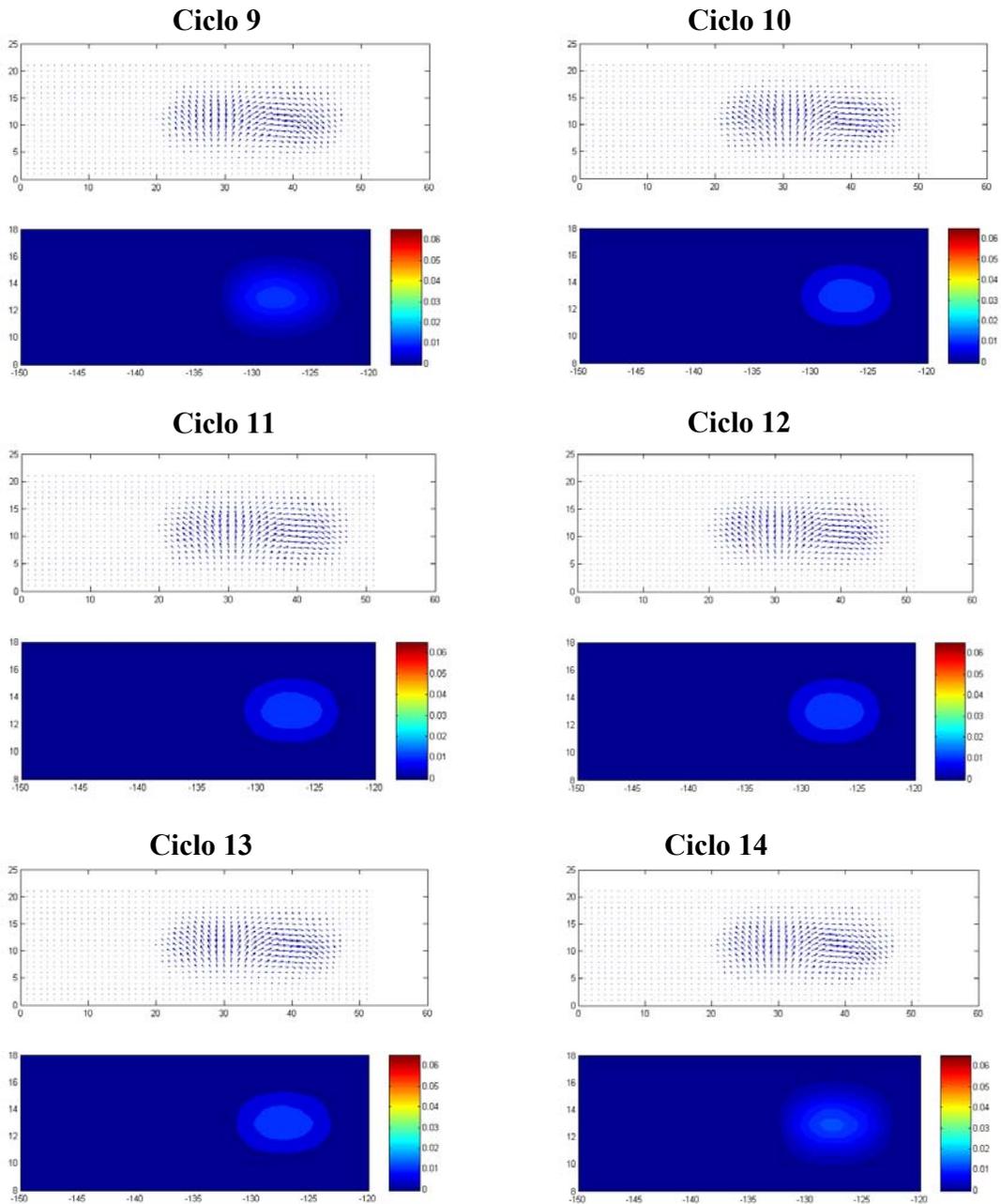


Figura 6.7 (Continuación).

La interpolación realizada con un solo modo (figura 6.7) recupera la forma circular de la función gaussiana a diferencia de la que utiliza todos los modos (figura 6.6). En los ciclos 1, 9 y 14 se pueden observar hasta 7 anillos de la función original, no con los valores adecuados pero sí con la forma debida, y se observa en las gráficas de la fase que hay continuidad en el cambio de dirección.

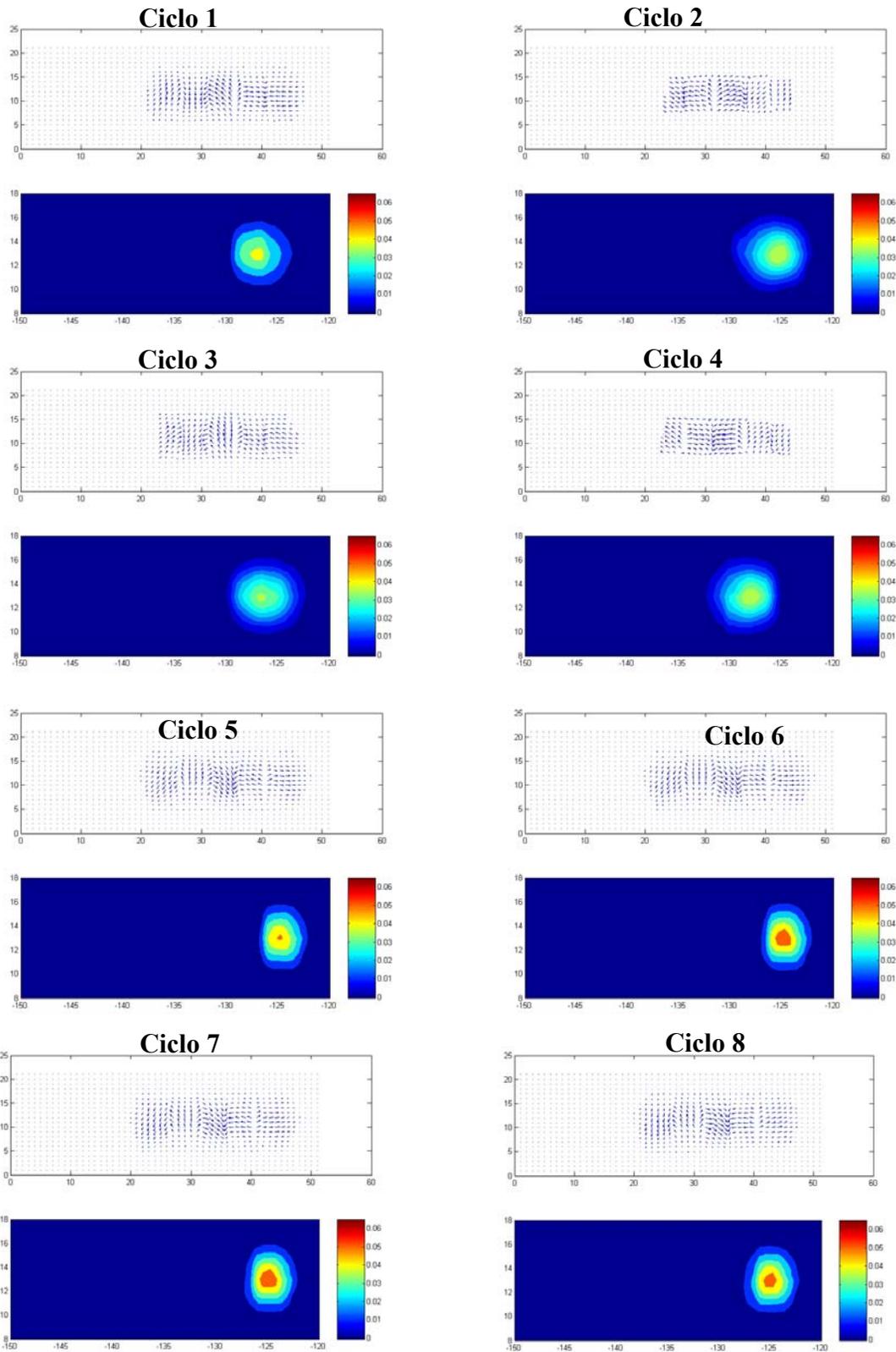


Figura 6.8 Similar a la figura 6.7 pero incluyendo 5 modos en la interpolación.

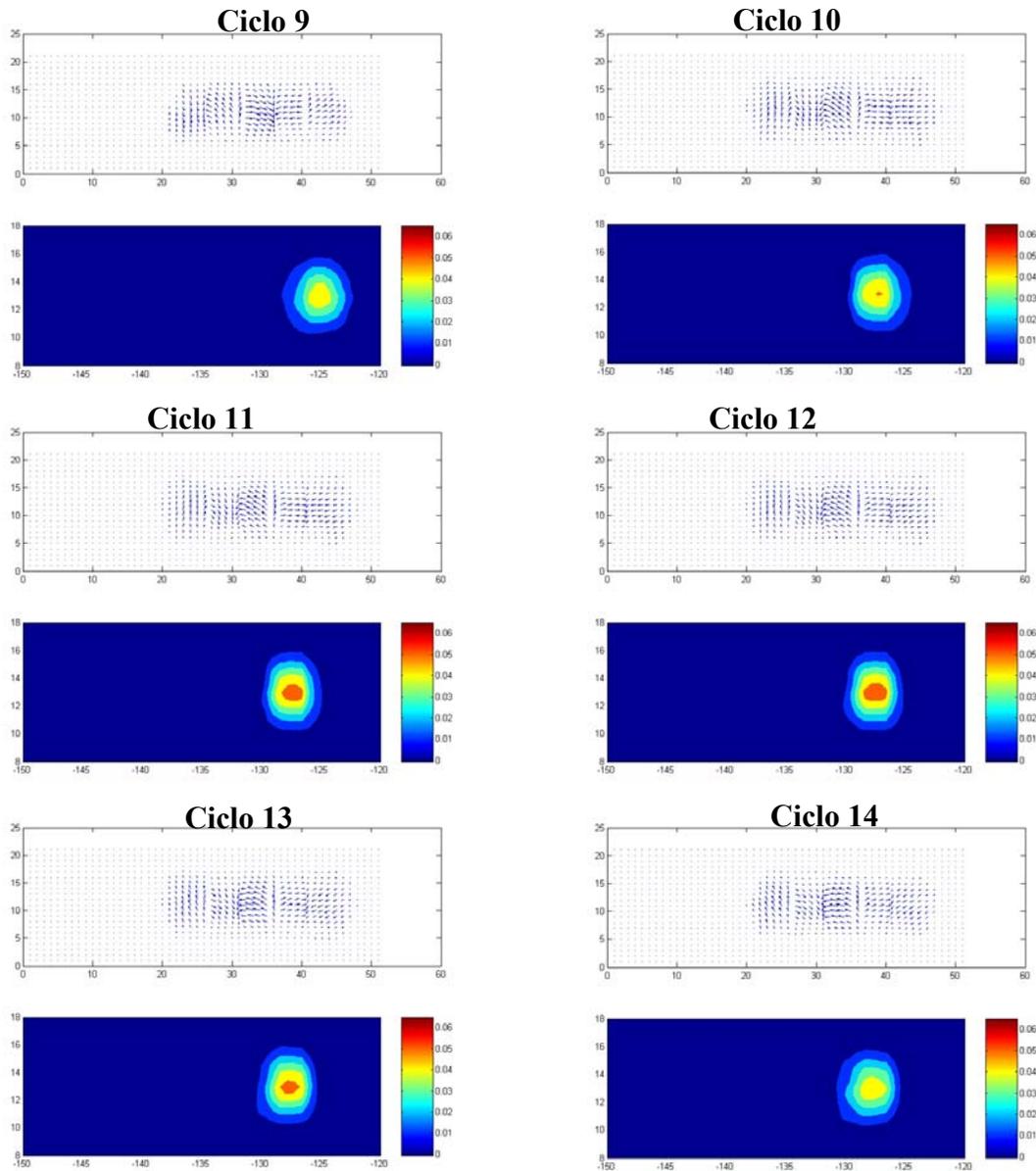


Figura 6.8 (Continuación).

Al utilizar 5 modos (figura 6.8) la fase comienza a tener discontinuidades, algunas de más de 90° . Con este experimento se concluyó que a partir de un cierto número de modos la interpolación no mejora sino que resulta más deficiente.

En el experimento 2 se empleó una función oscilatoria en dos dimensiones y otra malla de valores. Las imágenes producidas en el experimento son presentadas en la figura 6.9, donde el recuadro superior es la reconstrucción de la función con la interpolación lineal y el recuadro inferior es la función reconstruida con FEOCs. La figura 6.10 es una gráfica

de la fase ampliada en donde se observa que no presenta los saltos de las imágenes en la figura 6.8, y como resultado la interpolación con FEOCs generó una mejor reconstrucción de la función de estudio con respecto a la interpolación lineal. Implementando mejoras la interpolación de la fase de los datos, la continuidad en los incrementos de la fase pudo ser corregida parcialmente, es decir, esta corrección permite que la función de fase tenga un incremento constante en el ángulo y así minimizar el error al interpolar.

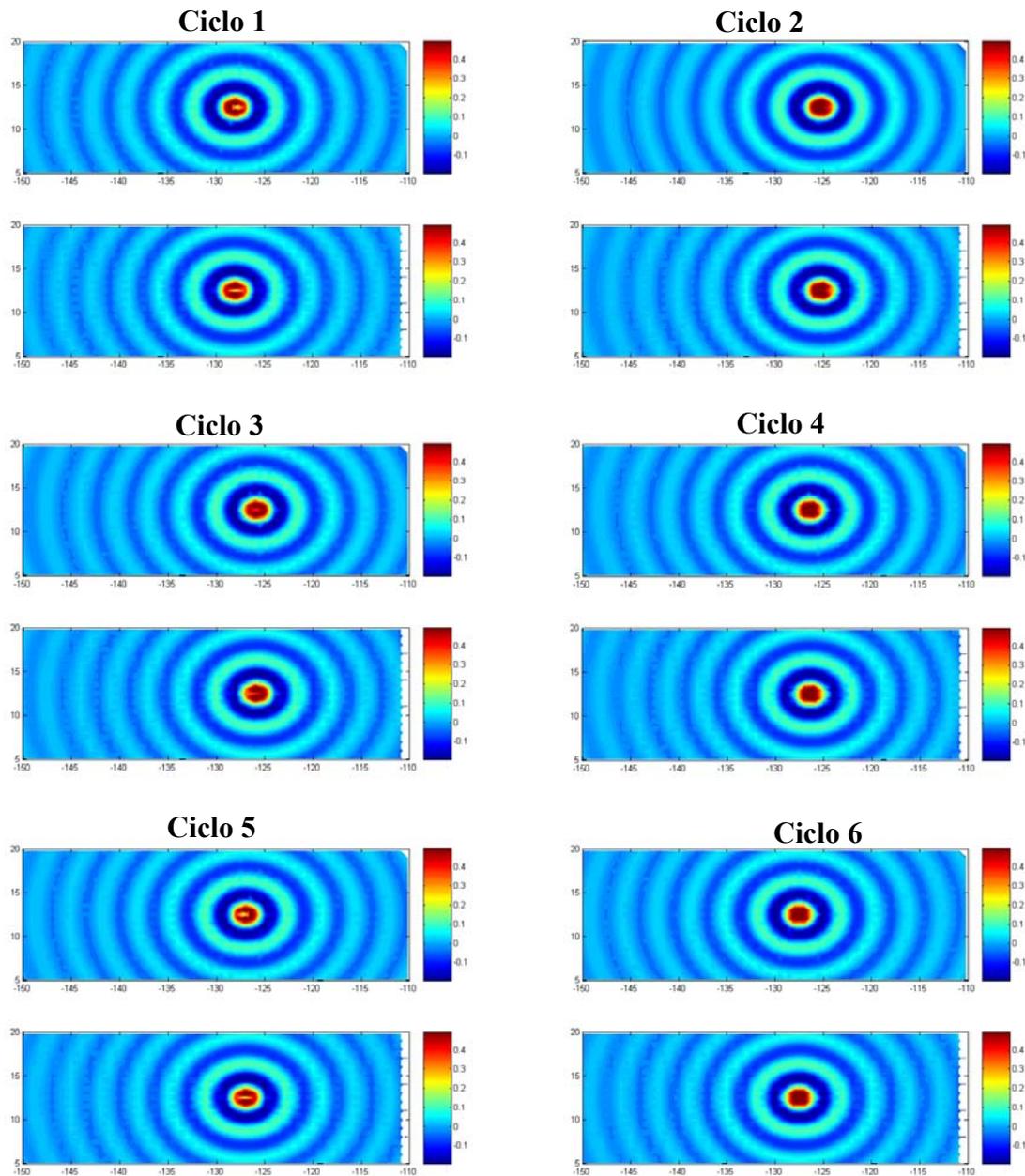


Figura 6.9 Reconstrucción de una función de rango oscilatorio. Interpolación lineal (recuadro superior) e interpolación con FEOCs (recuadro inferior).

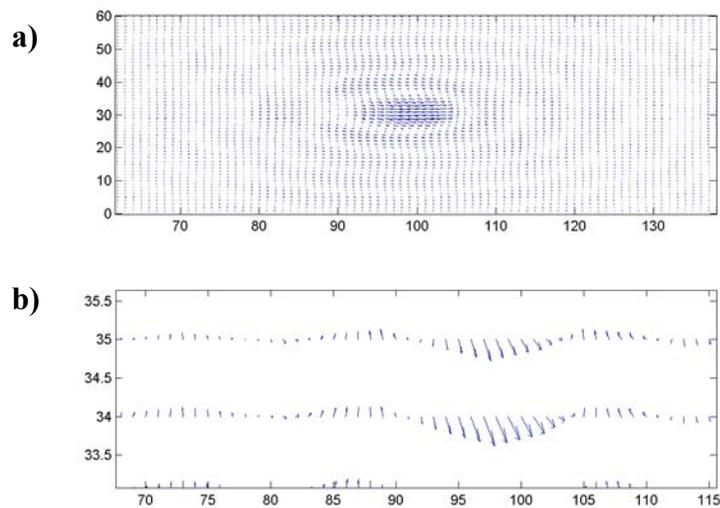


Figura 6.10 a) Función de fase corregida, y b) Acercamiento a la función de fase.

Con este experimento se concluye que al presentar continuidad en la horizontal o en la vertical en la malla de muestreo, es posible tener mejores representaciones de los datos con las FEOCs que con el método común de interpolación lineal.

Las trayectorias de los satélites no presentan la estructura anterior, pero al utilizar los datos de los dos satélites e interpolar posteriormente se busca generar una malla más densa en la región de estudio. Las imágenes generadas se presentan en la figura 6.11, donde el recuadro superior es la interpolación lineal y el recuadro inferior es la interpolación con FEOCs.

Debido a la estructura de las trayectorias de los satélites y a la deficiencia en las rutinas de cómputo utilizadas para la interpolación de la fase de los datos en mallas irregulares, la interpolación con FEOCs presenta un error cuadrático medio ligeramente más alto comparado con el de la interpolación lineal. Sin embargo, la interpolación lineal promedia valores para obtener sus resultados mientras que las FEOCs utilizan la variabilidad de los datos conservando la varianza en los datos interpolados y así recuperan la información de forma mucho más real que la interpolación lineal.

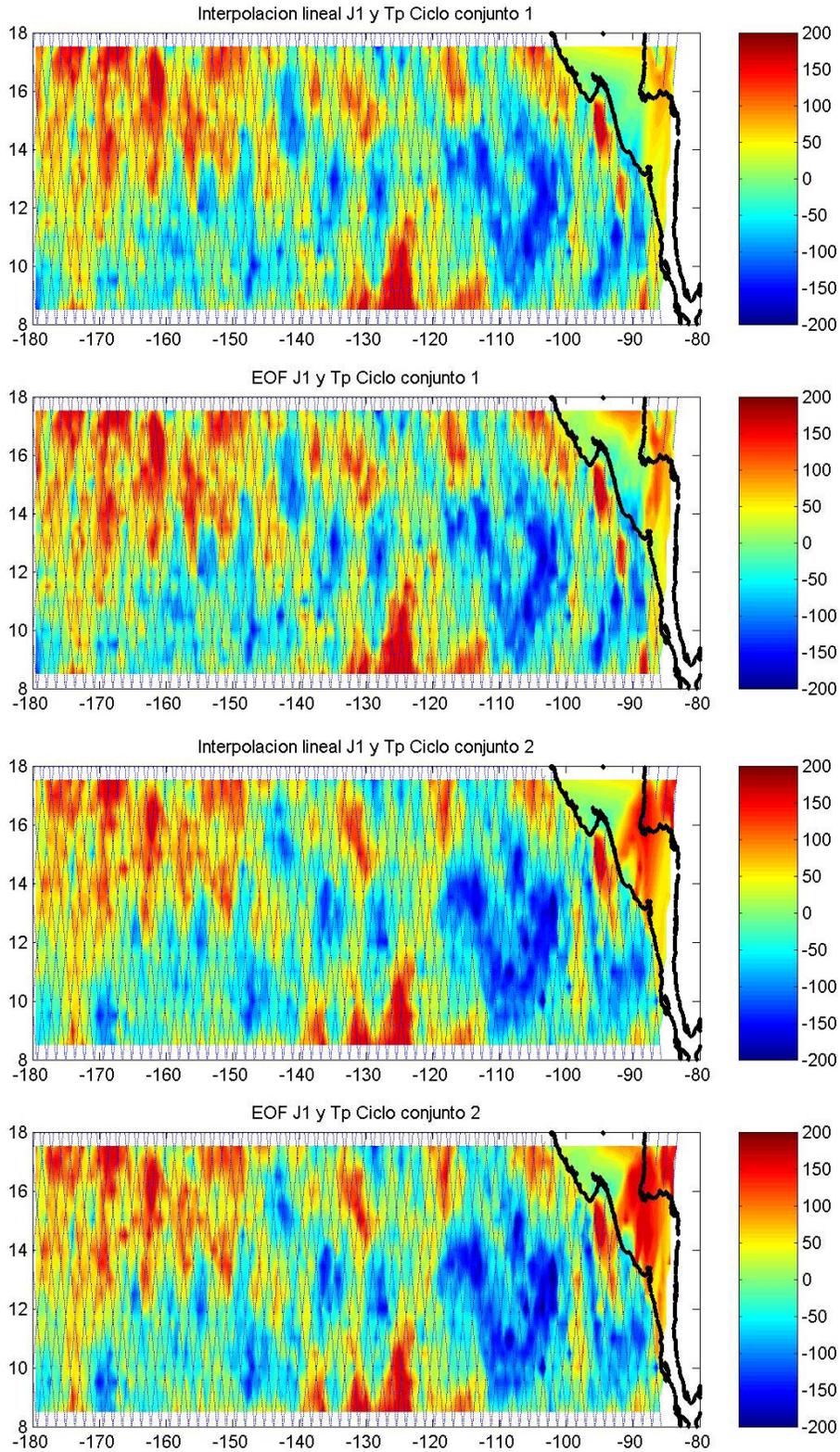


Figura 6.11 Reconstrucción del campo de altimetría utilizando datos de los satélites T/P y J-1. Utilizando el método de interpolación lineal (recuadro superior) y con FEOCs (recuadro inferior).

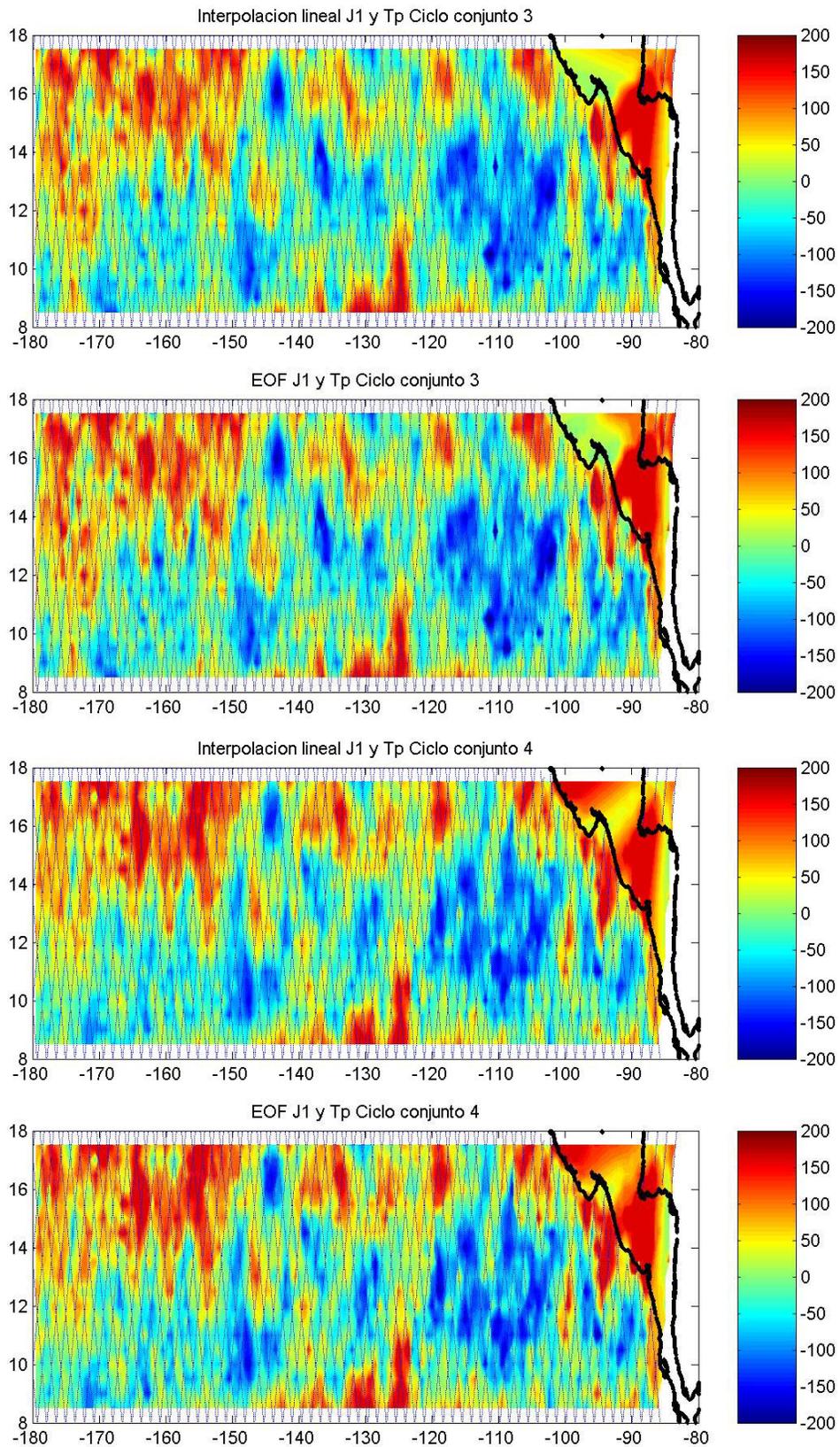


Figura 6.11 (Continuación).

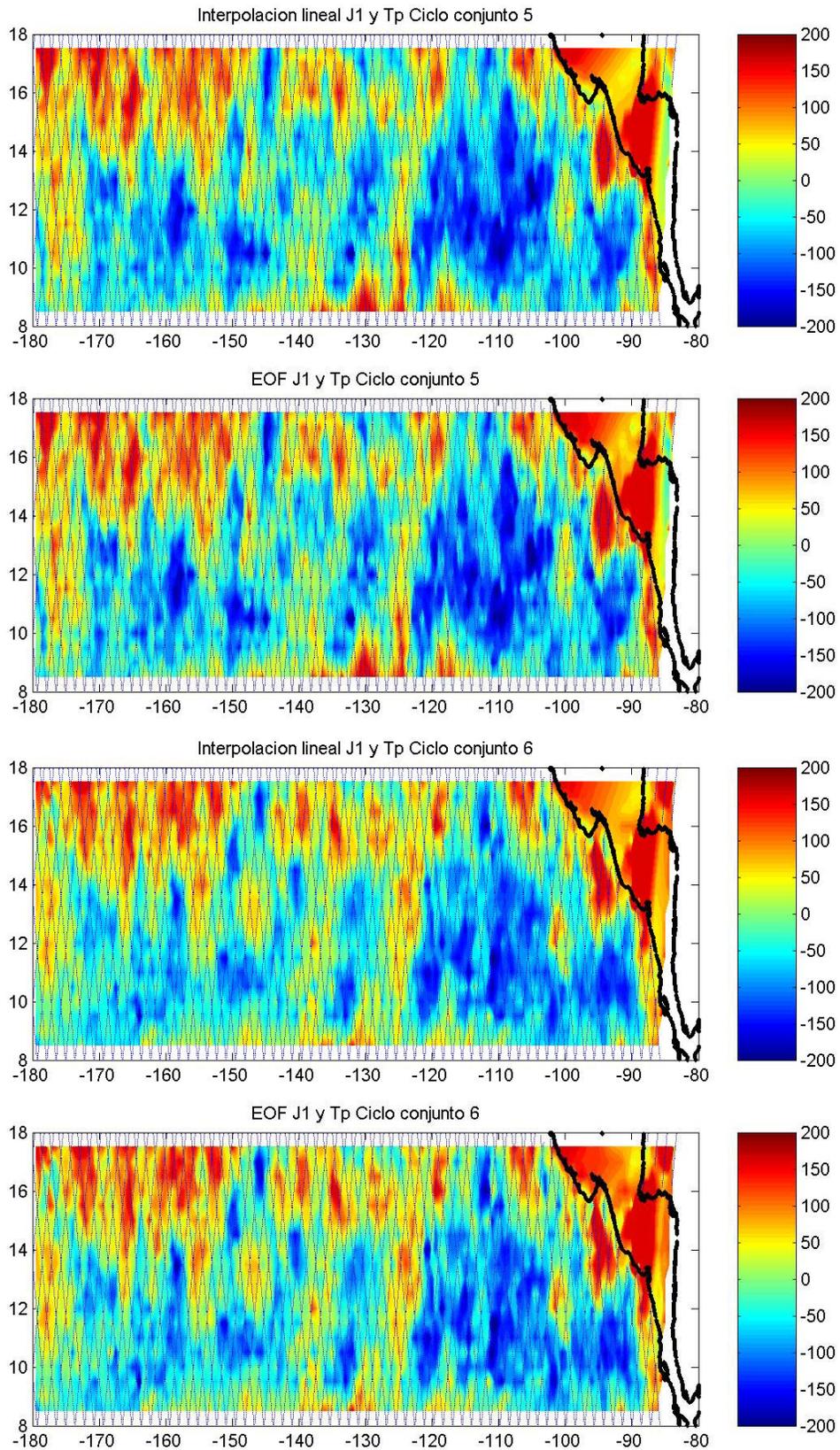


Figura 6.11 (Continuación).

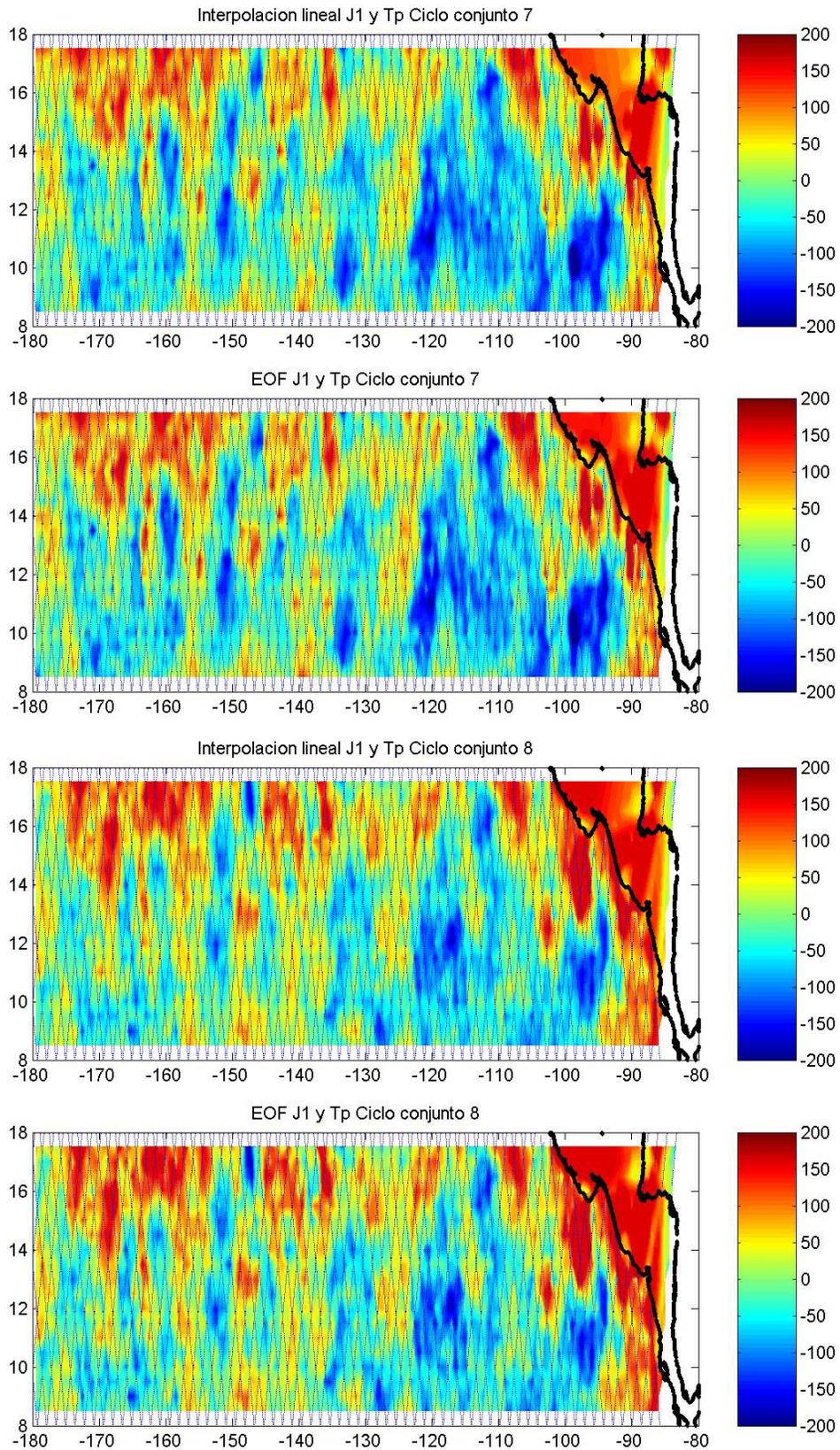


Figura 6.11 (Continuación).

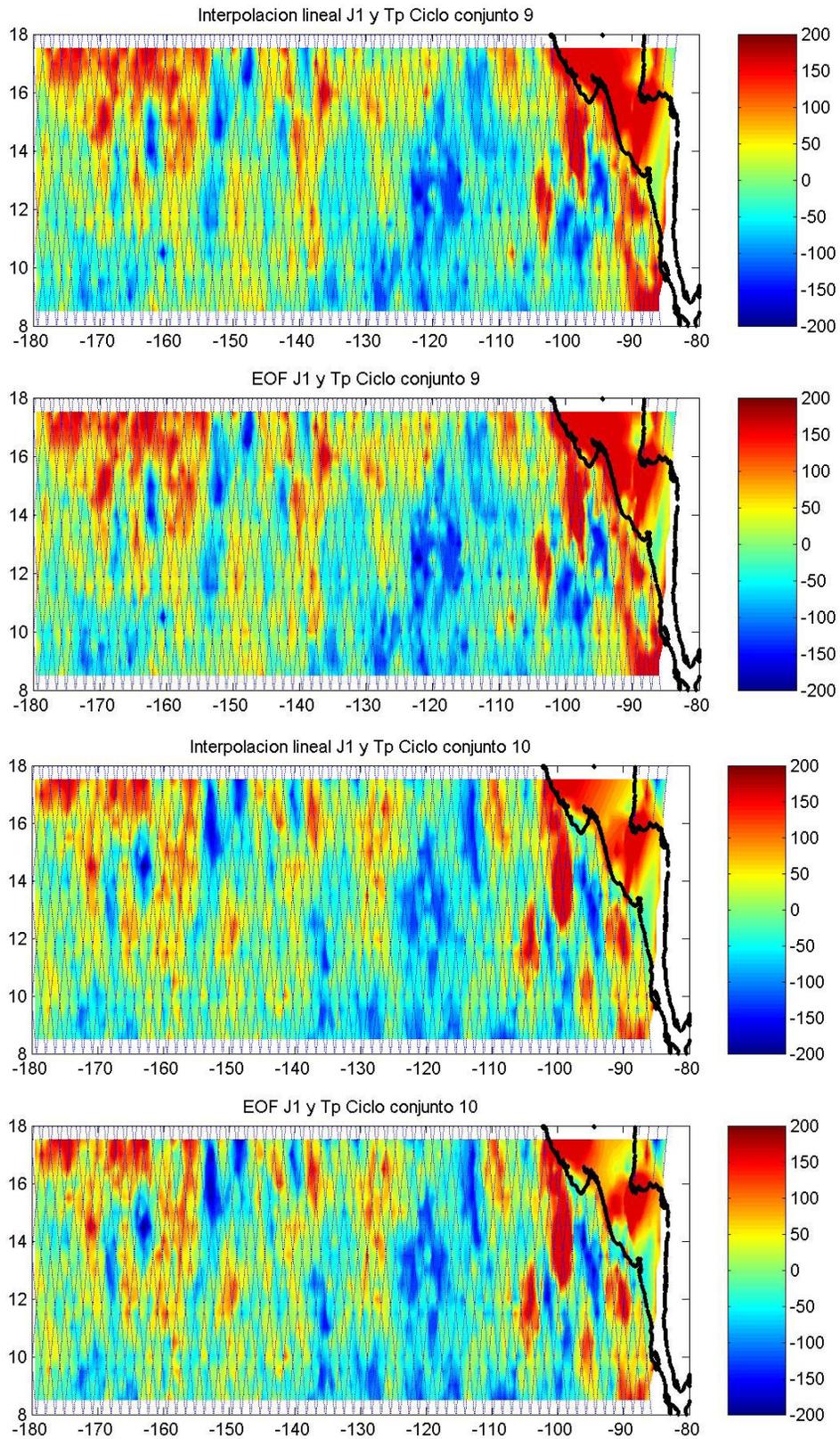


Figura 6.11 (Continuación).

Conclusiones

La interpolación espacial de datos haciendo uso de las funciones empíricas ortogonales complejas (FEOCs) es un proceso que utiliza la información de la fase y amplitud de una matriz compleja generada a partir de la transformada de Hilbert de los datos originales. El método interpola los dos parámetros por separado para después obtener una matriz corregida y al final recupera solamente la parte real de la matriz.

El método de interpolación mediante FEOCs requiere continuidad en la amplitud y fase en las posiciones interpoladas, pero en los datos de altimetría que se analizaron en este trabajo, la fase generada con los algoritmos utilizados presentaba discontinuidades entre posiciones cercanas. La deficiencia en la interpolación de fase de las rutinas numéricas obligó a implementar algoritmos de corrección, antes de llevar a cabo comparaciones entre distintas interpolaciones.

Las limitaciones en los algoritmos para la interpolación mediante FEOCs, hacen de éste un problema que puede ser abordado en trabajos futuros desde la perspectiva de una mejora computacional, así, la interpolación con FEOCs genera un camino para otras posibles aplicaciones.

La necesidad de tener un muestreo más denso de los datos de altimetría satelital es un problema que se ha resuelto parcialmente mediante la puesta en órbita de nuevos satélites. Sin embargo, aunque en un futuro se pueda muestrear todo el océano, la información histórica puede ser analizada y estudiada con métodos como el propuesto en este trabajo.

La interpolación espacial es sólo una aplicación de la interpolación con FEOCs y, como se vió en el experimento 5 de este trabajo, también puede ser utilizada en la generación de un muestreo con mayor frecuencia en tiempo. Asimismo, puede proponerse el estudio del método para la interpolación de datos en espacio y en tiempo conjuntamente.

El método de interpolación mediante FEOCs es bueno cuando la malla de muestreo presenta continuidad en los puntos tanto en la horizontal como en la vertical, pero si no se tiene esta estructura es deseable desarrollar algoritmos computacionales para lograr un mejor desempeño.

Referencias

AVISO/Altimetry, 1996: AVISO User Handbook for Merged TOPEX/POSEIDON products, AVI-NT-02-101, Edition 3.0.

Björck, Åke, 1996: Numerical methods for least squares problems. Editorial Siam, Philadelphia, Estados Unidos.

Case, K. et al., 2003: Topex/Poseidon and Jason-1: Surveyors of the oceans. NASA & CNES.

Casella, G., and R. Berger, 1990: Statistical Inference. Duxbury Press, California, E. U. A.

Emery, W. J., and R. E. Thomson, 1998: Data analysis methods in physical oceanography. California, E. U. A.

Goult, R. J., 1978: Applied linear algebra. Ellis Horwood publishers. England.

Hartmann, D. L., 2003: Matrix methods for analysis of structure in data sets. Notes. http://www.atmos.washington.edu/~dennis/552_Notes_4.pdf.

Hoffman, K., 1973: Álgebra Lineal. Prentice Hall Hispanoamericana. México.

Horel, J. D., 1984: Complex Principal Component Analysis: Theory and Examples. Journal of Climate and Applied Meteorology, Vol. 23, 1660-1673.

Hsu, H. P., 1973: Análisis de Fourier. Fondo Educativo Interamericano. México.