

La investigación reportada en esta tesis es parte de los programas de investigación del CICESE (Centro de Investigación Científica y de Educación Superior de Ensenada, B.C.).

La investigación fue financiada por el CONACYT (Consejo Nacional de Ciencia y Tecnología). Todo el material contenido en esta tesis está protegido por la Ley Federal del Derecho de Autor (LFDA) de lo 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 o titular de los Derechos Autor.

CICESE © 2022, Todos los Derechos Reservados, CICESE

Centro de Investigación Científica y de Educación Superior de Ensenada, Baja California



Maestría en Ciencias en Oceanografía Física

Asimilación de datos en un modelo regional del Pacífico nororiental mexicano

Tesis

para cubrir parcialmente los requisitos necesarios para obtener el grado de
Maestro en Ciencias

Presenta:

Eduardo Ashida Hernández

Ensenada, Baja California, México

2022

Tesis defendida por

Eduardo Ashida Hernández

y aprobada por el siguiente Comité

Dr. Alejandro Francisco Parés Sierra
Director de tesis

Dr. Julio Sheinbaum Pardo

Dr. Rubén Castro Valdez

Dr. Reginaldo Durazo Arvizu

Dra. Bertha Eugenia Lavaniegos Espejo



Dra. María Tereza Cavazos Pérez
Coordinadora del Posgrado en Oceanografía Física

Dr. Pedro Negrete Regagnon
Director de Estudios de Posgrado

Resumen de la tesis que presenta Eduardo Ashida Hernández como requisito parcial para la obtención del grado de Maestro en Ciencias en Oceanografía Física.

Asimilación de datos en un modelo regional del Pacífico nororiental mexicano

Resumen aprobado por:

Dr. Alejandro Francisco Parés Sierra
Director de tesis

Observaciones de temperatura superficial del mar (SST) y la altura del nivel del mar (SSH) del año 2003 se asimilan al modelo oceánico de circulación general CROCO (Coastal and Regional Ocean COMMunity Model) implementado en el Pacífico nororiental mexicano con el objetivo de estimar el impacto de la asimilación sobre las propiedades de la columna de agua (temperatura y salinidad) y sobre el nivel del mar. Los datos de SST asimilados provienen del producto Multi-scale Ultra-High Resolution (MUR) producido por JPL/NASA y los datos de SSH consisten en la suma de las anomalías del nivel del mar, proporcionadas por CMEMS/C3S, con una topografía dinámica media. Para validar y evaluar el desempeño del modelo se utilizan observaciones independientes de datos hidrográficos de temperatura y salinidad de las campañas del proyecto IMECOCAL. Se utiliza un esquema de relajación newtoniano de coeficiente constante como primera aproximación para asimilar los datos del satélite. Como esquema de asimilación se utiliza la versión de base fija del filtro SEEK (Singular Evolutive Extended Kalman filter). Para evaluar el rendimiento de las simulaciones de asimilación con respecto a una simulación libre (sin asimilación), se calcula la raíz del error cuadrático medio (RMS) de cada simulación con respecto a las observaciones asimiladas e independientes. En general, el modelo representa adecuadamente la estructura termohalina y las principales masas de agua presentes en el sistema de corrientes de California. La simulación de control presentó una RMS de 0.96°C y 6.94 cm en la SST y SSH, y 0.56°C y 0.12 psu en la temperatura vertical y salinidad, respectivamente. En la simulación implementada con el filtro SEEK, la RMS se redujo en (i) un 5.85 % en SST; (ii) 6.78 % en SSH; (iii) 16.80 % en temperatura vertical; (iv) 6.64 % en salinidad vertical. El efecto de esta mejora se concentró en los primeros 500 m de la columna de agua. Estos diagnósticos demuestran el impacto positivo de la asimilación de datos SST y SSH para estimar la propiedades termohalinas del océano.

Palabras clave: Asimilación de datos, filtro SEEK, Pacífico mexicano, nudging

Abstract of the thesis presented by Eduardo Ashida Hernández as a partial requirement to obtain the Master of Science degree in Physical Oceanography.

Data assimilation in a regional model of the Mexican Northeast Pacific

Abstract approved by:

Dr. Alejandro Francisco Parés Sierra

Thesis Director

Sea surface temperature (SST) and sea level height (SSH) from the year 2003 are assimilated into the oceanic model of general circulation CROCO (Coastal and Regional Ocean COMMunity Model) implemented in the northeastern Mexican Pacific with the objective to estimate the impact of assimilation on the properties of the water column (temperature and salinity) and on sea level. The assimilated SST data comes from the Multi-scale Ultra-high Resolution (MUR) product produced by JPL/NASA and the SSH data consists of the sum of sea level anomalies, provided by CMEMS/C3S, with a mean dynamic topography. In order to validate and evaluate the performance of the model, independent observations of temperature and salinity hydrographic data of the IMECOCAL project campaigns are used. A constant coefficient Newtonian relaxation scheme is used as a first approximation to assimilate the satellite data. As an assimilation scheme, the fixed base version of the SEEK filter (Singular Evolutive Extended Kalman filter) is used. In order to evaluate the performance of the assimilation simulations with respect to a free simulation (without assimilation), the root mean square (RMS) error of each simulation is calculated with respect to the assimilated and independent observations. In general, the model adequately represents the thermohaline structure and the main water masses present in the California current system. The control simulation presented an RMS of 0.96°C and 6.94 cm in the SST and SSH, and 0.56°C and 0.12 psu in the vertical temperature and salinity, respectively. In the simulation implemented with the SEEK filter, the RMS was reduced by (i) 5.85 % in SST; (ii) 6.78 % in SSH; (iii) 16.80 % in vertical temperature; (iv) 6.64 % in vertical salinity. The effect of this improvement was concentrated in the first 500 m of the water column. These diagnoses demonstrate the positive impact of assimilating SST and SSH data to estimate the termohaline ocean structure.

Keywords: Data assimilation, SEEK filter, Mexican Pacific, nudging

Dedicatoria

**A todas las personas que buscan
y comparten conocimiento.**

Agradecimientos

Al Centro de Investigación Científica y de Educación Superior de Ensenada, Baja California (CICESE) por darme la oportunidad de estudiar un posgrado.

Al Consejo Nacional de Ciencia y Tecnología (CONACyT) por brindarme el apoyo económico para realizar mis estudios de maestría. No. de becario: 994523

Al mi director Alejandro Parés por abrirme una puerta a nuevo conocimiento y por su gran sentido del humor, su comentarios y pláticas siempre fueron enriquecedoras y amenas de escuchar.

A los miembros de mi comité, Julio, Rubén, Reginaldo y Bertha por sus comentarios y aportes a este trabajo.

A los profesores del posgrado, gracias por el conocimiento que me han compartido.

A mi familia por apoyarme desde lejos, siempre los llevo cerca de mí.

A todos mis amigos y personas que quiero y que me acompañaron durante estos años. A Citlalli, Giovanni e Iván, por su amistad y compartir momentos buenos y divertidos este tiempo. A Albert, Mary Mar y Elliot, que desde lejos, sigo disfrutando de sus elocuentes y acertadas conversaciones y comentarios.

Tabla de contenido

	Página
Resumen en español	ii
Resumen en inglés	iii
Dedicatoria	iv
Agradecimientos	v
Lista de figuras	vii
Lista de tablas	x
Capítulo 1. Introducción	
1.1. Antecedentes	2
1.2. Hipótesis	4
1.3. Objetivos	4
1.3.1. Objetivo general	4
1.3.2. Objetivos específicos	4
Capítulo 2. Metodología	
2.1. Datos	5
2.1.1. Temperatura superficial (SST)	5
2.1.2. Altura del nivel del mar (SSH)	6
2.1.3. Datos hidrográficos	6
2.2. Diseño experimental	7
2.2.1. Implementación numérica	7
2.2.2. Esquema de relajación (Nudging)	9
2.2.3. Esquema de asimilación de datos (filtro SEEK)	10
2.3. Métodos de análisis	19
Capítulo 3. Resultados	
3.1. Simulación libre	21
3.2. Relajación Newtoniana	25
3.3. Asimilación de datos	30
3.4. Comparación entre experimentos	36
Capítulo 4. Discusión y conclusiones	
Literatura citada	44

Lista de figuras

Figura	Página
1. Estaciones de muestreo de las campañas de IMECOCAL de 2003. Los números indican la línea de muestreo.	7
2. Dominio espacial en la implementación numérica. Se muestra 1 de cada 5 puntos de la malla horizontal.	8
3. Diagrama del algoritmo recursivo del filtro de Kalman extendido.	15
4. Diagrama del algoritmo recursivo del filtro SEEK.	18
5. Varianzas explicadas por número de modos de SST (A) y SSH (B).	18
6. Promedios temporales de la simulación de control en la frontera norte para la componente v (panel izquierdo), temperatura (panel central) y salinidad (panel derecho). Las líneas blancas indican las isopicnas de 25.4 y 26.3 kg m ⁻³	22
7. Promedios temporales de la simulación de control en la frontera oeste para la componente v (panel superior), temperatura (panel central) y salinidad (panel inferior). Las líneas blancas indican las isopicnas de 25.4 y 26.3 kg m ⁻³	22
8. Promedios temporales de la simulación de control en la frontera sur para la componente v (panel izquierdo), temperatura (panel central) y salinidad (panel derecho). Las líneas blancas indican las isopicnas de 25.4 y 26.3 kg m ⁻³	23
9. Diagramas T-S de las campañas de IMECOCAL (A) y simulación libre (B). Diagramas de dispersión de temperatura (C) y salinidad (D) entre datos de la simulación libre e IMECOCAL. Los recuadros negros y siglas en lo diagramas T-S indican los límites y nombres de las masas de agua prevalentes en el SCC: ASA = Agua Subártica, ATS = Agua Tropical Superficial, ATr = Agua Transicional, AStS = Agua Subtropical Superficial, AESs = Agua Ecuatorial Subsuperficial, AIP = Agua Intermedia del Pacífico y APP = Agua Profunda del Pacífico. En los diagramas de dispersión se muestra el coeficiente de determinación r ² , RMS, y número de datos (N) utilizados para calcular los estadísticos. La línea negra y roja continua representa el ajuste perfecto y el de la regresión lineal respectivamente.	24
10. Evolución temporal de la RMS (panel superior) de SST de la simulación de control y de cada experimento con diferentes coeficientes de relajación (señalado como líneas continuas de color) comparadas contra observaciones de MUR de toda la región de estudio. Evolución temporal de la RMSD (panel inferior).	26
11. Evolución temporal de la RMS (panel superior) de SSH de la simulación de control y de cada experimento con diferentes coeficientes de relajación (señalado como líneas continuas de color) comparadas contra observaciones de toda la región de estudio. Evolución temporal de la RMSD (panel inferior)	27
12. Perfiles de Sesgo (A), RMS (B) y RMSD (C) de temperatura de cada simulación con distintos coeficientes de relajación respecto a datos de perfiles hidrográficos de IMECOCAL.	28
13. Perfiles de Sesgo (A), RMS (B) y RMSD (C) de salinidad de cada simulación con distintos coeficientes de relajación respecto a datos de perfiles hidrográficos de IMECOCAL.	28

Figura	Página
14. RMS de cada simulación con diferentes coeficientes de relajación (marcado como puntos negros) respecto a mediciones satelitales de SST (A), SSH (B), y mediciones hidrográficas de temperatura (C) y salinidad (D). Los valores de RMS de correspondientes a la simulación de control son marcados como cruces negras. Se muestra valores de γ de referencia como líneas punteadas rojas y se indica el intervalo adecuado de selección del coeficiente de relajación como un recuadro verde.	29
15. Modos dominantes del análisis de FEOs derivados de una simulación libre de 5 años (A, B y C) y de una serie larga de datos de SST del producto MUR (D, E, F). Se indica la Varianza Explicada (VE) de cada modo.	31
16. Modos dominantes del análisis de FEOs derivados de una simulación libre de 5 años (A, B y C) y de una serie larga de datos de SSH del altimetría (D, E, F). Se indica la Varianza Explicada (VE) de cada modo.	31
17. Evolución temporal de la RMS (panel superior) de SST de la simulación de control y de cada experimento del filtro SEEK con diferente número de modos (señalado como líneas continuas de color) comparadas contra observaciones de MUR de toda la región de estudio. Evolución temporal de la RMSD (panel inferior).	32
18. Evolución temporal de la RMS (panel superior) de SSH de la simulación de control y de cada experimento del filtro SEEK con diferente número de modos (señalado como líneas continuas de color) comparadas contra observaciones de toda la región de estudio. Evolución temporal de la RMSD (panel inferior).	33
19. Perfiles de Sesgo (A), RMS (B) y RMSD (C) de temperatura de cada experimento del filtro SEEK con diferente número de modos respecto a datos de perfiles hidrográficos de IMECOCAL.	34
20. Acercamiento de los primeros 1,000 m de la Figura 19. Perfiles de Sesgo (A), RMS (B) y RMSD (C) de temperatura de cada experimento del filtro SEEK con diferente número de modos respecto a datos de perfiles hidrográficos de IMECOCAL.	34
21. Perfiles de Sesgo (A), RMS (B) y RMSD (C) de salinidad de cada experimento del filtro SEEK con diferente número de modos respecto a datos de perfiles hidrográficos de IMECOCAL.	35
22. Acercamiento de los primeros 1,000 m de la Figura 21. Perfiles de Sesgo (A), RMS (B) y RMSD (C) de salinidad de cada experimento del filtro SEEK con diferente número de modos respecto a datos de perfiles hidrográficos de IMECOCAL.	35
23. RMS de cada experimento del filtro SEEK con diferente número de modos (marcado como puntos negros) respecto a mediciones satelitales de SST (A), SSH (B), y mediciones hidrográficas de temperatura (C) y salinidad (D). Los valores de RMS correspondientes a la simulación de control son marcados como cruces negras.	36

Figura	Página
24. Distribución de la RMS respecto a observaciones (SST y SSH) de las simulaciones de control (A, D), esquema de relajación (B, E) y asimilación (C, F). El valor total de la RMS para cada simulación es mostrado en la esquina inferior izquierda de cada mapa, además del aumento o decremento de la RMS total relativa a la simulación de control. .	37
25. Distribución del Sesgo respecto a observaciones (SST y SSH) de las simulaciones de control (A, D), esquema de relajación (B, E) y asimilación (C, F).	37
26. Evolución temporal de la RMS (panel superior) de SST de la simulación de control, esquema de relajación y asimilación (señalado como líneas continuas de color) comparadas contra observaciones de MUR de toda la región de estudio. Evolución temporal de la RMSD (panel inferior).	38
27. Evolución temporal de la RMS (panel superior) de SSH de la simulación de control, esquema de relajación y asimilación (señalado como líneas continuas de color) comparadas contra observaciones de toda la región de estudio. Evolución temporal de la RMSD (panel inferior).	39
28. Perfiles de Sesgo (A), RMS (B) y RMSD (C) de temperatura de la simulación de control, esquema de relajación y asimilación respecto a datos de perfiles hidrográficos de IMECOCAL.	40
29. Perfiles de Sesgo (A), RMS (B) y RMSD (C) de salinidad de la simulación de control, esquema de relajación y asimilación respecto a datos de perfiles hidrográficos de IMECOCAL.	40

Lista de tablas

Tabla	Página
1. Coeficientes de relajación utilizados (γ), incluyendo su valor numérico respecto a un Δt de 10 min ($\tilde{\gamma}$).	10
2. Número de modos usados para reconstruir la covarianza y sus respectivos porcentajes de varianza explicada de SST y SSH.	19

Capítulo 1. Introducción

Las observaciones satelitales han modificado el entendimiento de la circulación y el estado del océano y se han vuelto una herramienta indispensable para su estudio y monitoreo. Mediciones satelitales de temperatura superficial y anomalías del nivel del mar (SST y SSH, respectivamente por sus siglas en inglés) proporcionan información sinóptica y continua en tiempo real del estado del océano, además de proveer datos en regiones con mediciones in situ inexistentes. El uso de métodos avanzados de asimilación de datos (AD) que combinan mediciones hidrográficas o sinópticas superficiales como la SST y SSH en modelos de circulación oceánica constituye una forma de interpolar datos valiosos, pero a veces inconsistentes, en el tiempo y para extrapolar la información superficial al resto de la columna de agua. La AD se puede definir como el uso sistemático de datos para restringir un modelo matemático (Hofmann y Friederichs, 2001). La idea es combinar observaciones (estaciones meteorológicas, datos satelitales, derivadores) y los principios dinámicos subyacentes (modelos teóricos) que gobiernan el estado del océano para proporcionar una mejor estimación respecto a lo que podría obtenerse usando sólo las observaciones o el modelo.

Tradicionalmente, el problema de asimilación fue formulado como un problema de estimación de mínimos cuadrados. Desde la década de los 60, el llamado filtro de Kalman (FK) ha sido estudiado y ampliamente utilizado para estimar de forma estadística el estado de un sistema (Brasseur, 2006). Una de las razones para el uso del FK ha sido la simplicidad de su marco teórico. En resumen, el FK es un algoritmo eficiente que minimiza la distancia entre la trayectoria de un modelo dinámico (sujeto a ciertas suposiciones) y modelo observacional (Ghil y Malanotte-Rizzoli, 1991). Bajo la suposición de que todas las fuentes de errores son Gaussianas y que tanto los modelos de observación como los dinámicos son lineales, el FK es la solución exacta (óptima) para estimar recursivamente el estado de un sistema con su incertidumbre asociada (Carrassi et al., 2018). Esta aproximación es secuencial, lo que significa que las observaciones son asimiladas en orden cronológico. Existen varias modificaciones para el FK, en particular el filtro extendido de Kalman (FEK) se deriva para modelos no lineales, donde la AD procede por linealizaciones sucesivas en el tiempo de la dinámica no lineal (Ghil y Malanotte-Rizzoli, 1991). Un método prometedor y que proviene de la teoría del FK, es el filtro SEEK (singular evolutive extended Kalman)(Pham et al., 1998), que ha sido aplicado con éxito para asimilar observaciones (Carrassi et al., 2018).

El filtro SEEK es un método avanzado de asimilación diseñado para sistemas dinámicos de dimensiones muy grandes (Brasseur y Verron, 2006). En la práctica, el filtro SEEK consiste en no hacer correcciones durante la AD en aquellas direcciones para las cuales el error es el más atenuado por el sistema, de tal forma que reduce así el costo de implementación computacional y mejora la estabilidad del filtro (Pham et al., 1998). Existen métodos diseñados para estimar la trayectoria del modelo que se ajuste a todas las observaciones dentro de una ventana espacio-temporal. Esta aproximación es conocida como método variacional y tienen un notable costo computacional. Comparados con otros métodos, como los variacionales de 3D-VAR y 4D-VAR, los algoritmos de estimación estadística (como el FK, FEK, SEEK) requieren menos inversión inicial en términos de programación y son diseñados para incorporar mejoras graduales (Brasseur, 2006).

1.1. Antecedentes

El filtro SEEK ha sido aplicado en diferentes campos de dinámica oceánica, climática, biogeoquímica y ecosistémica. Testut et al. (2003) asimilaron simultáneamente datos satelitales de anomalía del nivel del mar y temperatura del mar en la cuenca del norte del Atlántico. El estudio demostró el impacto positivo de los datos satelitales para reconstruir la variabilidad de la circulación en el norte del océano Atlántico. Penduff et al. (2002) asimilaron datos de temperatura de la superficie del mar y datos de altura de la superficie del mar utilizando una versión similar del filtro SEEK en un modelo de ecuaciones primitivas en el sur del Atlántico. La circulación obtenida en este estudio mostró mejoras significativas en la estimación del estado medio de la circulación y la variabilidad de la mesoescala, en comparación con las simulaciones de modelos sin asimilación. Brusdal et al. (2003) realizaron una intercomparación entre el filtro SEEK y otros métodos secuenciales de asimilación (Ensemble Kalman Filter (EnKF), the Ensemble Kalman Smoother (EnKS)). En dicho estudio se asimilaron datos de anomalías del nivel del mar y temperatura del mar en una implementación del modelo MICOM en el norte del Atlántico. Se descubrió que los tres sistemas tienen rendimientos similares y pueden considerarse factibles para desarrollarse como prototipos operativos de predicción oceánica y experimentos en tiempo real.

Respecto a la región de estudio del PNM, la circulación oceánica está dominada por el sistema de la corriente de California (SCC), que a su vez es parte del gran giro anticiclónico del norte del Pacífico. La zona del PNM posee una gran importancia ecológica y económica. El SCC está caracterizado por una alta productividad asociada a surgencias (Zaytsev et al, 2003), lo que vuelve a las costas de Baja California (BC) un ambiente de gran biodiversidad y óptimo para aprovechar sus recursos. Por lo tanto, estudiar la dinámica de la zona es necesario para mejorar las estrategias de manejo de pesca, así como

otras aplicaciones climáticas.

La dinámica oceánica del SCC responde a la variabilidad de la intensidad y la posición del Centro de Alta Presión del Pacífico Norte y está dominada por meandros y remolinos a lo largo de la costa. El SCC tiene dos componentes principales de circulación, la Corriente Superficial de California o Corriente de California (CC) y la Contracorriente Costera de California (CCC). La CC viaja superficialmente (0 a 200 m) hacia el ecuador, acarrea agua subártica de baja temperatura (10°C a 20°C), baja salinidad (32.5 a 33.8) y alta concentración de oxígeno (Durazo, 2015). De manera opuesta, la CCC viaja subsuperficialmente pegada a la costa (100 a 300 m) hacia el norte, llevando agua de alta temperatura y salinidad y baja concentración de oxígeno. La CC abarca en promedio ~300 km al oeste de la costa y es más intensa y estrecha durante primavera e inicios de verano mientras que la CCC, es más intensa en los meses de junio-julio y diciembre-enero (Gómez-Valdivia, 2017). Las estaciones de primavera y verano están caracterizadas por la generación de meandros y remolinos cerca de la costa, mientras que invierno es la estación donde los remolinos se desprenden de la península y viajan hacia el oeste (Soto et al., 2004). Utilizando mediciones *in situ* se ha logrado observar estructuras persistentes ciclónicas en la circulación superficial durante verano y otoño (Durazo, 2015), y estructuras persistentes anticiclónicas durante invierno (Soto et al., 2004), ambas frente a Punta Eugenia. Estudios donde se analiza la circulación del PNM al implementar modelos numéricos también encuentran estructuras persistentes. Sin embargo, la posición y tiempo de algunos eventos en estudios numéricos pueden presentar un desfase temporal o espacial. Una forma de mejorar el posible desfase de los modelos numéricos es utilizando AD, donde básicamente se ajustan los eventos de circulación con mediciones *in situ* de donde se presenta el evento.

Las investigaciones realizadas sobre la aplicación de AD para el SCC se han enfocado en la región del Oeste de Estados Unidos (EU) (Broquet et al., 2009; Chao et al., 2017). En dichos estudios han empleado el sistema de asimilación del modelo de circulación general ROMS (Regional Ocean Model System) utilizando métodos variacionales (4D-VAR y 3D-VAR) en las líneas de estaciones de CalCOFI (California Cooperative Oceanic Fisheries Investigations) pertenecientes a la zona de EU, y logran reproducir con éxito la variabilidad prevista para esta región. Por otro lado, no se han implementado modelos con AD en la región del SCC en el PNM.

El propósito de este trabajo es estudiar e implementar un esquema de asimilación de datos basado en la teoría del filtro de Kalman en un modelo de circulación general en la región del Pacífico Nororiental mexicano. Dicho método es el filtro SEEK, donde se asimilan datos de SST y SSH en el modelo de circulación general CROCO para reconstruir la variabilidad del Sistema de Corriente de California (SCC) para el año 2003. Para poder tener un punto de referencia entre la simulación libre y el esquema de

asimilación, se utiliza un esquema de relajación hacia las mismas observaciones. Las bases de datos asimiladas, la configuración del modelo, el método y derivación de los esquemas de relajación y asimilación, así como los métodos de análisis y diseño experimental, son descritos en el Capítulo 2. El impacto estadístico y físico en las variables observadas y no observadas de los esquemas de relajación, asimilación y su comparación es presentado en el Capítulo 3. Finalmente, en el Capítulo 4 se discute y dan las conclusiones a este trabajo.

1.2. Hipótesis

La implementación del esquema de asimilación de datos del filtro SEEK en el modelo CROCO proporcionará una mejor estimación del estado del océano en la región del SCC respecto a lo que podría obtenerse usando sólo las observaciones o la simulación libre del modelo.

1.3. Objetivos

1.3.1. Objetivo general

Asimilar mediciones satelitales de SST y SSH en un modelo de circulación oceánica implementado con el filtro SEEK y estimar el impacto de la asimilación en las propiedades de la columna de agua (temperatura y salinidad) y en el nivel del mar.

1.3.2. Objetivos específicos

- Implementar una configuración en el modelo de circulación general CROCO que se asemeje adecuadamente a la circulación del SCC.
- Implementar un esquema de relajación newtoniana (nudging) y el filtro SEEK como esquema de asimilación de datos en una configuración del CROCO en el SCC.
- Comparar y describir los cambios que ocurren en la circulación y propiedades de la columna de agua respecto a cada una de las implementaciones.

Capítulo 2. Metodología

En este capítulo se describen las bases de datos usadas para asimilar en el modelo. Seguido se presenta el esquema experimental, donde se muestra la configuración del modelo CROCO utilizada y la derivación de los esquemas de relajación y asimilación. Finalmente se presentan los métodos de análisis utilizados.

2.1. Datos

El periodo de estudio escogido para este trabajo es el año 2003, donde se tienen disponibles observaciones satelitales continuas de SST y SSH, además de tener un año bien muestreado de datos independientes de lances hidrográficos de las campañas de IMECOCAL.

2.1.1. Temperatura superficial (SST)

La base de datos de SST consiste en la versión 4.1 del producto Multi-scale Ultra-high Resolution (MUR) producido por Jet Propulsion Laboratory (JPL) y distribuido por el Physical Oceanography Distributed Active Archive Center (PODAAC; <https://podaac.jpl.nasa.gov/dataset/MUR-JPL-L4-GLOB-v4.1>). MUR es un análisis compuesto nivel 4 de Group for High Resolution Sea Surface Temperature (GHRSSST; <https://www.ghrsst.org/>) basado en observaciones de noche (skin y subskin) de GHRSSST L2P derivados de diferentes instrumentos. Utiliza wavelets como función base en una interpolación óptima para generar un producto con resolución temporal diaria, desde 2002 hasta la fecha, y con una resolución espacial (longitud-latitud) de 0.01 grados (~ 1 km) (Chin et al., 2017). Para conseguir esa resolución, se combinan datos de sensores de microondas e infrarrojos, incluyendo el Advanced Microwave Scanning Radiometer-EOS (AMSR-E) de NASA, el Moderate Resolution Imaging Spectroradiometer (MODIS) en las plataformas de Aqua y Terra de NASA, el US Navy microwave WindSat radiometer, el Advanced Very High Resolution Radiometer (AVHRR) en varios satélites de la National Oceanic and Atmospheric Administration (NOAA), así como observaciones *in situ* del proyecto iQuam de la NOAA. Los datos resultantes son un estimado de la "temperatura de fundación", es decir, la temperatura cercana a la superficie ($\sim 1 - 5m$) por debajo del grado de fluctuación diurna debida al calentamiento solar (Nencioli, 2018). La ventaja de este producto sobre otros es que provee de observaciones diarias y espacialmente completas (sin ausencia de datos por presencia de nubes) que pueden ser asimiladas secuencialmente.

2.1.2. Altura del nivel del mar (SSH)

La altimetría satelital ha permitido obtener campos continuos en el espacio y tiempo del nivel del mar. Este trabajo tiene como objetivo asimilar altura del nivel del mar (SSH) derivada de anomalías de nivel del mar (SLA por sus siglas en inglés) ajustadas en el modelo. Para comparar observaciones de SLA de altimetría en el nivel del mar de un modelo, se requiere de una topografía dinámica promedio (MDT por sus siglas en inglés). Los datos de SLA observados se agregan a la MDT antes de compararlos con el nivel del mar del modelo. La elección de la MDT que se agrega a los datos de SLA es importante, ya que el nivel del mar promedio del modelo tiende a diferir de la MDT generada por observaciones de altimetría (debido a diversos factores, entre ellos errores del modelo). Por lo tanto, la MDT considerada en este trabajo es calculada del promedio del nivel del mar de una simulación libre de 10 años del modelo CROCO (implementado con la configuración descrita en la sección 2.2.1). La base de datos de SLA consiste en mallas con resolución temporal diaria y espacial de $0.25^\circ \times 0.25^\circ$ proporcionadas por el Copernicus Marine Environment Monitoring Service (CMEMS; <https://marine.copernicus.eu/>) y Copernicus Climate Change Service (C3S; <https://cds.climate.copernicus.eu/>). Con esto, la base de datos de SSH para asimilar consiste de mapas de SLA (interpolados a la malla del modelo) agregados a una MDT calculada de una simulación libre.

2.1.3. Datos hidrográficos

Las simulaciones del modelo pueden ser validadas con observaciones independientes cuando estas estén disponibles. En este trabajo solo se asimilan observaciones satelitales superficiales, se utilizan datos de temperatura y salinidad de lances hidrográficos de la campaña de IMECOCAL para validar los resultados de la simulación libre y de la asimilación. Esta comparación también permite ver el efecto de la proyección de la información asimilada en la superficie al resto de la columna de agua.

El proyecto de IMECOCAL (Investigaciones Mexicanas de la Corriente de California) es un programa dedicado a monitorear el comportamiento del SCC a lo largo de una red de estaciones en el área mexicana frente a la Península de Baja California. Desde 1997, el proyecto cubre esta red de estaciones hasta cuatro veces al año durante los meses de enero, abril, julio y octubre donde se obtienen muestras de agua para la determinación de nutrientes y oxígeno disuelto, y mediciones de biomasa y estructura de la comunidad de zooplancton, temperatura, salinidad, fluorescencia superficiales y velocidad de las corrientes. En este trabajo, se utilizan los datos de temperatura y salinidad de las campañas de enero, abril, julio y octubre del año 2003 en la red de estaciones mostrada en la Figura 1. Este año se seleccionó por tener una red de estaciones con gran cobertura y ser muestreadas de forma uniforme en las cuatro estaciones del año.

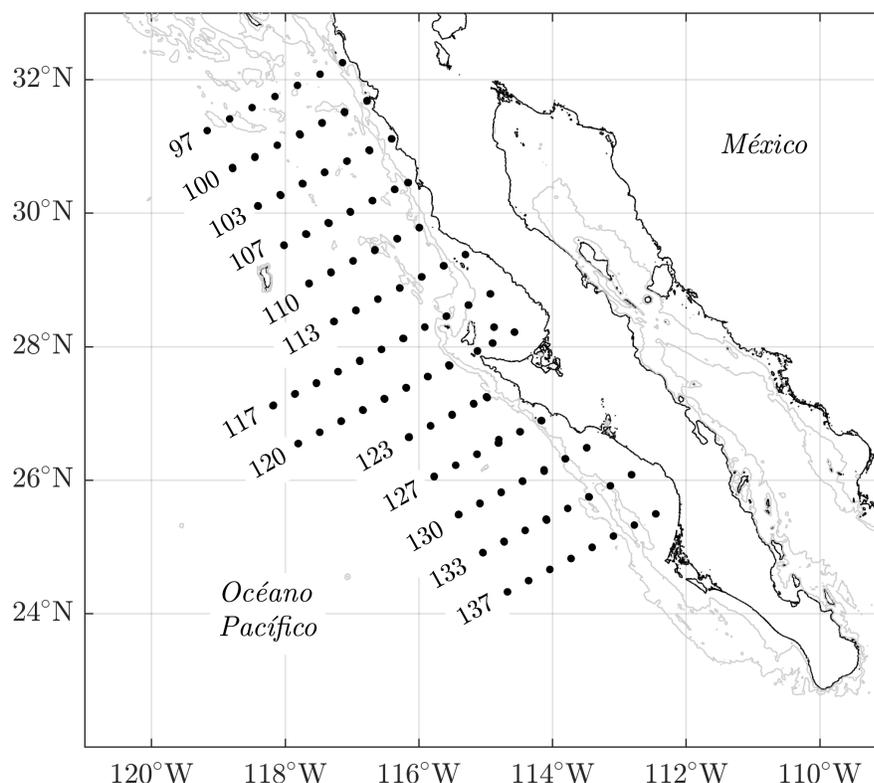


Figura 1. Estaciones de muestreo de las campañas de IMECOCAL de 2003. Los números indican la línea de muestreo.

2.2. Diseño experimental

El diseño experimental consta de tres series de experimentos numéricos del SCC en el año 2003 que parten de una simulación libre de 10 años. El primero, corresponde a una simulación libre solo con forzamientos y condiciones de frontera, llamada desde ahora simulación control, el segundo y tercer experimento consisten en las simulaciones con relajación y asimilación respectivamente. La configuración del modelo es descrita en la sección 2.2.1, y la derivación de los esquemas de relajación y asimilación es descrita en las secciones 2.2.2 y 2.2.3.

2.2.1. Implementación numérica

Los experimentos numéricos en este trabajo han sido realizados utilizando el modelo oceánico de circulación general CROCO (Coastal and Regional Ocean Community; Debreu et al., 2012). CROCO es un modelo desarrollado en torno al modelo ROMS_AGRIF (Shchepetkin y McWilliams, 2005), que resuelve las ecuaciones primitivas discretizadas en coordenadas curvilíneas en la horizontal, de seguimiento al terreno en la vertical, de superficie libre con pasos de tiempo explícitos, bajo la aproximación de Boussinesq y dinámica hidrostática (en este trabajo).

El modelo avanza con un paso de tiempo de 10 minutos y su dominio es mostrado en la Figura 2. Consiste en una malla curvilínea construida a lo largo de la península de Baja California con una resolución horizontal de 100×200 (~ 8.5) km y 40 niveles en la vertical discretizados acorde al sistema de coordenadas sigma. La topografía fue construida utilizando la base de datos de ETOPO2 (Earth Topography) del National Geophysical Data Center (NGDC; <https://www.ngdc.noaa.gov/mgg/global/etopo2.html>). El modelo fue forzado en la superficie con vientos interanuales derivados del reanálisis atmosférico de ERA-Interim con una resolución espacial de $0.25^\circ \times 0.25^\circ$ y temporal de 6 hrs, producido por el European Centre for Medium-Range Weather Forecasts (ECMWF). Las condiciones iniciales y de frontera fueron derivadas del reanálisis GLORYS (Global Ocean reanalysis and Simulation), con resolución espacial de $0.25^\circ \times 0.25^\circ$ y temporal de 1 día. La malla del modelo fue construida de tal forma que tuviera fronteras abiertas al norte, oeste y sur, y una cerrada al este con la península. Se utilizó esta configuración para correr una simulación de 10 años, y partir de ahí para los experimentos propuestos en este trabajo.

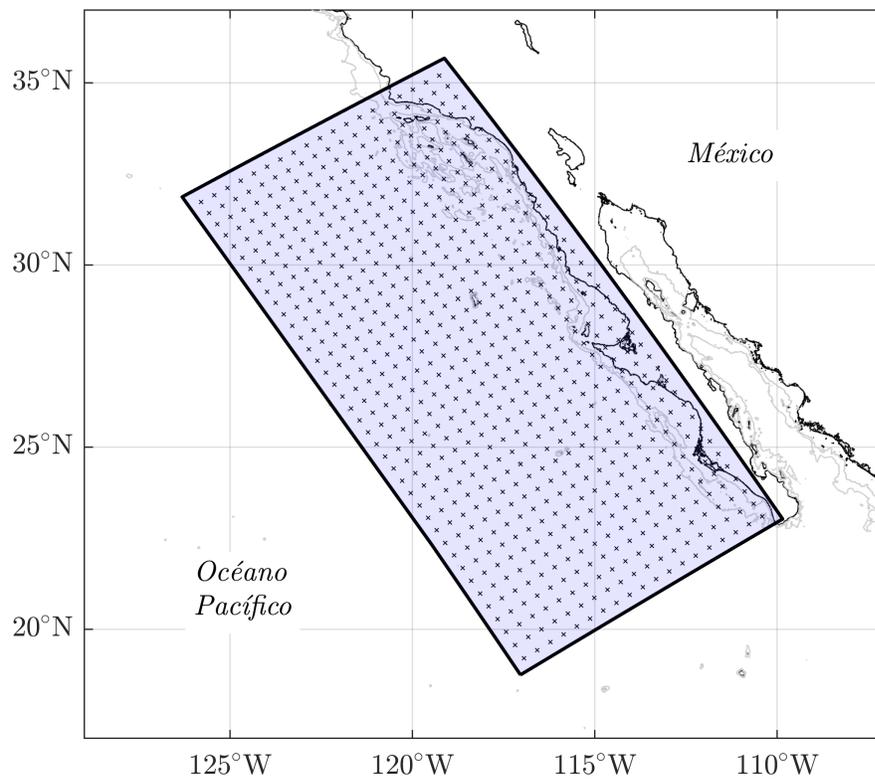


Figura 2. Dominio espacial en la implementación numérica. Se muestra 1 de cada 5 puntos de la malla horizontal.

2.2.2. Esquema de relajación (Nudging)

Una manera simple de integrar datos a un sistema físico es mediante la corrección del análisis del modelo con las observaciones que haya disponibles. Sea x_i el vector de estado del sistema, el cual contiene todas las variables en todo el dominio para un tiempo i dado, y M_i un operador que responde a la dinámica del modelo, la evolución del sistema está dada por:

$$x_{i+1}^f = M_i x_i^a \quad (1)$$

donde los superíndices a y f indican el análisis del modelo y la estimación a tiempo futuro respectivamente. La ecuación anterior indica de forma directa que para obtener el estado futuro, se requiere aplicar la dinámica del modelo a un previo análisis. Para corregir esta predicción se realiza lo siguiente:

$$x_i^a = x_i^f + \tilde{\gamma} (x_i^o - x_i^f) \quad (2)$$

donde el superíndice o indica las observaciones disponibles, y $\tilde{\gamma}$ es un peso llamado coeficiente de relajación. La ecuación anterior muestra que para obtener el análisis (corrección), se suma un término, llamado innovación, multiplicado por un coeficiente de relajación. Este esquema es llamado relajación Newtoniana o nudging en inglés. Consiste en corregir el análisis cada paso del tiempo para forzar al sistema a que se asemeje a las observaciones. La rapidez con la cual el sistema se asemeja a las observaciones depende del coeficiente de relajación, donde:

- Si $\tilde{\gamma} = 1$, entonces $x_i^a = x_i^o$ i.e., el análisis es igual a la observación
- Si $\tilde{\gamma} = 0$, entonces $x_i^a = x_i^f$ i.e., el análisis es igual a la predicción

Por lo tanto, la elección del coeficiente de relajación modifica en gran medida el análisis que se realice.

Usualmente, los modelos oceánicos contienen términos de relajación en su implementación numérica con el propósito de que las estimaciones no se alejen de climatologías conocidas. En este trabajo, se utiliza este esquema para acercar las estimaciones de SST y SSH del modelo a observaciones conocidas. Realizar esta corrección a la trayectoria del modelo durante la simulación prácticamente no requiere añadir un

costo computacional adicional, por lo que es una herramienta útil para tener una primera aproximación de un sistema de asimilación.

En la implementación numérica del modelo, $\tilde{\gamma}$ es calculado como:

$$\tilde{\gamma} = W \frac{\Delta t}{\gamma} \quad (3)$$

donde W es un operador que indica el espacio donde será aplicada la relajación, y Δt es el paso de tiempo (en esta implementación 10 min). El operador W generalmente ocasiona que la corrección se realice en todo el dominio, o solo cercano a las fronteras. En particular para este trabajo, se definió W para que solo se relaje hacia las observaciones superficiales. Sin embargo, aún se requiere seleccionar un γ en la ecuación anterior. Cuando se trata al coeficiente de relajación de forma constante en el espacio y tiempo, como se hace en este esquema, la elección de γ es algo arbitraria. Esto es debido a que el término de innovación de la ecuación 2 no responde a ningún aspecto físico del sistema, y es más bien un forzante numérico añadido. En este trabajo se estudió el efecto que tienen distintos coeficientes de relajación, con el propósito de definir un intervalo de valores que permitan al modelo corregirse hacia observaciones sin descartar la dinámica de la predicción. Los coeficientes usados son mostrados en la siguiente tabla:

Tabla 1. Coeficientes de relajación utilizados (γ), incluyendo su valor numérico respecto a un Δt de 10 min ($\tilde{\gamma}$).

γ	360 días	180 días	30 días	10 días	5 días	2 días	1 días	1000 min	12 hrs	200 min	100 min	20 min
$\tilde{\gamma}$	2×10^{-5}	4×10^{-5}	2×10^{-4}	7×10^{-4}	0.001	0.003	0.007	0.01	0.014	0.05	0.1	0.5

2.2.3. Esquema de asimilación de datos (filtro SEEK)

Las observaciones satelitales son asimiladas secuencialmente en un ciclo de corrección-predicción cada que hay observaciones disponibles para ajustar la trayectoria del modelo. El vector de estado en este trabajo esta compuesto de los campos tridimensionales de la velocidad horizontal, temperatura y del campo bidimensional de la superficie libre.

Para poder comprender el esquema de asimilación, en esta sección, primero se resume la derivación del filtro de Kalman, para después presentar su versión extendida, y finalmente el filtro SEEK.

- Filtro de Kalman

Los esquemas de asimilación secuencial buscan estimar el estado verdadero de un sistema físico en cada tiempo $t = i$ utilizando todas las observaciones disponibles en ese instante. Asumiendo que se quiere conocer el estado, el sistema físico se describe como:

$$x_{i+1}^t = M_i x_i^t + \epsilon_i^m \quad (4)$$

donde x_i^t es el vector de estado verdadero que contiene todas las variables del modelo en todo el dominio en el tiempo i , M_i es el operador del modelo, o matriz de transición de estado del tiempo i al tiempo $i + 1$, corresponde a las ecuaciones y dinámica del modelo, y ϵ_i^m es el error asociado al modelo originado de diferentes parametrizaciones, discretizaciones numéricas, y otros procesos.

Las observaciones disponibles son modeladas como:

$$x_i^o = H_i x_i^t + \epsilon_i^o \quad (5)$$

donde x_i^o es el vector observacional que incluye todas las variables observadas en el tiempo i , y ϵ_i^o es su error asociado. H_i es el operador observacional, que convierte el vector de estado al espacio de las observaciones disponibles. El tamaño de las observaciones disponibles (N_o) es usualmente de mucho menor tamaño que el vector de estado (N_x), i.e., $N_o \ll N_x$. Por lo general se obtienen datos en algunas regiones del dominio y en el tiempo, tratando de que abarquen grandes escalas temporales y espaciales con la intención de conseguir una mejor estimación con esas observaciones.

Se asume lo siguiente acerca de los errores del modelo y de las observaciones:

- Error del modelo insesgado, centrado: $\mathbf{E}[\epsilon^m] = 0$
- Error observacional insesgado, centrado: $\mathbf{E}[\epsilon^o] = 0$
- Errores no correlacionados: $\mathbf{E}[\epsilon_i^m \epsilon_i^o \mathbf{T}] = 0$
- Covarianza del error del modelo conocida: $\mathbf{Q}_i = \mathbf{E}[\epsilon_i^m \epsilon_i^m \mathbf{T}]$
- Covarianza del error observacional conocida: $\mathbf{R}_i = \mathbf{E}[\epsilon_i^o \epsilon_i^o \mathbf{T}]$

donde \mathbf{E} es la esperanza del vector o matriz y el superíndice T indica el vector transpuesto.

Con esto, sea x_i^a (el superíndice a indica análisis) una estimación del estado verdadero del sistema x_i^t para cada tiempo i , con un error asociado $\epsilon_i^a = x_i^a - x_i^t$ centrado con una covarianza \mathbf{P}_i^a , de tal forma que:

$$\mathbf{P}_i^a = \mathbf{E}[\epsilon_i^a \epsilon_i^{aT}] = \mathbf{E}[(x_i^a - x_i^t)(x_i^a - x_i^t)^T] \quad (6)$$

Asumiendo una estimación (predicción) *a priori* x_i^f con su respectivo error $\epsilon_i^f = x_i^f - x_i^t$ y su covarianza \mathbf{P}_i^f conocida, es posible escribir una ecuación para actualizar el estado del modelo combinando linealmente las observaciones y estado anterior:

$$x_i^a = x_i^f + \mathbf{K}_i (x_i^o - H_i x_i^f) \quad (7)$$

donde \mathbf{K} es una matriz de pesos llamada matriz de ganancia de Kalman y el término $(x_i^o - H_i x_i^f)$ por el cual es multiplicado es conocido como vector de innovación.

Ahora, sustituyendo la ecuación 5 y 7 en 6, y tras manipulación de la expresión, se consigue una ecuación para la covarianza del error del análisis:

$$\mathbf{P}_i^a = (\mathbf{I} + \mathbf{K}_i H_i) \mathbf{P}_i^f (\mathbf{I} + \mathbf{K}_i H_i)^T + \mathbf{K}_i \mathbf{R}_i \mathbf{K}_i^T \quad (8)$$

donde \mathbf{I} es la matriz identidad.

El filtro de Kalman es un filtro óptimo, esto quiere decir, que la matriz de ganancia de Kalman minimiza la covarianza del error en la estimación del estado. La diagonal principal de la matriz de covarianza \mathbf{P}_i^a (ec. 6) corresponde a la media de los errores al cuadrado. La suma de la diagonal principal de una matriz cuadrada es la traza de la matriz. Entonces, con el fin de minimizar la covarianza del error en la estimación, se requiere minimizar la traza de \mathbf{P}_i^a . Para lograr esto, primero se deriva respecto a \mathbf{K}_i la traza de la ecuación 8 ($d(tr(\mathbf{P}_i^a))/d\mathbf{K}_i$), luego se iguala a 0 y tras manipulación de la expresión, se obtiene una ecuación para \mathbf{K}_i :

$$\mathbf{K}_i = \mathbf{P}_i^f H_i^T (H_i \mathbf{P}_i^f H_i^T + \mathbf{R}_i)^{-1} \quad (9)$$

La ganancia de Kalman puede ser interpretada como la razón entre la covarianza del error en la estimación previa con la covarianza del error total. Entre mayor el error en la predicción con respecto al error de las observaciones, mayor corrección en el análisis. Con la ecuación del análisis (7) y la de la ganancia de Kalman (9) se observa que:

- Si la observación es perfecta $\mathbf{R}_i = 0$, entonces $\mathbf{K}_i = 1$ y $x_i^a = x_i^o$
- Si la predicción es perfecta $\mathbf{P}_i^f = 0$, entonces $\mathbf{K}_i = 0$ y $x_i^a = x_i^f$

La ecuación de la matriz de covarianza del error actualizada \mathbf{P}_i^a es usualmente encontrada en la literatura de forma simplificada. Para llegar a dicha expresión, se sustituye la ecuación 9 en la forma expandida de 8, se manipula y se llega a:

$$\mathbf{P}_i^a = (\mathbf{I} - \mathbf{K}_i H_i) \mathbf{P}_i^f \quad (10)$$

Ahora, la estimación actualizada con la ganancia de Kalman se proyecta utilizando la dinámica del modelo para obtener una nueva estimación. Esta estimación puede ser definida como una estimación del estado futuro, predicción, pronóstico, o conocida comúnmente en la literatura en inglés, como background, forecast o first guess. De aquí en adelante, se referirá a esta estimación del estado futuro como predicción, y se modela como:

$$x_{i+1}^f = M_i x_i^a \quad (11)$$

Para terminar el ciclo de iteración del filtro, se requiere conseguir una expresión para poder proyectar la covarianza del error del análisis hacia el futuro. Para esto, se sustituye la expresión del estado verdadero (ec. 4) y la ecuación para proyectar el estado (ec. 11) en el error de la predicción $\epsilon_i^f = x_i^f - x_i^t$, obteniendo:

$$\epsilon_i^f = M_i \epsilon_i^a - \epsilon^m \quad (12)$$

Y ahora, sustituyendo la ecuación anterior, y expandiendo en la expresión de $\mathbf{P}_i^f = \mathbf{E}[\epsilon_i^f \epsilon_i^{fT}]$, y considerando que los errores del análisis y del modelo no están correlacionados ($\mathbf{E}[\epsilon_i^f \epsilon_i^m T] = 0$), se obtiene:

$$\mathbf{P}_{i+1}^f = M_i \mathbf{P}_i^a M_i^T + \mathbf{Q}_i \quad (13)$$

Con esto, se termina el ciclo de asimilación del esquema tradicional del filtro de Kalman que consiste en las ecuaciones en secuencia 11, 13, 9, 7 y 10. Una vez iniciado el ciclo (ya sea con una condición inicial o estimación previa), el filtro predice el estado a un tiempo futuro con su covarianza del error. Utilizando las covarianzas de los errores en la predicción y de las observaciones el filtro calcula pesos donde actualizar y corrige para conseguir un análisis. Finalmente partiendo de este nuevo análisis y de su covarianza del error, se inicia un nuevo ciclo.

- Filtro de Kalman extendido

El filtro de Kalman es formulado para el caso lineal y esto es una limitación para los problemas y modelos que se usan en el campo de la oceanografía. No obstante, puede ser usado prácticamente en modelos no lineales con algunas modificaciones. En el paso de la predicción (ec. 11), el operador M_i corresponde al modelo no lineal. El problema recae en la propagación de la covarianza del error (ec. 13), que depende de las no linealidades del modelo. Por lo tanto M_i es remplazado por \mathbf{M}_i , que es el Jacobiano, o gradiente de M_i evaluado en x_i^a :

$$\mathbf{M}_i = \left. \frac{dM_i}{dx} \right|_{x_i^a} \quad (14)$$

Y de forma análoga, el operador observacional H_i es remplazado por una versión no lineal \mathbf{H}_i en todo el algoritmo

$$\mathbf{H}_i = \left. \frac{dH_i}{dx} \right|_{x_i^f} \quad (15)$$

Con todo esto, el algoritmo para el filtro de Kalman extendido puede ser resumido como se observa en el siguiente esquema:

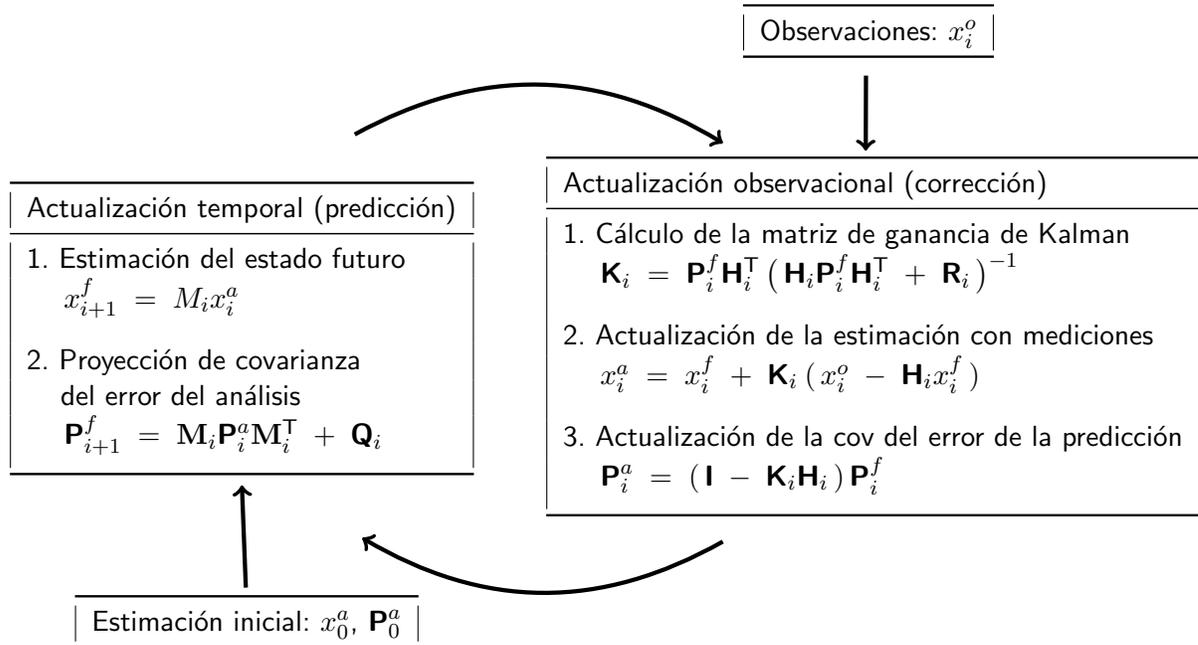


Figura 3. Diagrama del algoritmo recursivo del filtro de Kalman extendido.

- Filtro SEEK (Singular Evolutive Extended Kalman filter)

La mayor problemática de implementar el filtro de Kalman en oceanografía radica en que es tan computacionalmente costoso que no puede ser usado para la mayoría de los sistemas de grandes dimensiones. Particularmente, la ecuación 9 requiere de invertir una matriz cuadrada del tamaño del vector de estado, lo que conlleva un costo computacional grande. Además, la ecuación 13 involucra propagar el número de columnas de la covarianza, por lo que en un modelo típico en oceanografía, es poco práctico o intratable. Existen diferentes formas de abordar el tamaño de la matriz de covarianza, todas con el propósito de disminuir la carga computacional del algoritmo del filtro de Kalman preservando sus características esenciales. Por ejemplo, en el filtro de Kalman de ensamble, se aproxima la matriz de covarianza generando un ensamble de errores y se obtiene una estimación empírica. En este caso, se emplea el filtro SEEK, que trabaja con el concepto de reducción de orden, que en esencia, usa una covarianza del error en su versión reducida como una matriz de menor rango.

El filtro SEEK puede ser definido como una adaptación del Filtro de Kalman extendido, cuyas matrices de covarianza son de menor rango ("Singular") y que permite la propagación del error ("Evolutive"). Para reducir el rango de las matrices en el algoritmo, se consigue un estado y covarianza del error

inicial realizando un análisis de Funciones Empíricas Ortogonales (FEOs) a una simulación libre. Se utilizan FEOs como primera aproximación ya que se asume que una simulación libre es suficiente para poder describir la variabilidad real. De esta descomposición se obtiene $\mathbf{P}_0^a = \mathbf{L}_0 \Lambda_0 \mathbf{L}_0^T = \mathbf{S}_0 \mathbf{S}_0^T$ con $\mathbf{S}_0 = \mathbf{L}_0 \sqrt{\Lambda_0}$. Los r eigenmodos dominantes forman las columnas de la matriz \mathbf{L}_0 con dimensión $N_x \times N_r$ (donde N_x es el tamaño del vector de estado, y N_r es el número de modos dominantes), y sus eigenvalores asociados se contienen en la matriz diagonal Λ_0 de dimensión $N_r \times N_r$, de tal forma que el rango de \mathbf{P}_0^a sea del tamaño de r .

Utilizando esta reducción, la covarianza del error del análisis es:

$$\mathbf{P}_i^a = \mathbf{S}_i^a \mathbf{S}_i^{aT} \quad (16)$$

Sustituyendo la ecuación anterior en la ecuación de la propagación del error 13 se obtiene:

$$\mathbf{P}_{i+1}^f = \tilde{\mathbf{S}}_i^f \tilde{\mathbf{S}}_i^{fT} + \mathbf{Q}_i \quad \text{con} \quad \tilde{\mathbf{S}}_{i+1}^f = \mathbf{M}_i \mathbf{S}_i^a \quad (17)$$

A diferencia de la versión anterior, ahora el costo operacional de la propagación del error depende del rango r , donde $N_r \ll N_x$. La ecuación 17 corresponde a la formulación original del filtro SEEK, y es conocida como una "base evolutiva extendida" o base continua. Dicha base se consigue utilizando el gradiente del operador M , el cual es complicado de estimar. Por lo tanto, otra versión para propagar el error es mediante la siguiente ecuación:

$$\alpha \left\{ \tilde{\mathbf{S}}_{i+1}^f \right\}_j = M_i \left(x_i^a + \alpha \left\{ \mathbf{S}_i^a \right\}_j \right) - x_i^f \quad (18)$$

donde $\left\{ \right\}_j$ indica el número de la columna, y α es un parámetro ajustable que determina la escala de las perturbaciones en cada dirección. Esta propagación del error es conocida como "base evolutiva interpolada" y tiene como ventaja que no se requiere calcular el gradiente \mathbf{M} . Sin embargo, esta última base requiere de propagar el error mediante la aplicación del operador M para cada columna, que corresponde a cada modo dominante hasta r . Esto puede ser computacionalmente demandante si se usa un gran número de modos.

Por tanto, se puede simplificar aún más la propagación considerando:

$$\tilde{\mathbf{S}}_{i+1}^f = \mathbf{M}_i \mathbf{S}_i^a \approx \mathbf{I} \mathbf{S}_i^a \quad (19)$$

En donde los efectos que pudiera tener el modelo en modificar las direcciones a donde se propaga el error no son considerados. Esta propagación es conocida como "base fija" y tiene la ventaja práctica de no tener que propagar el error en cada dirección, lo que se traduce a un costo computacional equivalente al de la simulación libre, con algunas operaciones algebraicas del resto del algoritmo.

Para obtener por completo la covarianza del error de la predicción, se requiere sumar la propagación del error mediante alguna de las bases en las ecuaciones 17, 18 o 19 y la covarianza del error del modelo \mathbf{Q}_i . Esta última suma trae un problema ya que \mathbf{P}_i^f debería preservar el mismo rango que \mathbf{P}_i^a para mantener las ventajas de la reducción de orden. Para solucionar esto, \mathbf{Q}_i se parametriza como:

$$\mathbf{Q}_i = \frac{1}{\rho} (1 - \rho) \times \left(\tilde{\mathbf{S}}_i^f \tilde{\mathbf{S}}_i^{fT} \right) \quad (20)$$

donde ρ ($0 < \rho < 1$) es llamado "factor de atenuación" (forgetting factor en inglés), que amplifica la estimación del error sin modificar el rango.

Con esto, sustituyendo la parametrización del error del modelo (ec. 20) en la ecuación 13, y sustituyendo y manipulando algebraicamente la \mathbf{P}_i^f de rango reducido en las expresiones para conseguir la ganancia de Kalman (\mathbf{K}_i) y la actualización de la covarianza del error (\mathbf{P}_i^a), el algoritmo del filtro SEEK se resume con el esquema de la figura (4).

Y como se puede observar en el algoritmo de asimilación (Figura 4), el filtro SEEK mantiene el esquema recursivo del filtro de Kalman original (Figura 3). Sin embargo, como ya se discutió, el filtro SEEK es más eficiente. Durante la proyección de covarianza, al utilizar una base reducida, el número de operaciones y costo computacional se reduce también al utilizar cualquiera de las versiones de las bases. Además, en el cálculo de la matriz de ganancia de Kalman, el tamaño del problema de inversión se reduce de N_x a N_r (si se asume que la matriz \mathbf{R}_i es conocida o parametrizada como una matriz diagonal), lo que disminuye aún más la carga computacional.

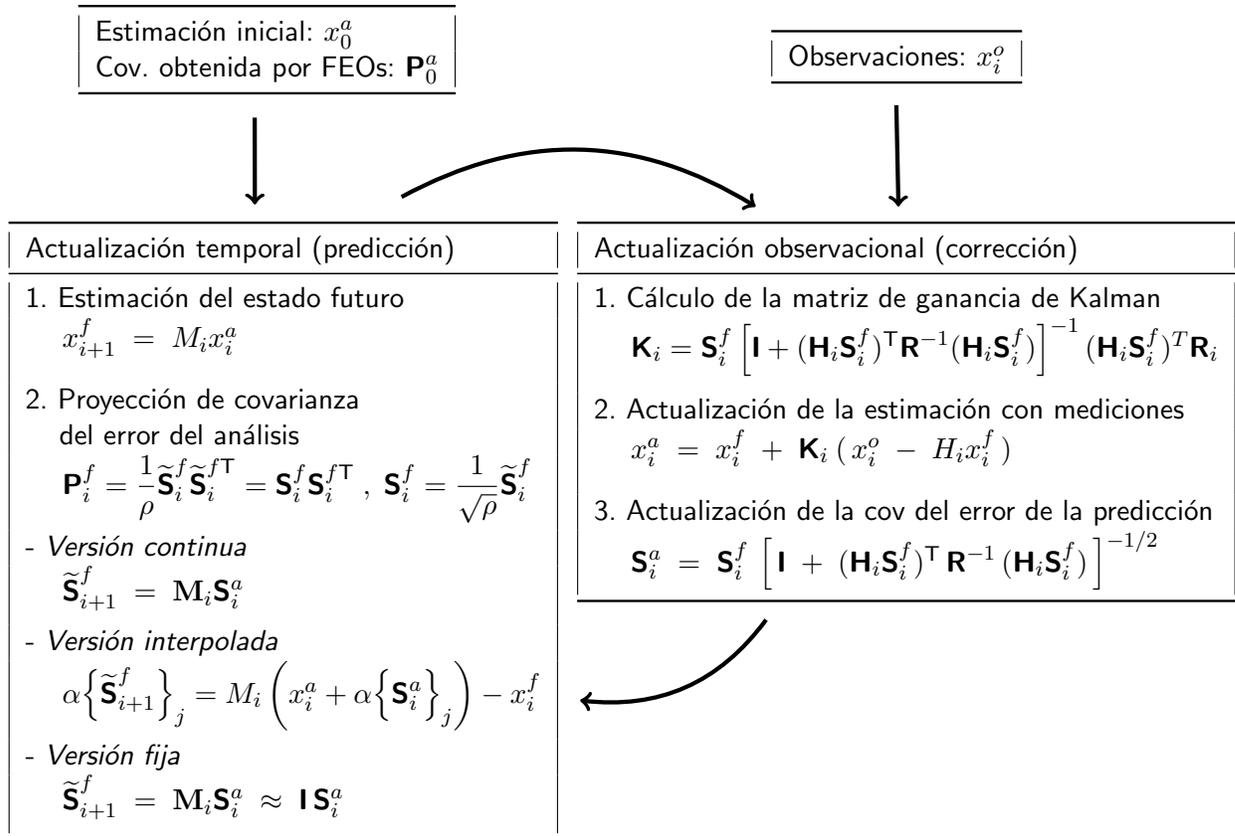


Figura 4. Diagrama del algoritmo recursivo del filtro SEEK.

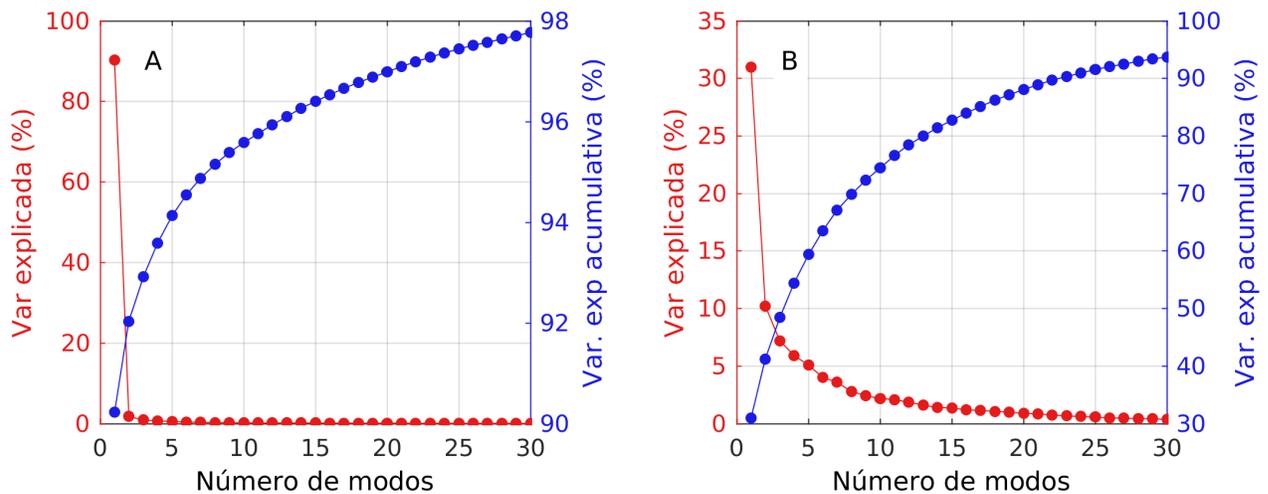


Figura 5. Varianzas explicadas por número de modos de SST (A) y SSH (B).

La estimación de la primera aproximación utilizada al iniciar los ciclos de asimilación, tanto como en el filtro de Kalman y el filtro SEEK, es de suma importancia ya que esa información perdura al propagarse en cada iteración de la secuencia del algoritmo. Como ya se vio, esta aproximación se consigue del análisis de FEOs de una simulación libre, y que tan bien represente esta aproximación la variabilidad real

del estado del océano depende del número de modos dominantes que se consideren. En la Figura 5 se observa que para la SST, los primeros 5 modos corresponden a cerca del 95 % de la varianza explicada, mientras que en la SSH, se requieren alrededor de 20 modos para conseguir cerca del 90 % de la varianza explicada.

Por lo tanto, para estudiar cómo evoluciona y qué o cuales efectos tiene en el estado del modelo distintas aproximaciones de la covarianza inicial, se realizaron diferentes experimentos utilizando los modos dominantes mostrados en la siguiente tabla:

Tabla 2. Número de modos usados para reconstruir la covarianza y sus respectivos porcentajes de varianza explicada de SST y SSH.

Número de modos	1	2	3	4	5	10	15	20
Varianza explicada de SST (%)	90.2	92.0	92.9	93.6	94.1	95.6	96.4	97.0
Varianza explicada de SSH (%)	31.0	41.2	48.5	54.4	59.5	74.5	82.8	88.1

2.3. Métodos de análisis

Para evaluar el desempeño de cada experimento se utiliza la raíz del error cuadrático medio (RMS, por sus siglas en inglés):

$$\text{RMS} = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i^m - x_i^o)^2} \quad (21)$$

donde x^m es la variable del estado del modelo, x^o es la variable observada o de referencia, y N es el número de elementos de la variable. La RMS entre las variables asimiladas y las predichas por el modelo, a diferencia de los errores del modelo con respecto a observaciones independientes, son desajustes del modelo respecto a las observaciones y no son errores en el sentido estricto de la definición. Sin embargo, por practicidad, se usará el término RMS tanto en los desajustes del modelo con las observaciones asimiladas como en los errores contra datos independientes.

Además de los errores del modelo respecto a las observaciones, también se examina el rendimiento de las simulaciones con relajación y asimilación respecto a la simulación de control. Para esto, se calcula la diferencia entre raíces del error cuadrático medio relativas (RMSD, por sus siglas en inglés) como:

$$\text{RMSD} = \text{RMS}^A - \text{RMS}^C = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i^A - x_i^o)^2} - \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i^C - x_i^o)^2} \quad (22)$$

donde x^A y x^C son las variables de estado de la simulación con asimilación o relajación y la simulación de control respectivamente. De la ecuación anterior se ve que, si $\text{RMSD} < 0$, el experimento con relajación o asimilación tiene un mejor desempeño que la simulación de control, y viceversa si $\text{RMSD} > 0$.

Otro parámetro útil de conocer es el Sesgo del error, que se define en este trabajo como el promedio aritmético de los errores entre las observaciones y las variables del modelo con asimilación o relajación:

$$\text{Sesgo} = \frac{1}{N} \sum_{i=1}^N (x_i^A - x_i^o) \quad (23)$$

El sesgo indica si el modelo está sobrestimado o subestimado respecto a las observaciones.

Capítulo 3. Resultados

En este capítulo se presentan los resultados obtenidos para las tres series de experimentos propuestos para este trabajo. Primero, se muestra una validación de la simulación de control. Seguido se describe los resultados de las series de experimentos de los esquemas de relajación y asimilación. Finalmente se comparan salidas seleccionadas de los tres experimentos.

3.1. Simulación libre

Para examinar si el modelo por si solo se desempeña de forma correcta, se valida cualitativa y cuantitativamente. En las Figuras 6, 7 y 8 se muestran los promedios temporales de temperatura, salinidad y la componente meridional de la velocidad de corrientes de las fronteras norte, oeste y sur de la malla del modelo (Figura 2). Usando los promedios en las fronteras, se describe y compara cualitativamente lo observado en la simulación contra los patrones espaciales conocidos del SCC. Las isopicnas de 25.4 y 26.3 kg m^{-3} mostradas en los promedios de las fronteras fueron seleccionadas como menciona por Durazo (2015) para indicar el límite del mínimo de salinidad subsuperficial de la CC y la profundidad de la isopicna correspondiente al núcleo de la CCC.

- Frontera norte

Los patrones espaciales de los promedios en la frontera norte (Figura 6) son semejantes a lo encontrado por Todd et al. (2012). Las isohalinas e isotermas emergen mientras se acerca la costa, característico de la región por la existencia de surgencias costeras debido al viento continuo paralelo a la costa. La salinidad y temperatura promedio muestran la estructura termohalina persistente de gran escala de la CC. La señal de baja salinidad de ~ 33 psu y temperatura de $\sim 14-18$ °C de la CC se encuentra en los entre los primeros 200 m como se esperaría. La isopicna de 25.4 kg m^{-3} define de forma clara el límite de la CC subsuperficial, que se extiende hasta los 100 m de profundidad lejos de la costa y más somera (~ 20 a 50 m) cerca de la península. La CC se exhibe de igual forma en el promedio de la componente meridional, con un fuerte núcleo ~ -0.2 m s^{-1} que se dirige al ecuador sobre la isopicna de 25.4 m^{-3} . La CCC por su parte se muestra correctamente posicionada entre las isopicnas de 25.4 y 26.3 kg m^{-3} pegada a la costa, con un fuerte núcleo de velocidad meridional de ~ 0.1 m s^{-1} que se dirige al polo.

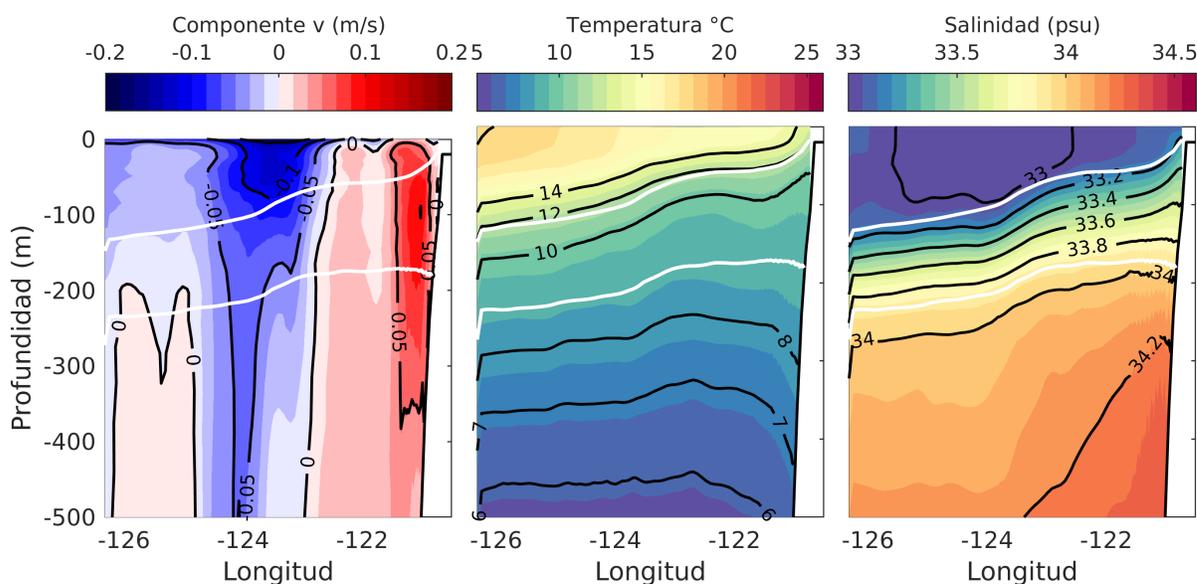


Figura 6. Promedios temporales de la simulación de control en la frontera norte para la componente v (panel izquierdo), temperatura (panel central) y salinidad (panel derecho). Las líneas blancas indican las isopicnas de 25.4 y 26.3 kg m^{-3} .

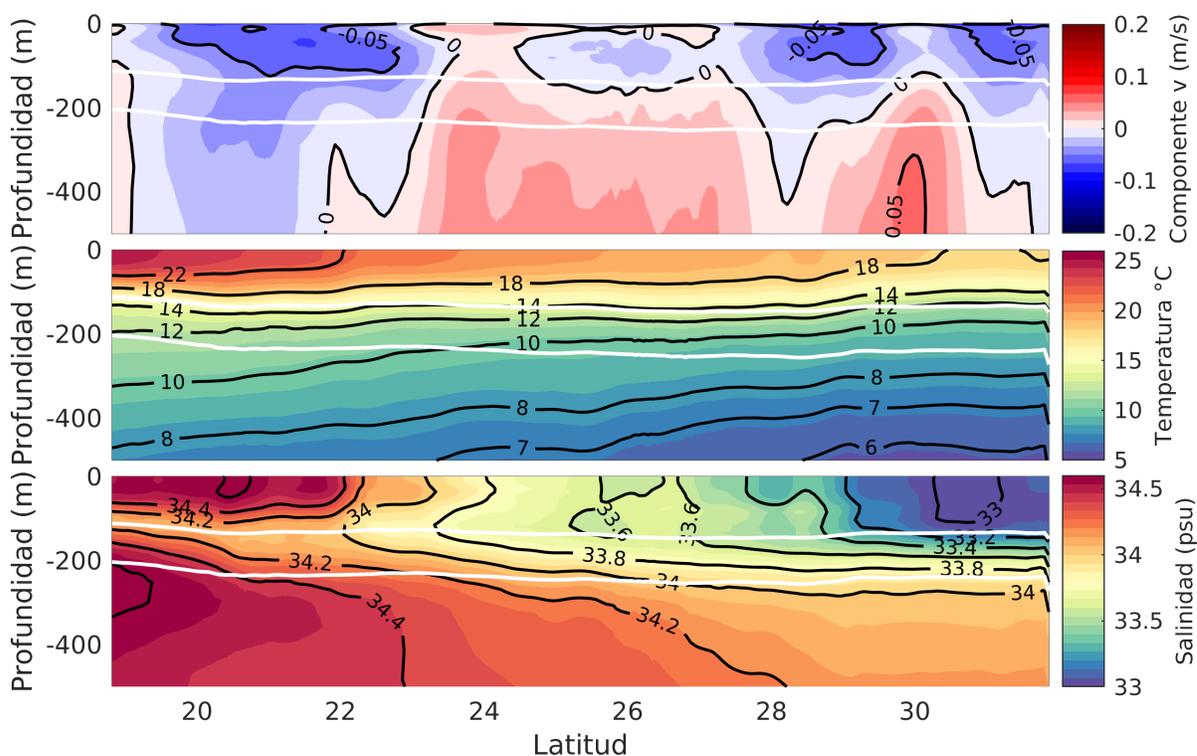


Figura 7. Promedios temporales de la simulación de control en la frontera oeste para la componente v (panel superior), temperatura (panel central) y salinidad (panel inferior). Las líneas blancas indican las isopicnas de 25.4 y 26.3 kg m^{-3} .

- Frontera oeste

Los promedios temporales de la frontera oeste (Figura 7) reflejan la transición de la CC que acarrea aguas polares de baja temperatura y salinidad ($\sim 14\text{-}15^\circ\text{C}$ y $\sim 33\text{-}33.2$ psu en los primeros 100 m) hacia aguas subtropicales cálidas y saladas ($\sim 18\text{-}25^\circ\text{C}$ y $\sim 34.2\text{-}34.6$ psu en los primeros 100 m). En general,

la influencia de la CC se identifica de forma correcta con velocidades meridionales promedio negativas (hacia el ecuador) sobre la isopicna de 25.4 m^{-3} . El Agua Subártica (ASA) que acarrea la CC se sumerge mientras se traslada hacia el sur, esto se observa como un núcleo de salinidad relativamente más baja ($\sim 34 \text{ psu}$) que se hunde debajo de una masa de agua más cálida y salada alrededor de los 23°N . Esta masa de agua en donde se hunde el ASA corresponde al Agua Subtropical Superficial (ASTs) y Agua Transicional (ATr) en la superficie, y se hunde sobre el Agua Ecuatorial Subsuperficial (AESs).

- Frontera sur

Los promedios temporales de la frontera sur (Figura 8) muestran, de igual forma que en la frontera norte (Figura 6), una estratificación continua en la estructura termohalina. Las isotermas se alinean a lo largo de la frontera, con una termoclina marcada alrededor de los 100 m. La señal de baja temperatura de la ASA que es acarreada por la CC no es visible como en la frontera norte. Sin embargo, la señal de baja salinidad si es mostrada bien definida con un núcleo de $\sim 34 \text{ psu}$ de salinidad que se sumerge alrededor de los 100 m y se encuentra cercana a la costa. El promedio de la velocidad meridional aún muestra una señal débil de la CC sobre la piconclina de 25.4 m^{-3} . Por su lado, es notable la CCC que viaja pegada a la costa con un núcleo de $\sim 0.12 \text{ m}^{-1}$ a una profundidad entre 200 a 500 m.

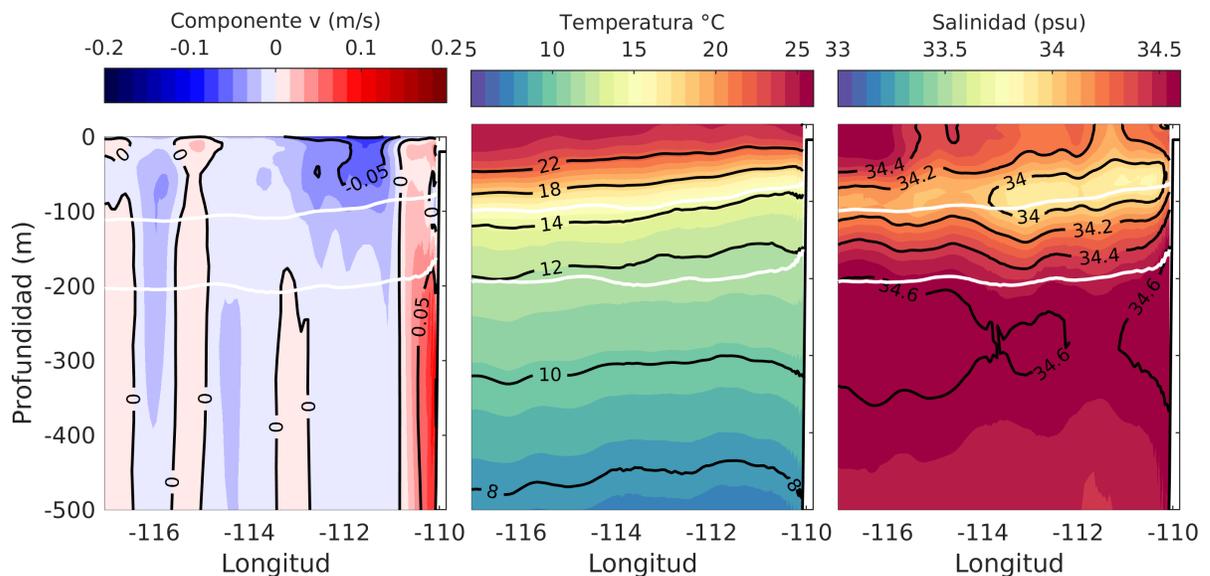


Figura 8. Promedios temporales de la simulación de control en la frontera sur para la componente v (panel izquierdo), temperatura (panel central) y salinidad (panel derecho). Las líneas blancas indican las isopicnas de 25.4 y 26.3 kg m^{-3} .

En general, los promedios temporales en las fronteras exhiben los patrones espaciales esperados para el SCC. Una corriente somera (0 a 100 m) que viaja hacia el sur con baja temperatura y salinidad, y una contracorriente subsuperficial (100 a 500 m) que viaja hacia el norte con alta temperatura y salinidad.

Para examinar el desempeño cuantitativo de la simulación de control, se comparan contra los datos independientes de la campaña de IMECOCAL. Para esto, se comparan los datos de temperatura potencial y salinidad absoluta de los perfiles hidrográficos contra su contraparte del modelo interpolados a las posiciones geográficas y a profundidad de cada lance.

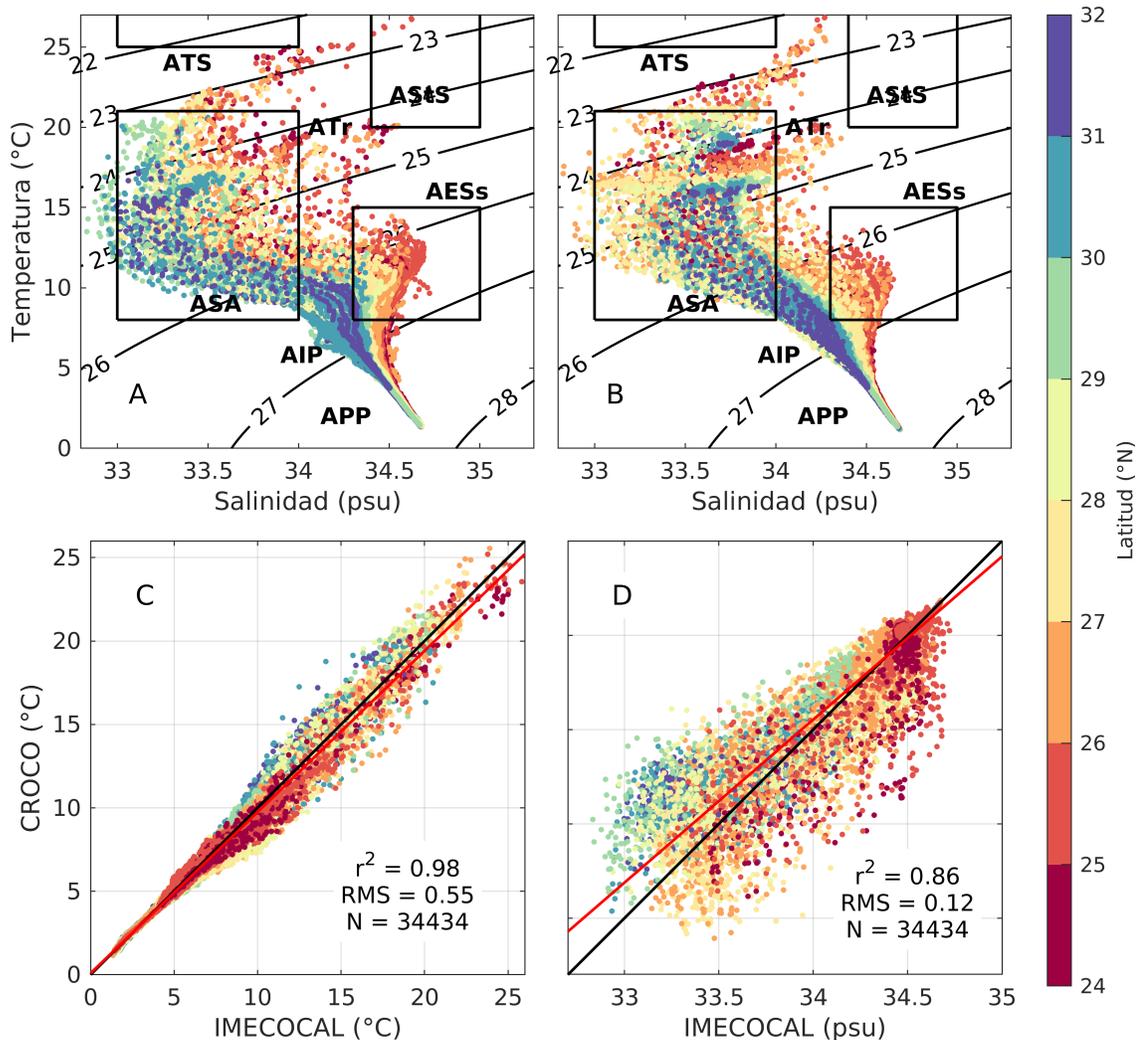


Figura 9. Diagramas T-S de las campañas de IMECOCAL (A) y simulación libre (B). Diagramas de dispersión de temperatura (C) y salinidad (D) entre datos de la simulación libre e IMECOCAL. Los recuadros negros y siglas en lo diagramas T-S indican los límites y nombres de las masas de agua prevalentes en el SCC: ASA = Agua Subártica, ATS = Agua Tropical Superficial, ATr = Agua Transicional, ASSts = Agua Subtropical Superficial, AESs = Agua Ecuatorial Subsuperficial, AIP = Agua Intermedia del Pacífico y APP = Agua Profunda del Pacífico. En los diagramas de dispersión se muestra el coeficiente de determinación r^2 , RMS, y número de datos (N) utilizados para calcular los estadísticos. La línea negra y roja continua representa el ajuste perfecto y el de la regresión lineal respectivamente.

Se observa tanto en el modelo como en las observaciones, que las masas de agua esperadas para la región del SCC están bien representadas (Figura 9A y 9B). Con el fin de localizar el origen meridional del agua, y de reconocer las diferencias entre el modelo y las observaciones, los diagramas T-S y de dispersión son señalados con un mapa de color que indica el valor de la latitud donde se encuentra la

observación.

En la superficie la ASA es la masa de agua más frecuente con un mínimo salino de salinidad y temperatura ~ 33 psu y ~ 10 °C y se encuentra ubicado en la isopicna de ~ 25.4 kg m⁻³, característico del flujo de la CC. En la mayoría de las observaciones, el mínimo de la ASA proviene de latitudes altas (mayores a 29°N), sin embargo, el agua más norteña del modelo presenta un mínimo de salinidad mayor al esperado (>33.2 psu). Además, el modelo muestra agua con el mismo mínimo termohalino que la ASA en latitudes menores (27-29°N).

A medida la CC viaja superficialmente hacia el sur, esta se mezcla con agua subtropical, lo que genera la masa de Agua Transicional (ATr). Esta agua, es representada de forma adecuada en los datos como filamentos que parten de la ASA hacia aguas más cálidas y saladas (Agua Tropical Superficial y Agua Subtropical Superficial). Por otro lado, el modelo parece mostrar una subestimación en los máximos salinos del ASA y el ATr, alrededor de los 18°C, 33.57 psu e isopicna de 24.5 kg m⁻³.

Debajo de la superficie, existe un máximo de salinidad de ~ 34.72 psu y alrededor de la isopicna de de 26.5 kg m⁻³, característico de la presencia de la CCC (~ 150 -200 m). En ambos casos de las observaciones y en el modelo este máximo es visible, aunque en el caso del modelo no es tan pronunciado el máximo ni llega a la misma magnitud ~ 34.65 psu. A mayor profundidad, se ubican las masas de Agua Intermedia del Pacífico y Agua Profunda del Pacífico, y en ambos casos, es similar su distribución en los diagramas T-S.

En general, las masas están bien representadas en el modelo, sin embargo, su distribución meridional no parece ser la correcta. La simulación tiene valores de salinidad mayores y menores en latitudes altas (29-32°N) y medias (27-29°N) respectivamente, además de tener diferencias en la distribución del ATr y en realzar la presencia de la señal de la CCC y el AESs que contiene. Así mismo, se observa de los diagramas de dispersión (Figura 9C y 9D) que el modelo tiende a diferir en los valores de salinidad en mayor medida que en la temperatura. No obstante, para ambas variables, los valores de RMS y del coeficiente de determinación son los deseables para sugerir que la simulación de control retrata de forma adecuada las características termohalinas del SCC.

3.2. Relajación Newtoniana

Con el fin de revisar el desempeño de las simulaciones con el esquema de relajación primero se examina la evolución temporal de la RMS en todo el dominio superficial. Es importante mencionar que en este ejercicio se comparan las estimaciones del modelo de SST y SSH contra las mismas observaciones a

las cuales se relaja. El hacer esto permite extraer información valiosa de las estimaciones del modelo:

- Validar la simulación de control y experimentos con relajación al analizar la serie de tiempo del error (RMS), así como conocer sus patrones temporales;
- Evaluar el desempeño que tienen los experimentos con relajación respecto a la simulación de control examinando la evolución temporal de las diferencias en los errores (RMSD);
- Confirmar que el esquema de relajación funcione como es esperado.

La evolución temporal de la RMS de la SST (Figura 10) muestra que la relajación reduce el error a lo largo de todo el periodo de estudio, e incrementa esta mejoría conforme se utiliza un coeficiente de relajación menor. La RMS de la simulación de control promedio de $0.95\text{ }^{\circ}\text{C}$ es un valor adecuado para modelos numéricos. Este promedio reduce gradualmente hasta llegar a un mínimo de $0.18\text{ }^{\circ}\text{C}$ para la simulación con $\gamma = 20\text{ min}$. El desempeño de los experimentos fue mejor respecto a la simulación de control (RMSD) para todos los coeficientes excepto para los mayores coeficientes durante los últimos meses del periodo de estudio.

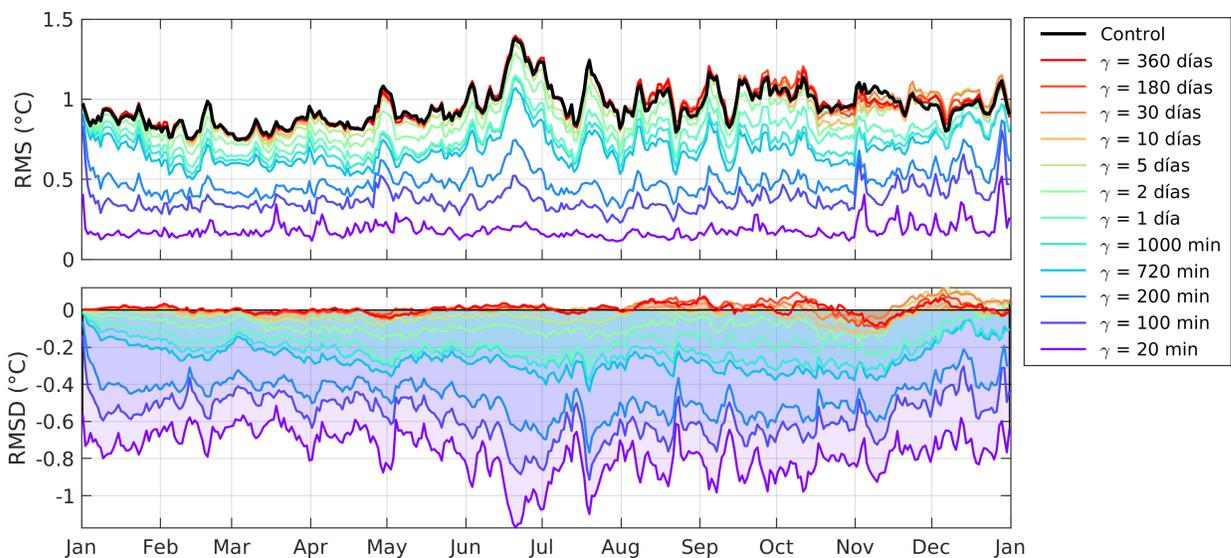


Figura 10. Evolución temporal de la RMS (panel superior) de SST de la simulación de control y de cada experimento con diferentes coeficientes de relajación (señalado como líneas continuas de color) comparadas contra observaciones de MUR de toda la región de estudio. Evolución temporal de la RMSD (panel inferior).

La evolución temporal de la RMS de la SSH (Figura 11) por su parte exhibe un comportamiento similar al de la SST. La RMS disminuye gradualmente conforme se usan coeficientes más pequeños, y se aproxima hacia el cero para el coeficiente de $\gamma = 20\text{ min}$. A diferencia del caso de la SST, los experimentos con coeficientes mayores a 5 días tuvieron un peor desempeño respecto a la simulación de control (RMSD) durante los meses de junio a enero.

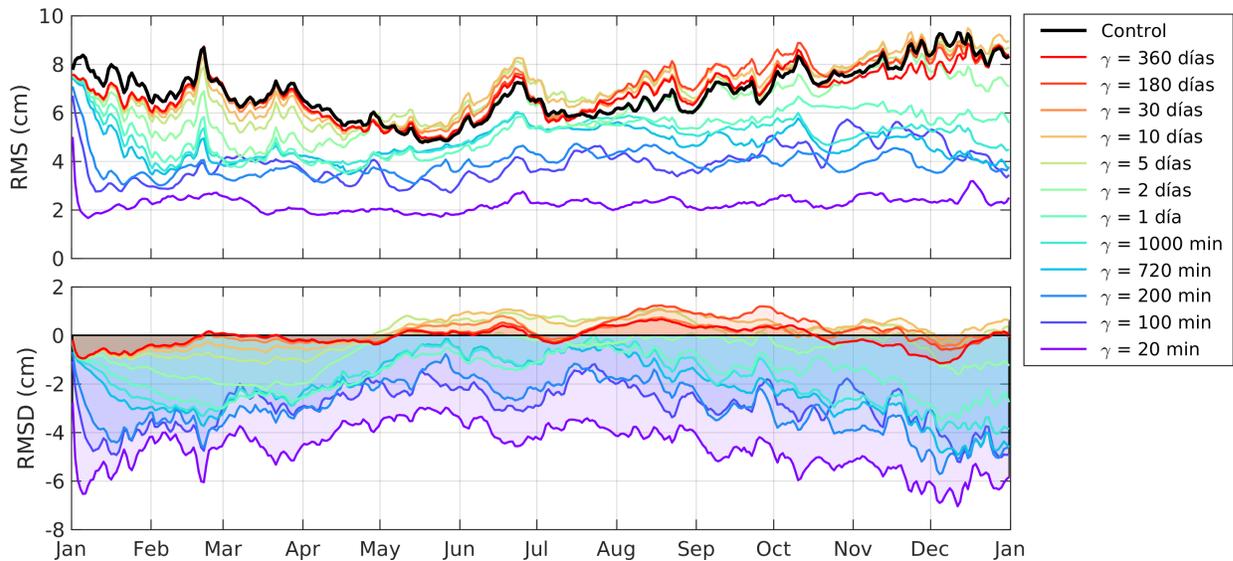


Figura 11. Evolución temporal de la RMS (panel superior) de SSH de la simulación de control y de cada experimento con diferentes coeficientes de relajación (señalado como líneas continuas de color) comparadas contra observaciones de toda la región de estudio. Evolución temporal de la RMSD (panel inferior)

Como era lo esperado, los experimentos con relajación muestran una disminución en el error de SST y SSH al compararse contra las mismas observaciones. Ahora, para validar y examinar el efecto de la relajación en la columna de agua, los datos de los experimentos se comparan contra las observaciones independientes de temperatura y salinidad de las campañas de IMECOCAL. Esta comparación es presentada en las Figuras 12 y 13, donde se muestra el Sesgo, RMS y RMSD de los experimentos respecto a las observaciones. El Sesgo es particularmente útil para mostrar si el modelo sub o sobrestima respecto a los lances hidrográficos, es decir, si el modelo cuenta con menor o mayor temperatura o salinidad y a cuáles profundidades. Por su parte, la RMS y RMSD son útiles de la misma manera que en el análisis anterior.

La RMS de la temperatura (Figura 12B) de la simulación de control y de los experimentos de relajación muestran un patrón espacial similar. Se localizan dos máximos de error en la simulación de control, uno de $\sim 1.35^\circ\text{C}$ a ~ 60 m, y otro de $\sim 1^\circ\text{C}$ a ~ 220 m. Ambas profundidades corresponden a las posiciones de la CC y CCC respectivamente. A mayor profundidad, el error disminuye hasta llegar casi a 0. Por el contrario de lo observado en los errores superficiales, la RMS de la temperatura aumenta en toda la columna de agua conforme se utiliza un coeficiente de relajación menor. Esto es más notable con el experimento de $\gamma = 20$ min. La RMSD (Figura 12C) muestra que las correcciones realizadas con coeficientes de relajación menores a 2-5 días no parecen tener efecto en la vertical. Estos errores, en particular entre los 100 a 700 m son debido a una subestimación del modelo (Figura 12A).

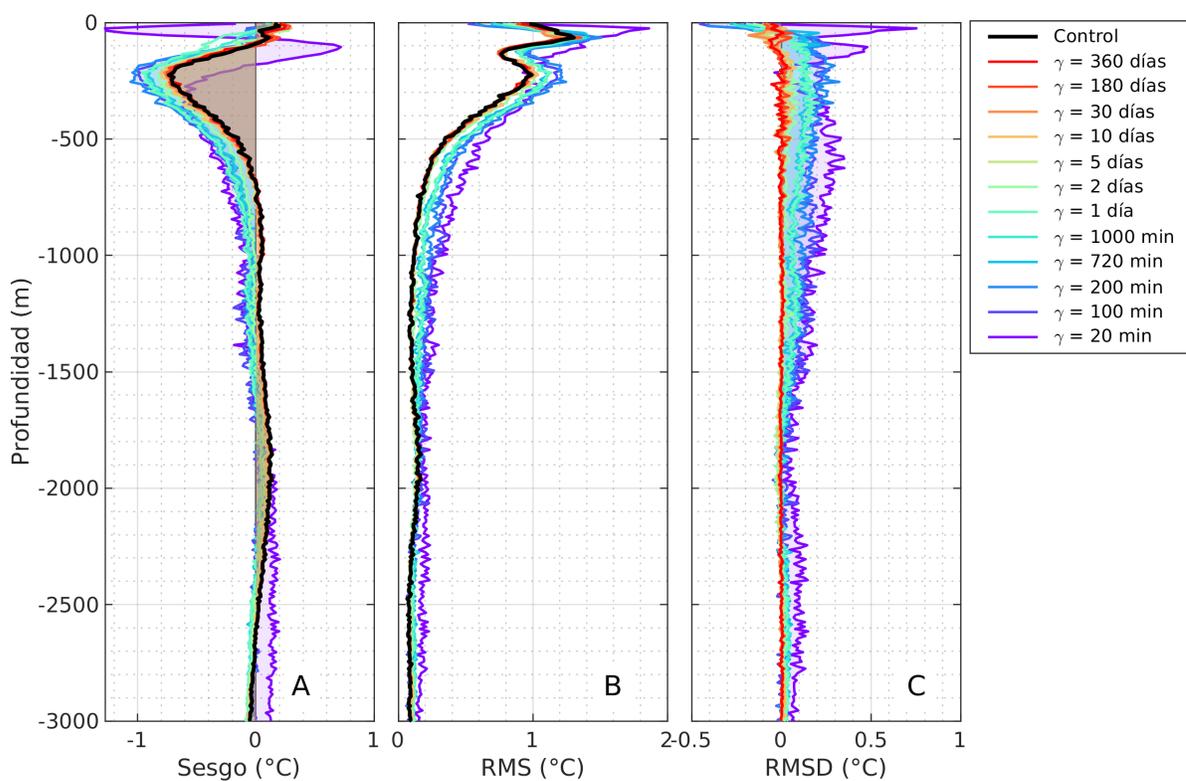


Figura 12. Perfiles de Sesgo (A), RMS (B) y RMSD (C) de temperatura de cada simulación con distintos coeficientes de relajación respecto a datos de perfiles hidrográficos de IMECOCAL.

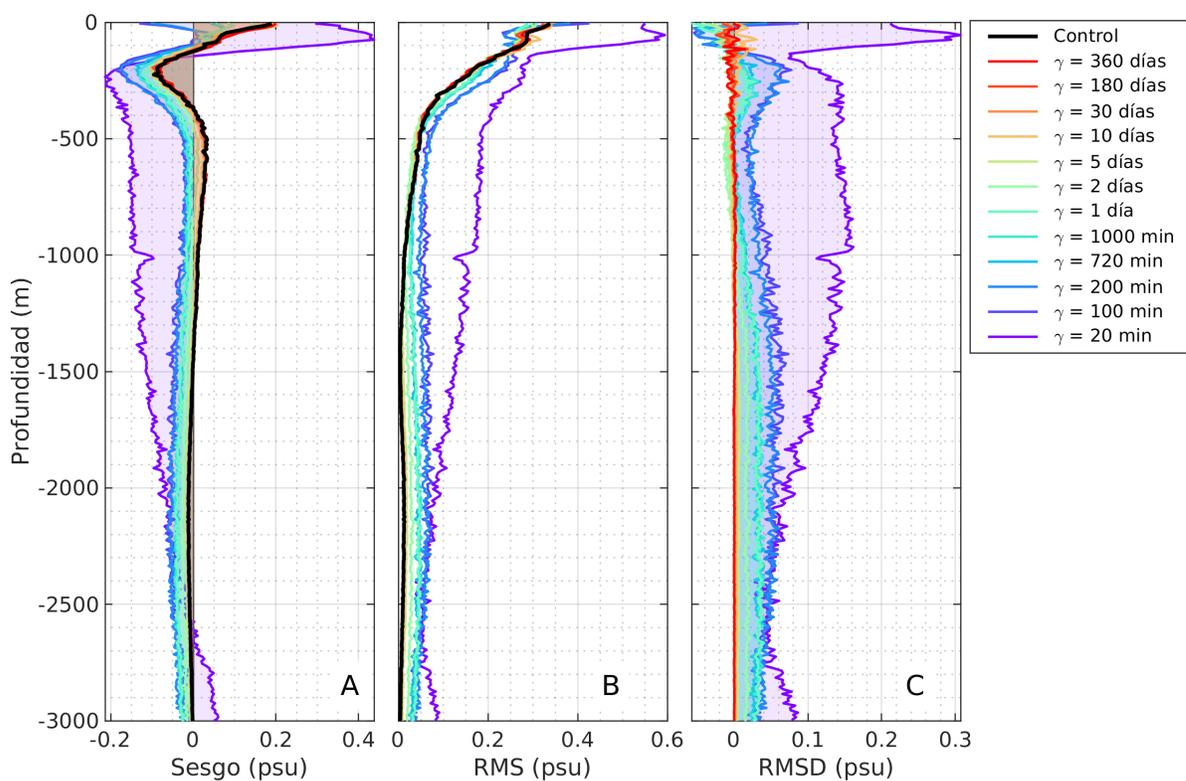


Figura 13. Perfiles de Sesgo (A), RMS (B) y RMSD (C) de salinidad de cada simulación con distintos coeficientes de relajación respecto a datos de perfiles hidrográficos de IMECOCAL.

De forma análoga a la temperatura, la RMS de la salinidad (Figura 13B) de la simulación de control y de los experimentos de relajación muestran un patrón similar. El mayor error en la simulación de control se encuentra en la superficie ~ 0.34 psu, para después decrecer conforme aumenta la profundidad, hasta llegar a valores mínimos (~ 0.005) a los 1500 m. Nuevamente, se tiene un peor desempeño en toda la columna de agua (Figura 13C) respecto a la simulación de control para los coeficientes de relajación menores a 2-5 días. Para el caso de la salinidad, el sesgo del modelo (Figura 13A) muestra una sobrestimación alrededor de los primeros 100 m y subestimación alrededor de 100-400 m.

La Figura 14 resume el efecto de la selección de γ , donde los errores de cada simulación respecto a los datos superficiales y a profundidad son mostrados con relación a sus γ . Se percibe que existe un intervalo del coeficiente de relajación ($\gamma > 5$ días) donde las correcciones son lo suficientemente pequeñas para no modificar el estado del modelo, y por consiguiente, no se muestra un cambio en el error relativo a la simulación de control. Y de forma inversa, existe un intervalo ($\gamma < 2$ días) en donde las modificaciones son suficientemente grandes para forzar al modelo a que se asemejen a las observaciones satelitales y así disminuir el error superficial, pero con ello aumentar el error en el resto de la columna de agua. Por ello, el intervalo entre 2 a 5 días parece ser el indicado para relajar en esta configuración al modelo con el fin de disminuir el error superficial sin ocasionar un aumento en la vertical.

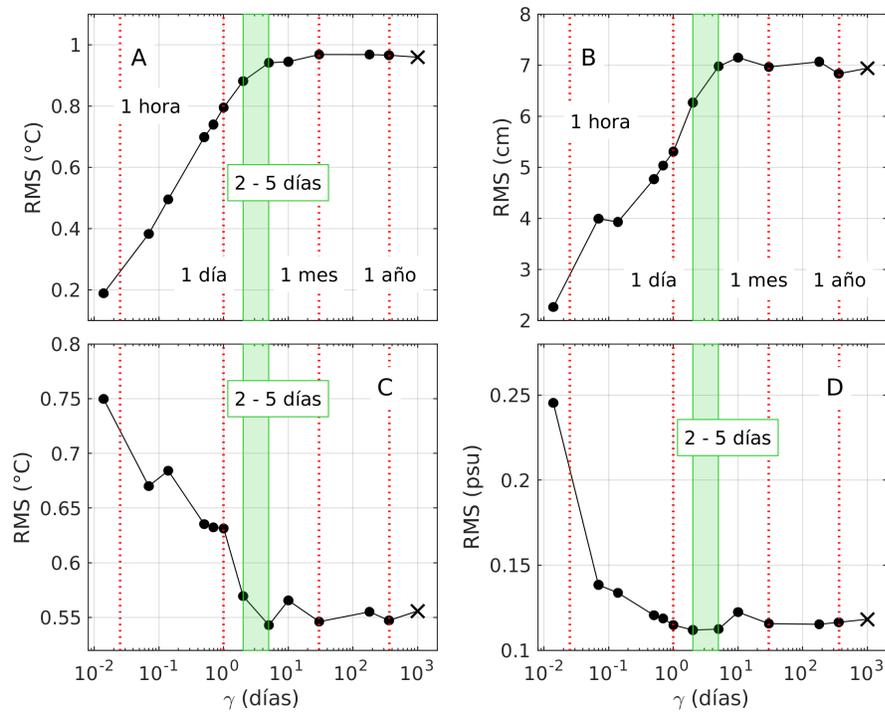


Figura 14. RMS de cada simulación con diferentes coeficientes de relajación (marcado como puntos negros) respecto a mediciones satelitales de SST (A), SSH (B), y mediciones hidrográficas de temperatura (C) y salinidad (D). Los valores de RMS de correspondientes a la simulación de control son marcados como cruces negras. Se muestra valores de γ de referencia como líneas punteadas rojas y se indica el intervalo adecuado de selección del coeficiente de relajación como un recuadro verde.

3.3. Asimilación de datos

Este trabajo utiliza una base fija (Ecuación 19) para parametrizar la matriz de covarianza. El análisis de FEOs requerido para construir la matriz de menor rango fue obtenido de una simulación libre de 5 años muestreada cada día con la configuración mostrada en la sección 2.2.1. Los primeros tres modos dominantes de la SST y SSH son mostrados en las figuras 15 y 16 respectivamente. Con el fin de verificar que la aproximación de la matriz de covarianza sea realista, se comparan los modos con mayor varianza explicada de la simulación libre contra los modos extraídos de series largas de tiempo de SST del producto MUR y SSH derivada de altimetría.

En general el modelo presenta los mismos patrones espaciales de temperatura superficial que las observaciones para los tres primeros modos (Figura 15). El modo 1 y 2 muestran una señal al sur de la península, aunque en menor medida en el modelo. El modo 3 es similar en ambos casos, donde explica $\sim 1\%$ de la varianza y presenta una señal más fuerte al norte del dominio en la región de las Islas del Canal al oeste de California. Para ambos casos, los tres primeros modos contienen cerca del 92% de la Varianza Explicada (VE). Sin embargo, la VE no se distribuye de igual forma entre los modos. El primer modo de la simulación del modelo explica $\sim 4\%$ más que el modo de las observaciones. Y de forma opuesta, el segundo modo de las observaciones explica $\sim 2.6\%$ más que el modo de la simulación. Esto sugiere que los procesos que ocurren en el sur de la península que hacen variar la SST no se reflejan de forma adecuada en el modelo.

Respecto a los modos dominantes de la SSH (Figura 16), los modos del modelo y observaciones explican porcentajes de la varianza similares y muestran patrones espaciales comparables, aunque a diferencia de la SST, la similitud no es tan clara en algunas regiones. El modo 1 y 2 muestran una señal fuerte pegada a la costa, particularmente esto es más intenso para el modo 2. El modelo parece tener discrepancias respecto a las observaciones lejos de la costa en los tres modos. Esto tiene mayor peso en el primer modo, donde en la región del norte se observan patrones de mayor intensidad en el modelo respecto a las observaciones. Si bien la VE del modo 3 del modelo y de las observaciones es similar, su patrón espacial es distinto, aunque con pesos similares, en todo el dominio.

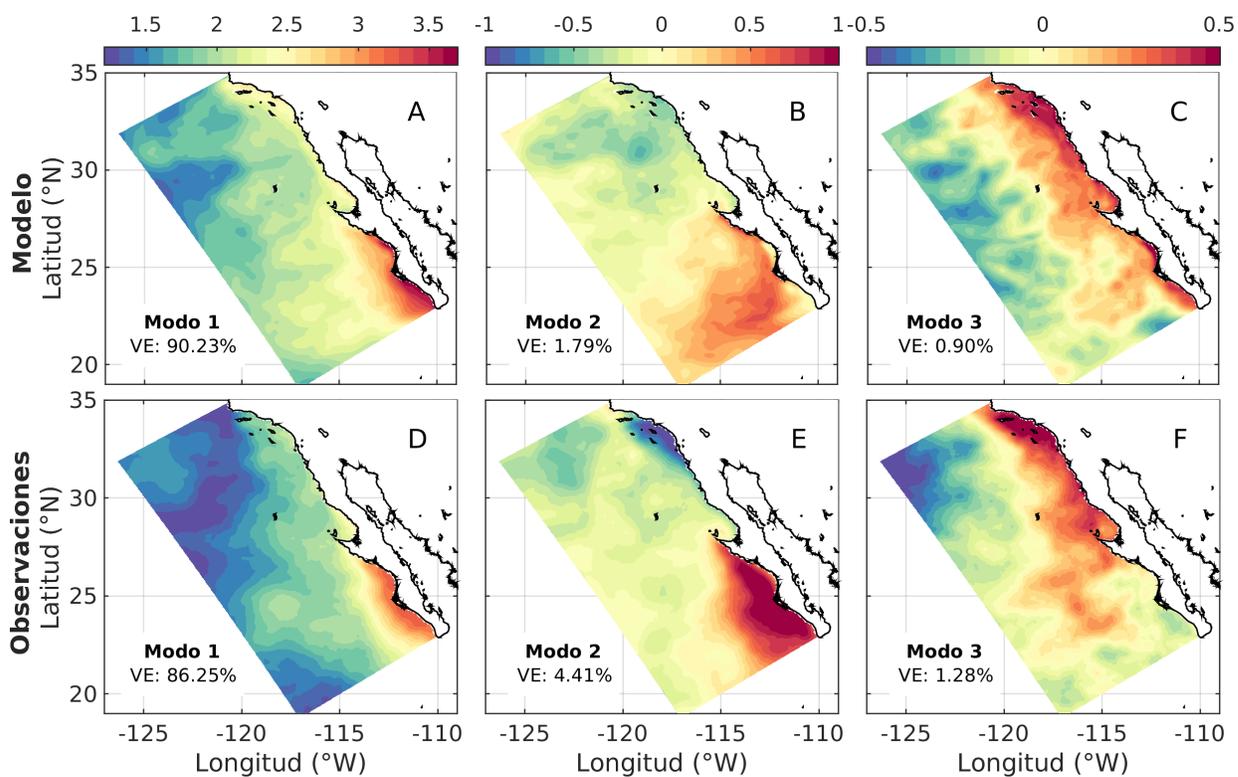


Figura 15. Modos dominantes del análisis de FEOs derivados de una simulación libre de 5 años (A, B y C) y de una serie larga de datos de SST del producto MUR (D, E, F). Se indica la Varianza Explicada (VE) de cada modo.

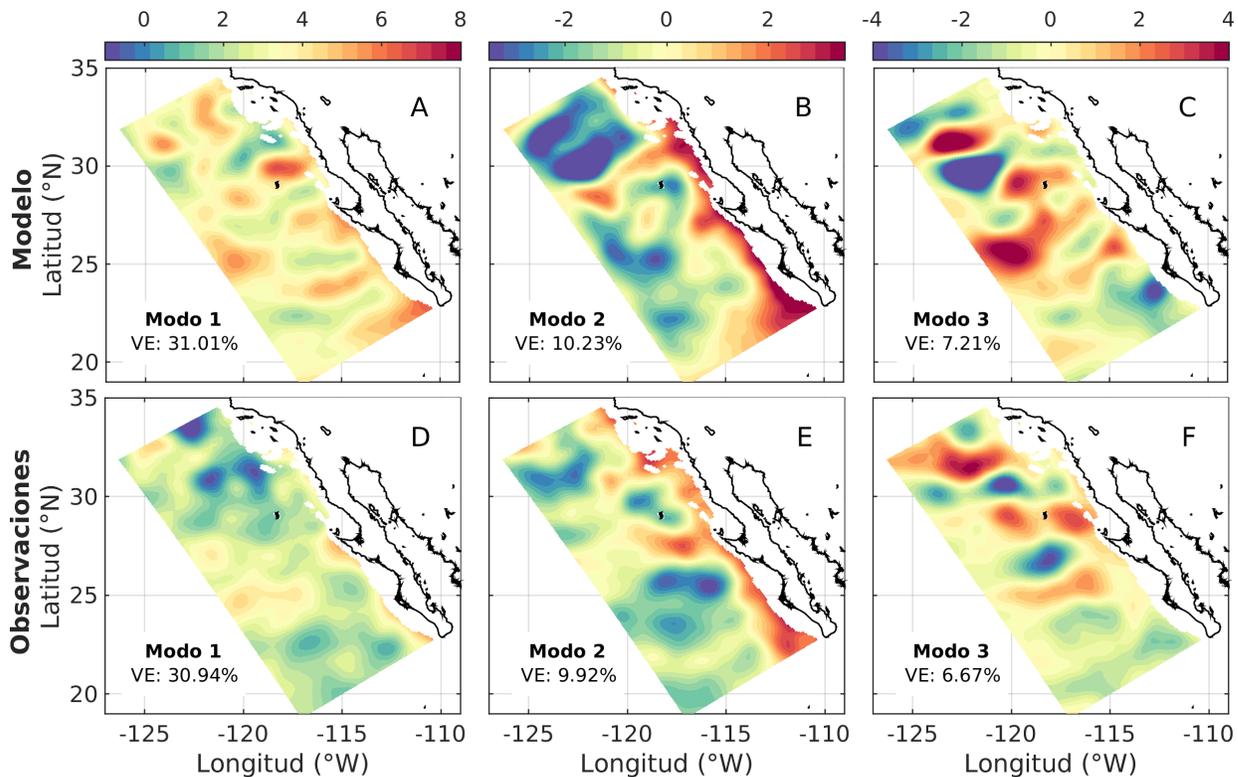


Figura 16. Modos dominantes del análisis de FEOs derivados de una simulación libre de 5 años (A, B y C) y de una serie larga de datos de SSH del altimetría (D, E, F). Se indica la Varianza Explicada (VE) de cada modo.

De forma análoga al análisis de los experimentos de relajación, se revisa el desempeño de las simulaciones del filtro SEEK con diferente número de modos examinando la evolución temporal de la RMS en todo el dominio superficial en las Figuras 17 y 18.

La evolución temporal de la RMS de SST para todos los experimentos presentó el mismo comportamiento previamente visto. A diferencia de los experimentos con relajación, el error de la SST en los experimentos con filtro SEEK es muy similar al de la simulación de control. Sin embargo, la RMSD muestra que los experimentos con mayor número de componentes tienen mejor desempeño respecto a la simulación de control. Esta mejoría es mayor durante los primeros meses del año.

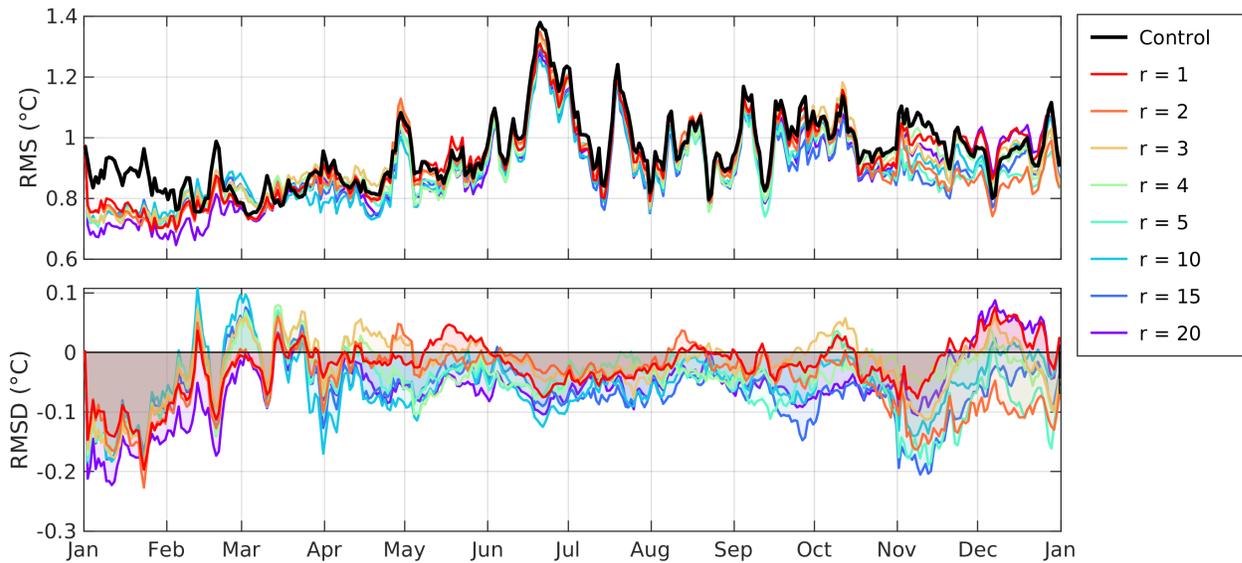


Figura 17. Evolución temporal de la RMS (panel superior) de SST de la simulación de control y de cada experimento del filtro SEEK con diferente número de modos (señalado como líneas continuas de color) comparadas contra observaciones de MUR de toda la región de estudio. Evolución temporal de la RMSD (panel inferior).

Por otro lado, la evolución temporal de la RMS de SSH (Figura 18) muestra mayores cambios entre las simulaciones con asimilación y la simulación de control. La RMSD muestra que durante los primeros meses del año (enero a mayo) el desempeño de las simulaciones con asimilación es mejor que la de control, y durante los siguientes meses (junio a enero), es similar. El mejor desempeño, al igual que en el caso de la SST es el de las simulaciones con mayor número de modos utilizados para aproximar la matriz de covarianza. En particular con la SSH, las simulaciones con menor número de modos ($r < 5$) exhiben peor desempeño respecto a la simulación de control durante los meses de junio a enero.

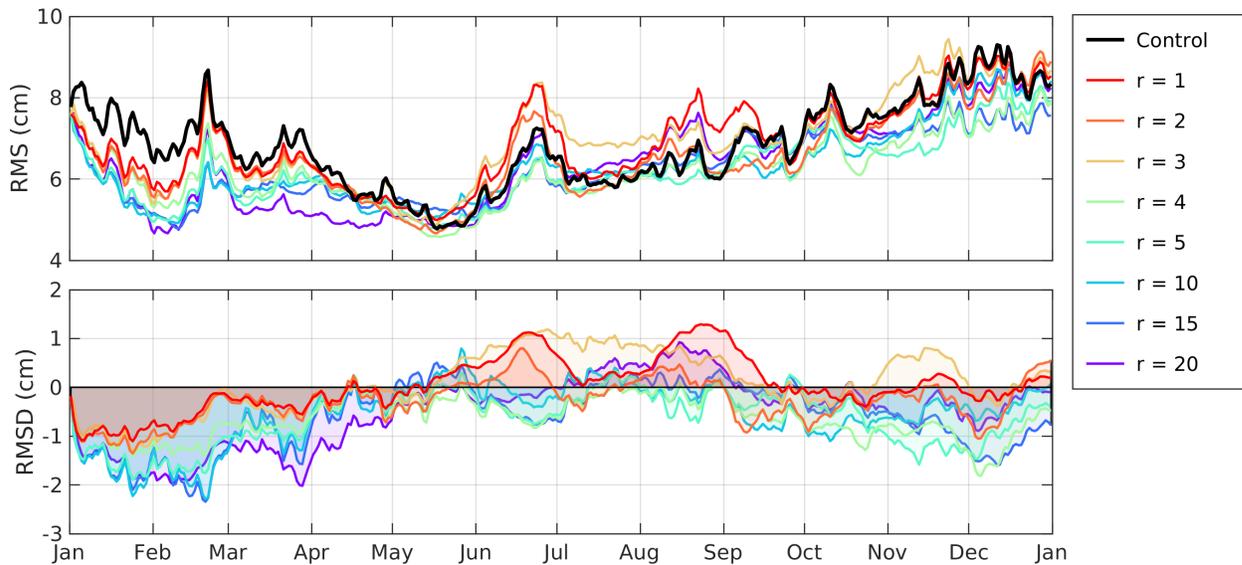


Figura 18. Evolución temporal de la RMS (panel superior) de SSH de la simulación de control y de cada experimento del filtro SEEK con diferente número de modos (señalado como líneas continuas de color) comparadas contra observaciones de toda la región de estudio. Evolución temporal de la RMSD (panel inferior).

Al igual que con el análisis de los experimentos con el esquema de relajación, se validan y comparan los experimentos con el filtro SEEK contra datos de temperatura y salinidad de la campaña de IMECOCAL.

El diagrama de la RMS de la temperatura (Figura 19B y 20B) a profundidad muestra el mismo patrón observado anteriormente (dos máximos a ~ 60 m, y otro ~ 220 m). Sin embargo, a diferencia de los experimentos con relajación, la asimilación no altera prácticamente nada el estado del océano por debajo de los 1,000 m. Esto se observa de forma más clara con la RMSD (Figura 19C), donde las simulaciones con asimilación presentan un mejor desempeño respecto al control en los primeros 500 m, y conforme aumenta la profundidad, se mantiene cerca de 0. Particularmente, la mejoría en los primeros 500 m es reducida en los 100 m, y hasta es invertida conforme se utilizan un menor número de modos. En el acercamiento a los primeros 1,000 m del perfil del sesgo (Figura 20) se observa más claro que la asimilación aumenta la temperatura en los primeros 500 m, teniendo de referencia la simulación de control.

En cuanto a la RMS de la salinidad a profundidad (Figura 21 y 22), el efecto de la asimilación no es tan notable. No obstante, los experimentos presentaron un mejor desempeño en los primeros 50 m, respecto a la simulación de control. Similar a la temperatura, el sesgo de los primeros 200 m (Figura 22A) muestra un aumento en la salinidad en las simulaciones con asimilación.

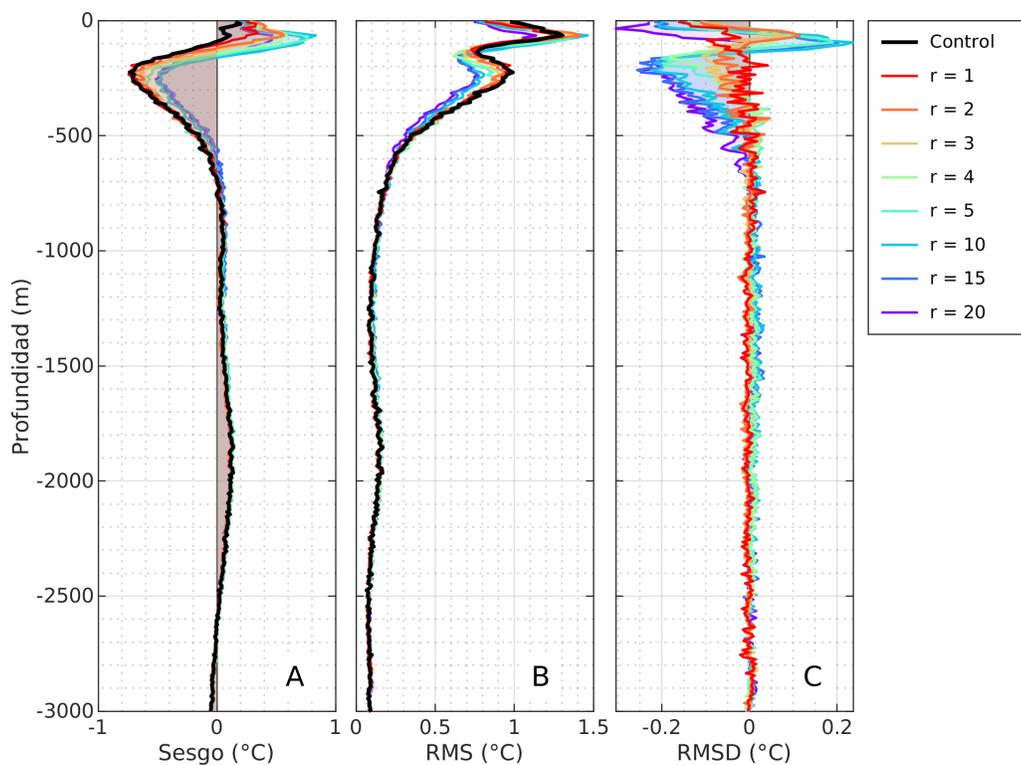


Figura 19. Perfiles de Sesgo (A), RMS (B) y RMSD (C) de temperatura de cada experimento del filtro SEEK con diferente número de modos respecto a datos de perfiles hidrográficos de IMECOCAL.

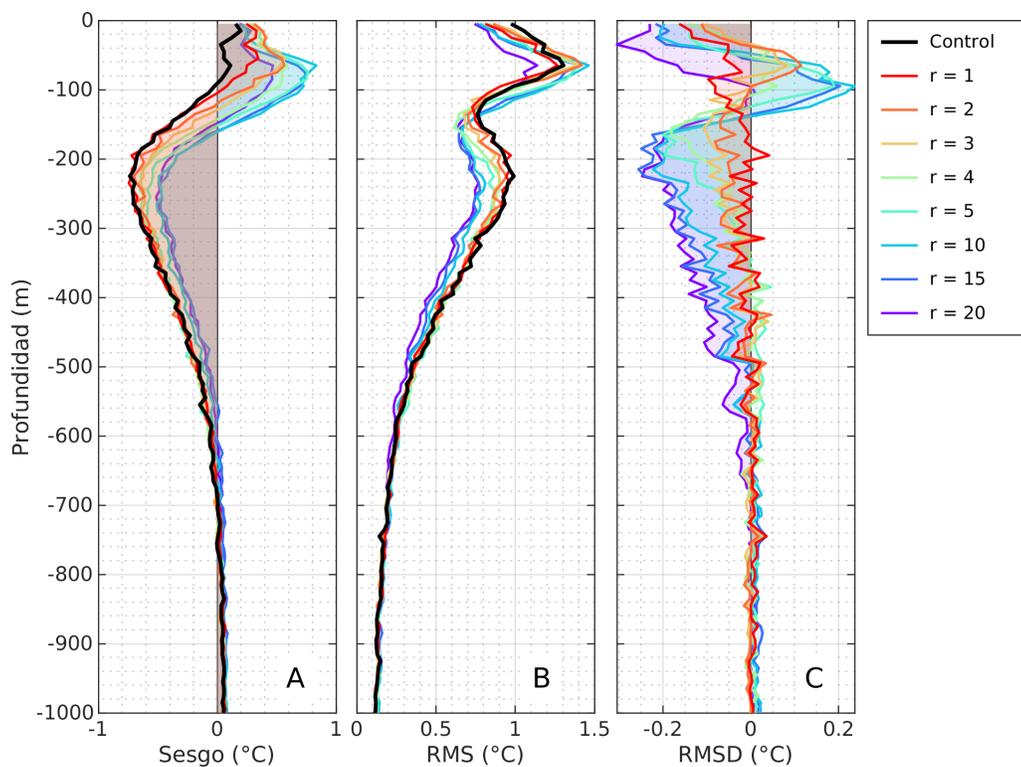


Figura 20. Acercamiento de los primeros 1,000 m de la Figura 19. Perfiles de Sesgo (A), RMS (B) y RMSD (C) de temperatura de cada experimento del filtro SEEK con diferente número de modos respecto a datos de perfiles hidrográficos de IMECOCAL.

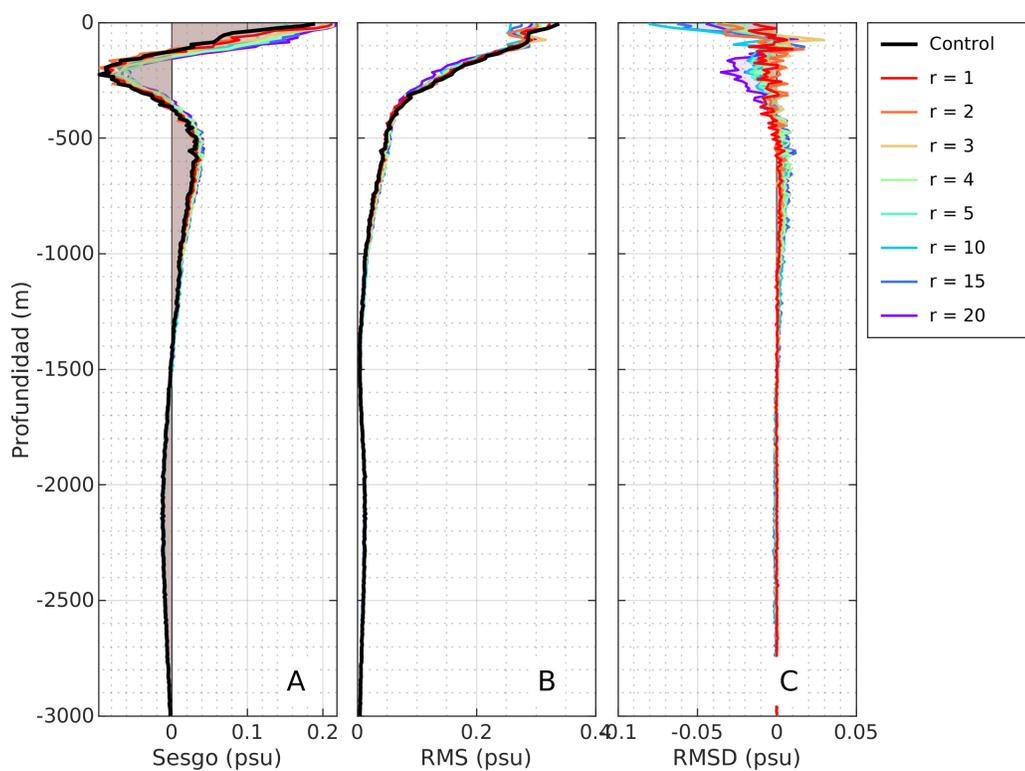


Figura 21. Perfiles de Sesgo (A), RMS (B) y RMSD (C) de salinidad de cada experimento del filtro SEEK con diferente número de modos respecto a datos de perfiles hidrográficos de IMECOCAL.

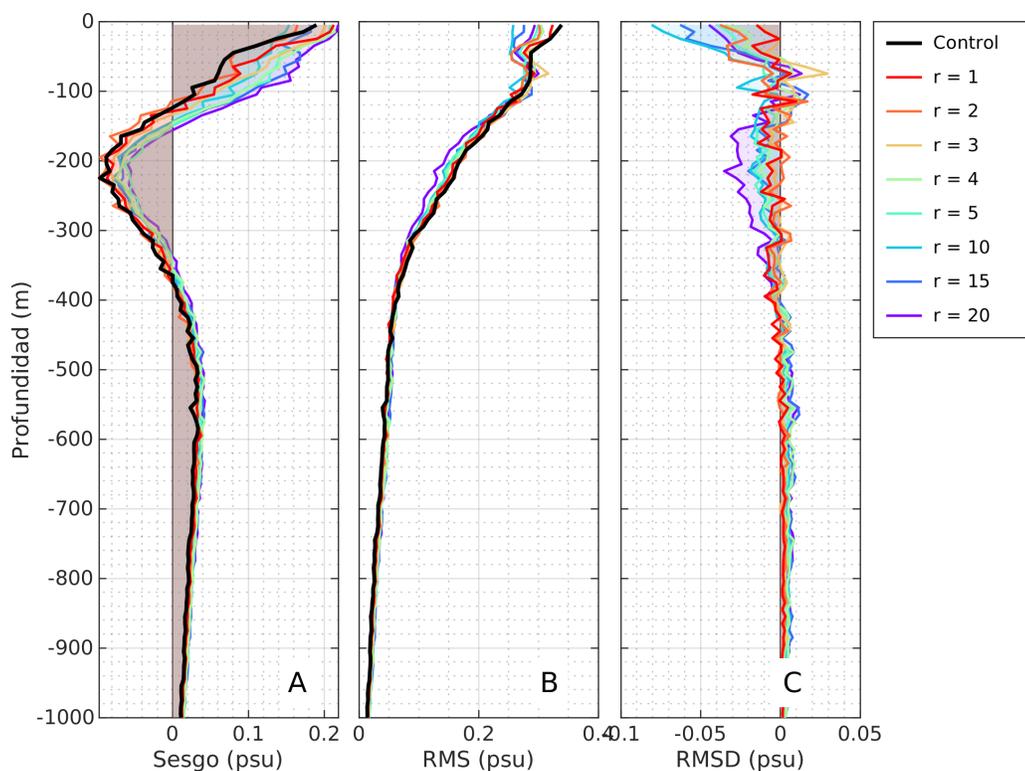


Figura 22. Acercamiento de los primeros 1,000 m de la Figura 21. Perfiles de Sesgo (A), RMS (B) y RMSD (C) de salinidad de cada experimento del filtro SEEK con diferente número de modos respecto a datos de perfiles hidrográficos de IMECOCAL.

Los diagramas de RMS de la Figura 23 muestran el resumen del comportamiento de los experimentos con asimilación. Conforme se utiliza un mayor número de modos, el error disminuye tanto en la superficie (SST y SSH), como en la vertical (temperatura y salinidad).

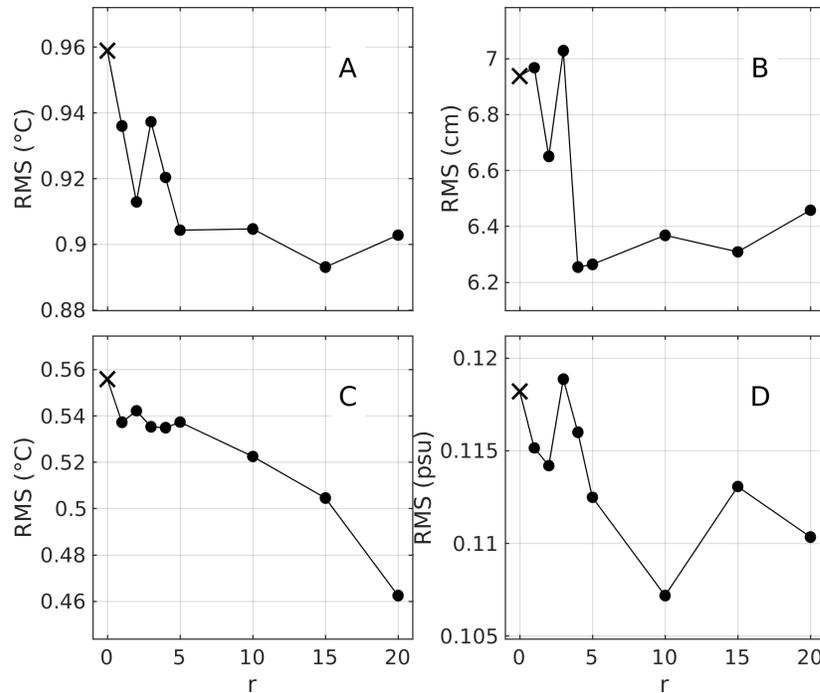


Figura 23. RMS de cada experimento del filtro SEEK con diferente número de modos (marcado como puntos negros) respecto a mediciones satelitales de SST (A), SSH (B), y mediciones hidrográficas de temperatura (C) y salinidad (D). Los valores de RMS correspondientes a la simulación de control son marcados como cruces negras.

3.4. Comparación entre experimentos

En esta sección se compara el desempeño de los experimentos con relajación, asimilación, y la simulación de control. Se utiliza como experimento de relajación la simulación con el coeficiente $\gamma = 2$ días, ya que por como se muestra en la Figura 14, se encuentra en el intervalo en el cual se acerca el modelo hacia las observaciones superficiales sin provocar un aumento del error en el resto de la columna de agua. Por otro lado, se selecciona como experimento de asimilación el experimento en donde se utilizan 20 número de modos para reconstruir la matriz de covarianza de menor rango, ya que como se observó, es la simulación con menor error.

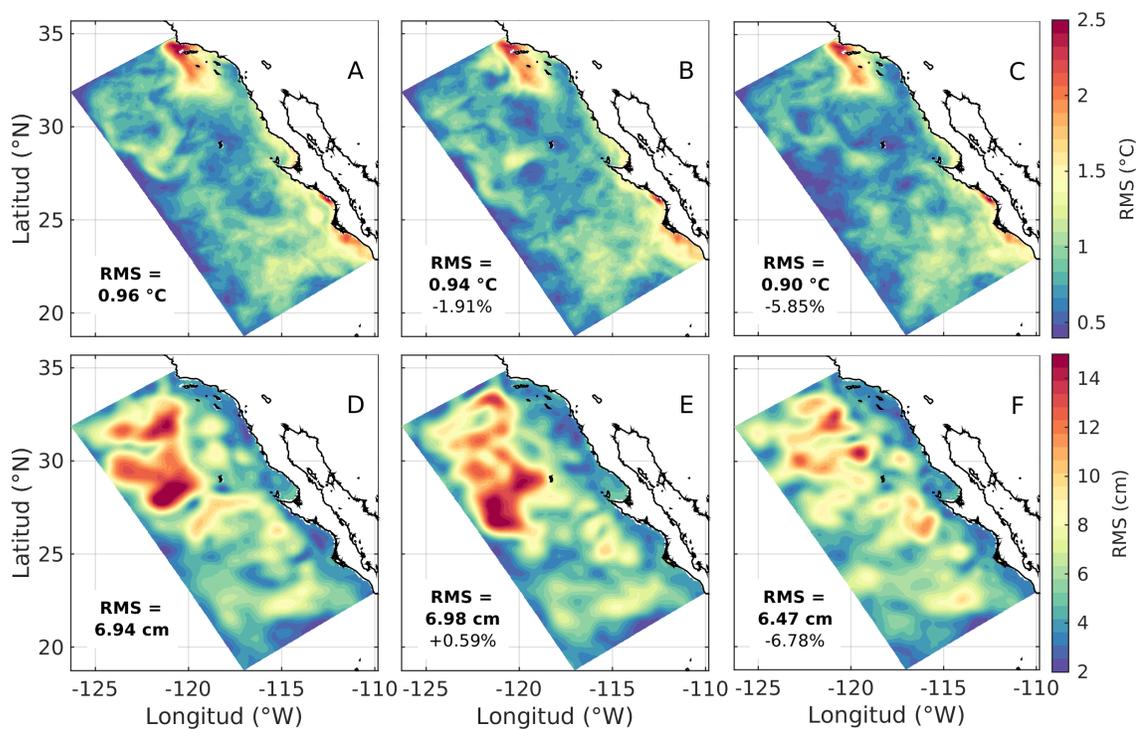


Figura 24. Distribución de la RMS respecto a observaciones (SST y SSH) de las simulaciones de control (A, D), esquema de relajación (B, E) y asimilación (C, F). El valor total de la RMS para cada simulación es mostrado en la esquina inferior izquierda de cada mapa, además del aumento o decremento de la RMS total relativa a la simulación de control.

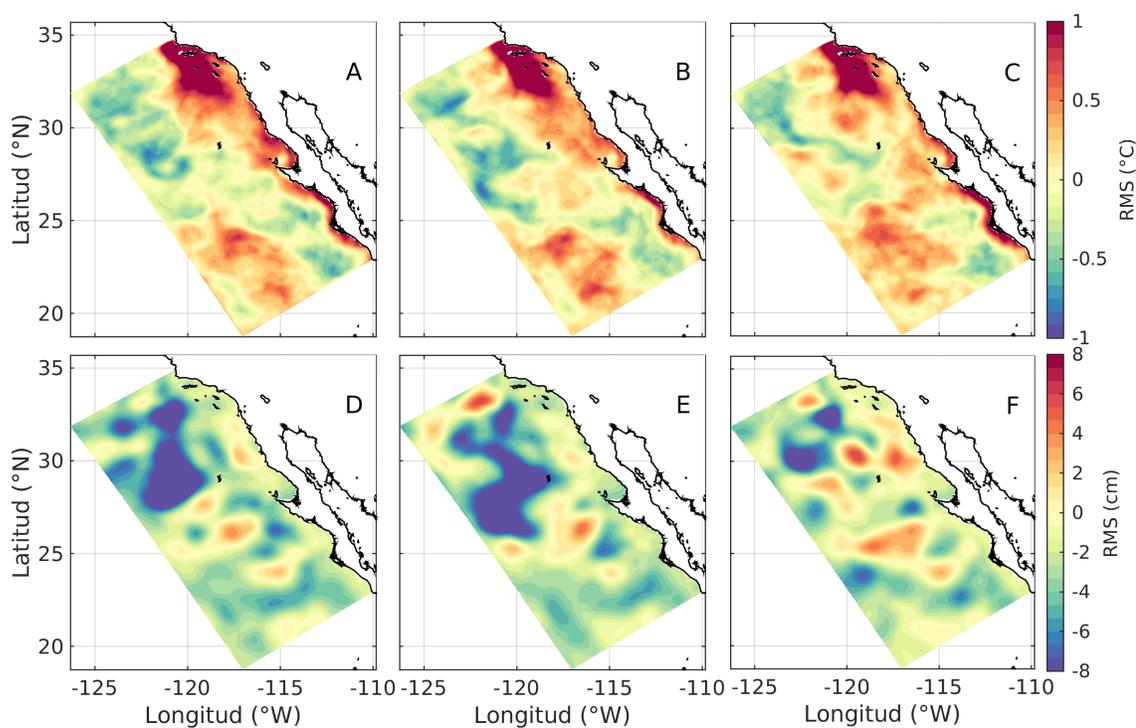


Figura 25. Distribución del Sesgo respecto a observaciones (SST y SSH) de las simulaciones de control (A, D), esquema de relajación (B, E) y asimilación (C, F).

La distribución espacial de la RMS de la temperatura (Figura 24A, B y C) muestra que para todas las simulaciones, el error se concentra al norte y sur del dominio, pegado a la costa donde se encuentran los máximos. Las simulaciones con relajación y asimilación presentan una disminución relativa (teniendo de referencia a la simulación de control) del 1.91 % y 5.85 % respectivamente. El sesgo de la temperatura (Figura 25A, B y C) muestra que los errores en la mayoría del dominio es debida a sobrestimación por parte del modelo, en particular en la región de las Islas del Canal al norte y al sur en el Golfo de Ulloa.

Respecto a la distribución espacial de la RMS de la SSH (Figura 24D, E y F), los mayores errores se localizan alejados de la costa y cercano a las fronteras del dominio. A diferencia del caso de la SST, el error aumentó 0.59 % en la simulación con relajación, mientras que disminuyó 6.78 % en la simulación con asimilación.

La series de la evolución temporal de la RMS (Figura 26) de las simulaciones de relajación y asimilación se mantienen relativamente cercanas a la de la simulación de control. La principal diferencia entre estas simulaciones es durante los primeros meses del periodo de estudio, donde la simulación con el filtro SEEK presenta un mejor desempeño. Por otro lado, la RMS de la SSH (Figura 18) exhibe una evolución temporal distinta a la de la SST. El desempeño de las simulaciones de relajación y asimilación fue mejor, relativo a la de control, durante los meses de enero a mayo al principio del periodo de estudio, para después aumentar el resto del año.

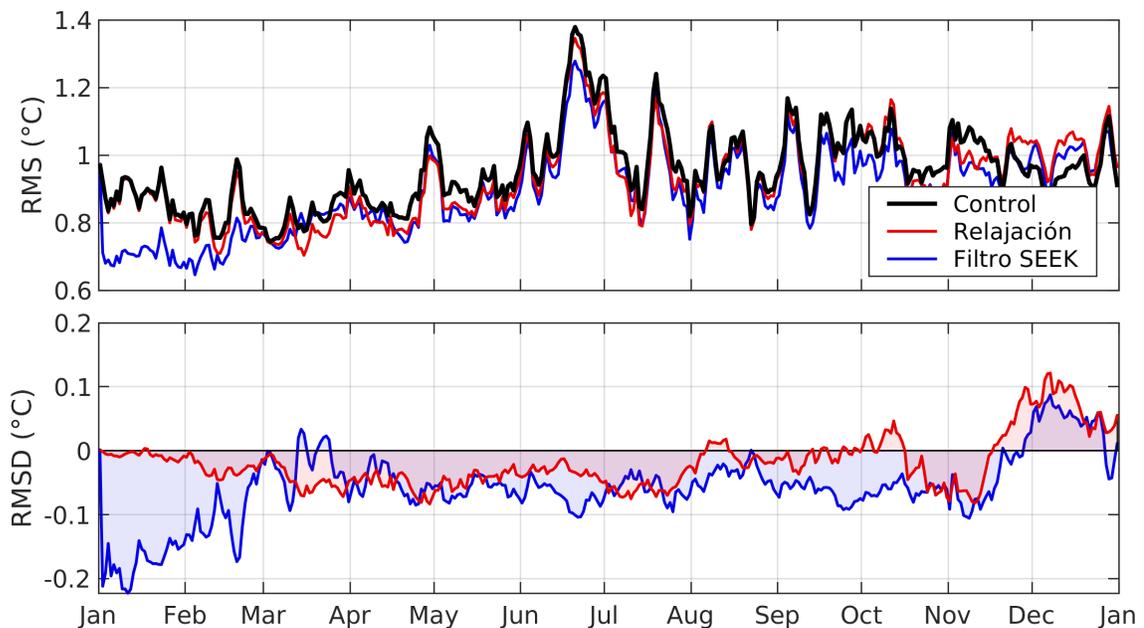


Figura 26. Evolución temporal de la RMS (panel superior) de SST de la simulación de control, esquema de relajación y asimilación (señalado como líneas continuas de color) comparadas contra observaciones de MUR de toda la región de estudio. Evolución temporal de la RMSD (panel inferior).

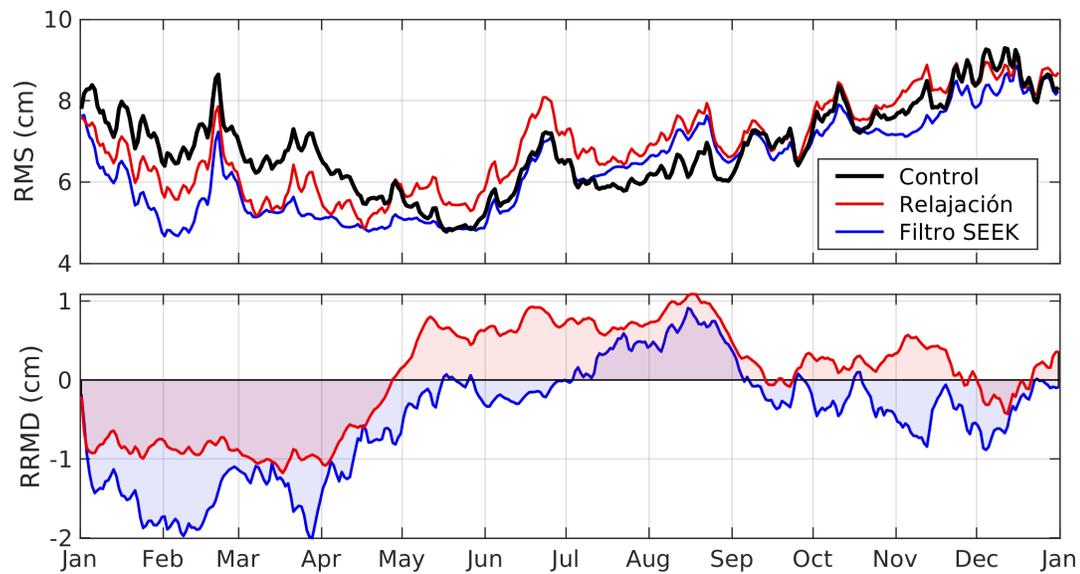


Figura 27. Evolución temporal de la RMS (panel superior) de SSH de la simulación de control, esquema de relajación y asimilación (señalado como líneas continuas de color) comparadas contra observaciones de toda la región de estudio. Evolución temporal de la RMSD (panel inferior).

La distribución en la vertical de la RMS de temperatura (Figura 28B) muestra que la simulación con asimilación es la que tiene mejor desempeño (RMSD) en toda la columna de agua. El sesgo de la simulación del filtro SEEK (Figura 28A) muestra un aumento de temperatura de los 50 a 500 m. En la simulación con relajación, se encuentra una mejoría en los primeros 50 m, no obstante, no es observado algún efecto en el resto de la vertical (relativo a la simulación de control).

Los efectos en la salinidad en la vertical son menores que los vistos en la temperatura. La RMS de la salinidad (Figura 29B) muestra que los errores son similares entre las simulaciones. De igual forma que en la temperatura, la simulación con relajación no muestra efectos significativos en la columna de agua mas que en los primeros 50-100 m. Por otro lado, el sesgo de la salinidad (Figura 29A) indica un aumento por parte de la simulación con el filtro SEEK.

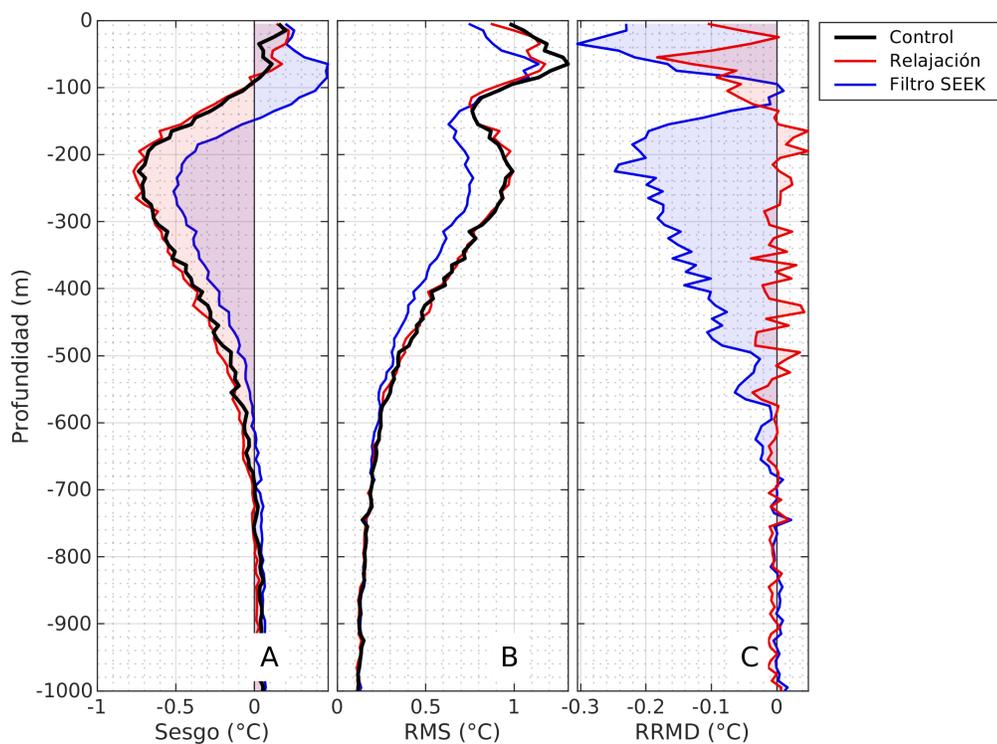


Figura 28. Perfiles de Sesgo (A), RMS (B) y RMSD (C) de temperatura de la simulación de control, esquema de relajación y asimilación respecto a datos de perfiles hidrográficos de IMECOCAL.

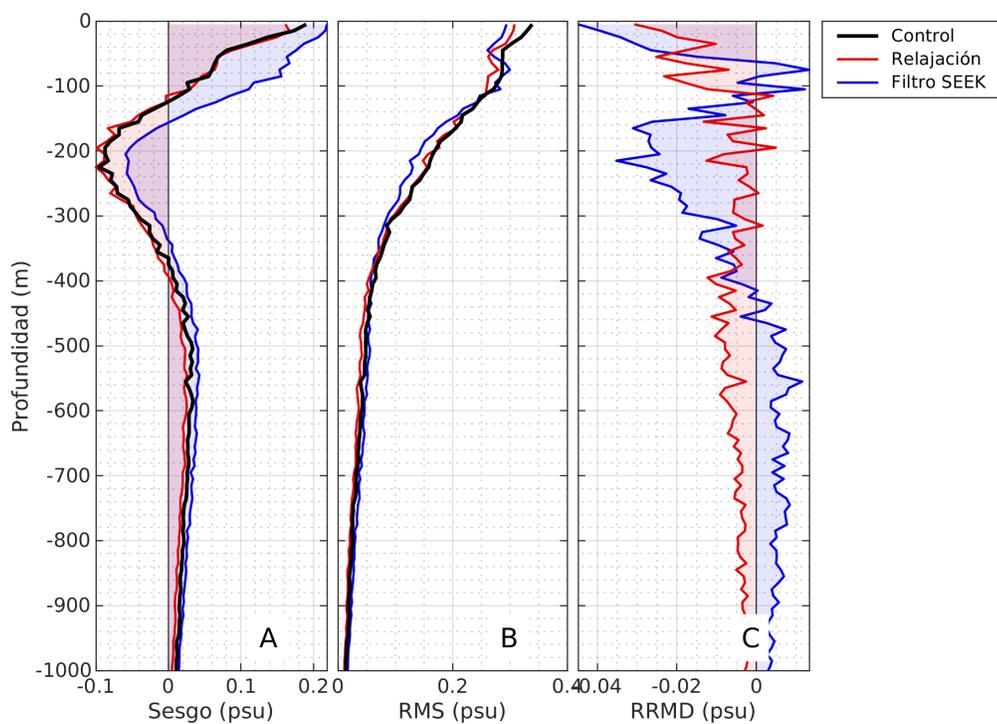


Figura 29. Perfiles de Sesgo (A), RMS (B) y RMSD (C) de salinidad de la simulación de control, esquema de relajación y asimilación respecto a datos de perfiles hidrográficos de IMECOCAL.

Capítulo 4. Discusión y conclusiones

En este trabajo se presentó el impacto que tienen los esquemas de relajación y asimilación tanto en superficie como a profundidad. Los experimentos utilizaron la misma configuración del modelo, forzamientos y condiciones de frontera idénticos y por lo tanto se logró examinar exclusivamente el rendimiento de los esquemas.

- Respecto a la simulación de control

El modelo libre presentó valores de RMS de SST y SSH de 0.96°C y 6.94 cm , y en la temperatura y salinidad vertical de 0.56°C y 0.12 psu respectivamente. Cabe resaltar que estos valores representan una buena estimación del estado termohalino del océano. Esto se logró mediante la configuración realista implementada en este trabajo, la cual incluye los forzamientos atmosféricos del reanálisis de ERA-Interim y las condiciones de frontera del reanálisis GLORYS.

En general, el modelo presenta dos errores máximos de temperatura en la vertical, uno de $\sim 1.35^{\circ}\text{C}$ a $\sim 60\text{ m}$, y otro de $\sim 1^{\circ}\text{C}$ a $\sim 220\text{ m}$. Además, se encuentra un sesgo máximo de salinidad de -0.1 psu a $\sim 200\text{ m}$. Estas profundidades corresponden a las posiciones de la CC y CCC respectivamente.

Latitudinalmente y mediante los diagramas T-S se observó que: el agua superficial (0 a 100 m) en el norte del dominio ($29\text{-}32^{\circ}\text{N}$) presentó mayor salinidad que la esperada; el agua en latitudes medias del dominio ($27\text{-}29^{\circ}\text{N}$) la salinidad es menor a la esperada; no se percibe de forma tan clara el máximo termohalino de la CCC en el modelo. La distribución espacial del error en la temperatura mostró máximos en la región cercana a la costa del norte y sur del dominio. Estas regiones corresponden a las Islas del Canal y al golfo de Ulloa respectivamente.

Todo esto sugiere que, en el norte, el modelo sobrestima las características de la ASA entre los 0 a 100 m. Esto podría ser por un aporte de aguas profundas de mayor salinidad a la superficie, o bien, que no se resuelve de forma adecuada la región topográficamente complicada de las Islas del Canal. En latitudes medias y al sur del modelo, los errores parecen ser localizados entre los 100 y 500 m, debidos a una subestimación de las características (salinidad y temperatura) de la AESs. Esto podría sugerir que el modelo está sobre estimando la extensión de la CC hacia latitudes más sureñas, o bien, que el modelo no refleja adecuadamente la entrada de la CCC hacia aguas del norte.

- Respecto al esquema de relajación

Se demostró que, como era de esperarse, entre menor sea el coeficiente de relajación que se utilizó, mayor semejanza se tendrá hacia las observaciones superficiales, aunque significativamente, con un costo en el aumento del error en el resto de la columna de agua. Esto era lo esperado. La ecuación 3 muestra el cálculo computacional que se utiliza para hacer la corrección en la trayectoria del modelo. Es claro entonces que mientras se utilizó un coeficiente de relajación (γ) cercano al paso del tiempo (Δt), $\tilde{\gamma}$ se aproxima a 1, y por lo tanto se descarta la predicción y el análisis es igual a la observación.

Particularmente, se encontró que el intervalo para el coeficiente de relajación entre 2 a 5 días parece ser el adecuado para acercar las estimaciones del modelo hacia las observaciones sin aumentar el error en la vertical. Cabe resaltar que este intervalo surgió debido a la configuración del modelo utilizada en este trabajo, particularmente la selección del operador W . En esta implementación el operador W ocasiona que solo se relaje hacia observaciones superficiales. Si se escogiera, por ejemplo, un W que permita que se relaje hacia observaciones en todo el dominio, la influencia de la relajación sería mucho mayor y por respuesta se modificarían los límites óptimos del coeficiente.

Otra observación interesante es que las simulaciones con coeficientes grandes tienen peor desempeño relativo a la simulación de control para este periodo de estudio. Este efecto es más notable en la SSH durante los meses de mayo a enero, pero se alcanza también a apreciar en menor medida en la SST durante los meses de septiembre a enero. En teoría, al utilizar un coeficiente de relajación γ grande (orden de meses a años), las correcciones que se realizan en el análisis son pequeñas y por ende la simulación tendría que ser similar al control. Si bien estas correcciones son pequeñas, son además constantes para cada paso del tiempo, por lo que estas ligeras correcciones a lo largo de un periodo largo de tiempo se podrían acumular para modificar negativamente el estado del océano.

- Respecto al esquema de asimilación

Este trabajo ha demostrado que la asimilación secuencial de datos de SST y SSH en el modelo CROCO mejora la estimación de las propiedades termohalinas del océano, no solo en las variables superficiales asimiladas, si no también en profundidad.

Comparado contra la simulación libre de asimilación y la simulación con relajación ($\gamma = 2$ días), la simulación con el filtro SEEK utilizando 20 modos tiene mejor desempeño. La RMS total de la SST y SSH disminuyó un 5.85 % y 6.78 % respectivamente. Como contraste, en el caso de la simulación con relajación, disminuyó un 1.91 % y aumentó un 0.59 %. El mejor desempeño de la simulación con

asimilación se mantuvo, y aumentó, respecto a los datos independientes de lances hidrográficos de IMECOCAL, donde la RMS total de la temperatura y salinidad disminuyó un 16.80% y 6.64%. Como contraste, en el caso de la simulación con relajación, disminuyó un 2.33% y 4.89%.

Una de las grandes ventajas examinadas del filtro SEEK fue la habilidad de extrapolar la información de las observaciones superficiales a la vertical mediante la asimilación. Esto se observó como una mejora gradual desde la superficie hasta una profundidad de 500 m, sin modificar el resto de la columna de agua donde el error es menor por el mismo sistema (modelo). No obstante, se observó una disminución en la mejora en las profundidades de 50 a 150 m. Esta es la posición promedio de la termoclina estacional, y es donde los gradientes verticales de temperatura y salinidad son mayores. Cualquier desajuste en la posición vertical de la termoclina puede dar lugar a grandes diferencias de temperatura y/o salinidad. El sesgo muestra que durante la asimilación, la temperatura y salinidad aumentan en los primeros 500 m. Esto puede aumentar la capa de mezcla, y con ello hundir la termoclina y alterar las estimaciones a esta profundidad.

La salinidad no parece ser modificada en gran medida por la asimilación, o por lo menos no en la misma magnitud que en la temperatura. Esto es de esperarse ya que la salinidad no se utiliza en el vector de estado en este trabajo, por lo tanto no es corregida en el paso del análisis del algoritmo y su evolución esta sujeta a la dinámica del modelo después de haber sido corregido.

Adicionalmente se encontró que conforme se utiliza un mayor número de modos, el error disminuye tanto en la superficie (SST y SSH), como en la vertical (temperatura y salinidad). Esto es de gran importancia ya que el número de modos dominantes define que tan bien se aproxima la matriz de covarianza inicial y con ello la evolución del filtro. Respecto a esta matriz, los modos dominantes mostrados en las Figuras 15 y 16 muestran que, por lo menos, el patrón espacial es bien retratado. Sin embargo, se observó también que la mayor parte del tiempo en donde el esquema de asimilación presenta una mejoría, se concentra en los primeros meses del periodo de estudio. En este trabajo, se utilizó la base fija para parametrizar la matriz de covarianza. Usando esta base se simplifica el cálculo computacional, pero a cambio de no evolucionar la covarianza del error con la dinámica del modelo. Esto puede provocar que la primera aproximación de la matriz de covarianza del error no sea suficiente para modificar el estado del océano en los meses de mayo a septiembre, donde el filtro parece disminuir su eficiencia. Por lo tanto, consideramos importante investigar el uso de otros esquemas de parametrización de la matriz de covarianza así como el número óptimo de modos a utilizar en la reducción de orden.

Literatura citada

- Brasseur, P., Baharel, P., Bertino, L., Birol, F., Brankart, J.M., Ferry, N., Losa, S., Remy, E., Schroeter, J., Skachko, S., Testut, C.E., Tranchat, B., Van Leeuwen P., Verron J. 2005. Data assimilation for marine monitoring and prediction: The Mercator operational assimilation systems and the Mersea developments. *Quarterly Journal of the Royal Meteorological Society*, 131, 3561-3582.
- Brasseur P. 2006. Ocean Data Assimilation Using Sequential Methods Based on the Kalman Filter. *Ocean Weather Forecasting*, 271-316.
- Brasseur, P., Verron, J. 2006. The SEEK filter method for data assimilation in oceanography: a synthesis. *Ocean Dynamics*, 56(5-6), 650-661.
- Broquet, G., Edwards, C. A., Moore, A. M., Powell, B. S., Veneziani, M., Doyle, J. D. 2009. Application of 4D-Variational data assimilation to the California Current System. *Dynamics of Atmospheres and Oceans*, 48(1-3), 69-92.
- Brusdal, K., Brankart, J. M., Halberstadt, G., Evensen, G., Brasseur, P., van Leeuwen, P. J., Dombrowsky, E., y Verron, J. 2003. A demonstration of ensemble-based assimilation methods with a layered OGCM from the perspective of operational ocean forecasting systems. *Journal of Marine Systems*, 40-41, 253-289.
- Carrassi, A., Bocquet, M., Bertino, L., y Evensen, G. 2018. Data assimilation in the geosciences: An overview of methods, issues, and perspectives. *Wiley Interdisciplinary Reviews: Climate Change*, 9(5), e535.
- Chao, Y., Farrara, J.D., Zhang, H., Armenta, K.J., Centurioni, L., Chavez, F., Girton, J.B., Rudnick, D., y Walter, R.K. 2017. Development, implementation, and validation of a California coastal ocean modeling, data assimilation, and forecasting system. *Deep Sea Research Part II: Topical Studies in Oceanography*.
- Chin, T., M., Vazquez-Cuervo, J., y Armstrong, E.M. 2017. A multi-scale high-resolution analysis of global sea surface temperature. *Remote Sensing of Environment* 200 (2017) 154-169
- Debreu, L., Marchesiello, P., Penven, P., y Cambon, G. 2012. Two-way nesting in split-explicit ocean models: Algorithms, implementation and validation. *Ocean Modelling* 49-50, 1-21.
- Durazo, R. 2015. Seasonality of the transitional region of the California Current System off Baja California. *Journal of Geophysical Research: Oceans*, 120(2), 1173-1196.
- Ghil, M., Malanotte-Rizzoli, P. 1991. Data Assimilation in Meteorology and Oceanography. *Advances in Geophysics*, 141-266.
- Gómez-Valdivia, F., Parés-Sierra, A., y Flores-Morales, A.L. 2015. The Mexican Coastal Current: A sub-surface seasonal bridge that connects the tropical and subtropical Northeastern Pacific. *Continental Shelf Research*, 110, 100-107.
- Gómez-Valdivia, F., Parés-Sierra, A., y Flores-Morales, A.L. 2017. Semiannual variability of the California Undercurrent along the Southern California Current System: A tropical generated phenomenon. *Journal of Geophysical Research: Oceans*, 122(2), 1574-1589.

- Hofmann, E.E., Friederichs, y M.A.M. 2001. Biogeochemical Data Assimilation. *Encyclopedia of Ocean Sciences (Second Edition)*, 364-370.
- Nencioli, F., Dall'Olmo, G., y Quartly, G. D. 2018. Agulhas ring transport efficiency from combined satellite altimetry and Argo profiles. *Journal of Geophysical Research: Oceans*, 123: 5874– 5888.
- Pantoja, D.A., Marinone, S.G., Parés-Sierra, A., Gómez-Valdivia, F. 2012. Numerical modeling of seasonal and mesoscale hydrography and circulation in the Mexican Central Pacific. *Ciencias Marinas*, 38(2), 363-379.
- Penduff, T., Brasseur, P., Testut, C.E., Barnier, B., Verron, J. 2002. Assimilation of sea-surface temperature and altimetric data in the South Atlantic Ocean: impact on basin-scale properties. *Journal of Marine Research* 60:805–833
- Shchepetkin, A. F., and J. C. McWilliams. 2005. The Regional Ocean Modeling System (ROMS): A split-explicit, free-surface, topography-following-coordinate oceanic model, *Ocean Modell.*, 9(4), 347–404,
- Soto-Mardones, L., Parés-Sierra, A., Garcia, J., Durazo, R., Hormazabal, S. 2004. Analysis of the mesoscale structure in the IMECOCAL region (off Baja California) from hydrographic, ADCP and altimetry data. *Deep Sea Research Part II: Topical Studies in Oceanography*, 51(6-9), 785-798.
- Testut, C.E., Brasseur, P., Brankart, J.M., Verron, J. 2003. Assimilation of sea-surface temperature and altimetric observations during 1992–1993 into an eddy permitting primitive equation model of the North Atlantic Ocean. *Journal of Marine Systems*, 40-41, 291–316.
- Pham, T.D., Verron, J., Christine Roubaud, M. 1998. A singular evolutive extended Kalman filter for data assimilation in oceanography. *Journal of Marine Systems*, 16(3-4), 323-340.
- Todd, R. E., Rudnick, D. L., Mazloff, M. R., Cornuelle, B. D., y Davis, R. E. 2012. Thermohaline structure in the California Current System: Observations and modeling of spice variance, *Journal of Geophysical Research*, 117: C02008. doi:10.1029/2011JC007589.
- Zaytsev, O., Cervantes-Duarte, R., Montante, O., Gallegos-Garcia, A. 2003. Coastal Upwelling Activity on the Pacific Shelf of the Baja California Peninsula. *Journal of Oceanography*, 59(4), 489-502.