



**INSTITUTO POLITECNICO NACIONAL**  
**CENTRO INTERDISCIPLINARIO DE CIENCIAS MARINAS**  
**Departamento de Plancton y Ecología Marina**



**RECONSTRUCCIÓN Y MODELACIÓN CLIMÁTICA A PARTIR DEL USO DE  
DIATOMEAS PLANCTÓNICAS**

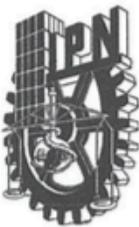
**TESIS**

Para Obtener el Grado de Doctor en Ciencias Marinas

**PRESENTA:**

**M.C. JUAN DAVID ACEVEDO ACOSTA**

La Paz, B.C.S. México; diciembre 2021



# INSTITUTO POLITÉCNICO NACIONAL

## SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

### ACTA DE REVISIÓN DE TESIS

En la Ciudad de **La Paz, B.C.S.,** siendo las **12:00** horas del día **03** del mes de **Noviembre** del **2021** se reunieron los miembros de la Comisión Revisora de la Tesis, designada por el Colegio de Profesores de Posgrado de: **CENTRO INTERDISCIPLINARIO DE CIENCIAS MARINAS** para examinar la tesis titulada: **“RECONSTRUCCIÓN Y MODELACIÓN CLIMÁTICA A PARTIR DEL USO DE DIATOMEAS PLANCTÓNICAS”** del (la) alumno (a):

Apellido Paterno:	ACEVEDO	Apellido Materno:	ACOSTA	Nombre (s):	JUAN DAVID
-------------------	---------	-------------------	--------	-------------	------------

Número de registro: **B 1 7 0 7 7 5**

Aspirante del Programa Académico de Posgrado: **DOCTORADO EN CIENCIAS MARINAS**

Una vez que se realizó un análisis de similitud de texto, utilizando el software antiplagio, se encontró que el trabajo de tesis tiene **3** % de similitud. **Se adjunta reporte de software utilizado.**

Después que esta Comisión revisó exhaustivamente el contenido, estructura, intención y ubicación de los textos de la tesis identificados como coincidentes con otros documentos, concluyó que en el presente trabajo **SI**  **NO**  **SE CONSTITUYE UN POSIBLE PLAGIO.**

#### JUSTIFICACIÓN DE LA CONCLUSIÓN:

~~El % de similitud se localiza en metodologías adecuadamente referidas a fuente original~~

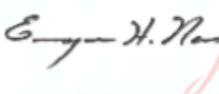
**\*\*Es responsabilidad del alumno como autor de la tesis la verificación antiplagio, y del Director o Directores de tesis el análisis del % de similitud para establecer el riesgo o la existencia de un posible plagio.**

Finalmente y posterior a la lectura, revisión individual, así como el análisis e intercambio de opiniones, los miembros de la Comisión manifestaron **APROBAR**  **SUSPENDER**  **NO APROBAR**  la tesis por **UNANIMIDAD**  o **MAYORÍA**  en virtud de los motivos siguientes:

**“SATISFACE LOS REQUISITOS SEÑALADOS POR LAS DISPOSICIONES REGLAMENTARIAS VIGENTES”**

#### COMISIÓN REVISORA DE TESIS

  
DRA. AÍDA MARTÍNEZ LÓPEZ  
Director de Tesis  
Nombre completo y firma

  
DR. ENRIQUE HIPARCO NAVA SÁNCHEZ  
Nombre completo y firma

Firmado digitalmente por Enrique H. Nava Sanchez  
Nombre de reconocimiento (DN):  
cn=Enrique H. Nava Sanchez, o=Instituto Politécnico Nacional, ou=CICIMAR, email=enava@ipn.mx, c=MX  
Fecha: 2021.11.04 18:56:02 -07'00'

  
DR. GUILLERMO MARTÍNEZ FLORES  
Nombre completo y firma

  
DRA. MIRTHA ORALIA ALBÁÑEZ LUCERO  
Nombre completo y firma

  
DR. TOMAS MORALES ALCOLTZI  
Nombre completo y firma

DR. SERGIO HERNÁNDEZ TRUJILLO  
Nombre completo y firma  
**PRESIDENTE DEL COLEGIO DE PROFESORES**



**INSTITUTO POLITÉCNICO NACIONAL  
SECRETARÍA DE INVESTIGACIÓN Y POSGRADO**

**CARTA CESIÓN DE DERECHOS**

En la Ciudad de La Paz, B.C.S., el día 17 del mes de Noviembre del año 2021

El (la) que suscribe M en C. JUAN DAVID ACEVEDO ACOSTA Alumno (a) del Programa

DOCTORADO EN CIENCIAS MARINAS

con número de registro B170775 adscrito al CENTRO INTERDISCIPLINARIO DE CIENCIAS MARINAS

manifiesta que es autor(a) intelectual del presente trabajo de tesis, bajo la dirección de:

DRA. AÍDA MARTÍNEZ LÓPEZ

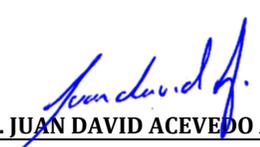
y cede los derechos del trabajo titulado:

“RECONSTRUCCIÓN Y MODELACIÓN CLIMÁTICA A PARTIR DEL USO  
DE DIATOMEAS PLANCTÓNICAS”

al Instituto Politécnico Nacional, para su difusión con fines académicos y de investigación.

Los usuarios de la información no deben reproducir el contenido textual, gráficas o datos del trabajo sin el permiso expreso del autor y/o director del trabajo. Éste, puede ser obtenido escribiendo a la siguiente dirección: jacevedoa1300@alumno.ipn.mx - amartin@ipn.mx -

Si el permiso se otorga, el usuario deberá dar el agradecimiento correspondiente y citar la fuente del mismo.

  
M en C. JUAN DAVID ACEVEDO ACOSTA

*Nombre y firma del alumno*



Acevedo-Acosta Tesis\_doctorado.pdf  
3 nov 2021  
32252 palabras/168474 caracteres

Mario Alberto Castillo Soria

## Acevedo-Acosta Tesis\_doctorado.pdf

### Resumen de fuentes

3%

SIMILITUD GENERAL

1	Juan David Acevedo-Acosta, Aída Martínez-López, Tomás Morales-Acoltzi, Mirtha Albáñez-Lucero, Gerardo Verdugo-Díaz. "Self-organization maps (SOM) in the definition of a "transfer f...	<1%
	CROSSREF	
2	tesis.ipn.mx	<1%
	INTERNET	
3	www.cienciasmarinas.com.mx	<1%
	INTERNET	
4	www.repositoriodigital.ipn.mx	<1%
	INTERNET	
5	www.elfuturodelpasado.com	<1%
	INTERNET	
6	repositoriodigital.ipn.mx	<1%
	INTERNET	
7	www.coursehero.com	<1%
	INTERNET	
8	cal.ucf.edu	<1%
	INTERNET	
9	es.scribd.com	<1%
	INTERNET	
10	repositorio.lamolina.edu.pe	<1%
	INTERNET	
11	vbook.pub	<1%
	INTERNET	
12	glyumiel2009.blogspot.com	<1%
	INTERNET	
13	www.researchgate.net	<1%
	INTERNET	
14	www.tandfonline.com	<1%
	INTERNET	
15	www.agbjarn.blog.is	<1%
	INTERNET	
16	portal.ucoj.mx	<1%
	INTERNET	
17	repository.unad.edu.co	<1%
	INTERNET	
18	www.ncbi.nlm.nih.gov	<1%
	INTERNET	
19	www.biblioteca.cicimar.ipn.mx	<1%
	INTERNET	
20	Martín E De La Cruz-Orozco, Eliana Gómez-Ocampo, Luis E Miranda-Bojórquez, Jushiro Cepeda-Morales et al. "Phytoplankton biomass and production off the Baja California Peninsula: ...	<1%
	CROSSREF	
21	hdl.handle.net	<1%
	INTERNET	
22	www.cec.uchile.cl	<1%
	INTERNET	
23	www.oaxaca.gob.mx	<1%
	INTERNET	
24	JUAN LUIS GÁMEZ MARTÍNEZ. "ANÁLISIS DE LA DEFORMACIÓN EN LA INYECCIÓN DE TERMOPLÁSTICOS BAJO VARIABLES DE FORMA DE LA PIEZA MEDIANTE RED NEURONAL Y S...	<1%
	INTERNET	

25	danyalejoana.blogspot.com	INTERNET	<1%
26	delfin.cicimar.ipn.mx	INTERNET	<1%
27	uvadoc.uva.es	INTERNET	<1%
28	verduras.consumer.es	INTERNET	<1%
29	www.darlingii.com	INTERNET	<1%
30	www.redalyc.org	INTERNET	<1%
31	www.remexmarbc.org	INTERNET	<1%
32	www.scielo.org.mx	INTERNET	<1%

**Se excluyeron los depósitos de búsqueda:**

Ninguno

**Excluido del Informe de Similitud:**

Bibliografía

Citas textuales

Citas

Coincidencias menores (10 palabras o menos)

**Se excluyeron las fuentes:**

Ninguno

El presente trabajo está dedicado a mi familia por haber sido mi motor y apoyo a lo largo de toda mi carrera profesional y vida.

A todas las personas especiales que me acompañaron en esta etapa, aportando a mi formación tanto profesional y como ser humano.

## AGRADECIMIENTOS

El desarrollo y culminación de esta tesis ha sido un esfuerzo en conjunto de múltiples actores que de una u otra manera aportaron a su elaboración. De mi parte no queda más que expresarles mi más sincera gratitud.

Esta tesis no habría sido posible sin el apoyo y asesoría de la Dra. Aída Martínez López, quien durante todo este tiempo me ha brindado su amistad, conocimiento e inspirado a crecer tanto en el ámbito profesional, como personal.

A mis sinodales la Dra. Mirtha Oralia Albañez Lucero, Dr. Enrique Hiparco Nava Sánchez, Dr. Guillermo Martínez Flores y Dr. Tomás Morales Acoltzi, por brindarme su apoyo y asesoría durante el transcurso del doctorado, así como también en las indicaciones e ideas brindadas que permitieron mejorar la tesis y llegar a la culminación de la misma. Del mismo modo, a partir de los diversos aportes, exigencias y gusto por la investigación, permitieron alcanzar un grado de excelencia de la tesis.

Al Dr. Tomás Morales Acoltzi y Dr. Leopoldo Carro Calvo por brindarme la oportunidad de realizar una estancia de investigación bajo su asesoría. En la cual me brindaron la oportunidad de trabajar bajo su asesoría y enriquecer mi conocimiento sobre diferentes campos de la ciencia climática y la modelación, así como experimentar el gusto por la enseñanza y la investigación.

A la Bio. Rocío Lara Rosales, M.c. Mauricio Beltrán y Cristian Hakspiel, por compartir ratos amenos en el grupo de investigación, así como también por su amistad, apoyo y sabiduría al momento de brindarme una voz de aliento y consejos en diferentes campos de la vida, gracias por compartir el gusto por la ciencia y en especial por el maravilloso mundo del fitoplancton.

A Nicole, Dianita, Mauricio, Carito, Lalo, a todos mis amigos y a mi familia por todo su apoyo, amistad, cariño y gratos momentos, y sobre todo por ser parte importante en mi vida.

Finalmente quiero agradecer al Instituto Politécnico Nacional, CONACyT, y sobre todo a CICIMAR por brindarme la oportunidad de estudiar el doctorado bajo la dirección y acompañamiento de tan excelentes investigadores.

## Contenido

GLOSARIO .....	10
RESUMEN .....	12
ABSTRACT .....	13
I. INTRODUCCIÓN .....	14
II. MODELO TEÓRICO DEL FUNCIONAMIENTO DE UNA RED NEURONAL..	17
III. HIPÓTESIS .....	22
IV. OBJETIVO GENERAL .....	22
V. OBJETIVOS PARTICULARES .....	22
VI. METODOLOGÍA.....	23
• <i>Área de estudio</i> .....	23
• <i>Muestras biológicas</i> .....	24
• <i>Datos instrumentales</i> .....	25
• <i>Procesamiento y análisis de las series de tiempo para la obtención de las funciones de transferencia y reconstrucción del clima</i> .....	26
Estandarización de datos biológicos y selección de especies .....	26
Evaluación de parámetros no lineales y eliminación del ruido de las series de tiempo	27
Identificación de las asociaciones de diatomeas como proxies de variables climáticas con potencial para reconstruir .....	28
Reconstrucción del clima (1417 – 2011 AD) a partir de diatomeas planctónicas .....	30
• <i>Predicción del clima a partir de las series de tiempo reconstruidas</i> .....	32
VII. RESULTADOS .....	34
• <i>Estandarización de las bases de datos biológicas y selección de especies</i> .....	34
• <i>Evaluación de parámetros no lineales y eliminación del ruido de las series de tiempo</i> .....	34
• <i>Definición de la arquitectura de la RNA tipo SOM a partir de los datos instrumentales y biológicos</i> .....	39
Variabilidad de las diatomeas planctónicas en el núcleo de sedimentos .....	42
• <i>Reconstrucción del clima a partir de diatomeas planctónicas</i> .....	44
• <i>Predicción de las variables climáticas a partir de las series de tiempo reconstruidas</i>	51
VIII. DISCUSIÓN .....	57

•	<i>Procesamiento y análisis de las series de tiempo para la obtención de las funciones de transferencia y reconstrucción del clima</i> .....	58
•	<i>Reconstrucción climática de los últimos 600 años para el sur del Golfo de California</i> .....	63
•	<i>Predicción de las condiciones climáticas de las próximas décadas</i> .....	71
IX.	CONCLUSIONES.....	76
X.	REFERENCIAS BIBLIOGRÁFICAS .....	78

## Relación de Tablas

Tabla 1. Recopilación de parámetros descriptivos de las series de tiempo instrumentales; dimensión de correlación (D), dimensión de embedding (m), el retraso (d), distancia euclidiana (e), error cuadrático media (RMSE) y relación señal-ruido (SNR). Rapidez del viento (Rap. Viento), temperatura media del aire (Temp. Media), profundidad de la zona eufótica ( $Z_{eu}$ ), radiación fotosintéticamente activa (PAR), presión atmosférica (P-atm), profundidad de la capa de mezcla (PCM), vector del viento U (VU) y V (VV). ..... 35

Tabla 2. Parámetros de rendimiento de las mejores dimensiones de las redes neuronales tipo SOM considerando el tamaño óptimo  $n= 5 * num. muestras$  propuesto por Vesanto (2000); EQ (Error de Quantization) y ET (Error Topológico). Negritas corresponden a las arquitecturas seleccionadas..... 39

Tabla 3. Resumen de los criterios considerados en cada uno de los modelos evaluados para la TSM y temperatura del aire; Criterio (GISOM: grupo 1 del SOM, *sp* abund: especies más abundantes, SOM: todas las asociaciones identificadas por el SOM); Esc. Temp: escala temporal; Técnica Vac. Info: Técnica para llenar vacíos de información; Capas: número de capas de la RNA; Proces: procesamiento de los datos (D.N: datos normalizados, D.S.N: datos sin normalizar) y Resul: resultado cualitativo de cada variante dentro de los modelos (P.S: poco satisfactorio, M.S: medianamente satisfactorio, S: satisfactorio). ..... 46

Tabla 4. Parámetros descriptivos de la eficiencia de las reconstrucciones a partir de redes neuronales artificiales *feedforward* para una topología de una y dos capas de neuronas en la capa intermedia. Regresión del modelo (Reg. Mod), RMSE del modelo (RMSE mod), RMSE de la reconstrucción (RMSE rec), correlación de Spearman para el promedio de las repeticiones de la modelación ( $\rho$ ), correlación para el promedio móvil de 3 pasos para la serie de tiempo promediada ( $\rho m3$ ). Negritas denotan las arquitecturas seleccionadas con el mejor desempeño. .... 47

Tabla 5. Recopilación de parámetros descriptivos no lineales evaluados para la reducción del ruido en las series de tiempo reconstruidas de la temperatura del aire y TSM. Dimensión de correlación (D), dimensión de embedding (m), error cuadrático medio (RMSE) y relación señal-ruido (SNR). Los valores en negrilla corresponden a los parámetros descriptivos seleccionados. .... 51

Tabla 6. Número de pasos que se puede avanzar en el tiempo con el menor error posible a partir de los máximos exponentes de Lyapunov ( $\lambda$ ) para la temperatura del aire y TSM. Donde: M= pendiente que define el máximo exponente de Lyapunov ( $\lambda$ ),  $P= 1/M$  y Pasos= pasos hacia el futuro. .... 53

Tabla 7. Parámetros descriptivos de la eficiencia de las predicciones de la temperatura del aire a partir de RNA “*feedforward – step by step*” para una arquitectura de una y dos capas

de neuronas en la capa intermedia. Regresión del modelo (Reg. Mod), RMSE del modelo (RSME mod) y RSME de la predicción (RSME pred), correlación de Spearman para el promedio de las repeticiones de la modelación ( $\rho$ ) y para el promedio móvil de 3 pasos ( $\rho$  m3) para la serie de tiempo promediada. Negritas denotan las arquitecturas seleccionadas con el mejor desempeño ..... 54

Tabla 8. Parámetros descriptivos de la eficiencia de las predicciones de la TSM a partir de una RNA “*feedforward--step by step*” para una arquitectura de una y dos capas de neuronas en la capa intermedia. Regresión del modelo (Reg. Mod), RSME del modelo (RSME mod), RSME de la predicción (RSME pred), correlación de Spearman para el promedio de las repeticiones de la modelación ( $\rho$ ) y para el promedio móvil de 3 pasos ( $\rho$  m3) para la serie de tiempo promediada. Negritas denotan las arquitecturas seleccionadas con el mejor desempeño. .... 56

Tabla 9. Comparación de las correlaciones ( $r$ ) entre reconstrucciones y datos instrumentales (1901 - 1980) para diferentes regiones, así como los valores obtenidos al comparar las diferentes reconstrucciones de la temperatura del aire y superficial del mar con los valores reconstruidos para Cuenca Alfonso (C\_Alfonso). En las últimas tres columnas las correlaciones representan la misma comparación con C\_Alfonso pero acotadas a los periodos climáticos: Cálido actual (CA), Pequeña edad de hielo (PEH) y Periodo cálido medieval (PCM). Los valores en negritas resaltan las correlaciones significativas ( $p < 0.005$ ). ..... 66

Tabla 10. Correlaciones múltiples ( $r$ ) entre la temperatura superficial del mar (TSM pm 3) y temperatura del aire (TA pm 3) contra el forzante solar (FS) y volcánico (FV), para la Pequeña Edad de Hielo (PEH; 1500 – 1850 AD), el Cálido Actual (CA; 1850 – 2011 AD) y para toda la serie de tiempo reconstruida (T. serie; 1417 – 2011 AD). Pm 3; corresponde a un promedio móvil de 3 pasos; los valores son estadísticamente significativos ( $p < 0.05$ ); X: no hubo correlaciones significativas. Los forzantes climáticos fueron obtenidos de Mann et al. (2005). ..... 70

## Relación de Figuras

Figura 1. a) Estructura básica de una RNA multicapa y b) esquema de la función de actividad lineal (ponderación de pesos) y función de activación de una sola neurona. .... 19

Figura 2. Representación de la arquitectura de una red neuronal artificial tipo Self Organizing Maps (SOM) con 6 x 6 neuronas organizadas en una capa bidimensional (modificado de Kaltec *et al.*, 2008). Los Círculos amarillos hacen alusión a neuronas activas y los círculos verdes a neuronas inactivas. .... 20

Figura 3. Estructura general de una red neuronal artificial *Feedforward* de tres capas. (modificado de Bevilacqua *et al.*, 2007). .... 21

Figura 4. Localización del área de estudio. El símbolo (estrella negra) representa la posición geográfica de la trampa de sedimentos instalada a una profundidad de 360 m. .... 24

Figura 5. Recopilación de las series de tiempo instrumentales: a) Presión atmosférica, b) Irradiancia, c), Temperatura media del aire, d) Profundidad de la capa de mezcla, e) Profundidad de la zona eufótica, f) PAR, g) Rapidez del viento, h) vector U del viento. Datos originales (línea negra) y valores predictivos con mayor reducción de ruido (línea roja). .. 37

Figura 6. Relación entre la dimensión de correlación (D) y la dimensión de embedding (m). a) presión atmosférica, b) irradiancia, c), temperatura media del aire, d) profundidad de la capa de mezcla, e) profundidad de la zona eufótica, f) PAR, g) rapidez del viento, h) vector U del viento. Las líneas punteadas marcan el punto de saturación de la dimensión de correlación para las series de tiempo original (dato original) y con reducción de ruido (Dato 1, 2 y 3 definidos bajo diferentes parámetros descriptivos). .... 38

Figura 7. Matriz-U obtenida con la red neuronal SOM de 9 x 8 neuronas, representando las relaciones entre las abundancias relativas de las especies y las variables sin reducción de ruido. Las líneas punteadas marcan la separación entre cada grupo. Temperatura superficial del mar (TSM), temperatura media del aire (T. Aire), zona eufótica (Z.eu), precipitación (Precip), radiación fotosintéticamente activa (PAR), presión atmosférica (P. Atm). *Cyclotella* spp. (<20 µm) (Cy. sp <20 µm), *Thalassionema frauenfeldii* (T. fra), *Rhizosolenia bergonii* (Rh. ber\*), *Chaetoceros didymus* (Ch. did\*), *Nitzschia bicaudata* (N. bic), *Fragilariopsis doliolus* (F. dol\*), *Chaetoceros messanensis* (Ch. mes\*), *Thalassiosira oestrupii* (Th. oes), *Thalassiotrix* sp. (Thx.), *Rhizosolenia setigera* (Rh. set\*), *Actinopterychus senarius* (Ac. sen\*). (\*) Especies presentes en la columna de agua. .... 40

Figura 8. Matriz-U obtenida con la red neuronal SOM de 10 x 7 neuronas, representando las relaciones entre las abundancias relativas de las especies y las variables con reducción de ruido. Las líneas punteadas marcan la separación entre cada grupo. Temperatura superficial del mar (TSM), precipitación (Precip), irradiancia (Irrad), densidad (sigma-T), temperatura

media del aire (T. Aire), vector U del viento (Vect. U), profundidad de la capa de mezcla (PCM), profundidad de la zona eufótica (Z.eu), *Cyclotella striata* (*Cy. str*), *Rhizosolenia styliformis* (*Rh. sty*), asociación de taxa de afinidad templada (A. Tem), Costeras (A. Cost), Cosmopolitas (Cosmo), Cálida (A. Calid), *Cyclotella* spp. (<20 µm) (*Cy. sp\** <20 µm), *Nitzschia bicapitata* (*N. bic*), *Nitzschia interruptestriata* (*N. int* - *N. int\**), *Rhizosolenia setigera* (*Rh. set* - *Rh. set\**), *Actinoptychus senarius* (*Ac. sen* - *Ac. sen\**). (\*) Especies presentes en la columna de agua..... 41

Figura 9. Flujo total de valvas de diatomeas en el núcleo de sedimentos (C-36), acompañado del flujo de valvas de las asociaciones de diatomeas definidas por el análisis de la RNA tipo SOM. *Nitzschia bicapitata*, (*N. bic*), asociación de agua cálida (A. cali), costeras (A. cost), cosmopolitas (Cosmo), templada (A. temp), *Rhizosolenia* sp. (*Rh. sp.*), *Cyclotella striata* (*Cy. str*), *N. interruptestriata* (*N. int*), *Rh. setigera* (*Rh. set*) y *Actinoptychus senarius* (*Ac. sen*). Se marcan con sombreado rojo los periodos cálido actual (CA) y el periodo cálido medieval (PCM) y en azul la pequeña edad de hielo (PEH)..... 44

Figura 10. Reconstrucción de la temperatura del aire entre el 1417 y 2011 AD utilizando una RNA tipo “feedforward” de retro propagación, con topología de una capa de neuronas. RNA de: a) 5 neuronas, c) 6 neuronas, e) 7 neuronas. Periodo instrumental de la RNA de: b) 5 neuronas, d) 6 neuronas y f) 7 neuronas. El área sombreada corresponde a una desviación estándar con un promedio móvil de 3 pasos, la línea negra representa el promedio de 100 corridas de cada modelo. La línea roja representa el promedio móvil de 3 pasos y la línea azul los datos instrumentales (1941 – 2011) de la temperatura máxima del aire para los meses de invierno (enero-abril). Las líneas punteadas verticales delimitan un segmento del periodo cálido medieval (PCM) y los periodos climáticos de la pequeña edad de hielo (PEH) y el cálido actual (AC)..... 48

Figura 11. Reconstrucción de la TSM entre el 1417 y 2011 AD utilizando una RNA tipo “feedforward” de retro propagación, con una topología de dos capas de neuronas. Reconstrucciones con RNA de: a) 13 x 13 neuronas, c) 14 x 14 neuronas, e) 15 x 15 neuronas. Periodo instrumental de las RNA de: b) 13 x 13 neuronas, d) 14 x 14 neuronas y f) 15 x 15 neuronas. El área sombreada corresponde a una desviación estándar con un promedio móvil de 3 pasos, la línea negra representa el promedio de 100 corridas de cada uno de los modelos. La línea roja representa el promedio móvil de 3 pasos y la línea azul el valor promedio anual de los datos instrumentales de la TSM (1951 – 2011). Las líneas punteadas verticales delimitan un segmento del periodo cálido medieval (PCM) y los periodos climáticos de la pequeña edad de hielo (PEH) y el cálido actual (AC)..... 50

Figura 12. Dimensión de correlación contra la dimensión de embedding de: a), Temperatura del aire, b) Temperatura superficial del mar. Las líneas punteadas marcan el punto de saturación de la dimensión de correlación para las series de tiempo con ruido y con reducción de ruido. Datos del 1 al 5, hacen referencia a las series de tiempo a las cuales se les redujó el ruido citadas en la Tabla 5..... 52

Figura 13. Máximo exponente de Lyapunov para la temperatura del aire, a) serie de tiempo original, b) serie de tiempo con reducción de ruido bajo los parámetros  $m=2$ ,  $D=3$  (Dato 1), c) máximo exponente de Lyapunov para la TSM (serie de tiempo original), d) TSM con reducción de ruido bajo los parámetros  $m=2$ ,  $D=3$  (Dato 2)..... 53

Figura 14. Predicción de la temperatura del aire ( $^{\circ}\text{C}$ ) para el periodo 2011 – 2032 utilizando una RNA tipo *feedforward* de retro propagación “*step by step*”. RNA alimentada con la serie de tiempo original y una arquitectura de a) 4 neuronas y b) de 4 x 4 neuronas. El área sombreada corresponde a una desviación estándar con un promedio móvil de 3 pasos. La línea negra representa el promedio de 100 corridas de cada modelo, la línea roja el promedio móvil de 3 pasos y la línea azul el promedio de la temperatura máxima de enero a abril (datos instrumentales de 1942 – 2020)..... 55

Figura 15. Predicción de la TSM ( $^{\circ}\text{C}$ ) para el periodo 2011 – 2035 utilizando una RNA tipo *feedforward* de retro propagación “*step by step*”. a) RNA alimentada con la serie de tiempo original y una arquitectura de 4 x 4 neuronas, b) RNA alimentada con la serie de tiempo con reducción de ruido (dato 2) y una arquitectura de 4 neuronas. El área sombreada corresponde a una desviación estándar con un promedio móvil de 3 pasos, la línea negra representa el promedio de 100 corridas de cada modelo, la línea roja el promedio móvil de 3 pasos y la línea azul el promedio anual de la TSM (datos instrumentales de 1951 – 2020)..... 57

Figura 16. Comparación de la reconstrucción de las anomalías de la temperatura del aire de los últimos 600 años para Cuenca Alfonso contra: a) reconstrucciones regionales de Norte América; b) reconstrucciones para el hemisferio norte. El área sombreada corresponde a una desviación estándar con un promedio móvil de 3 pasos..... 67

Figura 17. Comparación de: a) las anomalías de la TSM reconstruida para los últimos 600 años para Cuenca Alfonso, con reconstrucciones regionales de Norte América; b) la TSM de los últimos 600 años para Cuenca Alfonso (línea negra) y la Cuenca de San Lázaro; c) Cuenca de Guaymas y d) Cuenca de Santa Bárbara. El área sombreada corresponde a una desviación estándar con un promedio móvil de 3 pasos..... 69

Figura 18. Comparación de a) las anomalías de la temperatura del aire contra la Oscilación del Pacífico de Norte América (PNA); b) anomalías de la TSM y la Oscilación decadal del Pacífico (PDO). Línea oscura representa al valor promedio de la predicción, línea roja al promedio móvil de 3 pasos de la predicción, la línea azul corresponde al índice climático correspondiente a cada figura y el área sombreada corresponde a una desviación estándar con un promedio móvil de 3 pasos. (A partir del 2011 hasta el 2035 corresponde a la predicción). ..... 74

## GLOSARIO

**Certidumbre:** confiabilidad de las predicciones determinadas a partir de la distribución de probabilidad o de la desviación estándar de la modelación y cuanto coincide con la frecuencia relativa observada a partir de los datos instrumentales.

**Iteraciones:** aplicación repetitiva de una función (algoritmo de aprendizaje). El algoritmo de aprendizaje de manera iterativa va ajustando (reduciendo) el error entre los datos observados y esperados, modificando los parámetros internos (pesos) que regulan o inhiben las conexiones entre las neuronas de una capa y otra, hasta encontrar los parámetros adecuados que determinen el menor error posible.

**No-linealidad:** hace referencia a la ausencia de una relación proporcional o lineal entre un evento de causa y efecto. Un sistema no lineal no está sujeto al principio de superposición ni de proporcionalidad, por lo que pequeños cambios pueden tener efectos dramáticos, inesperados y difíciles de explicar de manera individual.

**Precisión:** en modelación hace referencia a la distancia o error promedio entre los valores modelados y las observaciones.

**Predicción climática:** corresponde a la capacidad de producir una descripción o estimación cuantitativa de la evolución de una variable climática o del sistema climático en el futuro.

**Previsibilidad:** se define por la tasa de amplificación de pequeños errores en términos de un tiempo de duplicación.

**Proxy:** indicador indirecto cuantificable que puede ser de diferente origen (biológico, geológico, químico) que se preservan de manera cronológica como parte de un registro natural. Las relaciones que presentan con diferentes variables climáticas y sus variaciones en el tiempo permiten reconstruir las condiciones climáticas del pasado.

**Red de proxies:** conjunto de registros naturales distribuidos estratégicamente a lo largo de una zona en particular, ya sea a escala regional, continental, de hemisferio o global que permite relacionar las señales climáticas reconstruidas en cada una de las localidades a una escala mayor. De esta forma es posible extrapolar la respuesta climática y reconstruir a nivel espacial el potencial comportamiento que pudo tener una variable climática en el pasado.

**Registros naturales:** elementos de la naturaleza (p. ej. anillos de crecimiento de árboles, núcleos de corales, sedimentos lacustres, estalactitas, etc.) con capacidad de preservar señales climáticas del pasado a partir de la conservación cronológica de varios

indicadores indirectos (elementos biológicos y geoquímicos) que tienen el potencial de relacionarse con el clima.

**Repeticiones:** número de veces que se repite el entrenamiento y validación de una red neuronal artificial, con el fin de obtener diferentes resultados para de este modo, poder promediar las modelaciones y obtener la desviación de las reconstrucciones o predicciones de las variables climáticas de interés.

**Sedimentos laminados (marinos y lacustres):** se componen por pares de láminas sedimentarias (claras y oscuras) que representan dos procesos que se alternan regularmente en el tiempo. Su formación ocurre en ambientes sedimentarios con bajas concentraciones de oxígeno, alta tasa de sedimentación y productividad primaria. Almacenan información con una alta resolución temporal sobre la variabilidad climática del pasado. Si los periodos de formación del par de láminas claras y oscuras ocurren en un ciclo anual se les llama varvas, mientras que, si el tiempo de formación es mayor, se les conoce como láminas.

**Variabilidad climática:** fluctuaciones climáticas o componente de esta, que ocurren alrededor del valor promedio histórico. La variabilidad climática actúa a diferentes escalas espaciales y temporales, siempre y cuando esté por encima de la escala meteorológica. La variabilidad puede deberse a forzantes naturales internos o externos.

## RESUMEN

Los modelos climáticos se utilizan para predecir la variabilidad climática, con el fin de prevenir y desarrollar planes de mitigación que ayuden a reducir sus efectos en la economía y sociedad. Sin embargo, el comportamiento no lineal del clima impone limitantes a la capacidad predictiva de los modelos climáticos tanto en las reconstrucciones como en las predicciones, de igual manera que no se dispone de series de tiempo de datos instrumentales lo suficientemente extensas que cubran toda la variabilidad climática. Por tal razón, se reconoce la importancia de acoplar estudios paleo-climáticos con modelos climáticos para hacer más robustas las predicciones. El objetivo del estudio fue reconstruir los últimos 600 años y predecir las condiciones climáticas de la próxima década para el sur del Golfo de California, utilizando un conjunto de series de tiempo de datos instrumentales, ambientales, además de datos actuales de las diatomeas en sedimentación y sus frústulas conservadas en un núcleo de sedimentos laminados de Cuenca Alfonso. Las redes neuronales artificiales (RNA) de mapas de autoorganización (SOM) y *feedforward* de retro propagación permiten la identificación de las asociaciones de las diatomeas idóneas como indicadoras indirectas (proxies) de las variables climáticas. El análisis para la primera etapa considera cuatro pasos: 1) estandarización y selección de bases de datos biológicas; 2) reducción de ruido de las series de tiempo; 3) definición de una arquitectura SOM y 4) reconstrucción de las condiciones climáticas utilizando RNA *feedforward*. Los resultados mostraron que las asociaciones de diatomeas son proxies apropiados para reconstruir la temperatura del aire (TA) y de la superficie del mar (TSM). Estas asociaciones se utilizaron para entrenar una RNA “*feedforward*” (6 y 14 x 14 neuronas en la capa oculta, respectivamente), las cuales mostraron un buen rendimiento. Las variables climáticas reconstruidas fueron validadas con datos instrumentales y reconstruidos para la región, observándose correlaciones significativas (Spearman, [24] = 0.62 y 0.57;  $p < 0.005$ , respectivamente). La segunda etapa consiste en tres pasos: 1) reducción del ruido de las series de tiempo reconstruidas; 2) calcular la previsibilidad y 3) predecir el clima mediante una RNA *feedforward* – “*step by step*”. Se realizó la predicción [2014 – 2032 (2035)] utilizando RNA de 4 y 4 x 4 neuronas para la TA y TSM, respectivamente. Las series temporales se validaron con datos instrumentales (1940 – 2020), obteniendo correlaciones significativas (Spearman, [27]= 0.59 y 0.62;  $p < 0.005$ , respectivamente). Se concluye que la metodología implementada presentó el potencial para mejorar la clasificación de datos, reducir la incertidumbre y el sesgo en la definición de la matriz biológica (diatomeas planctónicas); considerando la menor pérdida posible de información. Así, se definieron algunas asociaciones de diatomeas como proxies de la TA y TSM, y se realizó su reconstrucción. Las series de tiempo reconstruidas utilizadas como insumos para predecir el clima de la próxima década permitieron predecir el Hiatus en la tendencia del calentamiento global (1999 – 2012) y proyectar un escenario en el que la TA aumenta y disminuye la TSM, lo que sugiere un retraso en la respuesta del océano a la reducción de la temperatura del sistema.

### Palabras clave

Redes neuronales artificiales, Mapas de autoorganización (SOM), Redes multicapa “*Feedforward*”, Proxies climáticos, Diatomeas planctónicas, Núcleo de sedimentos laminados, Reconstrucción climática, Modelación climática.

## ABSTRACT

Climate models predict climate variability to prevent and develop mitigation strategies that help diminish their effects on the economy and society. However, the non-linear behavior of the climate imposes limitations on the predictive capacity of climate models in both reconstructions and predictions, while there are not sufficiently long time series of instrumental data to cover all the climatic variability. For this reason, the importance of coupling palaeo-climatic studies with climate models to make predictions more robust is recognized. The aim study was to reconstruct the climate of the last 600 years and predict the climatic conditions of the next decade for the southern Gulf of California, using a data-set of instrumental, environmental time-series data in addition to the modern settling diatoms and their frustules conserved in a marine laminate sediment core from the Alfonso Basin. Self-organization maps (SOM) and *feedforward* of artificial neural networks (ANN) analysis allow the identification of associations of suitable diatoms as indirect indicators (proxies) of climatic variables. The analysis includes four steps in the first stage: 1) standardization and selection of biological databases; 2) reduction noise in the time series; 3) definition of a SOM-ANN architecture, and 4) reconstruction of climatic conditions using *feedforward* ANN. The results showed that diatom associations are appropriate proxies for reconstructing air (AT) and sea surface (SST) temperatures. These associations were used to train a *feedforward* ANN (6 and 14 x 14 neurons in the hidden layer, respectively), which showed good performance. The reconstructed climatic variables were validated with instrumental and reconstructed data for the region, observing high correlations (Spearman, [24] = 0.62 and 0.57;  $p < 0.005$ , respectively). The second stage consists of three steps: 1) reduction of noise from the reconstructed time-series; 2) predictability calculation, and 3) predicting the climate using a *feedforward* – “*step by step*” ANN. The prediction [2014 – 2032 (2035)] was made using ANN from 4 and 4 x 4 neurons for AT and SST, respectively. The time series were validated with instrumental data (1940 – 2020), obtaining significant correlations (Spearman, [27] = 0.59 and 0.68;  $p < 0.005$ , respectively). It is concluded that the implemented methodology presented the potential to improve data classification, reduce uncertainty and bias in the definition of the biological matrix (planktonic diatoms); considering the least possible loss of information. Thus, some associations of diatoms were defined as proxies of the AT and SST, and their reconstruction were done. The reconstructed time series used as inputs to predict the climate of the next decade made it possible to model the slowdown in the global warming trend (1999 – 2012) and project a scenario in which the AT increase, and SST decreased, which suggest a delayed response of the ocean to the reduction in the temperature of the system.

## Keywords

Artificial Neural Networks, Self-Organization Maps (SOM), *Feedforward* Multilayer Networks, Climatic Proxies, Planktonic Diatoms, Laminated Sediment Core, Climate reconstruction, Climate modelling.

## I. INTRODUCCIÓN

El clima es la manifestación de interacciones complejas y multidimensionales entre la hidrósfera, atmósfera, litósfera y demás componentes del sistema climático que se manifiestan en diferentes escalas espaciotemporales. Se ha buscado explicar y definir estas complejas interacciones, por medio del uso de modelos climáticos con la intención de predecir a mediano y largo plazo la variabilidad climática y eventos climáticos extremos, con el fin de prevenir y desarrollar planes de mitigación que ayuden a reducir las afectaciones en el ámbito económico y social (IPCC, 2007). Se reconoce que el sistema climático presenta un comportamiento no lineal, en el cual, tanto la interacción con forzantes naturales, la variabilidad interna, diferencias en el punto de inicialización de los modelos, como la estructura intrínseca de los mismos, limitan la capacidad predictiva en las reconstrucciones, así como en la predicción de las condiciones climáticas futuras (Taylor *et al.*, 2012). Adicionalmente, una de las limitantes que presentan los modelos climáticos es que no cuentan con series de tiempo instrumentales lo suficientemente extensas que permitan conocer toda la variabilidad del sistema climático, lo que repercute en el grado de certidumbre y predictibilidad de los modelos (Braconnot *et al.*, 2012; Dunne *et al.*, 2013).

En las últimas décadas, bajo la iniciativa del proyecto de inter-comparación de modelos acoplados (CMIP-5, por sus siglas en inglés, <https://www.wcrp-climate.org/wgcm-cmip>) se ha venido resolviendo dichas debilidades, ya que integra estudios paleoclimáticos con modelos climáticos proyectándolos a una escala decadal. Esta visión ha permitido extender las series de tiempo hacia el pasado considerando los últimos mil años (Meehl *et al.*, 2007; Braconnot *et al.*, 2012; Harrison *et al.*, 2015), así como incrementar el número de datos para entrenar los modelos, permitiendo validar la influencia que tienen los forzantes naturales y antropogénicos en la variabilidad climática, para de este modo poder reducir la incertidumbre y mejorar la precisión de los modelos predictivos del clima (IPCC, 2007; Braconnot *et al.*, 2012; Taylor *et al.*, 2012; Harrison *et al.*, 2015).

A pesar de que los modelos predicen las tendencias climáticas a nivel global y reconocen los cambios más representativos en cada una de las escalas temporales, no reproducen bien las magnitudes, ni identifican los forzantes y procesos de retroalimentación, así como tampoco logran una buena precisión a nivel regional. Generalmente, los datos instrumentales y

reconstruidos con los que se alimentan los modelos son puntuales y interpretan la variabilidad regional, además de corresponder a una señal modificada de la variabilidad climática, lo que en conjunto limita la respuesta de los modelos (IPCC, 2007; Braconnot *et al.*, 2012; Bellenger *et al.*, 2014; Dufresne *et al.*, 2013; Harrison *et al.*, 2015). Autores como Neukom *et al.* (2018), hacen referencia a que existen ciertas discrepancias entre las reconstrucciones realizadas a partir de elementos biogeoquímicos preservados en registros naturales (anillos de árboles, núcleos de sedimentos de lagos o marinos, etc.) que son utilizados como indicadores indirectos de variables climáticas (proxies) o de modelos. Estas discrepancias pueden ser causadas por la deficiencia de los modelos al simular la influencia de los forzantes externos y la variabilidad interna en el sistema climático, lo cual lleva a diferencias a escalas regionales, continentales y hemisféricas, así como a diferencias temporales. Adicionalmente, estos autores reconocen que la calidad, distribución y calibración de los datos proxy, así como el procesamiento y métodos estadísticos, pueden llevar a incrementar el sesgo y discrepancias en las reconstrucciones.

Los modelos pueden clasificarse como estadísticos, estocásticos y probabilísticos, según sus bases y fundamentos. Sin embargo, cuando los modelos contienen componentes del sistema climático, son denominados “modelos climáticos”. Estos se diferencian de los demás, ya que los modelos estadísticos se basan en correlaciones empíricas y no se apoyan en un conjunto de leyes físicas como lo hacen los modelos climáticos (IPCC, 1997). Adicionalmente, con el avance tecnológico y la comprensión de que el sistema climático presenta interacciones multidimensionales y un comportamiento no lineal, en las últimas dos décadas se ha venido implementando herramientas de inteligencia artificial (IA) las cuales han mostrado gran versatilidad en la modelación y clasificación del sistema climático (Guiot *et al.*, 2005; Bevilacqua *et al.*, 2007; Carro-Calvo *et al.*, 2013). Dentro de las IA encontramos las redes neuronales artificiales (RNA), las cuales se han utilizado ampliamente para la predicción climática a nivel regional y global (Bevilacqua *et al.*, 2007), en la predicción de eventos El Niño para el Pacífico tropical (Dijkstra *et al.*, 2019), en la reconstrucción climática de las anomalías de la temperatura global anual (Pasini *et al.*, 2006), considerando fuentes de energía no renovables, forzantes naturales y antropogénicas (Bevilacqua *et al.*, 2007).

Las RNA se han usado en paleo-reconstrucciones de variables ambientales para la región de Europa (últimos 18,000 años) considerando la fenología de diversos tipos de polen (Peyron *et al.*, 1998). Así como para el Atlántico Norte utilizando las abundancias de dinoquistes (Peyron & Vernal, 2001). Metodológicamente, las RNA se han comparado con diferentes análisis estadísticos para la paleo-reconstrucción de la temperatura superficial del mar (TSM) a partir del censo de datos de foraminíferos planctónicos (Malmgren *et al.*, 2002), radiolarios (Gupta & Malmgren, 2009) y diatomeas para el Océano Pacífico y Atlántico (Malmgren *et al.*, 2002), donde han demostrado que se producen mejores resultados que múltiples métodos lineales.

Un enfoque vigente en la climatología es el conocer cómo eran las condiciones climáticas del pasado más reciente (últimos dos milenios), ya que esto permite tener una mejor comprensión de la variabilidad climática natural y ayuda a contextualizar el clima actual, el cual parece no tener precedente a causa de la actividad antropogénica a partir de la revolución industrial (Mann *et al.*, 2008; Ortega *et al.*, 2017). El limitado acervo de registros históricos y de datos instrumentales hace necesario el uso de registros naturales (anillos de árboles, núcleos de sedimentos marinos, lacustres, de corales y hielo, entre otros) que tienen el potencial para preservar cronológicamente elementos biológicos y/o geoquímicos que conservan en el registro natural su relación con el clima, para de esta manera poder reconstruir las condiciones climáticas del pasado (IPCC, 2007; Mann *et al.*, 2008; Braconnot *et al.*, 2012; Harrison *et al.*, 2015).

El análisis de registros naturales de alta resolución como los sedimentos laminados de cuencas oceánicas representa una alternativa promisoría para cubrir esta necesidad, ya que los eventos climáticos que ocurrieron en el pasado quedan registrados cronológicamente en los componentes geoquímicos y biogénicos que los conforman. Dentro de estos últimos las frústulas de las diatomeas planctónicas han demostrado ser un excelente indicador de la variabilidad climática, debido a que son organismos de ciclos de vida cortos, que pueden variar en su distribución geográfica, composición y abundancia en respuesta al clima (Kemp *et al.*, 2000). Sin embargo, es indispensable estimar el grado de fidelidad con que las diatomeas conservan esta “información” en el fondo del océano con base en el conocimiento de la respuesta regional de las diatomeas ante la variabilidad climática (Kemp *et al.*, 2000).

En general, los estudios paleo-climáticos y actuales constituyen líneas de investigación que se han desarrollado de forma paralela. Sin embargo, han empezado a confluir debido a la necesidad de mejorar los resultados en la modelación del clima. En el presente trabajo se pretende integrar a través del uso de RNA la información asociada a los eventos de producción en la zona eufótica, la fracción en hundimiento y la preservada en el registro sedimentario aportada por las diatomeas planctónicas con los datos instrumentales. Todo esto con la finalidad de validar la utilidad que tienen las diatomeas planctónicas como indicadores indirectas de variables climáticas (proxies). Para finalmente alimentar un modelo predictivo con las asociaciones de diatomeas, identificadas como proxies y conservadas en el registro sedimentario, para de este modo poder reconstruir y posteriormente predecir las condiciones climáticas, obteniendo un buen ajuste y reduciendo el grado de incertidumbre.

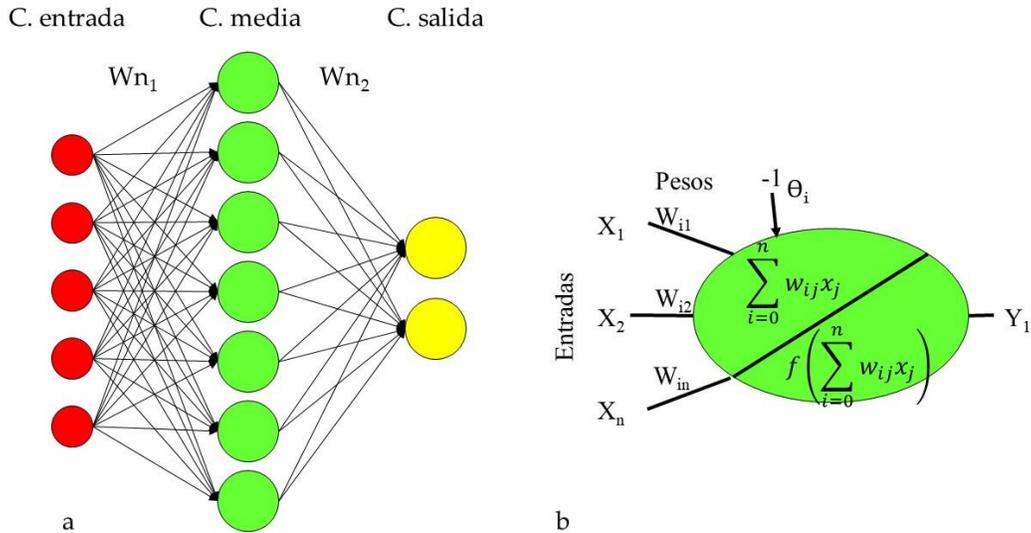
## **II. MODELO TEÓRICO DEL FUNCIONAMIENTO DE UNA RED NEURONAL**

En las últimas décadas las RNA han despertado gran interés en el medio científico, ingenierías, medicina, entre otros ámbitos, debido a que la estructura computacional de las RNA al estar compuestas por elementos simples (neuronas) interconectadas en paralelo, simulando las conexiones presentes en el cerebro, tienen la capacidad de interpretar y generar conocimiento de manera abstracta a partir de diferentes estímulos, con la particularidad de simplificar el funcionamiento y abstracción de los problemas (Gutiérrez *et al.*, 2004; Suzuki, 2011). Es decir, que las RNA no aprenden a asociar los datos de entrada a una salida determinada, sino que aprenden a reconocer las complejas relaciones entre los datos de entrada y las salidas (Maneta & Schnabel, 2003; Bevilacqua *et al.*, 2007). Son capaces de encontrar relaciones (patrones) de forma inductiva por medio de algoritmos de aprendizaje basado en los datos existentes, más que requerir de la ayuda de un modelador para especificar la forma funcional y sus interacciones (Albañez-Lucero, 2010). Adicionalmente, tienen la capacidad de resolver problemas de forma individual o combinada con otros métodos, además de tener la virtud de poder realizar clasificaciones, identificaciones, diagnósticos, optimizaciones y predicciones, así como la posibilidad de aprender en tiempo real. En estos casos las RNA se adaptan continuamente reajustando los parámetros (pesos) entre sus interconexiones (Lee *et al.*, 2003 Gutiérrez *et al.*, 2004; Albañez-Lucero, 2010).

Las RNA están construidas en principio por múltiples unidades lógicas de procesamiento de información conocidas como neuronas o nodos (Gutiérrez *et al.*, 2004; Bevilacqua *et al.*, 2007), las cuales están conectadas entre sí conformando una topología definida (red neuronal). Estas neuronas están organizadas por capas y la funcionalidad de la red está definida por el grado de conectividad y los “pesos” de conexión de unas neuronas con otras (Gutiérrez *et al.*, 2004; Bevilacqua *et al.*, 2007). La organización y conectividad de las neuronas definen la arquitectura, mientras que los pesos de las conexiones son los parámetros que ajustados a los datos de entrada permiten a la red aprender y generalizar el conocimiento. La estructura básica de una neurona está compuesta por un vector de entrada o estímulo “ $X$ ”, de un vector de pesos  $W_i = (W_{i1} \dots X_{in})$  equivalente a las conexiones sinápticas en una neurona real, de una función de activación “ $f$ ”, y de un umbral de activación de la neurona “ $\theta$ ”. Adicionalmente, el umbral de activación se puede incluir en la sumatoria considerando una neurona auxiliar ( $X_0 = -1$ ) conectada a  $Y_i$ , que corresponde a la salida de la unidad.

$$y_i = f\left(\sum_{j=1}^n w_{ij}x_j - \theta\right) = f\left(\sum_{j=0}^n w_{ij}x_j\right) \quad \text{Ecuación 1}$$

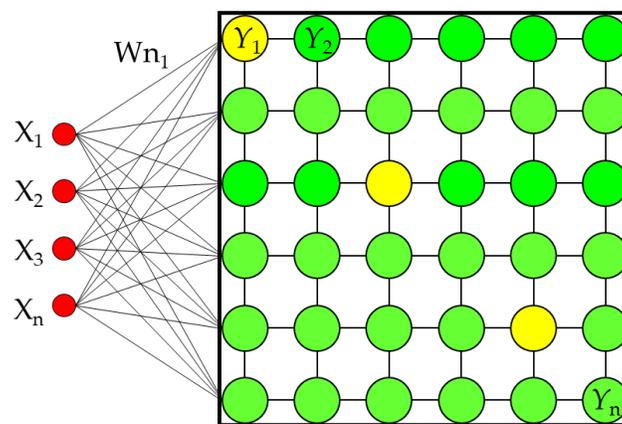
La arquitectura de una red típica consiste en una capa de entrada, una intermedia y una de salida (Bevilacqua *et al.*, 2007; Albañez-Lucero, 2010). Cada neurona recibe y procesa señales de entrada provenientes de neuronas en otras capas y produce un valor de salida ponderado, denominado valor activado (Figura 1a). Este valor es transformado utilizando una función (p. ej., sigmoidea) y transferido a otros elementos (nodos) interconectados. El nivel de actividad o suma ponderada de los valores activados (salida) de las neuronas en la capa de salida determina la solución de la red. Las RNA tienen la capacidad de utilizar diversos tipos de datos ya sean continuos, discretos, lineales o no lineales que favorecen su versatilidad ante diferentes escenarios (Bevilacqua *et al.*, 2007; Albañez-Lucero, 2010). La señal de salida se obtiene transformando la suma ponderada de las entradas que recibe la neurona usando la función de activación (Figura 1b).



**Figura 1.** a) Estructura básica de una RNA multicapa y b) esquema de la función de actividad lineal (ponderación de pesos) y función de activación de una sola neurona.

La modelación basada en RNA tiene dos etapas; entrenamiento y prueba. El primero se centra en que la red sea capaz de analizar, generalizar y reproducir un patrón con el menor error posible, todo ello lo realiza considerando un algoritmo de aprendizaje supervisado o no supervisado. Una vez definida la arquitectura y tipo de conexión entre las neuronas, se procede a ir modificando los pesos de las conexiones para codificar los datos de entrenamiento e ir incrementando el grado de ajuste de los datos. Este paso se realiza iterativamente con el objetivo de mejorar la funcionalidad de la red, generar cambios en los pesos específicos de la función de activación, lo que afecta el valor de salida y reparte el error entre todas las conexiones (Bevilacqua *et al.*, 2007). De manera estadística, el entrenamiento se puede considerar como la generalización de una regresión no lineal, donde el modelo es la propia actividad de la RNA y los parámetros de ajuste corresponden a los pesos de las conexiones (Gutiérrez *et al.*, 2004). Una vez entrenada la RNA, se procede a comprobar la eficiencia del modelo. En este punto si la red presenta un buen funcionamiento, podrá predecir adecuadamente los valores observados a partir de las variables independientes. Sin embargo, si la red tan sólo “memorizó” los datos de entrenamiento, las predicciones serán poco eficientes. La manera de evaluar la calidad de los modelos es determinar los errores entre los datos observados y los obtenidos por la RNA, dentro de los errores más utilizados resaltamos la raíz del error cuadrático medio (RMSE por sus siglas en inglés).

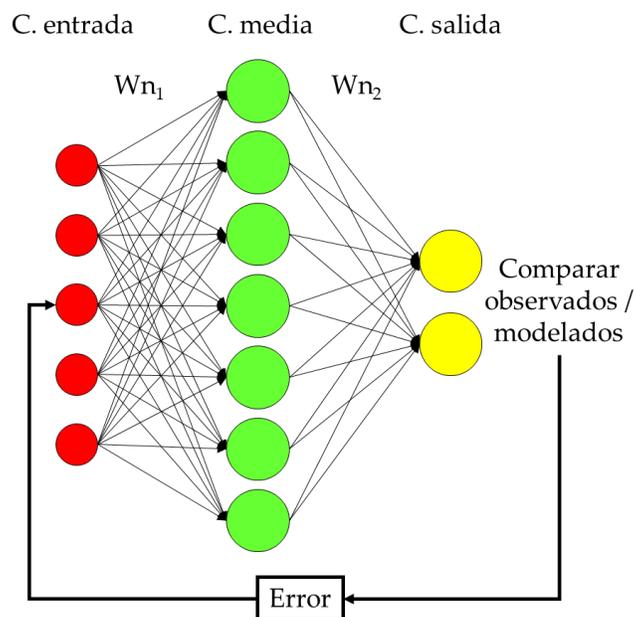
La red neuronal tipo Self Organizing Maps (SOM) es una red competitiva, especializada en reducir la dimensionalidad de los datos de entrada permitiendo agrupar y visualizar los datos de manera eficiente en un espacio bidimensional. Este método puede verse como una generalización de la técnica *K-means* (técnica de agrupamiento). Este tipo de red (Figura 2; **Error! No se encuentra el origen de la referencia.**) presenta una regla de aprendizaje no supervisado (autoorganización), por lo que es capaz de realizar análisis de agrupación (Kohonen, 2001; Gómez, 2014). Durante el proceso de aprendizaje la red debe descubrir por si misma rasgos comunes, patrones, correlaciones o categorías en los datos de entrada. Las redes SOM utilizan un aprendizaje competitivo, donde los nodos compiten por hacer una tarea dada. Cuando un nodo se activa o es el “ganador”, representaría al nodo más próximo a la entrada actual. En este contexto la distancia euclidiana se utiliza como criterio y cuanto más próximos estén los vectores, mejor será el resultado. Este criterio es también sensible a la magnitud de los vectores de entrada, por lo que los datos deben ser normalizados para evitar sesgos (Kohonen, 2001; Albañez-Lucero, 2010; Gómez, 2014). El análisis SOM es parcialmente análogo a un clúster (análisis de conglomerados), pero una de las ventajas del SOM es que genera una clasificación no lineal basada en un proceso iterativo, que premia a la clase ganadora y a sus vecinos más cercanos, mientras que inhibe a los más lejanos buscando descubrir patrones (Kohonen, 2001).



**Figura 2.** Representación de la arquitectura de una red neuronal artificial tipo Self Organizing Maps (SOM) con 6 x 6 neuronas organizadas en una capa bidimensional (modificado de Kaltec *et al.*, 2008). Los Círculos amarillos hacen alusión a neuronas activas y los círculos verdes a neuronas inactivas.

Otro tipo de RNA que comúnmente se utiliza para la modelación de eventos futuros, como del pasado, son las redes multicapas, donde las capas ocultas dan una mayor flexibilidad al

modelo resultante, llegándose a considerar como un modelo no paramétrico no lineal (Gutiérrez *et al.*, 2004). Dentro de los modelos multicapas encontramos las redes *Feedforward* (Figura 3), la cual transfiere la información entre las neuronas y capas hacia delante, invalidando la transferencia de manera lateral o hacia atrás. Este tipo de red requiere de un entrenamiento supervisado para reducir el error entre los datos observados y modelados, para lo cual es necesario utilizar un algoritmo de aprendizaje que ayude a ajustar los parámetros (pesos) de la red (Maneta & Schnabel, 2003; Bevilacqua *et al.*, 2007). El método más utilizado es el de retro-propagación “*backpropagation*” el cual busca reducir el error cuadrático total usando el método del descenso de gradiente. Este método se trata de un algoritmo iterativo que en cada etapa modifica los pesos incrementando sus valores de forma que se reduzca el error (inicialmente se toma un valor aleatorio en los pesos) (Maneta & Schnabel, 2003; Gutiérrez *et al.*, 2004). Entre cada iteración, los pesos obtenidos se utilizan para actualizar los pesos de la capa intermedia, propagando hacia atrás los errores anteriores y realizando reajustes en los pesos, para de esta manera dar lugar al entrenamiento de la red (Maneta & Schnabel, 2003; Gutiérrez *et al.*, 2004; Bevilacqua *et al.*, 2007).



**Figura 3.** Estructura general de una red neuronal artificial *Feedforward* de tres capas. (modificado de Bevilacqua *et al.*, 2007).

### **III. HIPÓTESIS**

La señal climática preservada durante el proceso de hundimiento de las diatomeas planctónicas (desde la columna de agua a la trampa de sedimentos), así como su variabilidad en composición y abundancia, serán indicadores indirectos (proxies) de la variabilidad del clima. Como consecuencia, al introducir los datos biológicos y ambientales como insumos a un modelo predictivo (reconstrucción), se podrá definir una función de transferencia que describirá las relaciones existentes entre los datos biológicos y ambientales, que al momento de ser alimentada con los flujos de diatomeas preservadas en un núcleo de sedimentos permitirá obtener una reconstrucción climática. Posteriormente, la información obtenida (variables climáticas reconstruidas) se podrá utilizar como insumos para un modelo predictivo que permitirá predecir el clima de la región de las próximas décadas con un buen ajuste y alto grado de certidumbre.

### **IV. OBJETIVO GENERAL**

- Demostrar la utilidad que tienen las diatomeas planctónicas como indicadores indirectas, tanto en la reconstrucción, como en la predicción del clima.

### **V. OBJETIVOS PARTICULARES**

- Entrenar y validar una red neuronal artificial para reconstruir el clima de los últimos 600 años para la región sur del Golfo de California.
- Realizar una predicción climática para las próximas décadas considerando las variables climáticas reconstruidas a partir de diatomeas planctónicas.

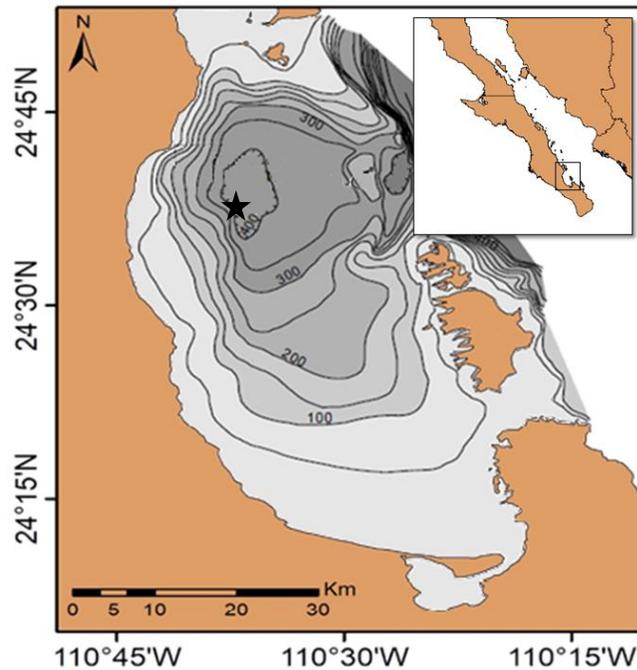
## VI. METODOLOGÍA

### Área de estudio

Cuenca Alfonso es una depresión marginal ubicada al interior de La Bahía de La Paz (Figura 4), en el margen suroccidental del Golfo de California y presenta un área aproximada de 37 km<sup>2</sup>. (24°41' N, 110°39' W y 24°33' N, 110°27' W) con una profundidad máxima de 415 m y un gradiente batimétrico que aumenta gradualmente de sur a norte (Nava-Sánchez *et al.*, 2001). Hacia el oriente de la bahía hay un umbral (Boca grande) a 275 metros de profundidad que restringe el ingreso de las aguas profundas del Golfo a la cuenca, manteniendo así bajas concentraciones de oxígeno (O<sub>2</sub> <0.1 ml l<sup>-1</sup>) en el fondo (Monreal-Gómez *et al.*, 2001). Esto limita la presencia de organismos bentónicos y favorece la depositación natural del material orgánico en forma de láminas, cuyo tiempo de formación de cada par (clara y oscura) representa un tiempo estimado de 3 años (Álvarez-Gómez, 2010).

La región sur del Golfo de California constituye una zona de transición ya que en ella convergen distintas masas de agua que cambian estacionalmente (Obeso-Nieblas *et al.*, 2008). Las condiciones hidrográficas prevaecientes son influenciadas por la circulación atmosférica regional. Para los meses de noviembre a mayo la región se encuentra influenciada por los vientos del noroeste, y de junio a octubre por los del sureste, con eventos esporádicos de depresiones tropicales o huracanes (Obeso-Nieblas *et al.*, 2008; Villegas-Aguilera, 2009; Hakspiel-Segura, 2014). Adicionalmente, la región se encuentra influenciada a una mayor escala por la circulación atmosférica y del Océano Pacífico a diferentes escalas temporales (Di Lorenzo *et al.*, 2013). Las condiciones hidrográficas de la zona se caracterizan por un periodo de mezcla modulado por la intensificación del viento del noroeste entre noviembre y marzo (Obeso-Nieblas *et al.*, 2008), seguido de la activación de un giro ciclónico (abril a julio) caracterizado por el calentamiento del sistema y la somerización de la termoclina y nutriclina, que provoca el aporte de nutrientes del fondo hacia la zona eufótica (Villegas-Aguilera, 2009; Hakspiel-Segura, 2014; Verdugo-Díaz *et al.*, 2014) y termina con un periodo de estratificación (agosto a octubre) en el que se alcanzan las mayores temperaturas (Villegas-Aguilera, 2009). Al ubicarse al sur del Golfo y al estar influenciada por el Océano Pacífico abierto, esta cuenca funciona como un registrador sensible de las variaciones regionales que ocurren a diferentes escalas temporales (alta y baja

frecuencia), lo cual se refleja en la composición y abundancia de las especies fitoplanctónicas conservadas en los sedimentos laminados de este sitio (Martínez López *et al.*, 2016).



**Figura 4.** Localización del área de estudio. El símbolo (estrella negra) representa la posición geográfica de la trampa de sedimentos instalada a una profundidad de 360 m.

#### *Muestras biológicas*

La base de datos biológica utilizada proviene de diferentes proyectos de investigación institucional, además de un par de tesis de maestría (Villegas-Aguilera, 2009; Acevedo-Acosta, 2015). En general está constituida por las abundancias específicas de diatomeas planctónicas colectadas en la zona central de Cuenca Alfonso (24° 35' N y 110° 36' W) procedentes de la zona eufótica (definida como la profundidad donde la irradiancia corresponde al 0.1% de su valor en superficie) entre diciembre del 2005 hasta junio del 2012. Adicionalmente, se consideran las diatomeas colectadas con una trampa de sedimentos (Figura 4), ubicada entre los 300 y 360 m de profundidad, de 2008 a 2012 con una resolución temporal de 7, 17 y 30 días. Además, se contó con un núcleo de sedimentos laminados de alta resolución colectado al interior de la cuenca en 2011.

El núcleo de sedimentos (C-36) fue obtenido con financiamiento del CONACYT y de la Universidad Nacional Autónoma de México (UNAM), a bordo del buque oceanográfico El

Puma, utilizando un nucleador de caja a una profundidad de 390 m. Posteriormente en superficie, se realizó un sub-muestreo del sedimento por el nucleador, utilizando una camisa de acrílico, la cual fue seccionada en losas longitudinales de aproximadamente 2 cm de espesor. Una de las losas se utilizó para la cronología, a partir de un fechado radiométrico con  $^{210}\text{Pb}$  y mediante conteo de láminas. Otra de las losas fue cortada siguiendo la estratificación y considerando el número de pares de láminas tratando de ser lo más fino y preciso posible, obteniendo 171 muestras que corresponden aproximadamente a 600 años (con alcance temporal entre 1417 y 2011 AD). El resto del núcleo fue resguardado en la Nucleoteca II del Programa Universitario de Perforaciones en Océanos y Continentes, del Instituto de Geofísica y Paleomagnetismo (UNAM).

Posterior al corte, el material fue procesado en el Laboratorio de Fitoplancton del CICIMAR, buscando eliminar la mayor cantidad del material litológico fino (limo y arcillas) presente en las muestras, mediante múltiples lavados con agua destilada (cada 36 horas). Después de la limpieza se procedió a eliminar la materia orgánica utilizando peróxido de hidrógeno ( $\text{H}_2\text{O}_2$ ). A partir del material limpio, se realizó un tamizado utilizando una luz de malla de 20  $\mu\text{m}$ , para separar el material en fracciones  $< 20$  y  $> 20$   $\mu\text{m}$ . Posteriormente, se utilizó un volumen conocido para cada una de las fracciones para el montaje del material en preparaciones permanentes con resina (Pleurax con alto índice de refracción IR: 1.65). Finalmente, las valvas de diatomeas fueron identificadas y cuantificadas hasta el mínimo taxón posible utilizando un microscopio de contraste de fase Olympus CH30 mediante transectos y barridos de toda la preparación. Se realizó un conteo de más de 300 especímenes, siguiendo los criterios de Schrader & Gersonde (1978).

#### ***Datos instrumentales***

La base de datos que comprendió el periodo 2005 – 2014, estuvo constituida por un conjunto de variables físico-químicas de la superficie de la columna de agua [temperatura ( $^{\circ}\text{C}$ ), salinidad (ups), sigma T ( $\text{kg m}^{-3}$ )], así como de la profundidad (m) de la zona eufótica ( $Z_{\text{eu}}$ ) y la profundidad (m) de la capa de mezcla (PCM) estimada acorde a Kara *et al.* (2000). Se contó con datos satelitales diarios de la TSM ( $^{\circ}\text{C}$ ), de 4x4 km de resolución espacial del sensor remoto de Terra-MODIS entre 2000 y 2012 ([https://coastwatch.pfeg.noaa.gov/erddap/griddap/erdMTsst48day.graph?sst%5B\(2013-10-](https://coastwatch.pfeg.noaa.gov/erddap/griddap/erdMTsst48day.graph?sst%5B(2013-10-)

12T00:00:00Z)%5D%5B(0.0)%5D%5B(18.150034730261623):(34.112062977541086)%5D%5B(235.90230350735038):(251.86248408380598)%5D&.draw=surface&.vars=longitud%7Clatitude%7Csst&.colorBar=%7C%7C%7C%7C%7C&.bgColor=0xffccccff&.click=?282,197), así como de los Vectores U y V del viento ( $m \cdot s^{-1}$ ) con una resolución espacial de 27x27 km derivados del proyecto CCMP winds (ATLAS FLK v1.1) ([https://coastwatch.pfeg.noaa.gov/erddap/griddap/jplCcmp35aWindPentad\\_LonPM180.graph](https://coastwatch.pfeg.noaa.gov/erddap/griddap/jplCcmp35aWindPentad_LonPM180.graph)). Los datos derivados de imágenes satelitales corresponden a el área que delimita Cuenca Alfonso. Los datos meteorológicos de la temperatura promedio diaria del aire ( $^{\circ}C$ ), presión atmosférica (mb), radiación fotosintética activa (PAR= fotones  $m^2 s^{-1}$ ), precipitación (mm), irradiancia ( $W m^{-2}$ ) de la base de datos climáticos del noroeste de México ( $24^{\circ} 30' N$  y  $111^{\circ} 21' W$ ) (<http://peac-bc.cicese.mx/datosclim/dcbc.php#>) y rapidez del viento ( $m s^{-1}$ ) (2001 – 2011) de la estación meteorológica El Comitán (La Paz, B.C.S.) (<http://intranet.cibnor.mx/personal/bmurillo/clima.php>), todas de acceso libre. Posteriormente, para la validación de las variables reconstruidas y modeladas hacia el futuro, se utilizó la temperatura del aire promedio anual, máxima y mínima para los meses de verano (julio a octubre) e invierno (enero a abril) derivados de la estación meteorológica de Baja California Sur (3074 - La Paz, DGE) a partir de 1940 hasta 2020 (<http://clicom-mex.cicese.mx/mapa.html>), se contó además con la temperatura promedio anual de la superficie del mar desde 1950 hasta 2017, para la región central de la Bahía de La Paz, incluidos los datos obtenidos en múltiples proyectos institucionales del Laboratorio de Fitoplancton del CICIMAR-IPN.

### ***Procesamiento y análisis de las series de tiempo para la obtención de las funciones de transferencia y reconstrucción del clima***

#### *Estandarización de datos biológicos y selección de especies*

Para la estandarización de las bases de datos biológicos correspondientes a los diferentes estratos de la columna de agua (zona eufótica, profundidad de la trampa de sedimentos) y aquellos provenientes del sedimento. Lo primero que se hizo fue estandarizar las unidades en # diatomeas  $m^2 d^{-1}$  siguiendo lo propuesto por Martínez-López *et al.* (2016). Posteriormente, se homogenizaron los taxones comunes presentes tanto en la trampa de

sedimentos, como en la columna de agua y el sedimento. Una necesidad debido a la pérdida de información que comúnmente ocurre por la presión de pastoreo y la disolución de las diatomeas durante el proceso de hundimiento (Martínez-López *et al.*, 2019). En segundo lugar, sobre la matriz biológica de la trampa de sedimentos y columna de agua previamente depuradas, se realizó una reducción del número de taxones, excluyendo los más abundantes (Las esporas de *Chaetoceros*, *Pseudo-nitzschia* spp., *Thalassionema nitzschioides*, y *T. nitzschioides* var. *parva*), ya que al encontrarse con abundancias altas todo el año, su presencia no constituía un aporte significativo a la identificación de señales climáticas de baja frecuencia. Asimismo, aquellos taxones que aportaron  $\leq 10\%$  de la abundancia relativa acumulada se agruparon en función a su afinidad a masas de agua, con el fin de reducir la pérdida de esta información.

#### *Evaluación de parámetros no lineales y eliminación del ruido de las series de tiempo*

Para la evaluación de los parámetros descriptivos no lineales de las series de tiempo se utilizó el programa *Visual Recurrence Analysis* (VRA). Se estimó el valor óptimo del tiempo de retraso ( $d$ ) para la reconstrucción del espacio de fase por medio de la información mutua (Grassberger *et al.*, 1991). Para distinguir el comportamiento aleatorio y caótico de las series de tiempo, se utilizó la dimensión de correlación ( $D$ ). El comportamiento aleatorio se detectó a través de la estimación del ruido blanco, en el cual la señal es completamente aleatoria y la serie de tiempo no es capaz de ser atrapada en el espacio de “envolvimiento” ( $m$ ; de ahora en adelante mencionado como embedding). Mientras que, para una señal caótica, su dimensión de correlación se saturará en el momento que alcance la máxima dimensión de embedding (Elshorbagy *et al.*, 2002; Elshorbagy, 2009). Para este estudio el parámetro ( $m$ ) se calculó a partir del teorema de Takens (1985), el cual representa la relación entre la dimensión del atractor y la dimensión de embedding.

Para la eliminación del ruido se utilizó el método propuesto por Grassberger *et al.* (1993) utilizando el software TISEAN de libre acceso (rutina ghkss.exe; Hegger *et al.*, 1999), para el cual fue necesario conocer los parámetros descriptivos de las series de tiempo previamente mencionados. Se calculó el RMSE, definido como:

$$RMSE = \sqrt{\sum_{i=1}^n \|y_o - y_i\|^2 / n}$$

Ecuación 2

Dónde:  $Y_o$ , corresponde a los datos observados;  $Y_i$ , los datos modelados y  $n$ , la cantidad de datos evaluados.

El RMSE mide las diferencias entre dos series de tiempo, cuanto más pequeño sea el RMSE, mejor será el rendimiento que se puede obtener por la eliminación del ruido (Tung *et al.*, 2011; Di *et al.*, 2014), y el SNR (relación señal-ruido) cuantifica cuanto ruido ha sido eliminado de una serie de tiempo (Ecuación 3).

$$SNR = 10 \log_{10} \frac{\sigma_x^2}{\sigma_n^2}$$

Ecuación 3

Dónde:  $\sigma_x^2$ , corresponde a la sumatoria de todos los valores de la serie original sobre la cantidad de puntos al cuadrado,  $\sigma_n^2$ , corresponde a la sumatoria de todos los valores de la serie a la cual se le redujo el ruido sobre la cantidad de puntos al cuadrado. Para la reducción del ruido, se optó por utilizar una estrategia de cascada en ambos sentidos hasta encontrar los valores y relación óptima entre el RMSE y SNR. Adicionalmente, se graficó la dimensión de correlación contra la dimensión de embedding, con la intención de identificar el punto de saturación de la dimensión de correlación para las series de tiempo original y sin ruido.

#### *Identificación de las asociaciones de diatomeas como proxies de variables climáticas con potencial para reconstruir*

Para la identificación de las asociaciones de diatomeas que tienen el potencial para ser utilizadas como proxies de variables climáticas, el primer paso fue definir la arquitectura de la red a partir de los datos instrumentales y biológicos. Para este propósito se usó una red neuronal tipo SOM, que tiene un aprendizaje no supervisado (autoorganización). Durante el proceso de aprendizaje la red define rasgos comunes, patrones o categorías en los datos de entrada, para finalmente transformar las complejas relaciones no lineales de los datos de entrada de elevada dimensión en relaciones geométricas expresadas en un mapa bidimensional (Deligiorgi *et al.*, 2014). Asimismo, presenta una función de activación

(sigmoidal), que cuando se activa una “neurona ganadora”, ésta representa al nodo más próximo a la entrada actual. La distancia euclidiana se utiliza como criterio para definir los nodos más próximos a las neuronas ganadoras, este criterio es sensible a la magnitud de los vectores de entrada, por lo que los datos deben ser normalizados para evitar sesgos y mejorar los resultados (Deligiorgi *et al.*, 2014; Li *et al.*, 2018).

La eficiencia de los modelos obtenidos con la RNA tipo SOM se realizó evaluando los datos instrumentales con y sin ruido. El número de neuronas fue determinado siguiendo la fórmula  $n = 5 * \sqrt{\text{num. muestras}}$  propuesta por Vesanto (2000). Se evaluaron las arquitecturas de las redes exploradas, con un número de neuronas cercano al óptimo estimado a partir de la fórmula anterior (Tabla 2). Los errores topológicos (TE) y de cuantificación (QE) se utilizaron para evaluar el rendimiento de los modelos. Para cada modelo, se evaluaron las relaciones entre las variables ambientales y las asociaciones de diátomeas, considerando la coherencia con las relaciones ecológicas. Finalmente, se eligió el modelo con el mejor rendimiento en relación con los errores más pequeños y las relaciones con el mayor significado ecológico. La base de datos con que se alimentó la arquitectura de la red neuronal se dividió en tres subconjuntos (fase de entrenamiento, verificación y modelación). El tamaño de cada subconjunto se determinó de manera aleatoria. Durante la fase de entrenamiento los pesos entre las sinapsis de las neuronas de la capa de entrada y salida se reajustaron hasta alcanzar el menor error posible de salida, así como el mayor equilibrio en la arquitectura de la red.

Se determinó el índice de estructura global (GSI, siglas en inglés) propuesto por Park *et al.* (2005), el cual define las especies y variables de mayor influencia en la estructura de la organización del mapa SOM. Su valor determina la importancia relativa de cada especie y/o variable, definiéndolas como indicadoras del clima. El GSI se calcula a partir de la suma de la proporción de las distancias entre los pesos (es decir, las intensidades de conexión) de todas las especies y/o variables en el SOM y la distancia topológica de dos unidades virtuales del SOM (Park *et al.*, 2005, Ecuación 4). Adicionalmente, a partir de la matriz de pesos de cada uno de los modelos, se realizó un análisis de conglomerados de distancias por correlaciones, para confirmar las agrupaciones definidas por los patrones.

$$GSI_i = \sum_{k=1}^s \sum_{j=1}^{k-1} \frac{|w_{ij} - w_{ik}|}{\|r_j - r_k\|} \quad \text{Ecuación 4}$$

Donde,  $GSI_i$  es un valor de la especie  $i$ ,  $w_{ij}$  y  $w_{jk}$  son los pesos de conexión de la especie  $i$  en la unidad  $j$  y  $k$  del resultado de la RNA SOM respectivamente,  $r_j$  y  $r_k$  son las coordenadas de las unidades  $j$  y  $k$ , y  $\|r_j - r_k\|$  es la distancia topológica entre las unidades  $j$  y  $k$ .  $S$  es el número total de unidades de salida del SOM y  $\sum$  es la sumatoria de todos los valores de la unidad  $j_1$  y  $k-1$  (Park *et al.*, 2005).

#### *Reconstrucción del clima (1417 – 2011 AD) a partir de diatomeas planctónicas*

Una función de transferencia con métodos lineales se define como una ecuación que describe el comportamiento de una variable climática en función a las abundancias y/o elementos preservados cronológicamente en un registro natural que presenta una relación con la variable climática (Birks *et al.*, 2010). En este estudio el término “función de transferencia” se refiere a las relaciones internas entre las neuronas que reconocen el comportamiento no lineal de una variable, lo que permitió reconstruir su comportamiento a partir de sus relaciones con los datos biológicos y geoquímicos derivados del registro natural. Para la reconstrucción del clima se utilizó una red neuronal multicapa “*feedforward*” con un algoritmo de aprendizaje de retro propagación. Se determinó una función de activación sigmoidea para la capa intermedia, y una función lineal en la capa de salida. La RNA en la capa de entrada se alimentó con las asociaciones de diatomeas de la trampa de sedimentos definidas previamente. Debido a que el flujo promedio en la trampa de sedimentos ( $10^6$  diatomeas  $m^{-2} d^{-1}$ ) fue mayor al del núcleo de sedimentos ( $10^4$  diatomeas  $m^{-2} d^{-1}$ ), se realizó un ajuste reduciendo dos órdenes de magnitud en los datos de la trampa como estrategia para reducir la sobreestimación en el entrenamiento de la red neuronal. Se utilizó el 70% de los datos en el entrenamiento, el 15% en la prueba y el 15% restante en la validación. La selección de los datos para cada una de las etapas se realizó de manera aleatoria, con la intención de poder captar toda la variabilidad posible de las series de tiempo.

Posteriormente, se evaluaron tres alternativas en el procesamiento de los datos del núcleo de sedimentos con relación a la resolución temporal; en la primera estrategia se respetó el fechado de la laminación del núcleo, mientras que para las siguientes se determinó una resolución temporal equidistante de 3 y 5 años. Tanto la base de datos de la trampa, como del núcleo de sedimentos presentaron vacíos de información por diferentes razones logísticas. Esto se resolvió evaluando dos procedimientos: el primero fue el uso de una interpolación de puntos adyacentes utilizando el programa de Statistica v10, el segundo, fue complementar los puntos con un suavizado de ponderación Gaussiana, considerándose una regresión polinomial de tercer grado, utilizando el programa SigmaPlot v11.0.

Para el entrenamiento, los flujos de diatomeas presentes en la trampa y el núcleo de sedimentos se evaluaron sin ningún procesamiento y posteriormente aplicándoles una transformación [ $y = (x - \bar{x})/\sigma$ ]. Para la capa intermedia se exploraron dos alternativas: una arquitectura de una y dos capas, respectivamente. Se consideró un aprendizaje supervisado, en el cual se vincularon las series de tiempo de las variables climáticas (Target) definidas a partir del análisis de la RNA tipo SOM, como estrategia para reconstruir las variables climáticas de interés. Finalmente se entrenaron múltiples combinaciones de parámetros de pre-procesamiento de los insumos para alimentar las redes neuronales (modelos) con un número creciente de neuronas en la(s) capa(s) intermedia(s) que fueron desde 5 (5 x 5) hasta 15 (15 x 15) neuronas, de las cuales se seleccionaron aquellas que presentaron la mejor varianza explicada (r) y rendimiento en función del RMSE.

Tras definir los criterios para el pre-procesamiento de los datos de entrada de cada uno de los modelos que se evaluaron, se definió una tasa de aprendizaje variable con un máximo de 1000 iteraciones durante el entrenamiento considerando una regulación Bayesiana. Posterior al entrenamiento, se evaluó el RMSE y la varianza explicada de cada modelo para determinar su eficiencia. En este punto, una vez elegida una topología, la funcionalidad de la red se determinó por la función concreta que realiza cada neurona y por los pesos de conexión entre neuronas, condición que teóricamente define la “función de transferencia”. Posteriormente, ésta se alimentó con los datos del núcleo de sedimentos para finalmente obtener las reconstrucciones de las variables de interés. Este proceso se repitió 100 veces, para de este modo obtener una reconstrucción promedio, a partir de la cual se obtuvo una serie suavizada

aplicando un promedio móvil de 3 pasos, así como  $\pm 1$  desviación estándar de todas las repeticiones. Las reconstrucciones se validaron considerando el RMSE y una correlación entre los datos reconstruidos y las series de tiempo de datos instrumentales (1940 – 2011) de la región. De igual manera las reconstrucciones se contrastaron contra reconstrucciones climáticas propuestas para la región de interés, Pacífico Norte, o para el hemisferio norte, así como también se realizaron correlaciones múltiples contra forzantes climáticos, las series de tiempo se obtuvieron de la plataforma de la NOAA (<https://www.ncdc.noaa.gov/paleo-search/>).

### ***Predicción del clima a partir de las series de tiempo reconstruidas***

Previo a la predicción de las variables climáticas se realizó un análisis exploratorio del valor teórico de la cantidad de pasos que se puede avanzar en el tiempo hacia delante con el menor error posible en la predicción. Para el desarrollo de este procedimiento, lo primero que se realizó fue calcular los parámetros descriptivos no lineales y la reducción de ruido de las series de tiempo reconstruidas, tal como fue explicado anteriormente. La información aquí obtenida fue utilizada como insumo para calcular el máximo exponente de Lyapunov ( $\lambda$ ) considerando la metodología propuesta por Kantz (1994) utilizando el software TISEAN (rutina lyap\_k.exe).

$$\lambda_{max} = \lim_{t \rightarrow \infty} \lim_{\varepsilon \rightarrow 0} \frac{1}{t} \ln \left( \frac{|x(t) - x_{\varepsilon}(t)|}{\varepsilon} \right) \quad \text{Ecuación 5}$$

Donde  $X$  corresponde a la evolución en el tiempo de una variable de interés, considerándose que  $X$  hace referencia a una trayectoria en la dimensión de embedding, se asume que la distancia entre el punto inicial ( $x(t)$ ) y el punto contiguo ( $x_{\varepsilon}(t)$ ) es una pequeña perturbación (valor muy cercano a cero) la cual crece exponencialmente en el tiempo (Kantz, 1994).  $\lambda$  describe en una escala logarítmica la tasa de crecimiento o contracción de las pequeñas perturbaciones en diferentes direcciones en el espacio de fase. Posterior a la obtención de  $\lambda$  para cada una de las dimensiones de correlación evaluadas, se calculó su comportamiento graficando su evolución al avanzar en el número de iteraciones consideradas. Se calculó la pendiente ( $m$ ) entre el punto más bajo y la asíntota de cada una de las curvas.

$$p = \frac{1}{\lambda}$$

Ecuación 6

Para la predicción ( $p$ , Ecuación 6) se calculó el valor inverso de lambda ( $\lambda = m$ ), posteriormente se calculó  $(p*n)*r$ ; donde;  $n$ : corresponde al número de datos de la serie de tiempo evaluada y  $r$ : la escala de longitud mínima para buscar un vecino (1000). Esta ecuación permite definir el número promedio de pasos que se puede avanzar hacia el futuro con el menor error posible (Fraedrich, 1986; Yingyi, 1993).

Para la predicción de las variables climáticas se utilizó una RNA *feedforward* con un algoritmo de aprendizaje de retro propagación y una modelación “*step by step*”. La RNA en la capa de entrada se alimentó con las variables climáticas reconstruidas (cada variable pronosticada se trabajó de manera individual). Durante la modelación, se utilizó el 70% de la serie de tiempo para el entrenamiento, el 15% para la prueba y el 15% restante para la validación. La selección de los datos para cada una de las etapas se realizó de manera aleatoria. Se exploraron dos alternativas para la configuración de la capa intermedia de las RNA (una topología de una y dos capas de neuronas). Adicionalmente, se consideró un aprendizaje supervisado, en el cual se utilizó la misma variable a predecir. Finalmente, se entrenaron múltiples redes neuronales con un número ascendente de neuronas en la(s) capa(s) intermedia(s); que fueron desde 3 (3 x 3) hasta 15 (15 x 15) neuronas. Se definió una tasa de aprendizaje variable con un máximo de 1000 iteraciones durante el entrenamiento considerando una regulación Bayesiana. Posterior al entrenamiento, se evaluó la eficiencia de los modelos considerando RMSE y la varianza explicada del modelo. Tras el entrenamiento y la definición de los pesos de conexión entre las neuronas, se procedió a realizar la predicción avanzando en el tiempo un paso a la vez “*step by step*”, se definió la cantidad de pasos que se puede ir hacia delante, a partir de calcular  $\lambda$  (Kantz, 1994). Inicialmente se hizo un análisis exploratorio de cada una de las series de tiempo obtenidas a partir de la reducción de ruido para cada una de las variables, para este punto se realizaron un total de 100 repeticiones de cada uno de los modelos para obtener un valor promedio. Todos los resultados se evaluaron a partir del rendimiento de la modelación y predicción, así como de las correlaciones obtenidas entre los datos modelados (predicción) y las series de tiempo instrumentales (1940 – 2020 temperatura del aire y 1950 – 2020 TSM) de la región. Adicionalmente, se evaluó la coherencia de las predicciones al comparar las series de tiempo

obtenidas contra índices climáticos de escala decadal que tuviesen influencia atmosférica y oceánica en la región ([https://psl.noaa.gov/gcos\\_wgsp/Timeseries/](https://psl.noaa.gov/gcos_wgsp/Timeseries/)). A cada serie se le realizó un pre-procesamiento en el cual se promediaron paquetes de 3 años para hacer comparables los valores obtenidos a partir de las series modeladas. Posteriormente se les realizó un promedio móvil de 3 pasos para hacer emerger las oscilaciones decadales con mayor claridad.

## VII. RESULTADOS

### *Estandarización de las bases de datos biológicas y selección de especies*

En la columna de agua se identificó un total de 270 taxones de diatomeas planctónicas, seguido de 305 en la trampa de sedimentos y 241 en el núcleo de sedimentos. Lo cual representa un 50 % de similitud entre el núcleo y la trampa de sedimentos, y un 40 % de similitud entre el núcleo y la columna de agua. Los valores del índice de similitud de Simpson entre los tres estratos reflejan una similitud intermedia. Tras identificar aquellas especies presentes en los tres estratos, la matriz se redujó a 112 taxones. Posteriormente, se realizó una reducción del número de taxones en función de la abundancia, seguida de la formación de 5 asociaciones de especies acordes con su afinidad a tipos de agua tales como: cálida, cosmopolita, fría, tropical, oceánica y una asociación de taxa por el tipo de ambiente bentónico. Obteniéndose 29 ítems (especies y ensambles).

### *Evaluación de parámetros no lineales y eliminación del ruido de las series de tiempo*

En la Tabla 1. se muestran los parámetros descriptivos de las series de tiempo no lineales, tales como la dimensión de correlación (D), dimensión de embedding (m) y el retraso (d) para las series originales y los parámetros que se utilizaron para definir la mejor reducción de ruido, a partir del RMSE y SNR. Para las series de tiempo de irradiancia, rapidez del viento, zona eufótica, presión atmosférica y profundidad de la capa de mezcla fue necesario reducir la dimensión de correlación, y con ello se obtuvieron las mejores relaciones entre

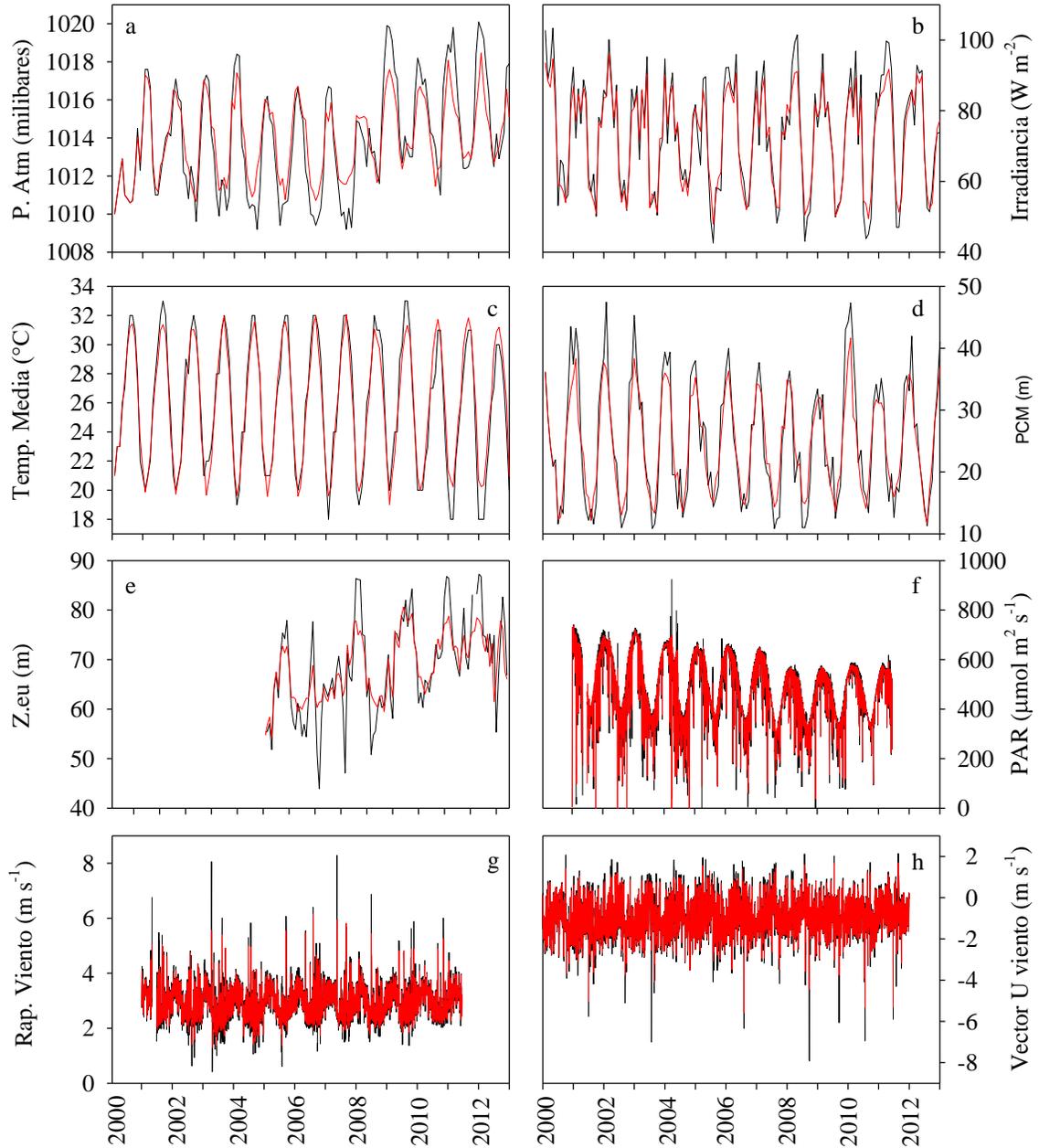
RMSE y SNR. Sin embargo, no en todos los casos se respetó este criterio, por ejemplo, con la serie de tiempo de la profundidad de la zona eufótica, se seleccionaron las series con el menor SNR y se le dio un mayor peso al RMSE, ya que al visualizar los resultados donde se evaluaron los diferentes parámetros descriptivos (datos no mostrados), estos no fueron coherentes visualmente (datos no mostrados) con lo esperado para la reducción del ruido con relación a la serie de tiempo original. En la Figura 5, se compilan las series de tiempo originales, con su respectiva reducción de ruido que fueron seleccionadas para posteriores análisis. Se observó una atenuación en los máximos y mínimos sin que se presente una pérdida de información del comportamiento original de la serie de tiempo.

**Tabla 1.** Recopilación de parámetros descriptivos de las series de tiempo instrumentales; dimensión de correlación (D), dimensión de embedding (m), el retraso (d), distancia euclidiana (e), error cuadrático media (RMSE) y relación señal-ruido (SNR). Rapidez del viento (Rap. Viento), temperatura media del aire (Temp. Media), profundidad de la zona eufótica ( $Z_{eu}$ ), radiación fotosintéticamente activa (PAR), presión atmosférica (P-atm), profundidad de la capa de mezcla (PCM), vector del viento U (VU) y V (VV).

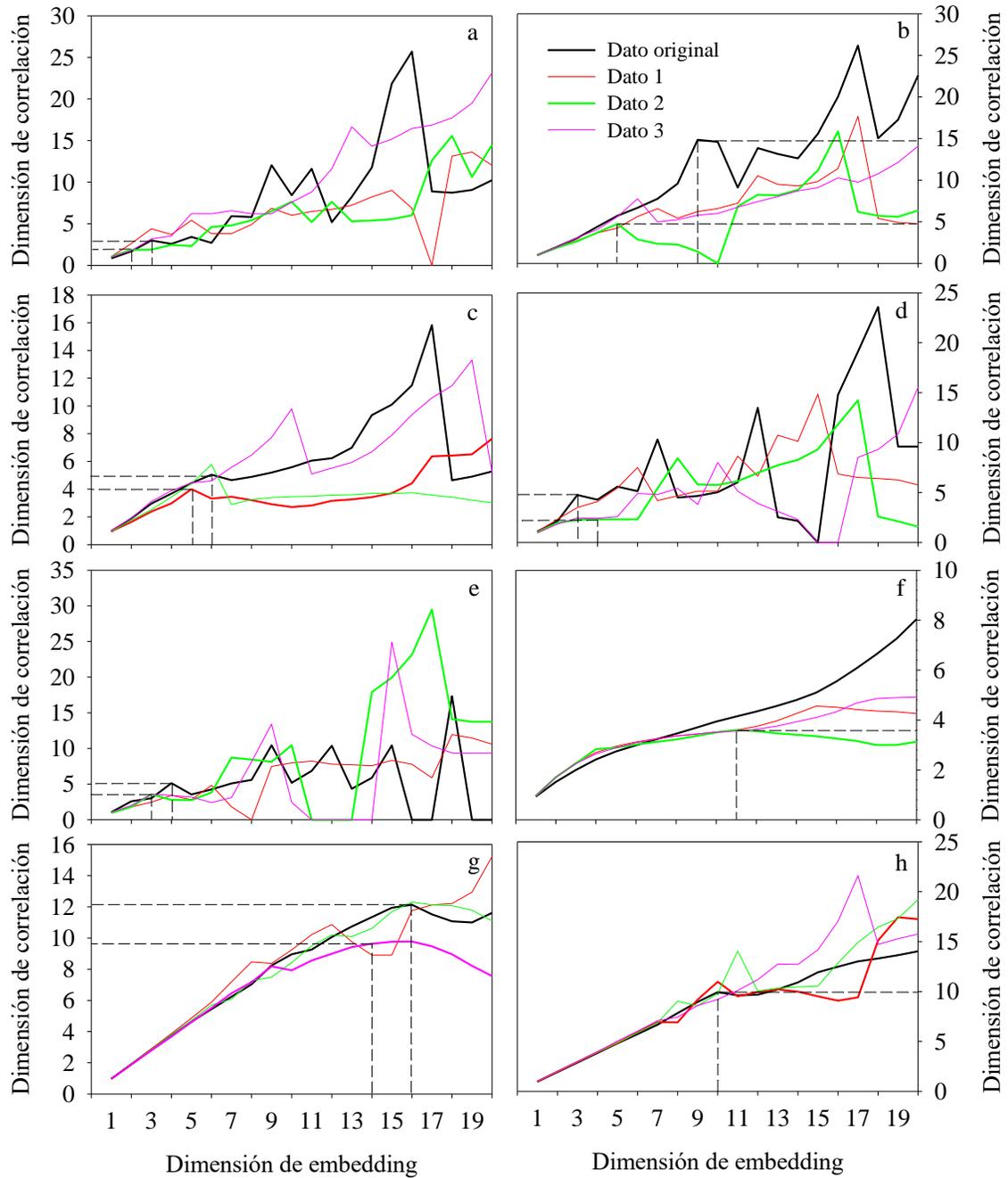
	datos originales			reducción de ruido			RMSE	SNR
	D	d	m	D	m	e		
Irradiancia	3	14.85	9	2	3		3.736	2.43
Rap. Viento	4	12.14	16	3	2		0.169	3.128
Temp. Media	4	5.4	6	3	2		0.297	1.34
$Z_{eu}$	2	5.13	4	3	2	e	8.882	7.95
PAR	8	5.53	14	9	2		16.23	1.025
P-atm	3	3	3	9	2		1.522	5.707
PCM	3	4.75	3	3	5		3.497	4.144
VU	5	9.93	10	6	2		0.196	2.008
VV	5	6.36	6	5	3	e	0.311	1.233

En la Figura 6, se compilan las relaciones entre la D y m como método para identificar la reducción del ruido. Por ejemplo, para las series de tiempo de la temperatura media del aire (Figura 6c), PAR (Figura 6f) y rapidez del viento (Figura 6g) se nota una reducción en el valor de D en función a la serie original. Por otro lado, encontramos que en el caso de las series de tiempo de la presión atmosférica (Figura 6a), irradiancia (Figura 6b), PCM (Figura 6d), profundidad de la zona eufótica (Figura 6e) y presión atmosférica (Figura 6a), aún se observó la presencia de un posible ruido aditivo que no pudo ser eliminado de las series de tiempo.

En el caso del PAR (Figura 6f), el RMSE y SNR definen la mejor eficiencia en la reducción del ruido, sin embargo, al evaluar  $D/m$  bajo los parámetros ( $m=2$ ;  $d=10$ ;  $RMSE=34.91$ ;  $SNR=1.01$ , datos no mostrados) encontramos una mayor eficiencia en la reducción del ruido, determinado por un menor valor en la dimensión de correlación ( $D = 3.6$ ;  $m = 11$ ). En el caso de PCM, se definió en función al mayor SNR, pero al evaluar el RMSE y la relación  $D/m$  (Figura 6d), encontramos que el mejor desempeño se obtuvo bajo los parámetros ( $m=5$ ;  $d=3$ ;  $RMSE=2.281$ ;  $SNR=1.32$ , datos no mostrados) y una  $D/m$  ( $D=2.32$ ;  $m=4$ ), respectivamente. Para el componente U del viento, encontramos que a pesar de seleccionar los mejores valores RMSE y SNR, al evaluar la relación  $D/m$  (Figura 6h), estas muestran un comportamiento de aleatoriedad en las series, lo cual hace difícil la selección de los mejores parámetros para la reducción del ruido en estas series de tiempo en particular.



**Figura 5.** Recopilación de las series de tiempo instrumentales: a) Presión atmosférica, b) Irradiancia, c), Temperatura media del aire, d) Profundidad de la capa de mezcla, e) Profundidad de la zona eufótica, f) PAR, g) Rapidez del viento, h) vector U del viento. Datos originales (línea negra) y valores predictivos con mayor reducción de ruido (línea roja).



**Figura 6.** Relación entre la dimensión de correlación ( $D$ ) y la dimensión de embedding ( $m$ ). a) presión atmosférica, b) irradiancia, c), temperatura media del aire, d) profundidad de la capa de mezcla, e) profundidad de la zona eufótica, f) PAR, g) rapidez del viento, h) vector  $U$  del viento. Las líneas punteadas marcan el punto de saturación de la dimensión de correlación para las series de tiempo original (dato original) y con reducción de ruido (Dato 1, 2 y 3 definidos bajo diferentes parámetros descriptivos).

✚ **Definición de la arquitectura de la RNA tipo SOM a partir de los datos instrumentales y biológicos**

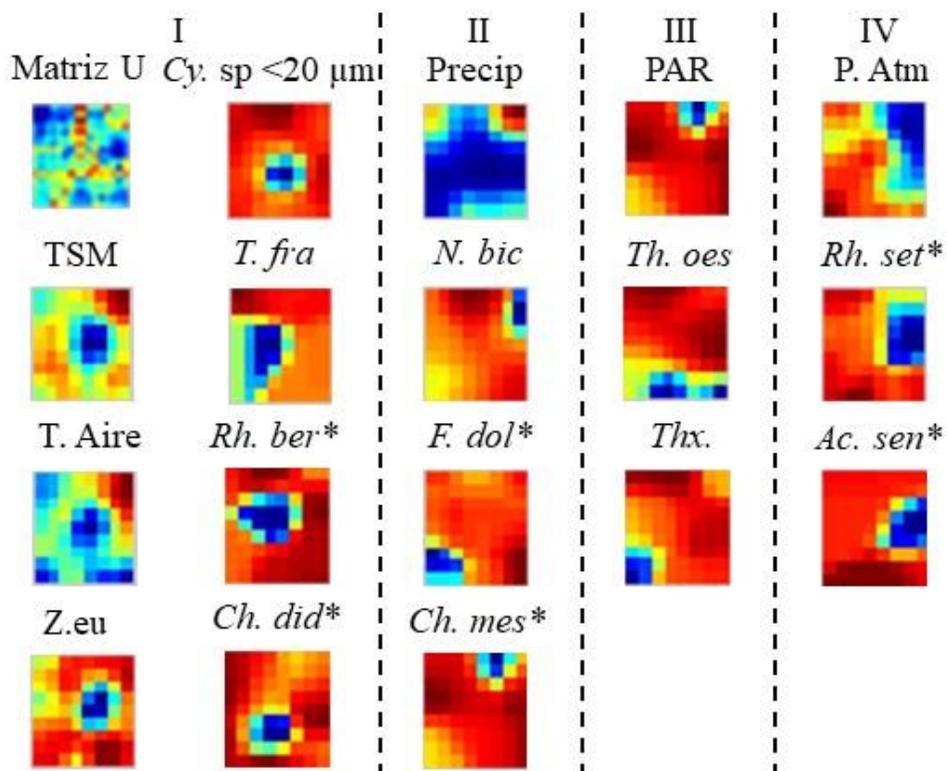
La base de datos biológica y ambiental se dividió de manera aleatoria en tres porciones, un 70% para el entrenamiento y el 30% restante para la evaluación y validación. El número de neuronas óptimas que se estimó fue  $n= 68$ . Con esta referencia y basándonos en el error topológico (TE) y de quantizacion (QE) se analizaron diferentes tamaños de las arquitecturas cercanas al número de neuronas óptimas. Para la base de datos con ruido, la arquitectura elegida fue de 72 (9 x 8) neuronas, con un TE= 0.060 y un QE= 3.096, mientras que para la base de datos sin ruido fue de 70 (10 X 7) neuronas, con un TE= 0.076 y QE= 3.211. En la Tabla 2 se muestra un resumen de diferentes arquitecturas evaluadas para la definición del tamaño de arquitectura.

**Tabla 2.** Parámetros de rendimiento de las mejores dimensiones de las redes neuronales tipo SOM considerando el tamaño óptimo  $n= 5* \sqrt{\text{num. muestras}}$  propuesto por Vesanto (2000); EQ (Error de Quantization) y ET (Error Topológico). Negritas corresponden a las arquitecturas seleccionadas.

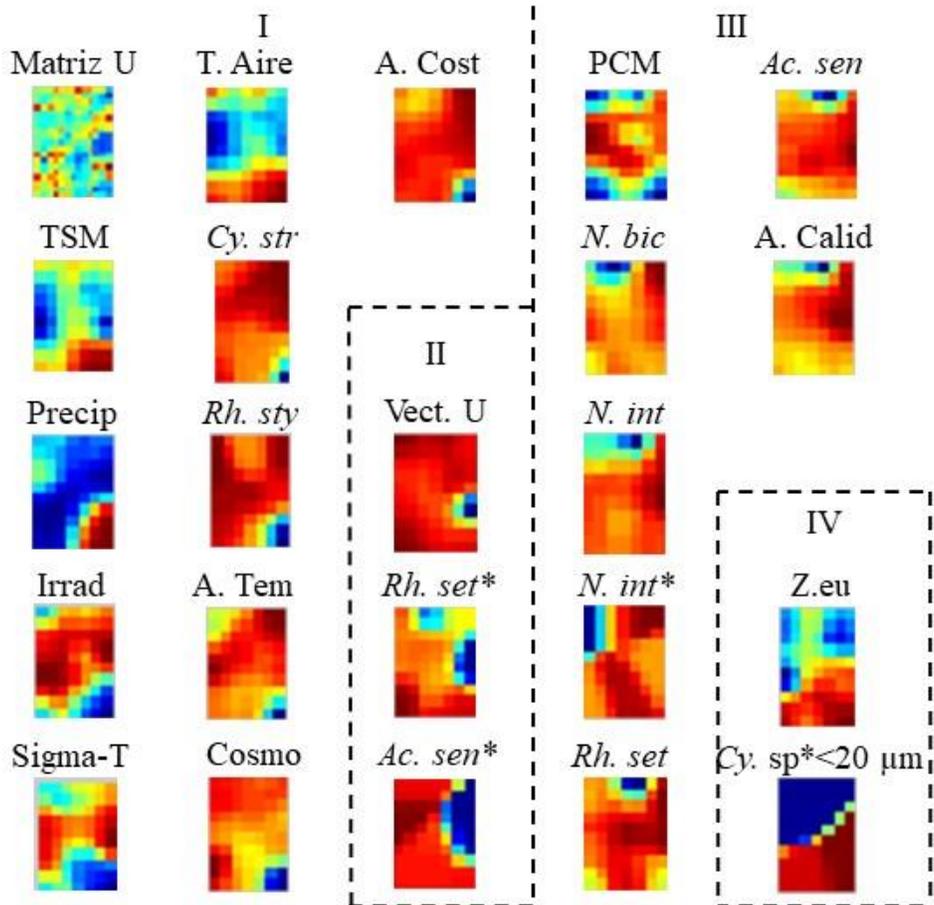
Arquitectura	Datos originales		Reducción de ruido	
	EQ	ET	EQ	ET
10 X 6	3.431	0.082	3.423	0.049
10 X 7	3.237	0.093	<b>3.211</b>	<b>0.076</b>
9 X 7	3.419	0.120	3.280	0.049
9 X 8	<b>3.096</b>	<b>0.060</b>	3.249	0.049
8 X 8	3.457	0.071	3.387	0.071

En las Figuras 7 y 8 se presentan las salidas de la red neuronal SOM. Los patrones de coloración similares simbolizan agrupaciones entre las abundancias relativas de las especies y las variables instrumentales para el modelo con y reducción de ruido. Los colores cálidos hacen referencia a valores altos y los colores fríos, valores bajos asignados a las unidades de las salidas del SOM. En el primer modelo (Figura 7) se formaron cuatro grupos. I) Constituido por la TSM, temperatura del aire y la profundidad de la zona eufótica, asociados a *Cyclotella* spp. ( $< 20 \mu\text{m}$ ), *Thalassionema frauenfeldii*, *Rhizosolenia bergonii* y *Chaetoceros didymus*. Dentro del grupo las especies presentaron un patrón similar con la profundidad de la zona eufótica, mientras que inverso con las demás variables. II) definido

por la precipitación asociada por un patrón de coloración inverso con *Nizschia bicapitata*, *Fragilariopsis doliolus*, *Chaetoceros messanensis*. III) definido por el PAR asociado con *Thalassiosira oestrupii*, y *Thalassiotrix* sp., presentando un patrón de coloración inverso al PAR. IV) la presión atmosférica presentó un patrón similar a *Rhizosolenia setigera* y *Actinoptychus senarius*.



**Figura 7.** Matriz-U obtenida con la red neuronal SOM de 9 x 8 neuronas, representando las relaciones entre las abundancias relativas de las especies y las variables sin reducción de ruido. Las líneas punteadas marcan la separación entre cada grupo. Temperatura superficial del mar (TSM), temperatura media del aire (T. Aire), zona eufótica (Z.eu), precipitación (Precip), radiación fotosintéticamente activa (PAR), presión atmosférica (P. Atm). *Cyclotella* spp. (<20 μm) (*Cy. sp <20 μm*), *Thalassionema frauenfeldii* (*T. fra*), *Rhizosolenia bergonii* (*Rh. ber\**), *Chaetoceros didymus* (*Ch. did\**), *Nizschia bicapitata* (*N. bic*), *Fragilariopsis doliolus* (*F. dol\**), *Chaetoceros messanensis* (*Ch. mes\**), *Thalassiosira oestrupii* (*Th. oes*), *Thalassiotrix* sp. (*Thx.*), *Rhizosolenia setigera* (*Rh. set\**), *Actinoptychus senarius* (*Ac. sen\**). (\*) Especies presentes en la columna de agua.



**Figura 8.** Matriz-U obtenida con la red neuronal SOM de 10 x 7 neuronas, representando las relaciones entre las abundancias relativas de las especies y las variables con reducción de ruido. Las líneas punteadas marcan la separación entre cada grupo. Temperatura superficial del mar (TSM), precipitación (Precip), irradiancia (Irrad), densidad (sigma-T), temperatura media del aire (T. Aire), vector U del viento (Vect. U), profundidad de la capa de mezcla (PCM), profundidad de la zona eufótica (Z.eu), *Cyclotella striata* (*Cy. str*), *Rhizosolenia styliformis* (*Rh. sty*), asociación de taxa de afinidad templada (A. Tem), Costeras (A. Cost), Cosmopolitas (Cosmo), Cálida (A. Calid), *Cyclotella* spp. (<20 μm) (*Cy. sp\** <20 μm), *Nitzschia bicapitata* (*N. bic*), *Nitzschia interruptestriata* (*N. int* - *N. int\**), *Rhizosolenia setigera* (*Rh. set* - *Rh. set\**), *Actinoptychus senarius* (*Ac. sen* - *Ac. sen\**). (\*) Especies presentes en la columna de agua.

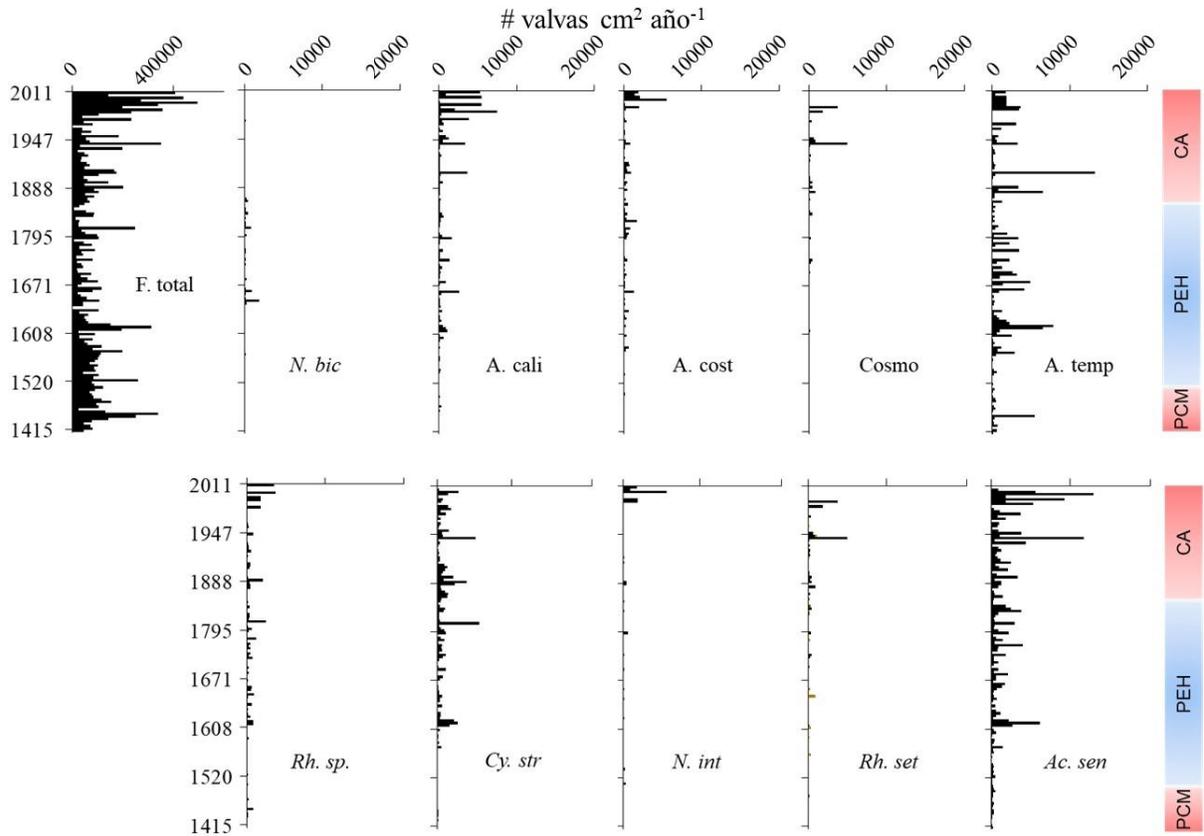
En la Figura 8 se observa la formación de 4 grupos, el primero se caracteriza por *Cy. striata*, *Rh. styliformis* y una asociación de diatomeas cosmopolitas, de ambientes templados y costeros, que presentan un patrón inverso a la TSM, temperatura del aire y precipitación, mientras que un patrón similar con la irradiancia y densidad (sigma-T). El segundo grupo se caracterizó por vincular *Rh. setigera* y *Ac. senarius* con el componente U del viento. El tercer grupo, definido por *N. bicapitata*, *Rh. setigera*, *Ac. senarius*, una asociación de taxones con

afinidad a aguas cálidas y *N. interruptestriata* presentes tanto en la columna de agua como de la trampa de sedimentos, asociados de manera inversa con el patrón de la profundidad de la capa de mezcla. El último grupo fue definido por *Cyclotella* spp. presente en la columna de agua compartiendo un patrón similar con la profundidad de la zona eufótica.

#### **Variabilidad de las diatomeas planctónicas en el núcleo de sedimentos**

Después de identificar las asociaciones de diatomeas planctónicas correspondientes a la columna de agua (zona eufótica y trampa de sedimentos) que se relacionaron con las diferentes variables ambientales descritas anteriormente por medio del uso de una RNA tipo SOM, se procedió a describir el flujo total y el flujo de las asociaciones de diatomeas previamente seleccionadas que quedaron preservadas en el registro sedimentario (Figura 9). De acuerdo con Matthews & Briffa (2005) y Diaz *et al.* (2011), el alcance temporal del núcleo permitió documentar cómo respondió la región de estudio a periodos climáticos seculares como el final del periodo cálido medieval (PCM: 1417 – 1500 AD), la pequeña edad de hielo (PEH: 1500 – 1850 AD) y el cálido actual (CA: 1850 – 2011 AD). El final del PCM se caracterizó por presentar un flujo promedio de  $1.19 \times 10^5$  valvas  $\text{cm}^2 \text{año}^{-1}$ , alcanzando un valor máximo de  $3.42 \times 10^5$  valvas  $\text{cm}^2 \text{año}^{-1}$  y mínimo de  $2.48 \times 10^4$  valvas  $\text{cm}^2 \text{año}^{-1}$ . Adicionalmente, entre 1547 y 1570 AD se registró un incremento en el flujo con un valor promedio de  $1.9 \times 10^5$  valvas  $\text{cm}^2 \text{año}^{-1}$ . Durante el periodo de la PEH el flujo promedio de valvas fue de  $7.6 \times 10^4$  valvas  $\text{cm}^2 \text{año}^{-1}$ . Se alcanzó un valor máximo de  $3.1 \times 10^5$  valvas  $\text{cm}^2 \text{año}^{-1}$  para principios de 1618 AD, mientras que el valor mínimo fue de  $8.3 \times 10^3$  valvas  $\text{cm}^2 \text{año}^{-1}$ . Entre 1500 y 1620 AD el flujo se caracterizó por registrar valores ( $\bar{X} = 1.1 \times 10^5$  valvas  $\text{cm}^2 \text{año}^{-1}$ ) por encima del promedio histórico (1417 – 2011 AD) con tres incrementos importantes en 1524, 1571 y 1618 AD. Entre 1626 y 1847 AD se registró una disminución en el flujo ( $\bar{X} = 1.1 \times 10^5$  valvas  $\text{cm}^2 \text{año}^{-1}$ ) con valores cercanos al promedio histórico. El periodo CA presentó un flujo promedio ( $1.35 \times 10^5$  valvas  $\text{cm}^2 \text{año}^{-1}$ ) encontrándose por encima del promedio de la serie de tiempo. Durante el CA se registró un valor máximo de  $4.96 \times 10^5$  valvas  $\text{cm}^2 \text{año}^{-1}$  y un mínimo de  $2.3 \times 10^4$  valvas  $\text{cm}^2 \text{año}^{-1}$ . Los periodos 1930 – 1955 AD y 1994 – 2011 AD se caracterizaron por presentar un incremento en el flujo total de valvas ( $1.1$  y  $2.6 \times 10^5$  valvas  $\text{cm}^2 \text{año}^{-1}$ , respectivamente).

De la asociación de diatomeas seleccionadas a partir del análisis con la RNA tipo SOM, se observó que entre 1417 y 1600 AD el aporte que tuvieron las asociaciones al flujo total fue bajo y no presentaron la misma tendencia. Entre 1600 y 2011 AD el flujo de las asociaciones de diatomeas presentó similitudes con el flujo total (Figura 9). Al analizar el comportamiento separado de las asociaciones de interés (análisis SOM) encontramos que *Cyclotella striata* se registró en todo el núcleo de sedimentos con un flujo promedio de  $6.2 \times 10^3$  valvas  $\text{cm}^2$   $\text{año}^{-1}$ , seguida de las diatomeas con afinidad a aguas templadas y *Actinoptychus senarius* ( $\bar{X}= 1.3$  y  $1.2 \times 10^3$  valvas  $\text{cm}^2$   $\text{año}^{-1}$ , respectivamente) estando un orden de magnitud por encima de las demás asociaciones. Los mayores aportes que presentó *Cy. striata* ocurrieron durante 1800 – 2011 AD, en cambio las diatomeas de afinidad templada presentaron pulsos importantes de manera esporádica entre la PEH y el CA, *Ac. senarius* presentó incrementos en pulsos en la segunda parte de la PEH y principio del CA, con los mayores aportes entre 1930 – 1947 AD y 2000 – 2011 AD. *Rhizosolenia setigera* y *Nitzschia bicapitata*, estuvieron presentes durante la PEH y CA con flujos bajos y pulsos ocasionales. De igual manera, *N. interruptestriata* y *Rhizosolenia* spp. presentaron incrementos esporádicos y sus mayores aportes ocurrieron durante el 2000 y 2011 AD. Las asociaciones de diatomeas de afinidad cálida, costeras y cosmopolita se encontraron con mayores abundancias entre 1999 y 2011 AD, y se registraron en el núcleo hasta 1500 AD aproximadamente.



**Figura 9.** Flujo total de valvas de diatomeas en el núcleo de sedimentos (C-36), acompañado del flujo de valvas de las asociaciones de diatomeas definidas por el análisis de la RNA tipo SOM. *Nitzschia bicapitata*, (*N. bic*), asociación de agua cálida (A. cali), costeras (A. cost), cosmopolitas (Cosmo), templada (A. temp), *Rhizosolenia* sp. (*Rh. sp.*), *Cyclotella striata* (*Cy. str*), *N. interruptestriata* (*N. int*), *Rh. setigera* (*Rh. set*) y *Actinoptychus senarius* (*Ac. sen*). Se marcan con sombreado rojo los periodos cálido actual (CA) y el periodo cálido medieval (PCM) y en azul la pequeña edad de hielo (PEH).

### 🌈 **Reconstrucción del clima a partir de diatomeas planctónicas**

A partir de la selección de las especies y asociaciones de diatomeas con el potencial para reconstruir las condiciones climáticas para la región sur del Golfo de California, se procedió a evaluar diferentes modelos (Tabla 3). Entre los 8 modelos evaluados, el que presentó los mejores resultados fue en el cual se consideraron todas las asociaciones definidas por el SOM, con una resolución temporal de 3 años, se complementaron los datos faltantes con un suavizado de ponderación gaussiana y los datos biológicos de la trampa, así como del núcleo de sedimentos se utilizaron sin normalizar (opción 8 de Tabla 3).

En la Tabla 4 se resumen en detalle los resultados para las diferentes topologías y número de neuronas evaluadas para el modelo # 8. Se ejemplifican los valores de la varianza explicada ( $r$ ), el RMSE de modelación y reconstrucción, así como sus respectivas correlaciones de Spearman para las series de tiempo promedio y de promedio móvil de 3 pasos. Para las redes de una sola capa de neuronas, la varianza explicada presentó valores entre 0.982 y 0.997. La variable que presentó los mejores resultados fue la temperatura del aire para las redes de 5, 6 y 7 neuronas en la capa oculta. El RMSE de los modelos fue 3.14, 2.88 y 2.72 respectivamente. Los RMSE de la reconstrucción fueron 1.5, 1.69 y 1.96, respectivamente. Se encontraron correlaciones significativas para la serie de tiempo promedio ( $r^2$  [24] = 0.5, 0.47 y 0.5;  $p < 0.005$ , respectivamente) y para el promedio móvil de 3 pasos ( $r^2$  [24] = 0.64, 0.62 y 0.66;  $p < 0.005$ , respectivamente). En el caso de la TSM, las mejores reconstrucciones se obtuvieron para una red de dos capas, con topologías de 13, 14 y 15 neuronas en la capa oculta, las cuales presentaron una varianza explicada con valores mayores a 0.99. El RMSE de los modelos fueron respectivamente de 0.74, 0.68 y 0.75. Mientras que los RMSE de la reconstrucción fueron de 2.49, 1.84 y 1.9, respectivamente. Se obtuvo correlaciones significativas para el valor promedio ( $r^2$  [21] = 0.36, 0.40 y 0.22;  $p < 0.005$ , respectivamente) y para el promedio móvil de 3 pasos ( $r^2$  [21] = 0.50, 0.57 y 0.46;  $p < 0.005$ , respectivamente).

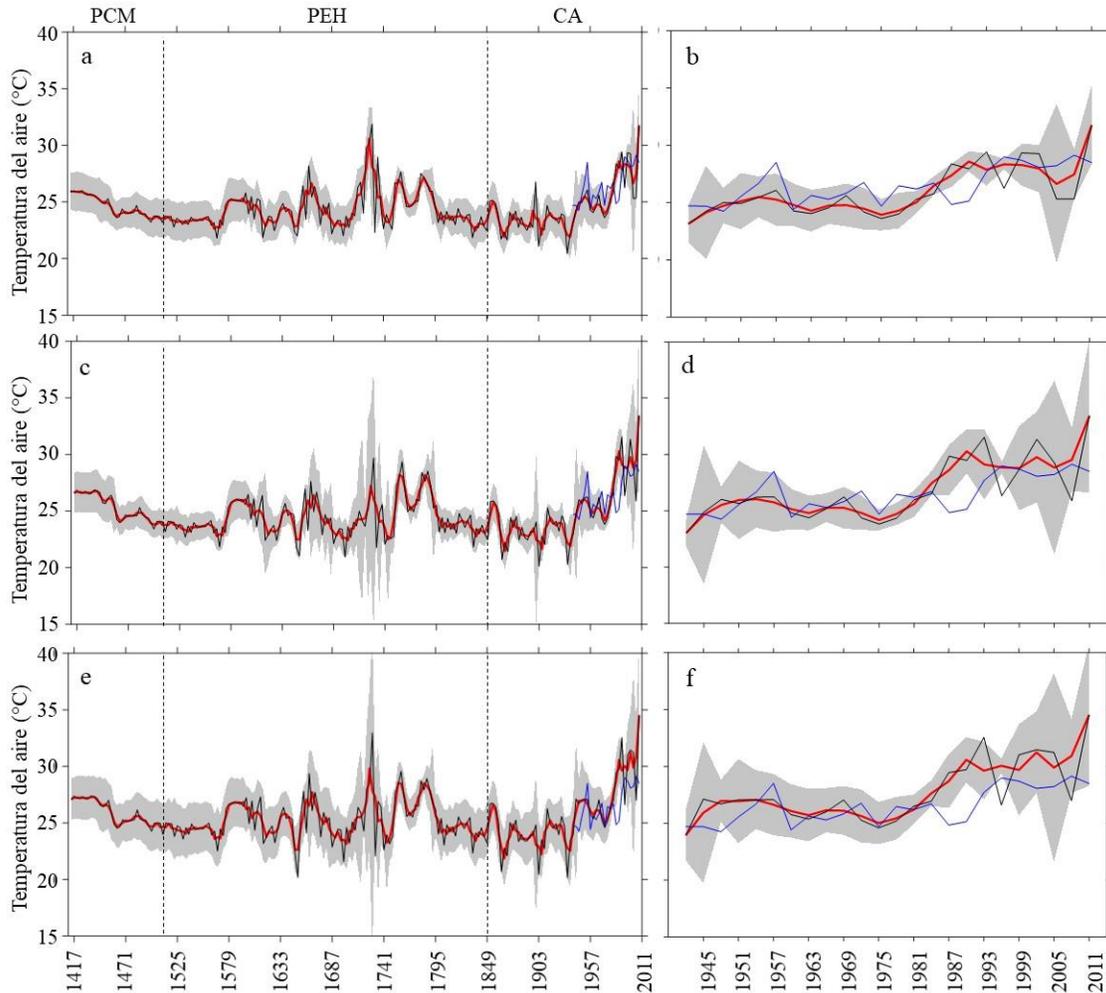
En la Figura 10 (a, c y e) se presentan las reconstrucciones de la temperatura del aire para los últimos 600 años utilizando diferentes topologías (5, 6 y 7 neuronas respectivamente), en estas se observa que tanto para el valor promedio como para el promedio móvil (3 pasos) se obtuvieron correlaciones significativas entre los datos reconstruidos y los datos instrumentales (Tabla 4). Para el periodo cubierto por la base de datos instrumental utilizada en este estudio (Figura 10b, d y f), se observó que los valores en su mayoría se distribuyeron dentro del área sombreada que muestra el intervalo de confianza de una desviación estándar de las reconstrucciones obtenidas a partir de cada modelo. Se seleccionó el modelo con una topología de 6 neuronas, el cual presentó una variabilidad intermedia (desviación estándar) en relación con las demás redes evaluadas para reconstruir la temperatura del aire.

**Tabla 3.** Resumen de los criterios considerados en cada uno de los modelos evaluados para la TSM y temperatura del aire; Criterio (GISOM: grupo 1 del SOM, *sp* abun: especies más abundantes, SOM: todas las asociaciones identificadas por el SOM); Esc. Temp: escala temporal; Técnica Vac. Info: Técnica para llenar vacíos de información; Capas: número de capas de la RNA; Proces: procesamiento de los datos (D.N: datos normalizados, D.S.N: datos sin normalizar) y Resul: resultado cualitativo de cada variante dentro de los modelos (P.S: poco satisfactorio, M.S: medianamente satisfactorio, S: satisfactorio).

Modelos	Criterio	Esc. Temp	Técnica Vac. Info	Capas	Proces	Número neuronas	Resul
1	GISOM	Original	I. puntos adyacentes	1 capa	D.N D.S.N	5 – 15 5 – 15	P.S P.S
				2 capas	D.N D.S.N	5 – 15 5 – 15	M.S P.S
2	GISOM + <i>sp</i> abun	Original	I. puntos adyacentes	1 capa	D.N D.S.N	5 – 15 5 – 15	P.S M.S
				2 capas	D.N D.S.N	5 – 15 5 – 15	M.S M.S
3	SOM	Sin procesar	I. puntos adyacentes	1 capa	D.N D.S.N	5 – 15 5 – 15	M.S M.S
				2 capas	D.N D.S.N	5 – 15 5 – 15	M.S M.S
4	SOM	Eq. 5 años	I. puntos adyacentes	1 capa	D.N D.S.N	5 – 15 5 – 15	M.S M.S
				2 capas	D.N D.S.N	5 – 15 5 – 15	M.S M.S
5	GISOM	Original	Pond. Gaussiana	1 capa	D.N D.S.N	5 – 15 5 – 15	P.S M.S
				2 capas	D.N D.S.N	5 – 15 5 – 15	P.S M.S
6	GISOM + <i>sp</i> abun	Original	Pond. Gaussiana	1 capa	D.N D.S.N	5 – 15 5 – 15	M.S M.S
				2 capas	D.N D.S.N	5 – 15 5 – 15	M.S M.S
7	SOM	Original	Pond. Gaussiana	1 capa	D.N D.S.N	5 – 15 5 – 15	M.S M.S
				2 capas	D.N D.S.N	5 – 15 5 – 15	M.S M.S
8	SOM	Eq. 3 años	Pond. Gaussiana	1 capa	D.N D.S.N	5 – 15 5 – 15	S S
				2 capas	D.N D.S.N	5 – 15 5 – 15	S S

**Tabla 4.** Parámetros descriptivos de la eficiencia de las reconstrucciones a partir de redes neuronales artificiales *feedforward* para una topología de una y dos capas de neuronas en la capa intermedia. Regresión del modelo (Reg. Mod), RMSE del modelo (RMSE mod), RMSE de la reconstrucción (RMSE rec), correlación de Spearman para el promedio de las repeticiones de la modelación ( $\rho$ ), correlación para el promedio móvil de 3 pasos para la serie de tiempo promediada ( $\rho$  m3). Negritas denotan las arquitecturas seleccionadas con el mejor desempeño.

TSM						Temperatura del aire			
# neuronas	Reg. Mod	RMSE mod	RMSE rec	$\rho$	$\rho$ m3	RMSE mod	RMSE rec	$\rho$	$\rho$ m3
5	0.982	2.630	1.380	0.397	0.629	3.144	1.506	<b>0.509</b>	<b>0.640</b>
6	0.985	2.466	1.668	0.336	0.583	2.889	1.690	<b>0.471</b>	<b>0.622</b>
7	0.987	2.288	1.997	0.425	0.677	2.725	1.964	<b>0.509</b>	<b>0.660</b>
8	0.990	2.086	2.285	0.447	0.685	2.472	2.775	0.535	0.676
9	0.992	1.767	2.681	0.355	0.638	2.183	3.583	0.427	0.555
10	0.994	1.550	2.317	0.172	0.463	1.960	3.496	0.234	0.308
11	0.995	1.339	2.038	-0.162	-0.136	1.738	3.431	-0.144	-0.148
12	0.996	1.192	1.962	0.023	0.160	1.564	3.614	-0.116	-0.098
13	0.997	1.118	2.003	-0.125	-0.087	1.478	3.750	-0.228	-0.233
14	0.997	1.023	1.847	-0.068	-0.028	1.341	3.510	-0.267	-0.302
15	0.997	0.980	2.250	-0.194	-0.214	1.272	4.110	-0.265	-0.316
5 x 5	0.990	1.969	2.209	0.359	0.472	2.250	2.399	0.404	0.550
6 x 6	0.994	1.554	2.979	0.352	0.537	1.799	3.338	0.332	0.485
7 x 7	0.996	1.252	2.010	0.184	0.419	1.490	3.547	0.020	0.082
8 x 8	0.998	0.969	1.885	0.320	0.436	1.253	3.082	0.119	0.302
9 x 9	0.998	0.879	2.509	0.206	0.393	1.141	4.038	0.052	0.187
10 x 10	0.999	0.784	1.911	0.300	0.428	0.955	3.052	-0.020	0.068
11 x 11	0.998	0.801	2.092	0.383	0.547	1.010	2.998	0.190	0.336
12 x 12	0.999	0.729	1.982	0.104	0.284	0.894	3.158	-0.221	-0.270
13 x 13	0.999	0.746	2.496	<b>0.366</b>	<b>0.509</b>	0.933	3.437	0.135	0.341
14 x 14	0.999	0.698	1.847	<b>0.403</b>	<b>0.576</b>	0.796	3.137	-0.120	-0.147
15 x 15	0.999	0.754	1.902	<b>0.223</b>	<b>0.465</b>	0.891	3.228	-0.156	-0.108

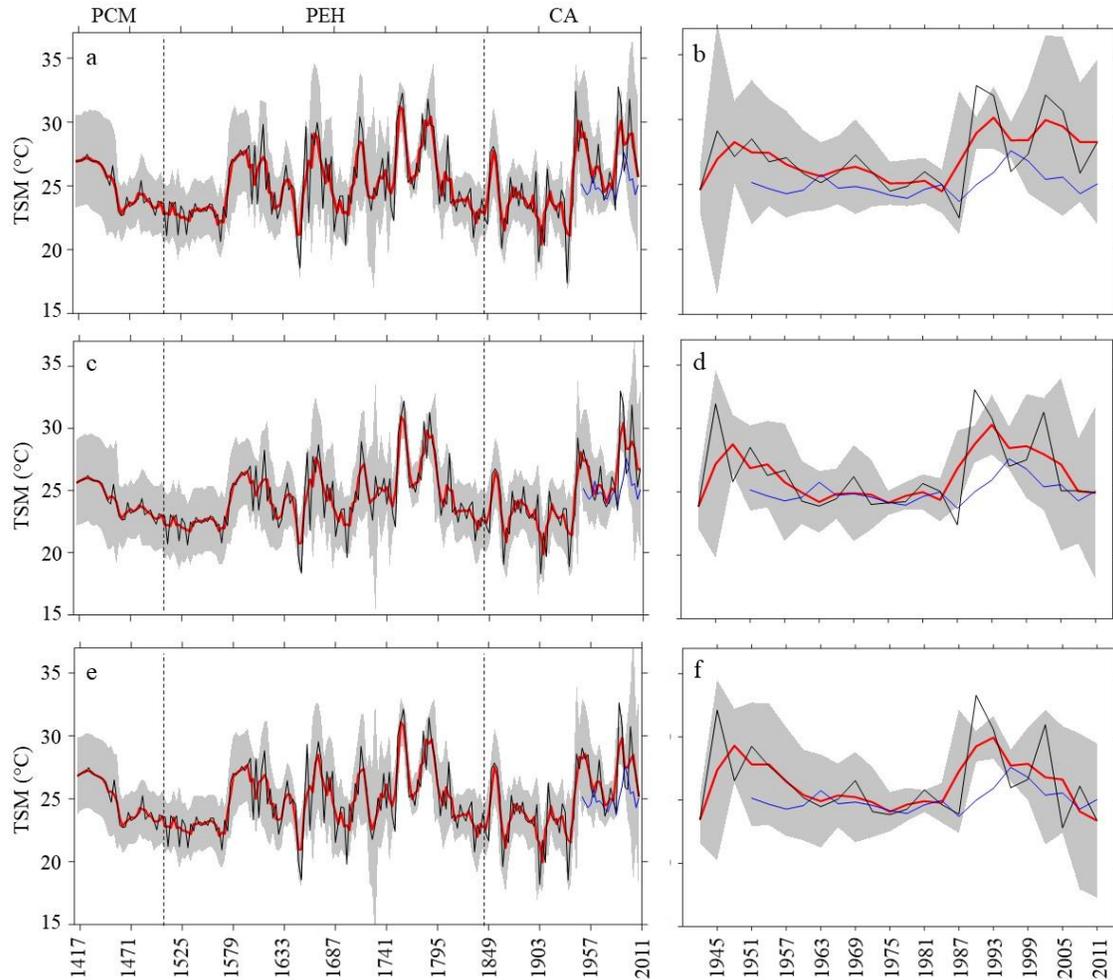


**Figura 10.** Reconstrucción de la temperatura del aire entre el 1417 y 2011 AD utilizando una RNA tipo “feedforward” de retro propagación, con topología de una capa de neuronas. RNA de: a) 5 neuronas, c) 6 neuronas, e) 7 neuronas. Periodo instrumental de la RNA de: b) 5 neuronas, d) 6 neuronas y f) 7 neuronas. El área sombreada corresponde a una desviación estándar con un promedio móvil de 3 pasos, la línea negra representa el promedio de 100 corridas de cada modelo. La línea roja representa el promedio móvil de 3 pasos y la línea azul los datos instrumentales (1941 – 2011) de la temperatura máxima del aire para los meses de invierno (enero-abril). Las líneas punteadas verticales delimitan un segmento del periodo cálido medieval (PCM) y los periodos climáticos de la pequeña edad de hielo (PEH) y el cálido actual (AC).

Las reconstrucciones presentaron buenos resultados dado que reconstruyen tanto las magnitudes, como las tendencias generales de la temperatura del aire. La serie de tiempo obtenida a partir de la RNA de 6 neuronas (Figura 10c y d) describe un descenso en la temperatura para el fin del PCM y principio de la PEH (~ 1580 AD). Durante la PEH se registró una alta variabilidad climática, donde se identificaron tres periodos de incremento

de temperatura (1579 – 1615 AD, 1663 – 1675 AD y 1759 – 1795 AD) que estuvieron en promedio 1.2 °C por encima del promedio histórico (24.67 °C). A partir de 1800 AD que corresponde a la transición entre la PEH y el CA se registró un descenso en la temperatura del aire de 1.22 °C por debajo del promedio histórico, el cual se mantuvo hasta 1940 AD, cuando se observó un incremento moderado de la temperatura hasta 1980 AD, que continuó con un calentamiento pronunciado alcanzando temperaturas de 3.7 °C por encima del promedio histórico y que se extendió hasta el 2011 AD (final de la serie de tiempo analizada).

En la Figura 11 se presentan las reconstrucciones de la TSM para los últimos 600 años utilizando las arquitecturas de dos capas que presentaron el mejor rendimiento (13 x 13, 14 x 14 y 15 x 15 neuronas). Para cada una de las reconstrucciones se observó un buen ajuste en los modelos [varianza explicada y RMSE (modelo y reconstrucción)], dando lugar a que las reconstrucciones promedio y promedios móviles (3 pasos) presentaran correlaciones significativas contra los datos instrumentales (Tabla 4). Para el periodo en el que se cuenta con datos instrumentales (Figura 11b, d y f) se validaron las reconstrucciones evaluándose tanto las correlaciones como el RMSE de cada una de las series de tiempo, así como el grado de desfase de los valores reconstruidos en comparación con los datos instrumentales. Las series de tiempo reconstruidas representaron con un margen de error bajo tanto las magnitudes, como las tendencias generales de la TSM, al mismo tiempo que las series de tiempo instrumentales se encontraron en su mayoría dentro del área sombreada que corresponde al intervalo de confianza de una desviación estándar de los datos reconstruidos. Bajo esta evaluación, se seleccionó el modelo de 14 x 14 neuronas (Figura 11c y d) como el que presentó la mejor reconstrucción debido a que presentó la menor amplitud, así como menos valores atípicos en relación a la distribución del área sombreada referente a la desviación estándar, además de que los valores reconstruidos durante el periodo que cuenta con datos instrumentales (1941 – 2011), se encontraron más cercanos a sus puntos de referencia en relación con los otros modelos (13 x 13 y 15 x 15 neuronas) los cuales tendieron a divergir y a presentar una desviación estándar atípica para el periodo entre 1700 – 1750 AD, a pesar de presentar un RMSE cercano y correlación mayor (Tabla 4).



**Figura 11.** Reconstrucción de la TSM entre el 1417 y 2011 AD utilizando una RNA tipo “*feedforward*” de retro propagación, con una topología de dos capas de neuronas. Reconstrucciones con RNA de: a) 13 x 13 neuronas, c) 14 x 14 neuronas, e) 15 x 15 neuronas. Periodo instrumental de las RNA de: b) 13 x 13 neuronas, d) 14 x 14 neuronas y f) 15 x 15 neuronas. El área sombreada corresponde a una desviación estándar con un promedio móvil de 3 pasos, la línea negra representa el promedio de 100 corridas de cada uno de los modelos. La línea roja representa el promedio móvil de 3 pasos y la línea azul el valor promedio anual de los datos instrumentales de la TSM (1951 – 2011). Las líneas punteadas verticales delimitan un segmento del periodo cálido medieval (PCM) y los periodos climáticos de la pequeña edad de hielo (PEH) y el cálido actual (AC).

La serie de tiempo reconstruida a partir de la RNA de 14x14 neuronas describe un descenso de la TSM durante la salida del PCM, el cual se extiende hasta la etapa temprana (~ 1580 AD) de la PEH, periodo en el cual se registraron valores por debajo del promedio histórico de toda la serie ( $\bar{X}=24.4$ ). Los periodos comprendidos entre 1580 – 1615 AD, 1660 – 1680 AD y 1710 – 1810 AD, se caracterizaron por ser cálidos, presentando temperaturas por

encima (0.77 °C) del promedio histórico. Entre 1820 hasta 1940 AD, periodo que corresponde al final de la PEH y principio del CA, se registró un descenso de la TSM un grado por debajo del promedio ( $\bar{X}= 23.4$ ). Posteriormente, durante el CA se registraron dos periodos con temperaturas altas (1940 – 1960 AD y 1987 – 2011 AD) y un periodo de relajación entre (1960 y 1987 AD) entre estos.

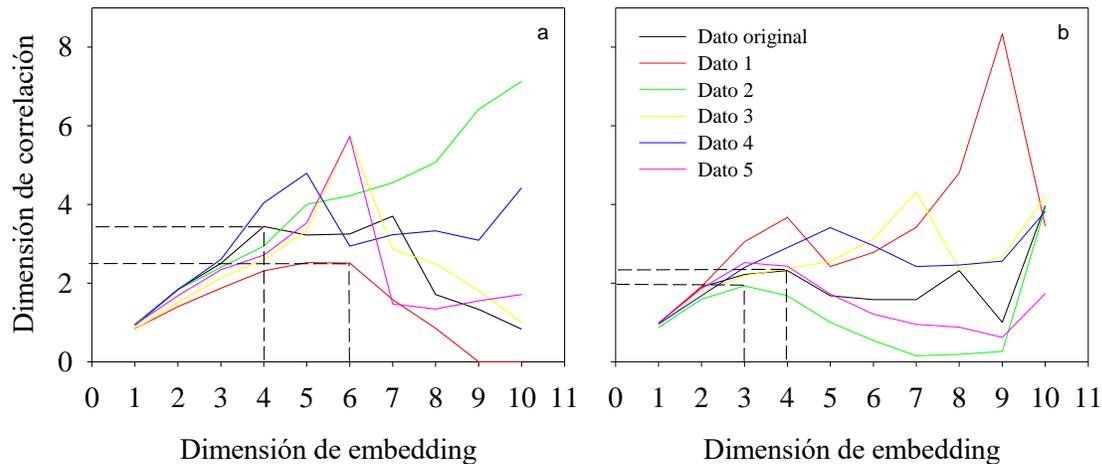
#### **Predicción de las variables climáticas a partir de las series de tiempo reconstruidas**

Las arquitecturas que se seleccionaron para reconstruir la temperatura del aire y TSM fueron de 6 y 14 x 14 neuronas, respectivamente. Inicialmente se evaluaron los parámetros descriptivos no lineales a las series de tiempo. Para la serie de tiempo original de temperatura del aire, se determinaron los siguientes valores: dimensión de correlación ( $D = 4$ ), dimensión de embedding ( $m = 8$ ) y retraso ( $d = 6$ ). En cambio, para la serie de tiempo original de la TSM los parámetros descriptivos fueron  $D = 3$ ,  $m = 7$  y  $d = 3$ . Posteriormente, para la reducción del ruido se exploraron diferentes parámetros cercanos a los valores iniciales para definir la mejor reducción de ruido. Se identificó que los parámetros que permitieron reducir el ruido fueron aquellos que presentaron los valores más bajos del RMSE y SNR, como se observa en la Tabla 5.

**Tabla 5.** Recopilación de parámetros descriptivos no lineales evaluados para la reducción del ruido en las series de tiempo reconstruidas de la temperatura del aire y TSM. Dimensión de correlación ( $D$ ), dimensión de embedding ( $m$ ), error cuadrático medio (RMSE) y relación señal-ruido (SNR). Los valores en negrilla corresponden a los parámetros descriptivos seleccionados.

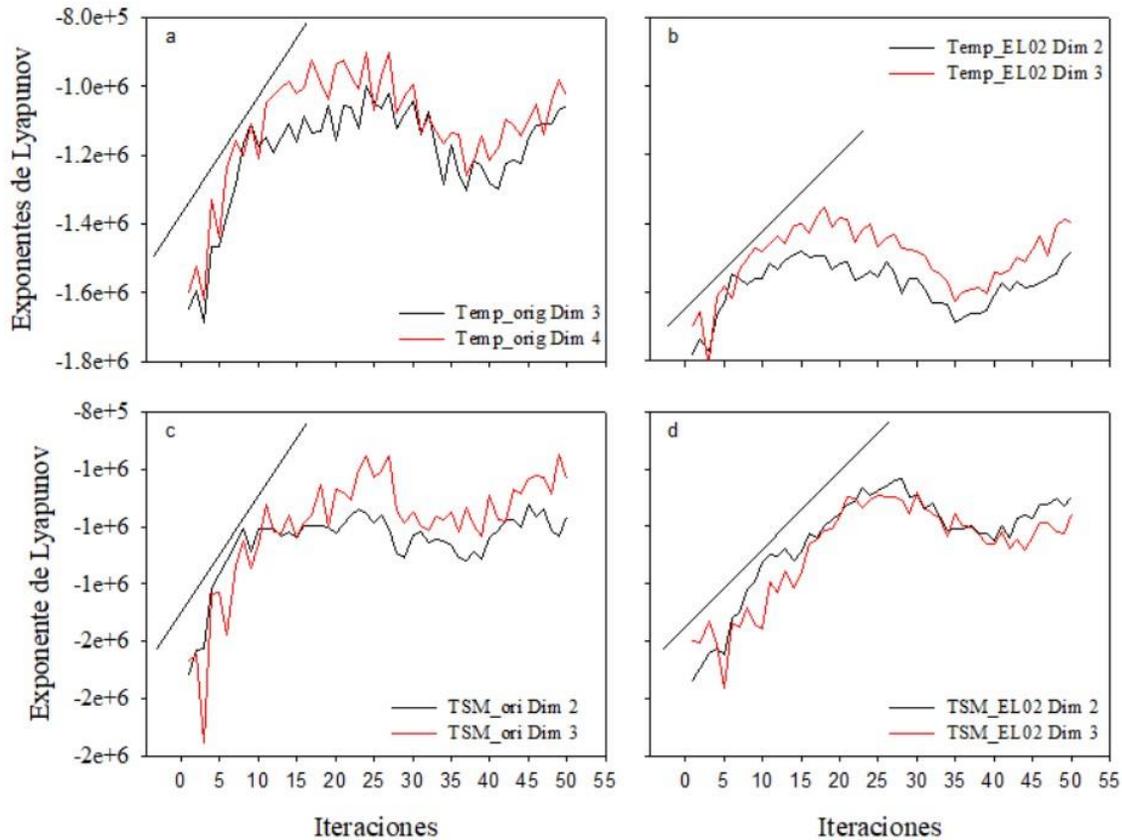
Dato	Temperatura del aire				Temperatura superficial de mar			
	m	D	RMSE	SNR	M	D	RMSE	SNR
1	<b>2</b>	<b>3</b>	<b>0.903</b>	<b>4.39</b>	2	2	1.04	5.14
2	2	4	0.929	4.67	<b>2</b>	<b>3</b>	<b>1.078</b>	<b>5.80</b>
3	2	5	0.924	4.18	2	4	1.085	6.52
4	2	6	0.941	4.99	2	5	1.135	7.12
5	2	7	0.9	4.62	3	3	1.185	8.17

En la Figura 12 se observa que tanto para la temperatura del aire (a) como para la TSM (b) se redujo el valor de D en función de la serie original, confirmándose la reducción de ruido en las series de tiempo reconstruidas (bajo los parámetros marcados en negrilla en la Tabla 5).



**Figura 12.** Dimensión de correlación contra la dimensión de embedding de: a), Temperatura del aire, b) Temperatura superficial del mar. Las líneas punteadas marcan el punto de saturación de la dimensión de correlación para las series de tiempo con ruido y con reducción de ruido. Datos del 1 al 5, hacen referencia a las series de tiempo a las cuales se les redujo el ruido citadas en la Tabla 5.

Se evaluaron los máximos exponentes de Lyapunov para las series de tiempo originales y para las series de tiempo a las cuales se les redujo el ruido con la finalidad de calcular la cantidad de pasos hacia el futuro que se puede avanzar considerando el mínimo error posible en la predicción (Tabla 6). En todas las series de tiempo de temperatura del aire y TSM se presentaron exponentes negativos (Figura 13). De ellas se resalta que la serie de tiempo original para la temperatura del aire y TSM en promedio pueden avanzar 7 pasos en el tiempo (Figura 13a y c), mientras que las series de tiempo a las cuales se les ha reducido el ruido presentan valores mayores, por ejemplo, para la temperatura del aire (Dato 1) puede avanzar 10 pasos (Figura 13b), mientras que para la TSM (Dato 2) puede avanzar 8 pasos en el tiempo (Figura 13d).



**Figura 13.** Máximo exponente de Lyapunov para la temperatura del aire, a) serie de tiempo original, b) serie de tiempo con reducción de ruido bajo los parámetros  $m= 2$ ,  $D= 3$  (Dato 1), c) máximo exponente de Lyapunov para la TSM (serie de tiempo original), d) TSM con reducción de ruido bajo los parámetros  $m= 2$ ,  $D= 3$  (Dato 2).

**Tabla 6.** Número de pasos que se puede avanzar en el tiempo con el menor error posible a partir de los máximos exponentes de Lyapunov ( $\lambda$ ) para la temperatura del aire y TSM. Donde:  $M=$  pendiente que define el máximo exponente de Lyapunov ( $\lambda$ ),  $P= 1/M$  y Pasos= pasos hacia el futuro.

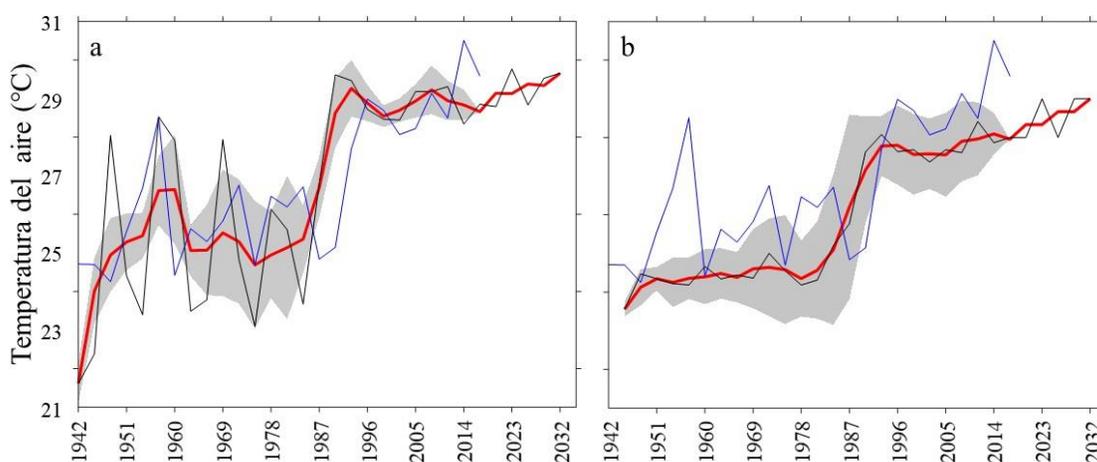
		Serie de tiempo	Dimensión	$M(\lambda)$	$P$	Pasos
T. del aire	Original	Dim 2	37229.94	2.68601E-05	5.37	
		Dim 3	27977.72	3.57427E-05	7.15	
	Dato 1	Dim 2	19836.40	5.04122E-05	10.08	
		Dim 3	19063.25	5.2457E-05	10.49	
TSM	Original	Dim 2	70648.83	1.41545E-05	2.83	
		Dim 3	27227.50	3.67276E-05	7.35	
	Dato 2	Dim 2	28846.95	3.46657E-05	6.93	
		Dim 3	24168.44	4.13763E-05	8.28	

**Tabla 7.** Parámetros descriptivos de la eficiencia de las predicciones de la temperatura del aire a partir de RNA “*feedforward – step by step*” para una arquitectura de una y dos capas de neuronas en la capa intermedia. Regresión del modelo (Reg. Mod), RMSE del modelo (RSME mod) y RSME de la predicción (RSME pred), correlación de Spearman para el promedio de las repeticiones de la modelación ( $\rho$ ) y para el promedio móvil de 3 pasos ( $\rho$  m3) para la serie de tiempo promediada. Negritas denotan las arquitecturas seleccionadas con el mejor desempeño

Arquitectura	RNA de una capa					RNA de dos capas									
	Reg. Mod	RSME mod	RSME pred	$\rho$	$\rho$ m3	Arq	Reg. Mod	RSME mod	RSME pred	$\rho$	$\rho$ m3				
Temperatura del aire	Serie original	3	0.94	0.26	2.42	0.55	0.75	3x3	0.94	0.25	2.83	0.59	0.76		
		4	<b>0.94</b>	<b>0.25</b>	<b>2.83</b>	<b>0.60</b>	<b>0.75</b>	4x4	0.95	0.23	3.13	0.59	0.76		
		5	0.95	0.23	2.86	0.60	0.75	5x5	0.95	0.23	4.69	0.59	0.77		
		6	0.95	0.23	4.11	0.61	0.75	6x6	0.92	0.29	5.10	0.59	0.78		
		7	0.95	0.23	3.64	0.61	0.76	7x7	0.76	0.62	6.98	0.59	0.80		
		8	0.95	0.24	3.48	0.63	0.77	8x8	0.95	0.22	13.28	0.59	0.76		
		9	0.95	0.24	3.55	0.61	0.76	9x9	0.95	0.24	13.26	0.59	0.78		
		10	0.95	0.23	3.45	0.64	0.77	10x10	0.94	0.26	21.68	0.59	0.80		
		11	0.95	0.23	3.09	0.63	0.76	11x11	0.95	0.24	25.67	0.59	0.82		
		12	0.95	0.23	3.71	0.63	0.77	12x12	0.95	0.24	23.33	0.59	0.77		
		13	0.95	0.23	3.68	0.61	0.76	13x13	0.94	0.25	30.08	0.59	0.77		
		14	0.95	0.23	3.80	0.61	0.76	14x14	0.93	0.32	61.33	0.59	0.79		
		15	0.94	0.24	3.33	0.60	0.76	15x15	0.94	0.26	38.11	0.59	0.77		
		Temperatura del aire	Dato 1	3	0.92	0.15	0.26	0.69	0.74	3x3	0.93	0.14	0.29	0.72	0.76
				4	0.93	0.14	0.25	0.72	0.76	<b>4x4</b>	<b>0.93</b>	<b>0.13</b>	<b>0.35</b>	<b>0.73</b>	<b>0.77</b>
5	0.92			0.15	0.21	0.70	0.73	5x5	0.91	0.14	0.60	0.76	0.80		
6	0.92			0.15	0.21	0.73	0.75	6x6	0.94	0.11	2.06	0.74	0.78		
7	0.92			0.15	0.24	0.71	0.76	7x7	0.83	0.26	1.93	0.73	0.79		
8	0.92			0.15	0.21	0.71	0.75	8x8	0.93	0.13	3.06	0.78	0.82		
9	0.92			0.15	0.22	0.70	0.74	9x9	0.89	0.19	8.90	0.74	0.78		
10	0.92			0.15	0.19	0.71	0.75	10x10	0.92	0.18	16.81	0.78	0.79		
11	0.92			0.15	0.25	0.71	0.75	11x11	0.93	0.13	14.67	0.65	0.72		
12	0.92			0.15	0.21	0.71	0.74	12x12	0.93	0.16	13.32	0.68	0.76		
13	0.92			0.15	0.24	0.69	0.74	13x13	0.93	0.16	28.65	0.77	0.76		
14	0.92			0.16	0.18	0.68	0.72	14x14	0.92	0.17	20.38	0.62	0.79		
15	0.92			0.16	0.18	0.70	0.73	15x15	0.92	0.16	31.25	0.68	0.75		

Tras definir los pasos hacia el futuro en la predicción para cada variable de interés a las cuales se les redujo el ruido, se procedió a realizar la modelación de cada una (datos originales y con reducción de ruido de la temperatura del aire y TSM). En las Tablas 7 y 8 se agrupan los resultados de las diferentes arquitecturas (1 y 2 capas) evaluadas para cada variable. Entre las predicciones destacadas para la temperatura del aire encontramos que la

arquitectura de 4 y 4 x 4 neuronas presentaron los mejores resultados (Tabla 7) para la serie con ruido y reducción de éste, respectivamente (Figura 14a y b). La serie de tiempo original (con reducción de ruido) presentó una varianza explicada de 0.94 (0.95), observándose que la serie de tiempo con menor ruido exhibió una menor dispersión de los puntos y línea de tendencia cercana al ajuste ideal. El RMSE del modelo para la serie de tiempo original fue de 0.25 (0.23), en cambio, el RMSE de la predicción fue de 2.83 (3.13). Ambas series de tiempo presentaron correlaciones significativas para el valor promedio ( $r^2 [27] = 0.6 (0.59)$ ;  $p < 0.005$ ) y para el promedio móvil de 3 pasos ( $r^2 [27] = 0.73 (0.77)$ ;  $p < 0.005$ ).



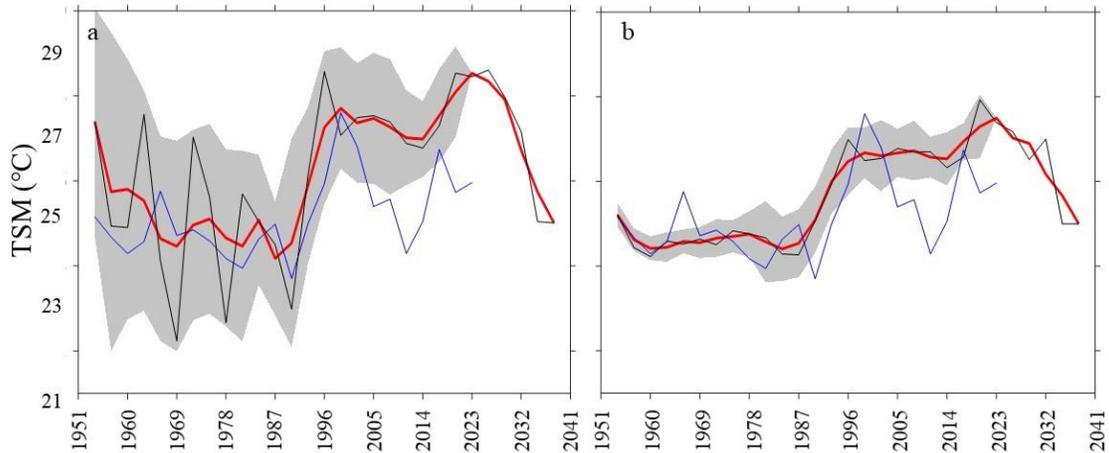
**Figura 14.** Predicción de la temperatura del aire (°C) para el periodo 2011 – 2032 utilizando una RNA tipo *feedforward* de retro propagación “*step by step*”. RNA alimentada con la serie de tiempo original y una arquitectura de a) 4 neuronas y b) de 4 x 4 neuronas. El área sombreada corresponde a una desviación estándar con un promedio móvil de 3 pasos. La línea negra representa el promedio de 100 corridas de cada modelo, la línea roja el promedio móvil de 3 pasos y la línea azul el promedio de la temperatura máxima de enero a abril (datos instrumentales de 1942 – 2020).

Dentro de los modelos evaluados (Tabla 8) para la TSM, los que presentaron los mejores resultados corresponden al de una arquitectura de 4 x 4 neuronas con la serie de tiempo original (Figura 15a), el cual presentó una varianza explicada de 0.94, el RMSE del modelo de 0.47 y de la predicción de 5.98. Presentó una correlación significativa para el valor promedio ( $r^2 [25] = 0.53$ ;  $p < 0.005$ ) y para el promedio móvil de 3 pasos ( $r^2 [25] = 0.62$ ;  $p < 0.005$ ). La Figura 15b corresponde a un modelo alimentado por una serie de tiempo con reducción de ruido (dato 2) y una arquitectura de 4 neuronas, modelo que presentó una varianza explicada de 0.96, se observó que la serie de tiempo con una reducción de ruido

exhibió una menor dispersión de los puntos entre los datos observados y modelados, presentando una línea de tendencia cercana al ajuste ideal. El RMSE del modelo fue de 0.15 y de la predicción de 0.29, se registró una correlación significativa para el valor promedio ( $r^2 [25] = 0.62$ ;  $p < 0.005$ ) y para el promedio móvil de 3 pasos ( $r^2 [25] = 0.65$ ;  $p < 0.005$ ).

**Tabla 8.** Parámetros descriptivos de la eficiencia de las predicciones de la TSM a partir de una RNA “*feedforward--step by step*” para una arquitectura de una y dos capas de neuronas en la capa intermedia. Regresión del modelo (Reg. Mod), RSME del modelo (RSME mod), RSME de la predicción (RSME pred), correlación de Spearman para el promedio de las repeticiones de la modelación ( $\rho$ ) y para el promedio móvil de 3 pasos ( $\rho$  m3) para la serie de tiempo promediada. Negritas denotan las arquitecturas seleccionadas con el mejor desempeño.

Arquitectura		Red neuronal de una capa					Red neuronal de dos capas							
		Reg. Mod	RSME mod	RSME pred	P	$\rho$ m3	Arq	Reg. Mod	RSME mod	RSME pred	$\rho$	$\rho$ m3		
Temperatura superficial del mar	Serie original	3	0.91	0.59	2.15	0.57	0.64	3x3	0.91	0.55	2.77	0.57	0.65	
		4	0.91	0.58	2.63	0.53	0.66	<b>4x4</b>	<b>0.94</b>	<b>0.47</b>	<b>5.98</b>	<b>0.55</b>	<b>0.68</b>	
		5	0.91	0.57	3.10	0.55	0.66	5x5	0.93	0.42	13.97	0.60	0.70	
		6	0.91	0.55	5.73	0.58	0.67	6x6	0.92	0.54	17.61	0.59	0.70	
		7	0.91	0.55	4.88	0.49	0.65	7x7	0.88	0.81	16.59	0.57	0.71	
		8	0.92	0.53	7.33	0.53	0.68	8x8	0.93	0.48	16.95	0.59	0.69	
		9	0.92	0.53	10.39	0.58	0.70	9x9	0.93	0.47	36.84	0.52	0.64	
		10	0.91	0.55	7.24	0.53	0.66	10x10	0.93	0.48	37.05	0.66	0.69	
		11	0.91	0.55	5.38	0.56	0.66	11x11	0.92	0.52	44.41	0.59	0.62	
		12	0.92	0.53	8.14	0.55	0.67	12x12	0.93	0.49	48.29	0.63	0.64	
		13	0.91	0.55	3.29	0.54	0.66	13x13	0.93	0.47	58.33	0.56	0.74	
		14	0.91	0.56	6.01	0.52	0.63	14x14	0.92	0.57	65.00	0.46	0.57	
		15	0.92	0.53	7.85	0.48	0.62	15x15	0.92	0.58	80.97	0.42	0.55	
		Dato 2	3	0.95	0.15	0.16	0.59	0.63	3x3	0.96	0.14	0.23	0.60	0.64
			<b>4</b>	<b>0.96</b>	<b>0.15</b>	<b>0.29</b>	<b>0.62</b>	<b>0.65</b>	4x4	0.96	0.13	0.59	0.56	0.61
	5		0.96	0.14	0.46	0.61	0.66	5x5	0.96	0.11	1.10	0.55	0.58	
	6		0.96	0.12	0.81	0.57	0.62	6x6	0.96	0.11	3.47	0.49	0.56	
	7		0.96	0.13	3.62	0.51	0.61	7x7	0.93	0.16	7.75	0.41	0.56	
	8		0.96	0.12	0.94	0.50	0.57	8x8	0.96	0.13	11.42	0.36	0.41	
	9		0.97	0.11	1.77	0.46	0.54	9x9	0.96	0.14	15.74	0.38	0.57	
	10		0.96	0.12	2.10	0.47	0.53	10x10	0.95	0.18	27.31	0.53	0.62	
	11		0.96	0.11	2.27	0.45	0.54	11x11	0.96	0.15	29.54	0.42	0.63	
	12		0.96	0.12	3.37	0.19	0.37	12x12	0.95	0.18	32.87	0.25	0.45	
	13		0.97	0.11	6.41	0.36	0.61	13x13	0.95	0.18	33.70	0.38	0.59	
	14		0.97	0.11	3.09	0.51	0.64	14x14	0.95	0.19	57.14	0.51	0.70	
	15		0.97	0.10	7.20	0.46	0.62	15x15	0.95	0.15	63.94	0.32	0.49	



**Figura 15.** Predicción de la TSM (°C) para el periodo 2011 – 2035 utilizando una RNA tipo *feedforward* de retro propagación “*step by step*”. a) RNA alimentada con la serie de tiempo original y una arquitectura de 4 x 4 neuronas, b) RNA alimentada con la serie de tiempo con reducción de ruido (dato 2) y una arquitectura de 4 neuronas. El área sombreada corresponde a una desviación estándar con un promedio móvil de 3 pasos, la línea negra representa el promedio de 100 corridas de cada modelo, la línea roja el promedio móvil de 3 pasos y la línea azul el promedio anual de la TSM (datos instrumentales de 1951 – 2020).

## VIII. DISCUSIÓN

El presente estudio proporciona la primera comparación cuantitativa entre los eventos de producción de la columna de agua, la fracción en hundimiento y la conservada en el registro sedimentario aportada por las diatomeas planctónicas con los datos instrumentales a partir del uso de RNA para el sur del Golfo de California. Así fue posible validar la utilidad que tienen las diatomeas planctónicas como proxies de la temperatura del aire y TSM para la región a partir del uso de la RNA tipo SOM como método de clasificación no lineal. También se demostró que el uso de las RNA tipo *Feedforward* tienen la capacidad de simplificar y de abstraer las relaciones no lineales que presenta el sistema climático con las asociaciones de diatomeas planctónicas identificadas como proxies, ejerciendo como función de transferencia para reconstruir las condiciones climáticas de los últimos 600 años a partir de las diatomeas conservadas en el registro fósil. Finalmente, se utilizaron las series de tiempo reconstruidas como insumos para realizar la predicción de las condiciones climáticas de las próximas décadas, obteniéndose un buen ajuste y reduciendo el grado de incertidumbre.

### ***Procesamiento y análisis de las series de tiempo para la obtención de las funciones de transferencia y reconstrucción del clima***

Debido que el sistema dinámico subyacente está constituido por relaciones multidimensionales y complejas, con un probable grado de aleatoriedad y ruido, es complicado trabajar con series de tiempo instrumentales. Esto hace que sea un reto el analizar correctamente las series de tiempo ambientales (Sivakumar, 2004; Lei & Meng, 2011). Una alternativa en el campo de la climatología y ciencias naturales es el uso de la teoría del caos (Sivakumar *et al.*, 1999; Elshorbagy *et al.*, 2002), donde las series de tiempo se consideran como datos irregulares generados por un sistema dinámico determinístico no lineal (sistema caótico). Estos sistemas evolucionan en el tiempo a partir de un estado inicial conocido que se describe a partir de un par de ecuaciones. Sin embargo, en sistemas reales no se conocen las ecuaciones, ni las condiciones iniciales. Por tal motivo, para entender el comportamiento de las series de tiempo, es necesario reconstruir el espacio de fase a partir de las mismas series de tiempo (Dhanya & Kumar, 2010) estableciendo un tiempo de retraso adecuado.

Una dificultad que se presentó en este estudio fue determinar el retraso adecuado de las series de tiempo. A pesar de contar con varios años de información (> 10 años), la resolución temporal de algunas de ellas fue baja. Millán *et al.* (2010) sugieren que los valores bajos en el retraso posiblemente están relacionados a series de tiempo relativamente cortas. También se reconoce que el retraso al ser pequeño, la poca información presente puede llevar a una subestimación de la dimensión de correlación (Sivakumar, 2000; Dhanya & Kumar, 2010), la cual se ha utilizado para determinar la presencia del caos, así como la exposición a señales aleatorias y ruido blanco (Sivakumar *et al.*, 1999; Elshorbagy *et al.* 2002; Sivakumar, 2004). Sivakumar (2004) documentó la presencia de caos para diversos procesos geofísicos (precipitación, temperatura, velocidad del viento, etc.) al documentar bajas dimensiones ( $1 > m < 5$ ), valores que se ajustan a nuestros resultados. En este trabajo se registraron valores entre 4 y 11 de la dimensión de embedding, que, a pesar de ser consideradas de baja dimensión, estuvieron por encima de las publicadas por Gutiérrez (2004) y Marwan *et al.* (2003), para series de tiempo climáticas.

El ruido se ha determinado como uno de los factores limitantes en el rendimiento de muchas técnicas de clasificación, modelado, predicción y control en sistemas determinísticos.

Autores como Grassberger *et al.* (1991; 1993), Schreiber (1993), Schreiber *et al.* (2002) han propuesto diversos métodos para reducir el ruido en series de tiempo teóricas e hidrológicas. Sin embargo, Kantz & Schreiber (1997), Elshorbagy *et al.* (2002) señalan que la mayoría de ellos reducen el ruido en cantidades similares y su rendimiento no difiere mucho. Asimismo, el alcance de la influencia del ruido en los datos, los métodos de reducción del ruido y el grado de mejora que puede ser logrado después de la reducción son difíciles de determinar (Grassberger *et al.*, 1991; Grassberger *et al.*, 1993; Schreiber, 1993; Sivakumar *et al.*, 1999; Sivakumar, 2000). Por tal motivo, se sugiere su uso con reserva, debido que los modelos tienen diferentes consideraciones y grado de sensibilidad que pueden hacer que pequeños cambios se magnifiquen, derivando en diferencias entre ellos. En este estudio se hizo frente a esta problemática utilizando como estrategia el uso de SNR y RSME como herramientas para evaluar la reducción del ruido.

Para la zona de estudio el uso integral de SNR y RMSE permitió definir de una manera adecuada las series que presentaron la mejor eficiencia y calidad en la reducción del ruido, así como las leves diferencias entre una y otra. Se identificó el ruido, ya que este modificó la autosimilitud del atractor y la reconstrucción del espacio de fases mostró una dimensión elevada. En la Figura 6, se resalta como se alcanza un segundo incremento en la dimensión de correlación en algunas series asociado con la presencia de ruido. Esto sugiere que la incertidumbre obtenida en el análisis pudo estar dada por el tamaño de muestra y la resolución temporal de la serie de tiempo, como es discutido por Dhanya & Kumar (2010).

En estudios ecológicos, las técnicas de ordenación y clasificación se usan comúnmente para la interpretación de un conjunto de datos complejos. Para lo cual, en muchos casos es conveniente reducir el número de variables para el análisis multivariado. Sin embargo, es imposible reducir el número de variables sin el riesgo de perder información (Park *et al.*, 2006). La estrategia utilizada permitió reducir significativamente el número de especies de diatomeas sin ocasionar una pérdida significativa de información ecológica y la complejidad de las relaciones existentes en el sistema, lo cual se pudo contrastar con trabajos ecológicos de la región (Verdugo-Díaz, 2004; Villegas-Aguilera, 2009; Hakspiel-Segura, 2014; Acevedo-Acosta, 2015).

Las RNA tipo SOM han demostrado ser útiles como alternativa a los métodos estadísticos tradicionales para tratar conjunto de datos complejos y no lineales (Tsai *et al.*, 2016). Son

tolerantes a la presencia de ruido y tienen la capacidad de manejar datos atípicos, así como faltantes (Li *et al.*, 2018). En el campo de la climatología su uso se ha centrado en identificar vínculos entre variables y regiones, así como identificar patrones espacio-temporales considerando diferentes escalas (Cavazos 2000; Liu *et al.*, 2010; Tsai *et al.*, 2016; Li *et al.*, 2018). Además, en este estudio se implementó para identificar las especies de diatomeas como proxy de variables climáticas, con el potencial para utilizarse como insumos para la determinación de una función de transferencia.

Se confirmó que el tamaño de la RNA tipo SOM tiene una fuerte influencia en la calidad de la clasificación y es un criterio importante. Dado que, si es tan grande o más que el número de muestras de entrenamiento, el patrón se torna muy detallado y hace complicado entender adecuadamente los patrones y las similitudes se vuelven difíciles de determinar (Park *et al.*, 2003), mientras que, si la arquitectura es muy pequeña, el patrón tiende a ser muy general y ciertamente las diferencias importantes no se pueden detectar (Li *et al.*, 2018). La metodología propuesta por Tsai *et al.* (2016) nos permitió reducir la dimensionalidad y el número de neuronas de manera significativa con relación a si solo considerábamos los errores como método determinante para el tamaño de la arquitectura como es sugerido por algunos autores (Céréghino *et al.*, 2001; Park *et al.*, 2003).

El número de neuronas seleccionadas nos permitió una interpretación y visión más clara de las relaciones entre las variables ambientales y las diatomeas, así como contrastar los valores de los TE y QE con los estimados en otros estudios (Park *et al.*, 2003; Astel *et al.*, 2007; Olawoyin *et al.*, 2013; Tsai *et al.*, 2016; Li *et al.*, 2018) encontrando rasgos similares en la definición del tamaño de las RNA e interpretación. Adicionalmente, los resultados de este trabajo al igual que los trabajos anteriormente citados sugieren que, aunque la metodología SOM no es principalmente un algoritmo de análisis de conglomerados, es una herramienta poderosa para clasificar las relaciones entre variables ambientales y biológicas, así como otros procesos presentes en la naturaleza.

El modelo con ruido (Figura 8) presentó un mayor número de variables relacionadas con una asociación de taxa de diferentes afinidades, pero principalmente asociadas a condiciones cálidas. Algunas presentaron patrones uni y bimodales que compartieron con las variables. Los gradientes no fueron del todo claros e inversos a la respuesta ecológica esperada, lo cual

sugiere que la magnitud de las contribuciones de las variables fue pequeña (Park *et al.*, 2003; Li *et al.*, 2018). La mayoría de las asociaciones estuvieron representadas por las diatomeas preservadas en la trampa de sedimentos, donde se reconoce una pérdida sustancial de información cualitativa y cuantitativa proveniente de la zona eufótica.

En el caso del modelo a los cuales se les redujó el ruido (figura 9), se observó que el rendimiento fue mayor, dado que las asociaciones de diatomeas presentes en cada grupo presentaron patrones más claros y congruentes con el comportamiento de las variables climáticas. Por consiguiente, se consideró que la contribución de las asociaciones de diatomeas fue significativa en función a las variables climáticas con las cuales se relacionaron (Park *et al.*, 2003; Astel *et al.*, 2007; Li *et al.*, 2018). La participación de *Cyclotella caspia* fue importante en ambos modelos, mientras que *Nitzschia interruptestriata* y una asociación de especies de agua templada fueron representativas del modelo con menor ruido. *Actinoptychus senarius* y *Cyclotella* sp., fueron representativas en la columna de agua como de la trampa de sedimentos relacionándose con las variables ambientales, lo que sugiere que la señal climática asociada a su producción dentro de la zona eufótica está siendo conservada en la trampa de sedimentos y potencialmente en el sedimento laminado. La preservación de estas señales posiblemente está siendo favorecida porque presentan frústulas fuertemente silicificadas que favorece su preservación durante el proceso de hundimiento (Olli *et al.*, 2001; Verity *et al.*, 2002; Lopes *et al.*, 2006; Vallina *et al.*, 2014).

En ambos modelos se identificó a la TSM y al componente U del viento como variables con potencial para ser reconstruidas a partir de una función de transferencia, definida por la asociación de diatomeas (proxy) establecida. Las diferencias entre los modelos se definen por el grado de eficiencia entre las relaciones, el número de variables y especies con las que compartieron similitudes.

Para los dos modelos con ruido y reducción de ruido, se identificó una asociación de taxa que respondieron como proxy de la TSM, temperatura del aire, precipitación, sigma-T, y profundidad de la zona eufótica. En este sentido Acevedo-Acosta (2015) resalta la presencia de *N. interruptestriata*, *Thalassionema nitzschioides* var. *parva*, *Nitzschia bicapitata*, *Fragilariopsis doliolus* y *Planktoniella sol*, relacionadas con el ingreso de agua superficial tropical (AST) al interior de la Bahía de La Paz. Lo cual permite sugerir que su presencia en el registro sedimentario estaría vinculado a condiciones cálidas. Esto coincide con la

reconstrucción realizada por Barron *et al.* (2010) donde a partir de inferencias identificaron que *Cy. caspia*, *Thalassionema frauenfeldii*, *Rhizosolenia bergonii*, *Chaetoceros didymus*, *Cy. striata* y diatomeas de afinidad templada, costera y cosmopolita, se relacionaron con condiciones cálidas-templadas y la influencia de masas de agua de afinidad tropical-subtropical en la región.

En el modelo con ruido, *N. bicapitata*, *F. doliolus* y *Chaetoceros mesanensis*, se asociaron como indicadoras de la precipitación (variable relacionada al periodo cálido del año) y la temperatura atmosférica. En este sentido, Barron *et al.* (2010), Esper *et al.* (2010) reconocen a *F. doliolus* de origen sub-tropical, idónea para la reconstrucción inferida del clima por ser común en el registro sedimentario, además que registra los cambios en la TSM. Además, las relaciones lineales que se han obtenido con esta especie y la TSM a pesar de ser débiles son significativas. Sin embargo, cuando se vincula con otras especies de afinidad cálida (semejante a nuestro caso), las relaciones que se obtienen son mejores (Barron *et al.*, 2010). De igual manera, Shimada *et al.* (2008) han relacionado estas especies como proxy de la TSM y la salinidad en estudios paleo-ecológicos y en reconstrucciones paleo-oceanográficas para el Atlántico Norte. Mientras que Kennington *et al.* (1999) al considerar la relación de diatomeas neríticas (incluyendo las bentónicas) y pelágicas, aunada a una función de transferencia reconstruye la TSM y la variabilidad en la influencia de los vientos alisios en el Pacífico Ecuatorial.

Ren *et al.* (2014) vinculan a *N. bicapitata* como parte del registro sedimentario superficial acompañado a otros taxa de origen sub-tropical, como indicadoras de la presencia de la masa de agua sub-tropical en el Pacífico sub-ártico, sugiriendo una relación cercana con la TSM, así como con la salinidad de la zona. En la Bahía de La Paz, *N. bicapitata*, *Ac. senarius*, *N. interruptestriata* (tanto en la columna de agua como en la trampa de sedimentos) se ha asociado a la profundidad de la capa de mezcla (Acevedo-Acosta, 2015). Sin embargo, estas especies (resaltando *N. bicapitata*), se han encontrado principalmente en la trampa de sedimentos, dado que su hundimiento y preservación esta mediado por el pastoreo y empaquetamiento en pelotillas fecales. Otros estudios han explicado cambios similares entre la composición de especies de la comunidad colectada en la columna de agua y en la trampa de sedimentos durante el periodo productivo como el resultado de la depredación selectiva por parte del zooplancton y eventos de producción elevada (Olli *et al.*, 2001; Verity *et al.*,

2002; Vallina *et al.*, 2014). Esto sugiere que la conservación de la señal tanto en la columna de agua como en la trampa por parte de *N. interruptestriata*, es eficiente en el proceso de hundimiento, durante los periodos productivos y de mezcla.

Así mismo *N. bicapitata*, *T. frauenfeldii*, asociadas a un ensamble de diatomeas en mal estado, frústulas débiles y fragmentadas, se han utilizado como indicadoras de periodos de baja productividad. Schrader & Sorknes (1991) reconstruyeron la paleo-productividad para los últimos 400,000 años para las costas de Perú utilizando el estado de preservación de las valvas de las diatomeas, como un indicador de periodos productivos de la zona eufótica. Asociaron la integridad de las valvas y altas abundancia de *Skeletonema* y *Delphineis* con periodos productivos. En cambio, durante periodos con baja productividad observaron taxa fuertemente silicificados en elevadas abundancias, como es el caso de *Azpeitia nodulifera* y *Fragilariopsis doliolus*.

#### **Reconstrucción climática de los últimos 600 años para el sur del Golfo de California**

A partir de la presente investigación las RNA presentaron la capacidad de modelar las complejas relaciones existentes entre la composición y abundancia de las diatomeas planctónicas presentes en la columna de agua y trampa de sedimentos (utilizadas como insumos), con los datos instrumentales (variables climáticas) contemporáneos a partir de la definición de una función de transferencia, permitiendo extraer dichas relaciones de la información preservada (proxies) en el registro sedimentario para reconstruir las condiciones climáticas del pasado (Bevilacqua *et al.*, 2007). Los resultados obtenidos coinciden con diversos estudios en que las RNA presentan mejores resultados que los métodos lineales en función a la predictibilidad, certidumbre y eficiencia (Malmgren *et al.*, 2001; Kucera *et al.*, 2005; Racca *et al.*, 2007; Gupta & Malmgren, 2009; Carro-Calvo *et al.*, 2013; Esper & Gersonde, 2014). Dentro de las virtudes que se resaltan, encontramos la flexibilidad en la definición de los parámetros internos (algoritmo de aprendizaje, el tipo de modelación) y el uso de diferentes tipos de datos de entrada (lineales, no-lineales, continuos o discretos). Dentro de los criterios a evaluar, se encuentra la definición del número adecuado de neuronas en la capa intermedia, como la topología de esta, la cantidad de taxones o variables a utilizar como insumos para evitar el sobre o sub-entrenamiento de las redes.

Las RNA tipo *Feedforward* seleccionadas para reconstruir la temperatura del aire (6 neuronas) y la TSM (13 x 13 neuronas) se consolidaron como modelos “funciones de transferencia” eficientes que arrojaron resultados satisfactorios en el ajuste de los modelos (varianza explicada), RMSE (fase de modelación) y la posterior validación de las reconstrucciones a partir del RMSE y la comparación con datos instrumentales, lo cual las definió como reconstrucciones más robustas comparadas con otros estudios paleoclimáticos realizados para la región (autores citados en Tabla 9). Al comparar nuestros resultados con otros estudios realizados para la región y el hemisferio norte, encontramos ciertas similitudes a resaltar. Por ejemplo, se observó que la mayoría de las reconstrucciones (regionales) presentan valores dentro del área de influencia de la desviación estándar de nuestra reconstrucción la mayor parte del tiempo (Figura 18 y 19), además de coincidir con algunos estudios de otras regiones (Tardif *et al.*, 2019). Asimismo, se encontraron discrepancias tanto en frecuencia como en amplitud para el noroccidente de Norte América y hemisferio norte (Tabla 9; Figura 18 y 19). Sin embargo, concordamos con otros estudios en que las diferencias obtenidas pueden ser atribuidas a varias causas, como: a) a la variabilidad interna de los modelos; b) a la resolución temporal de los datos utilizados para la calibración de los modelos, y c) a la sensibilidad climática regional y estacional que presentan los modelos (Harrison *et al.*, 2015; PAGES, 2017; Tardif *et al.*, 2019). En este caso creemos que la principal causa de las diferencias se debe a la sensibilidad y variabilidad interna de los diferentes modelos, debido a que generalmente tienden a basarse en modelos lineales las cuales no poseen la capacidad de captar todas las relaciones multidimensionales presentes en el sistema climático, cosa contraria a las RNA las cuales, al desempeñarse en el campo de la no linealidad, pueden captar las complejas relaciones que presenta el sistema climático, y como consecuencia obtener un resultado con mayor certidumbre.

Las diversas reconstrucciones realizadas para la región noroeste de Norte América y del hemisferio norte (Tabla 9), presentan correlaciones significativas en relación con los datos instrumentales. En este sentido, al comparar los resultados obtenidos para ambas variables [temperatura del aire ( $r^2 = 0.62$ ;  $p < 0.005$ ), TSM ( $r^2 = 0.5$ ;  $p < 0.005$ )] encontramos que nuestros valores son representativos y presentan cierta similitud con las tendencias de la mayoría de las series de tiempo reconstruidas, confirmando de esta manera la validez y coherencia que presentan nuestras reconstrucciones a nivel regional. Las diferencias

observadas con otros estudios del hemisferio norte pueden ser atribuidas a causas tales como: a) a que gran parte de la variabilidad a menor escala espacial se atenúa (promedia) como consecuencia del uso de un conjunto de proxies que se distribuyen por diferentes regiones y latitudes; b) al uso potencial de múltiples proxies, los cuales pueden presentar un mejor ajuste (correlacionados) considerando un periodo del año en particular; y c) a que posterior a obtener las reconstrucciones, múltiples autores tienden a utilizar promedios móviles de 10 a 30 años en las series reconstruidas con la finalidad de incrementar la significancia de las reconstrucciones y extraer la variabilidad multidecadal y secular (Christiansen & Ljungqvist, 2017).

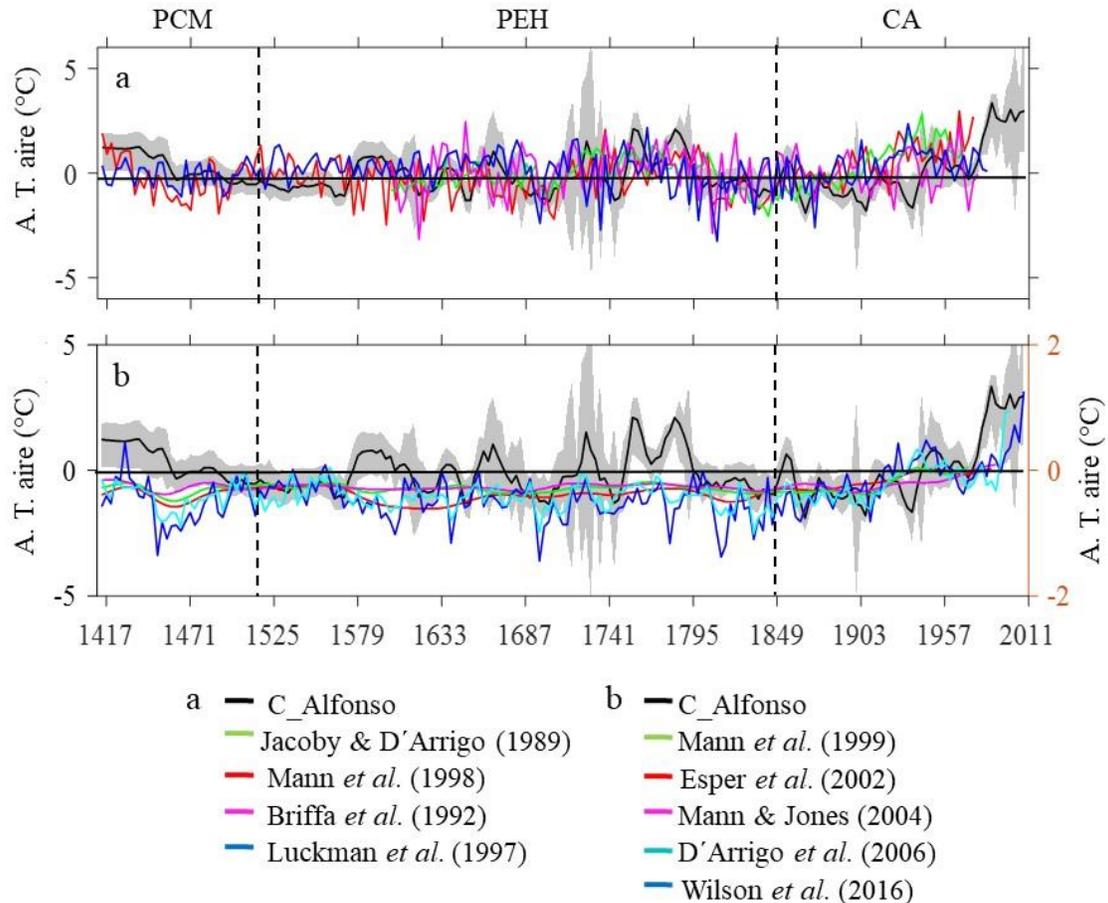
Las reconstrucciones de la temperatura del aire y TSM que se obtuvieron para el sur del Golfo de California exhiben condiciones cálidas para la primera parte del siglo XV con una tendencia hacia temperaturas bajas y estables hasta 1550 (siglo XVI), este periodo (PCM) se caracterizó por presentar un flujo promedio en el cual la participación de las asociaciones de las diatomeas planctónicas presentó una contribución reducida del flujo total. Sin embargo, *Actinoptychus senarius* y la asociación de diatomeas de agua templada fueron representativas durante este periodo, confirmándose que su presencia está relacionada a condiciones cálidas-templadas y a la influencia de masas de agua de afinidad tropical-subtropical en la región aspectos que se ajustan a lo observado por Barron *et al.* (2010) y Acevedo-Acosta (2015).

El comportamiento registrado durante PCM coincide con las reconstrucciones propuestas por Luckman *et al.* (1997), Mann *et al.* (1998) para Canadá y el oeste de Estados Unidos, respectivamente, así como para el hemisferio norte por Jones & Mann (2004), D'Arrigo *et al.* (2006), Wilson *et al.* (2016), aunque con ligeras diferencias hacia mayores anomalías positivas para la temperatura del aire (Figura 16). Para el caso de la TSM los resultados son más contrastantes, por ejemplo la reconstrucción para Santa Bárbara (Zhao *et al.*, 2000) presenta un periodo más corto con tendencia a enfriarse que el observado en Cuenca Alfonso, mientras que la reconstrucción propuesta por McCabe-Glynn *et al.* (2013) para la zona tropical presenta una alta variabilidad con valores positivos y sin una tendencia clara, aspecto que ha sido descrito por otros autores, los cuales sugieren que el enfriamiento para el Pacífico Tropical Central no ocurrió de manera inmediata y no fue tan pronunciado como lo ocurrido en latitudes altas a causa de la actividad volcánica e irradiación solar (Cobb *et al.*, 2003).

**Tabla 9.** Comparación de las correlaciones (r) entre reconstrucciones y datos instrumentales (1901 - 1980) para diferentes regiones, así como los valores obtenidos al comparar las diferentes reconstrucciones de la temperatura del aire y superficial del mar con los valores reconstruidos para Cuenca Alfonso (C\_Alfonso). En las últimas tres columnas las correlaciones representan la misma comparación con C\_Alfonso pero acotadas a los periodos climáticos: Cálido actual (CA), Pequeña edad de hielo (PEH) y Periodo cálido medieval (PCM). Los valores en negritas resaltan las correlaciones significativas ( $p < 0.005$ ).

		Periodo	C_Alfonso Vs. Regiones				
			Periodo	PCM	PEH	CA	
		1901-1980	1400-2011	1400-1500	1500-1850	1850-2011	
		AD	AD	AD	AD	AD	
Región							
Temperatura del aire	Mann <i>et al.</i> , 1998	O. EEUU	0.2	0.11	<b>0.65</b>	0.17	0.13
	Briffa <i>et al.</i> , 1992	O. EEUU	0.64	0.06		0.11	-0.06
	Jacoby & D'Arrigo, 1989	N. Canadá	0.36	<b>0.30</b>		<b>0.47</b>	0.29
	Luckman <i>et al.</i> , 1997	E. Canadá	0.44		<b>0.41</b>	0.03	0.01
	Esper <i>et al.</i> , 2002	HN		0.09	<b>0.50</b>	0.03	<b>0.56</b>
	Crowley <i>et al.</i> , 2003	HN		<b>0.43</b>	<b>0.80</b>	<b>0.24</b>	<b>0.72</b>
	Mann <i>et al.</i> , 1999	HN		<b>0.14</b>	<b>0.54</b>	<b>0.20</b>	<b>0.43</b>
	Jones & Mann, 2004	HN	0.65	<b>0.43</b>	0.27	<b>0.29</b>	<b>0.75</b>
	Wilson <i>et al.</i> , 2016	HN	0.78	<b>0.31</b>	<b>0.49</b>	0.02	<b>0.61</b>
	D'Arrigo <i>et al.</i> , 2006	HN		<b>0.32</b>	<b>0.54</b>	0.15	<b>0.55</b>
Temperatura superficial del mar	Wilson, 2006	Trópico	0.57	<b>0.59</b>			<b>0.59</b>
	McCabe-Glynn <i>et al.</i> , 2013	California		<b>0.23</b>		<b>0.25</b>	<b>0.28</b>
	Tierney <i>et al.</i> , 2015	Pacífico tropical		<b>0.24</b>		0.12	<b>0.39</b>
	Abella-Gutiérrez <i>et al.</i> , 2019	San Lázaro		<b>0.26</b>		0.25	<b>0.43</b>
	Goni <i>et al.</i> , 2006	Guaymas		<b>-0.45</b>		<b>-0.57</b>	-0.15
	Zhao <i>et al.</i> , 2000	Sta. Bárbara		<b>0.18</b>	-0.06	<b>0.2</b>	0.02

\*Briffa *et al.* (1992), Jacoby & D'Arrigo (1989), Luckman *et al.* (1997), Esper *et al.* (2002), Crowley *et al.* (2003), se encuentran citadas en Mann & Jones (2003).

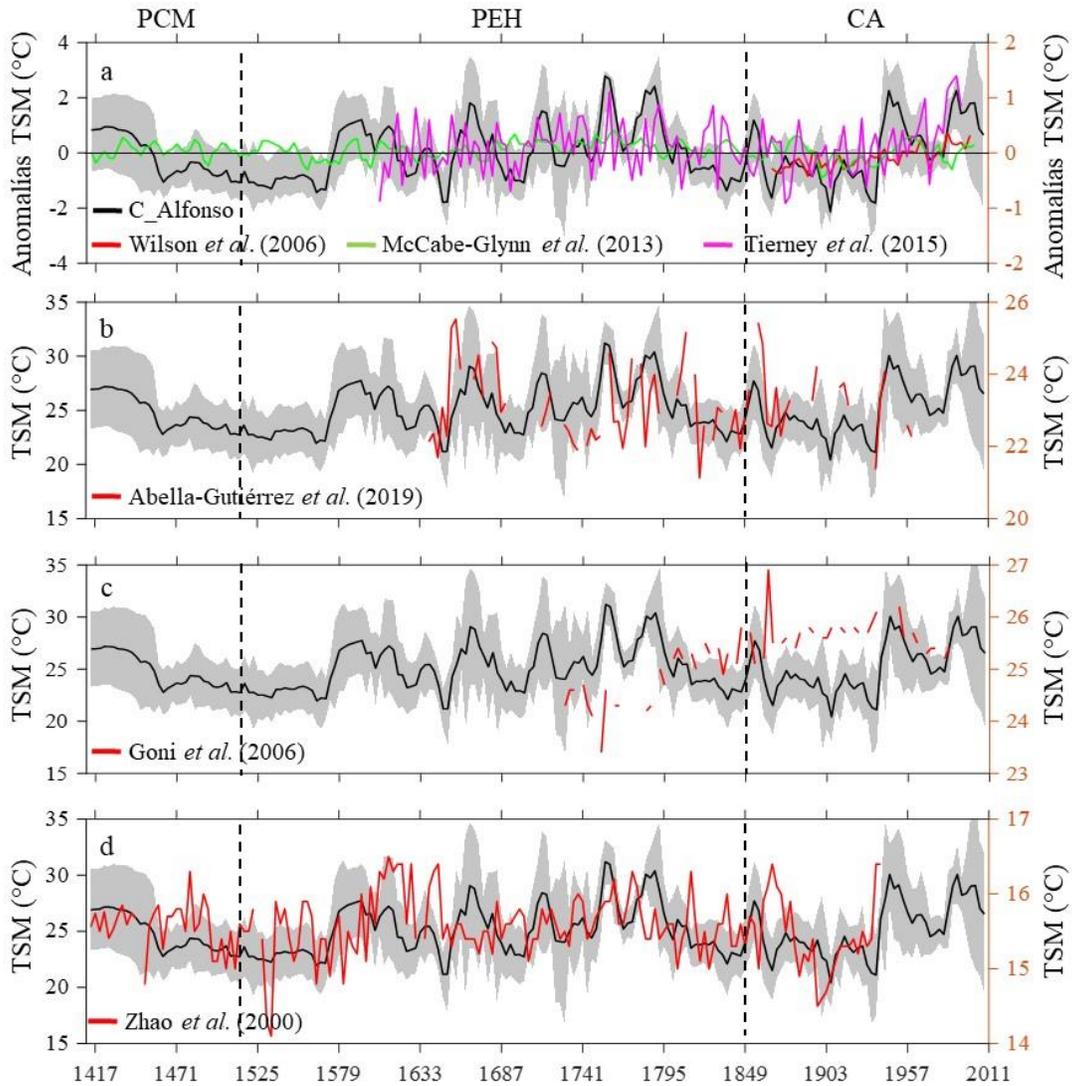


**Figura 16.** Comparación de la reconstrucción de las anomalías de la temperatura del aire de los últimos 600 años para Cuenca Alfonso contra: a) reconstrucciones regionales de Norte América; b) reconstrucciones para el hemisferio norte. El área sombreada corresponde a una desviación estándar con un promedio móvil de 3 pasos.

Durante la PEH se registró un desacople entre el flujo total y el flujo de las asociaciones de diatomeas de interés. Mientras, entre 1626 y 1847 AD se registró una disminución en el flujo con valores cercanos al promedio histórico, donde el flujo de las asociaciones de diatomeas fue similar al flujo total. Encontramos que las asociaciones y flujos de las diatomeas conservadas en el registro sedimentario durante la PEH sugieren que la señal preservada se ve influenciada a procesos tafonómicos (pastoreo y el empaquetamiento en pelotillas fecales) que durante el periodo productivo toman relevancia como resultado de la depredación selectiva por parte del zooplancton y eventos de producción elevada los cuales se sugieren, son recurrentes durante periodos fríos (Olli *et al.*, 2001; Verity *et al.*, 2002; Vallina *et al.*, 2014).

A pesar de que la PEH ha sido ampliamente estudiada, aún existen varios puntos controversiales a su alrededor tales como la amplitud y tiempo de duración de sus fluctuaciones, los forzamientos climáticos a gran escala, el inicio y fin del periodo, así como el grado de sincronía entre las diferentes regiones, la cual potencialmente estuvo modulada por cambios en la circulación atmosférica (Matthews & Briffa, 2005; González-Trueba *et al.*, 2007). La reconstrucción de este periodo denotó mayores discrepancias entre la temperatura del aire y la TSM con otros estudios (Figura 17). La reconstrucción de la temperatura del aire de Cuenca Alfonso presentó un comportamiento parcialmente semejante (Tabla 9) al registrado para el norte de Canadá (Jacoby & D'Arrigo, 1989) y para el hemisferio norte (Mann *et al.*, 1999; Crowley *et al.*, 2003; Mann & Jones, 2003) compartiendo tendencia y amplitud del rango de anomalías a lo largo de todo el periodo climático.

Para el caso de la TSM (Figura 19 b-d) se identificaron similitudes con la región de California (McCabe *et al.*, 2013), Santa Bárbara (Zhao *et al.*, 2000), Cuenca Alfonso (Herguera *et al.*, 2003) y una relación inversa con Cuenca de Guaymas (Goni *et al.*, 2006). Reconstrucciones como las de D'Arrigo *et al.* (2006), Wilson *et al.* (2016) y Tardif *et al.* (2019), a pesar de no presentar correlaciones significativas con Cuenca Alfonso, comparten similitudes durante los descensos de temperatura, que en algunos casos coinciden con valores extremos coetáneos con eventos volcánicos (Sigl *et al.*, 2015) como es el caso del volcán de Tambora (1815), Krakatoa (1883) y el Chichonal (1982), los cuales se ajustan a un descenso en la temperatura del aire y TSM a partir del mismo año en que ocurren o en los años siguientes (~ 3 años; Bronnimann *et al.*, 2019).



**Figura 17.** Comparación de: a) las anomalías de la TSM reconstruida para los últimos 600 años para Cuenca Alfonso, con reconstrucciones regionales de Norte América; b) la TSM de los últimos 600 años para Cuenca Alfonso (línea negra) y la Cuenca de San Lázaro; c) Cuenca de Guaymas y d) Cuenca de Santa Bárbara. El área sombreada corresponde a una desviación estándar con un promedio móvil de 3 pasos.

Jones & Mann (2004) reportan que los siglos XV, XVII y XIX fueron los siglos más fríos del último milenio para el hemisferio norte. Adicionalmente, el siglo XV fue uno de los siglos con mayor actividad volcánica del último milenio, el cual coincidió con una reducción en la actividad solar (mínimo de Maunder), expansión del hielo marino, alteración en la circulación del Atlántico norte (reducción en la temperatura), aunado a un incremento en el albedo durante los meses de verano, que redujeron la temperatura del aire de manera persistente en

el hemisferio norte (Miller *et al.*, 2012; Zanchettin *et al.*, 2012; Helama *et al.*, 2021). Para este estudio, durante estos mismos periodos se observó una relación significativa entre las anomalías negativas de la temperatura del aire (1650 – 1715 AD y 1790 – 1850 AD) y los mínimos solares de Maunder (1645 – 1715 AD) y Dalton (1790 – 1830 AD) indicando que tanto la PEH, como inicio del periodo CA se encontraron fuertemente influenciados por el forzamiento solar y en menor medida por el volcánico (Tabla 10), lo que coincide con lo propuesto por otros autores (Wilson *et al.*, 2006; Miller *et al.*, 2012; Helama *et al.*, 2021). Aunado a que las condiciones frías a principios del siglo XIX aparentemente son el resultado de la influencia combinada de un periodo de actividad volcánica, baja irradiancia solar, los modos de variabilidad climática a escala decadal del Atlántico norte, los cuales podrían haber influido el clima del Golfo de California a través de las teleconexiones (Wilson *et al.*, 2016; Lechleitner *et al.*, 2017; Bronnimann *et al.*, 2019).

Durante el período CA se observó que las asociaciones de diatomeas exhibieron un comportamiento similar al flujo total, donde las asociaciones de agua cálida, templada, *Ac. senarius* y *Cyclotella* sp. fueron las que tuvieron una mayor participación en el flujo. Se demuestra que la señal climática asociada a la producción dentro de la zona eufótica está siendo conservada durante el proceso de hundimiento y de enterramiento en el sedimento laminado como ha sido observado en otras regiones (Verity *et al.*, 2002; Lopes *et al.*, 2006; Vallina *et al.*, 2014), además de coincidir con las características ecológicas de las asociaciones de diatomeas planctónicas, donde sobresale su afinidad a condiciones cálidas-templadas y a la influencia de masas de agua de afinidad tropical-subtropical lo que define un periodo cálido en la región (Barron *et al.*, 2010; Acevedo-Acosta, 2015).

**Tabla 10.** Correlaciones múltiples (r) entre la temperatura superficial del mar (TSM pm 3) y temperatura del aire (TA pm 3) contra el forzante solar (FS) y volcánico (FV), para la Pequeña Edad de Hielo (PEH; 1500 – 1850 AD), el Cálido Actual (CA; 1850 – 2011 AD) y para toda la serie de tiempo reconstruida (T. serie; 1417 – 2011 AD). Pm 3; corresponde a un promedio móvil de 3 pasos; los valores son estadísticamente significativos ( $p < 0.05$ ); X: no hubo correlaciones significativas. Los forzantes climáticos fueron obtenidos de Mann et al. (2005).

	T. serie		PEH		CA	
	FS	FV	FS	FV	FS	FV
TSM pm 3	0.245	X	0.252	X	0.634	-0.296
TA pm 3	0.365	X	0.263	X	0.702	-0.378

El CA es considerado como el periodo climático más cálido del último milenio afectando alrededor del 98% del planeta (Neukom *et al.*, 2019). Los resultados de la reconstrucción realizada muestran durante la transición de la PEH hacia el CA, anomalías negativas de la temperatura del aire y TSM que se extendieron desde 1800 hasta principios de 1930 AD, pero conservando una tendencia acorde con el calentamiento global del planeta. Se reconoce que esta tendencia es resultado de la recuperación del sistema climático global después de una secuencia de erupciones volcánicas, además, de que la posible contribución en menor medida del incremento de los gases de efecto invernadero de origen antropogénico favoreció el incremento de la temperatura global (Bronnimann *et al.*, 2019). Para la temperatura del aire estas observaciones son consistentes con las tendencias observadas en las diferentes reconstrucciones realizadas para todo el hemisferio norte, mientras que para la TSM solo lo son para las regiones tropicales, de California y de Baja California (Figura 16b, 17a). Al comparar los parámetros no lineales obtenidos a partir de las series de tiempo reconstruidas encontramos que presentaron dimensiones de embedding con valores bajos, al igual que lo reportado por Gutiérrez (2004) y Marwan *et al.* (2003). El uso de SNR, RMSE y la reducción de la dimensión de correlación nos permitió definir de manera adecuada las series que presentaron la mejor respuesta y calidad ante la reducción del ruido, valores que se encuentran dentro del rango documentados por Sivakumar (2004). Finalmente, la validación de las reconstrucciones a partir de la comparación con datos instrumentales históricos, reconstrucciones de la región y del hemisferio norte, como la evaluación de la calidad y reducción del ruido de las series de tiempo a partir de parámetros descriptivos no lineales nos permitieron confirmar la validez y confiabilidad (certidumbre) de nuestras series de tiempo reconstruidas, que posteriormente nos brindaron la confianza de utilizarlas como insumos para la predicción del clima.

### ***Predicción de las condiciones climáticas de las próximas décadas***

Las relaciones no lineales que caracterizan el comportamiento del sistema climático, así como la interacción con forzantes naturales y externos, las pequeñas diferencias en el punto de origen de inicialización de los modelos, así como la estructura de los modelos hacen que la previsibilidad se vea limitada y sea un eje central de investigación en las predicciones

climáticas (Yingyi, 1993; Krishnamurthy, 2019). Al analizar los parámetros no lineales encontramos que las series de tiempo con ruido presentaron una mayor dimensión de correlación que las series a las cuales se les redujo el ruido, al igual que presentaron una menor previsibilidad, como es documentado por Yingyi (1993) y Zhang *et al.* (2004). Se observó que el incremento de la previsibilidad en las series de tiempo con reducción de ruido se relacionó con una disminución en la variabilidad y amplitud en la serie de tiempo, causando una subestimación de los valores modelados en relación con los datos instrumentales de las variables climáticas de interés (Figura 14 y 15).

La previsibilidad se ha explorado en diferentes escalas temporales (Fraedrich, 1986; Krishnamurthy, 2018; 2019) por medio de métodos no lineales, centrándose en reconocer las fuentes de predictibilidad climática en términos de procesos y fenómenos que actúan a diferentes escalas de una manera casi periódica [(TSM (respuesta lenta), los monzones, eventos interanuales (ENSO), oscilaciones decadales (PDO, NAO)]. Estos eventos se consideran como órbitas periódicas inestables incrustadas en un atractor caótico. Bajo esta visión, se ha demostrado que las predicciones mediante métodos de sistemas dinámicos no lineales han arrojado buenos resultados superando a los obtenidos con modelos climáticos (Krishnamurthy, 2018, 2019). Se resalta, además, el potencial que tienen los sistemas dinámicos no lineales en el mejoramiento de las predicciones a diferentes escalas temporales.

En cuanto a la previsibilidad de manera general se ha reportado valores mayores a 10 años para el Atlántico y menores para el Pacífico norte utilizando 6 modelos acoplados, para predecir el número de pasos que se podía avanzar con el menor error posible (Branstator *et al.*, 2012). Sun & Wang (2006) evalúan la previsibilidad de variables climáticas para el Pacífico a escala decadal (PDO y Oscilación Interdecadal del Pacífico) y proponen que la temperatura se puede predecir aproximadamente a 7 años. Mientras que Teng *et al.* (2011), encuentran que las dos primeras funciones ortogonales de la SST media anual presenta una previsibilidad de 6 a 10 años. Los valores para la temperatura del aire (8 años) y TSM (10 años) obtenidos en este estudio, se ajusta a los valores documentados para el Pacífico a escala decadal, confirmándose la coherencia en la capacidad de previsibilidad que presenta la región. Asimismo, están en línea con la propuesta, de que la previsibilidad de la TSM es mayor que de la temperatura del aire como consecuencia de la respuesta más lenta que

presenta el océano ante cambios en el sistema climático del planeta (Hermanson & Sutton, 2010).

Al predecir las condiciones climáticas para el sur del Golfo de California, como se menciona en la literatura, se identificó que las predicciones climáticas están sujetas a diferentes fuentes de incertidumbres, de las cuales podemos resaltar la variabilidad interna del sistema climático, la incertidumbre generada por los forzamientos naturales y antropogénicos durante la era industrial. En este sentido, al existir sesgos intrínsecos a las reconstrucciones de las variables climáticas (fechado, errores de observaciones, resolución de datos instrumentales, etc.), hacen que se magnifique la incertidumbre al momento de proyectar el comportamiento del sistema climático hacia el futuro. En el caso de las predicciones decadales, el principal factor de incertidumbre corresponde a la variabilidad interna, la dispersión y punto de inicialización del modelo (Li *et al.*, 2020).

Di *et al.* (2014) proponen un método de múltiples pasos para mejorar las predicciones climáticas en el cual, considera una reducción del ruido, descomposición de las series y posteriormente la predicción. En concordancia con la metodología propuesta por Di *et al.* (2014) se reconoce que la reducción de ruido tanto en la fase de reconstrucción (datos instrumentales e insumos para las reconstrucciones), como posteriormente en el pre-procesamiento para la predicción (series de tiempo reconstruidas) permiten mejorar el rendimiento de las predicciones (Tabla 7). Por otro lado, Li *et al.* (2020) describen que ni el aprendizaje automatizado ni los métodos estadísticos pudieron distinguir tanto los valores máximos, como los mínimos de los datos instrumentales. Caso cercano a lo obtenido a la serie de tiempo con (sin) reducción de ruido, que a pesar de registrar un buen desempeño no pudo capturar toda la variabilidad (amplitud y frecuencia) de los datos instrumentales, situación que es recurrente en diferentes escenarios predictivos y escalas temporales (Li *et al.*, 2020).

De manera general se identificó que los modelos definidos para predecir la TSM y la temperatura del aire obtuvieron coeficientes de determinación cercanos a la pendiente ideal, correlaciones significativas con relación a datos instrumentales de la región y valores de RMSE bajos, confirmando que el uso de RNA produce resultados satisfactorios. Con esto se ha demostrado que las RNA son ligeramente superiores en rendimiento a diferentes modelos

lineales (Trigo & Palutxikof, 1999; Deo *et al.*, 2015; Buckland *et al.*, 2019; Wu *et al.*, 2019; Tran *et al.*, 2020). Por otro lado, al evaluar las variables climáticas modeladas encontramos que exhiben fluctuaciones interdecadales significativas, sin embargo, las series de tiempo de los datos instrumentales se encuentran dentro del área de influencia de la desviación estándar de las mismas. La resolución temporal (intervalos de 3 años) que posee las series de tiempo utilizadas como insumos, permitió predecir y explicar el comportamiento de la TSM y temperatura del aire a una escala decadal.

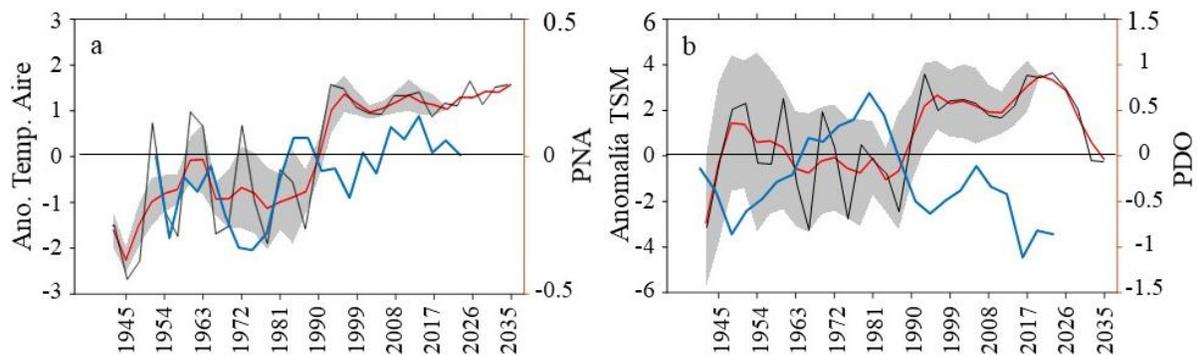


Figura 18. Comparación de a) las anomalías de la temperatura del aire contra la Oscilación del Pacífico de Norte América (PNA); b) anomalías de la TSM y la Oscilación decadal del Pacífico (PDO). Línea oscura representa al valor promedio de la predicción, línea roja al promedio móvil de 3 pasos de la predicción, la línea azul corresponde al índice climático correspondiente a cada figura y el área sombreada corresponde a una desviación estándar con un promedio móvil de 3 pasos. (A partir del 2011 hasta el 2035 corresponde a la predicción).

La relación encontrada entre las anomalías de la TSM con la Oscilación Decadal del Pacífico (PDO;  $r^2$ : -0.64,  $P < 0.05$ ), y la temperatura del aire con el patrón del Pacífico Norte América (PNA;  $r^2$ : 0.56,  $P < 0.05$ ) (Figura 18) es coherente a lo observado por Johnson *et al.* (2020) y Tseng *et al.* (2020), quienes sostienen que la región norte del patrón espacial de la Variabilidad Decadal del Pacífico (PDV) se encuentra influenciado por la ocurrencia del PDO, el cual se ve afectado de forma remota por la variabilidad del Pacífico Tropical y la propagación de ondas Rossby, al mismo tiempo que se encuentra vinculado con la variabilidad de baja frecuencia de las Aleutianas y el PNA relación que es confirmada para la región sur del Golfo de California. Adicionalmente, para el hemisferio norte, el Pacífico norte ha tenido una mayor contribución a la variabilidad de la temperatura media global del aire en la superficie que el Pacífico tropical y el Atlántico norte a una escala decadal (Huang *et al.*, 2017). Autores como Dong *et al.* (2014), Henley *et al.* (2017), Johnson *et al.* (2020) y

Tseng *et al.* (2020) describen que la PDV presenta un patrón espacial simétrico de las anomalías de la TSM, la cual se ha comprobado que tiene influencia más allá del Pacífico, impactando el clima regional y global a una escala multidecadal.

Huang *et al.* (2017) describen durante el periodo en el cual ocurrió la desaceleración del calentamiento global conocida como el “hiatus” (1999 – 2012) la existencia de un contraste térmico entre el océano y la tierra, bajo unas condiciones de un PDO (-). Esto aunado a la Oscilación del Atlántico Norte con una tendencia hacia una fase negativa, que indujeron conjuntamente a un enfriamiento en diferentes regiones del hemisferio norte, resaltando Norte América y norte de Eurasia (Iles & Heger, 2017). Por lo tanto, el aire frío proveniente de las zonas polares pudo enfriar las latitudes medias. Estas condiciones coinciden con la fase negativa del PNA, que favorecen la intensificación de los vientos del noroeste. Se reconoce que la incidencia de una condición negativa del PDO y PNA en la región ante un escenario de calentamiento global, favoreció la atenuación de la tendencia (Huang *et al.*, 2017), además que las teleconexiones tropicales a extratropicales son un mecanismo importante en la memoria y variación decadal en el Pacífico (Di Lorenzo *et al.*, 2013), todo esto nos lleva a que ante un aumento de la TSM desde principio del siglo XX hasta 1945 y desde 1980 hasta finales de la década de 1990. Estos periodos fueron interrumpidos por una atenuación o pausa en la tendencia (Hiatus) durante 1945 – 1980 y 1999 – 2012. Periodos que se ajustan a la transición de una fase positiva (+) a una negativa (-) del PDV, como es descrito por Johnson *et al.* (2020). Por el contrario, para la temperatura del aire observamos que se presentó un retraso en la atenuación de la tendencia a partir de 1965 hasta principios de 1990, comportamiento que se asemejó a la transición de una fase positiva a negativa de la Variabilidad Multidecadal del Atlántico (AMV). En este sentido, Johnson *et al.* (2020) y Monerie *et al.* (2021) describe una teleconexión circumglobal entre el PDV y AMV, donde el PDV (-) causa un retraso de una década del AMV (+), para lo cual se sugiere que el AMV puede utilizarse como fuente de previsibilidad para el PDV (Johnson *et al.*, 2020; Monerie *et al.*, 2021). Adicionalmente, se identificó que, bajo un AMV (+) se fortalecen los vientos alisios, causando un transporte anómalo de calor atmosférico hacia el oeste y un incremento en la temperatura en la superficie del océano Pacífico occidental. Estas condiciones estarían explicando el desfase entre las tendencias entre la temperatura del aire y TSM para la región sur del Golfo de California y él porque se pronostica que la temperatura del aire se mantenga

con una ligera tendencia de calentamiento en la próxima década (2021 – 2035), mientras que la TSM presente un enfriamiento del sistema, manifestándose como una respuesta retrasada de AMV (+) como sugiere Monerie *et al.* (2021).

## IX. CONCLUSIONES

- La implementación de una RNA tipo SOM y el análisis no lineal de series de tiempo, permitió reducir el ruido y comprender el comportamiento de las series de tiempo instrumentales. La metodología propuesta mejoró las clasificaciones ecológicas, redujó la incertidumbre y el sesgo en la definición de la matriz biológica (diatomeas planctónicas), considerando la menor pérdida de información posible. Esto contribuyó a confirmar que las asociaciones de diatomeas planctónicas tienen el potencial como proxies en la reconstrucción de la temperatura del aire y superficial del mar.
- Se corroboró que las RNA *Feedforward* tienen la capacidad de simplificar y de abstraer algunas relaciones multidimensionales, no lineales que presentaron las asociaciones de diatomeas planctónicas con el clima a través de las variables de la temperatura del aire y TSM. Con esto, fue posible definir las funciones de transferencia a partir de las RNA, reconstruyendo las series de tiempo de las variables de la temperatura del aire y TSM.
- A partir de la reconstrucción de las variables de la temperatura del aire y TSM para la región sur del Golfo de California durante los últimos 600 años, se identificó la influencia de los siguientes eventos climáticos seculares: finales del Periodo Cálido Medieval (PCM: 1400 – 1500 AD), la Pequeña Edad de Hielo (PEH: 1500 – 1850 AD) y el Cálido Actual (CA: 1850 al presente). Se demostró la capacidad predictiva que presentan las RNA como herramienta para reconstruir las condiciones climáticas del pasado, presentando un alto grado de certidumbre a partir de bajos niveles de error (RMSE), ruido (SNR) y congruencia en las reconstrucciones en relación con datos instrumentales históricos y reconstruidos por otros autores para la región y hemisferio norte.
- La PEH se caracterizó por ser un periodo altamente dinámico y controversial, donde se identificaron diferencias y similitudes en las tendencias y amplitud en las anomalías en relación con otras regiones. Se confirma que durante la PEH la actividad solar (mínimos

de manchas solares) fue el forzante de mayor relevancia como precursor del descenso de la temperatura del aire y TSM para el hemisferio norte. Para el CA la reconstrucción permitió documentar una transición lenta de una condición fría hacia una condición cálida para la región sur del Golfo de California, donde la actividad solar y volcánica fueron forzantes significativos.

- El preprocesamiento (reducción de ruido) de las series de tiempo reconstruidas permitió mejorar la previsibilidad y eficiencia de los modelos predictivos, registrándose coeficientes de determinación cercanos a la pendiente ideal, lo que se tradujo en un buen ajuste de las predicciones de las condiciones climáticas hasta el 2032 (2035). Sin embargo, a pesar de registrar un buen desempeño, las RNA no pudieron capturar toda la variabilidad (amplitud y frecuencia) de los datos instrumentales, debido a que las RNA tienden a sobre o subestimar la variabilidad natural de los datos.
- Las RNA *feedforward* demostraron la capacidad de reducir la incertidumbre de las predicciones de la temperatura del aire y TSM para la región sur del Golfo de California. Lo cual se vio reflejado en la capacidad de modelar las condiciones climáticas que pausaron la tendencia de calentamiento a nivel global durante el periodo (1999 – 2012). Se propone la influencia del Patrón de Norte América sobre las condiciones atmosféricas de la región (temperatura del aire) y de la Oscilación Decadal del Pacífico sobre la TSM. Reconociéndose que los modos de variabilidad climática por medio de teleconexiones tropicales a extratropicales son un mecanismo importante en la memoria del sistema climático y variación decadal en el Pacífico llegando a causar la atenuación o intensificación de la tendencia de calentamiento a nivel global.
- De manera general se reconoce que la sensibilidad que presentaron las diatomeas (actuales y paleodatos) ante la variabilidad climática, mostraron la capacidad de reconstruir y modelar las condiciones climáticas, con resultados coherentes y comparativamente mejores a otros estudios a nivel regional. Demostrándose además que, Cuenca Alfonso corresponde a un sitio sensible a la variabilidad del clima regional y global a diferentes escalas temporales y que actúa como un registrador del clima del pasado.

## X. REFERENCIAS BIBLIOGRÁFICAS

- Abella–Gutiérrez, J., Herguera, J.C., Mortyn, P.G., Kelly, C.S., Martínez-Botí, M.A., 2020. Multidecadal climate variability in the southern region of the California current system during the last 1,800 years. *Paleoceanography and Paleoclimatology*, 35(2). <https://doi.org/10.1029/2019PA003825>
- Acevedo–Acosta, J.D., 2015. Ecología y flujo vertical de las diatomeas planctónicas en Cuenca Alfonso: implicaciones para los estudios paleoclimáticos. Tesis de Maestría, Instituto Politécnico Nacional; centro interdisciplinario de ciencias marinas. México. 127 p.
- Albañez-lucero, M., 2010. Distribución espacial del mero (*Epinephelus morio*) del banco de Campeche, México. Tesis de doctorado, Instituto Politécnico Nacional; centro interdisciplinario de ciencias marinas. México. 110 p.
- Álvarez-Gómez, I., 2010. Reconstrucción de la variabilidad del clima a través de los silicoflagelados conservados en los sedimentos laminados de Cuenca Alfonso. Tesis de Maestría. Instituto Politécnico Nacional; centro interdisciplinario de ciencias marinas. México. 113 p.
- Astel, A., Tsakovski, S., Barbieri, P., Simeonov, V., 2007. Comparison of self-organizing maps classification approach with cluster and principal components analysis for large environmental data sets. *Water Research*, 41(19), 4566–4578. <https://doi.org/10.1016/j.watres.2007.06.030>
- Barron, J.A., Burkry, D., Field, D., 2010. Santa Barbara basin diatom and silicoflagellate response to global climate anomalies during the past 2200 years. *Quaternary International*, 215, 34 – 44. doi:10.1016/j.quaint.2008.08.007
- Bellenger, H., Guilyardi, E., Leloup, J., Lengaigne, M., Vialard, J., 2014. ENSO representation in climate models: From CMIP3 to CMIP5. *Climate Dynamics*, 42(7–8), 1999–2018. <https://doi.org/10.1007/s00382-013-1783-z>
- Bevilacqua, V., Intini, F., Orabona, V., 2007. A model of artificial neural network for the analysis of climate change. 1, 1–11.
- Birks, H. J. B., Heiri, O., Seppa, H., Bjune, A. E., 2010. Strengths and weaknesses of quantitative climate reconstructions based on Late-Quaternary biological proxies. *Open Ecology Journal*, 3, 68 – 110. DOI: 10.2174/1874213001003020068
- Braconnot, P., Harrison, S., Kageyama, M., Bartlein, P., Masson-Delmotte, V., Abe-Ouchi, A., Zhao, Y., 2012. Evaluation of climate models using palaeoclimatic data. *Nature Climate Change*, 2(6), 417–424. <https://doi.org/10.1038/nclimate1456>

- Branstator, G., Teng, H. Y., Meehl, G. A., Kimoto, M., Knight, J. R., Latif, M., Rosati, A., 2012. Systematic estimates of initial-value decadal predictability for six AOGCMs. *Journal of Climate*, 25, 1827–1846. <https://doi.org/10.1175/JCLI-D-11-00227.1>
- Brönnimann, S., Franke, J., Nussbaumer, S. U., Zumbuhl, H. J., Steiner, D., Trachsel, M., Heger, G. C., Schurer, A., Worni, M., Malik, A., Fluckiger, J., Raible, C. C., 2019. Last phase of the Little Ice Age forced by volcanic eruptions. *Nature Geoscience*, 12, 650–656. <https://doi.org/10.1038/s41561-019-0402-y>
- Buckland, C. E., Bailey, R. M., Thomas, D. S. G., 2019. Using artificial neural networks to predict future dryland responses to human and climate disturbances. *Scientific Reports*, 9(1). <https://doi.org/10.1038/s41598-019-40429-5>
- Carro-Calvo, L., Salcedo-Sanz, S., Luterbacher, J., 2013. Neural computation in paleoclimatology: General methodology and a case study. *Neurocomputing*, 113, 262–268. <https://doi.org/10.1016/j.neucom.2012.12.045>
- Cavazos, T., 2000. Using self-organizing maps to investigate extreme climate events: An application to wintertime precipitation in the Balkans. *Journal of Climate*, 13(10), 1718–1732. [https://doi.org/10.1175/1520-0442\(2000\)013<1718:USOMTI>2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013<1718:USOMTI>2.0.CO;2)
- Céréghino, R., Giraudel, J. L., Compin, A., 2001. Spatial analysis of stream invertebrates distribution in the Adour-Garonne drainage basin (France), using Kohonen self-organizing maps. *Ecological Modelling*, 146(1–3), 167–180. [https://doi.org/10.1016/S0304-3800\(01\)00304-0](https://doi.org/10.1016/S0304-3800(01)00304-0)
- Cobb, K. M., Charles, C. D., Cheng, H., Edwards, R. L., 2003. El Nino/Southern Oscillation and tropical Pacific climate during the last millennium. *Nature*, 424, 271–276. <https://doi.org/10.1038/nature01779>
- Crowley, T. J., Baum, S. K., Kim, K. Y., Hegerl, G. C., Hyde, W. T., 2003. Modeling ocean heat content changes during the last millennium. *Geophysical Research Letters*, 30(18), 1932. doi:10.1029/2003GL017801.
- Christiansen, B., Ljungqvist, F. C., 2017. Challenges and perspectives for large-scale temperature reconstructions of the past two millennia: *Reviews of Geophysics*, 55, 40–96. <https://doi.org/10.1002/2016RG000521>
- D'Arrigo, R., Wilson, R., Jacoby, G., 2006. On the long-term context for late twentieth century warming. *Journal Geophysical Research*, 111.
- Deligiorgi, D., Philippopoulos, K., Kouroupetroglou, G., 2014. An assessment of self-organizing maps and k-means clustering approaches for atmospheric circulation classification. In: International conference on environmental science and geoscience, Venice, Italy., 15–17 March 2014, pp 17–23.
- Deo, R.C., Şahin, M., 2015. Application of the extreme learning machine algorithm for the prediction of monthly Effective Drought Index in eastern Australia. *Atmospheric Research*, 153, 512–525. <https://doi.org/10.1016/j.atmosres.2014.10.016>

- Dhanya, C. T., Kumar, D., 2010. Nonlinear ensemble prediction of chaotic daily rainfall. *Advances in Water Resources journal*, 33(3), 327–347. <https://doi.org/10.1016/j.advwatres.2010.01.001>
- Di Lorenzo, E., Combes, V., Keister, J., Strub, P. T., Thomas, A., Franks, A., Parada, C., 2013. Synthesis of Pacific Ocean climate and ecosystem dynamics. *Oceanography*, 26(4), 68–81. <https://doi.org/10.5670/oceanog.2013.76>
- Di, C., Yang, X., Wang, X., 2014. A four-stage hybrid model for hydrological time series forecasting. *PLoS ONE*, 9(8). <https://doi.org/10.1371/journal.pone.0104663>
- Diaz, H., Trigo, R., Hughes, M., Mann, M., Xoplaki, E., Barriopedro, D., 2011. Spatial and temporal characteristics of climate in medieval times revisited. *American Meteorological Society*, 1487–1500.
- Dijkstra, H. A., Petersik, P., Hernández-García, E., López, C., 2019. The application of machine learning techniques to improve El Niño prediction skill. *Frontiers in Physics*, 7(October), 1–13. <https://doi.org/10.3389/fphy.2019.00153>
- Dong, L., Zhou, T., 2014. The formation of the recent cooling in the eastern tropical Pacific Ocean and the associated climate impacts: a competition of global warming, IPO, and AMO. *Journal of Geophysical Research: Atmospheres*, 119, 11272–11287 (2014).
- Dufresne, J., Foujols, M., Denvil, S., Caubel, A., Marti, O., Aumont, O., Vuichard, N., 2013. Climate change projections using the IPSL-CM5 Earth System Model: From CMIP3 to CMIP5. *Climate Dynamics*, 40.
- Dunne, J. P., John, J. G., Shevliakova, S., Stouffer, R. J., Krasting, J. P., Malyshev, S. L., Zadeh, N., 2013. GFDL's ESM2 global coupled climate-carbon earth system models. Part II: Carbon system formulation and baseline simulation characteristics. *Journal of Climate*, 26(7), 2247–2267. <https://doi.org/10.1175/JCLI-D-12-00150.1>
- Elshorbagy, A., 2009. noise reduction approach in chaotic hydrologic time series revisited. *Canadian Water Resources Journal*, 26(4), 537–550. <https://doi.org/10.4296/cwrj2604537>
- Elshorbagy, A., Simonovic, S. P., Panu, U. S., 2002. Noise reduction in chaotic hydrologic time series: Facts and doubts. *Journal of Hydrology*, 256(3–4), 147–165. [https://doi.org/10.1016/S0022-1694\(01\)00534-0](https://doi.org/10.1016/S0022-1694(01)00534-0)
- Esper, J., Cook, E. R., Schweingruber, F. H., 2002. Low-frequency signals in long tree-ring chronologies for reconstructing past temperature variability. *Science*, 295(5563), 2250–2253. <https://doi.org/10.1126/science.1066208>
- Esper, O., Gersonde, R., 2014. Quaternary surface water temperature estimations: New diatom transfer functions for the Southern Ocean. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 414, 1–19. doi:10.1016/j.palaeo.2014.08.008

- Esper, O., Gersonde, R., Kadagies, N., 2010. Diatom distribution in southeastern Pacific surface sediments and their relationship to modern environmental variables. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 287(1–4), 1–27. <https://doi.org/10.1016/j.palaeo.2009.12.006>
- Fraedrich, K., 1986. Estimating the dimensions of weather and climate attractors. *Journal of the Atmospheric Science*, 43 (5), 419–432.
- Gómez, A. M., 2014. Redes neuronales artificiales: The Self-Organizing Maps (SOM) para el reconocimiento de patrones. *Cuadernos de Estadística Aplicada*, I(1), 27–38.
- Goni, M. A., Thunell, R. C., Woodworth, M. P., Müller-Karger, F. E., 2006. Changes in wind-driven upwelling during the last three centuries: Inter-ocean teleconnections. *Geophysical Research Letters*, 33(15), 3–6. <https://doi.org/10.1029/2006GL026415>
- González-Trueba, J., Martín Moreno, R., Serrano Cañadas, E., 2007. El glaciario de la Pequeña Edad del Hielo en las Montañas Ibéricas: Síntesis y estado actual de conocimiento. *Revista C & G*, 21(1), 57–86.
- Grassberger, P., Hegger, R., Kantz, H., Schaffrath, C., Schreiber, T., 1993. “On noise reduction methods for chaotic data”. *Chaos*, 3(2), 127–141. <https://doi.org/10.1063/1.165979>
- Grassberger, P., Schreiber, T., Schaffrath, C., 1991. Nonlinear time sequence analysis. *International Journal Bifurcation Chaos Applied Science Engineering*, 1(3), 521–547. <https://doi.org/10.1142/S0218127491000403>
- Guiot, J., Nicault, A., Rathgeber, C., Edouard, J. L., Guibal, F., Pichard, G., Till, C., 2005. Last-millennium summer-temperature variations in western Europe based on proxy data. *The Holocene*, 15(4), 489–500.
- Gupta, S. M., Malmgren, B. A., 2009. Comparison of the accuracy of SST estimates by artificial neural networks (ANN) and other quantitative methods using radiolarian data from the Antarctic and Pacific Oceans. *Journal Earth Science India*, 2(II), 52–75.
- Gutiérrez, J. D., Riss, W., Ospina, R., 2004. Bioindicación de la calidad del agua en la Sabana de Bogotá - Colombia, mediante la utilización de la lógica difusa neuro-adaptativa como herramienta. *Caldasía*, 28(1), 45-56.
- Gutiérrez, R. M., 2004. Optimal nonlinear models from empirical time series: an application to climate. *International Journal Bifurcation Chaos Applied Science Engineering*, 14(6), 2041–2052. <https://doi.org/10.1142/s0218127404010400>
- Hakspiel–Segura, C., 2014. Rutas y procesos fisiológicos del ciclo del nitrógeno en Cuenca Alfonso, Golfo de California. Tesis de doctorado, Instituto Politécnico Nacional; centro interdisciplinario de ciencias marinas. México.

- Harrison, S., Bartlein, P., Izumi, K., Li, G., Annan, J., Hargreaves, J., Kageyama, M., 2015. Evaluation of CMIP5 palaeo-simulations to improve climate projections. *Nature Climate Change*, 5(8), 735–743.
- Hegger, R., Kantz, H., Schreiber, T., 1999. Practical implementation of nonlinear time series methods. The TISEAN package. *Chaos*, 9, 413.
- Helama, S., Stoffel, M., Hall, R. J., Jones, P. D., Arppe, L., Matskovsky, V. V., Oinonen, M., 2021. Recurrent transitions to Little Ice Age-like climatic regimes over the Holocene. *Climate Dynamics*, (0123456789). <https://doi.org/10.1007/s00382-021-05669-0>
- Henley, B.J., Meehl, G., Power, S.B., Folland, C.K., King, A.D., Brown, J.N., Karoly, D.J., Delage, F., Gallant, A.J., Freund, M., Neukom, R., 2017. Spatial and temporal agreement in climate model simulations of the Interdecadal Pacific Oscillation. *Environmental Research Letters*, 12.
- Herguera, J. C., Franco, G. B., Molina-Cruz, A., 2003. Decadal surface ocean variability in the lower Gulf of California: Records for the past 300 years. *Geofísica Internacional*, 42(3), 397–406.
- Hermanson, L., Sutton, R. T., 2010. Case studies in interannual to decadal climate predictability. *Climate Dynamics*, 35, 1169–1189.
- Huang, J., Xie, Y., Guan, X., Li, D., Ji, F., 2017. The dynamics of the warming hiatus over the Northern Hemisphere. *Climate Dynamics*, 48(1–2), 429–446. <https://doi.org/10.1007/s00382-016-3085-8>
- Iles, C., Hegerl, G., 2017. Role of the North Atlantic Oscillation in decadal temperature trends. *Environmental Research Letters*, 12(11). <https://doi.org/10.1088/1748-9326/aa9152>
- IPCC., 1997. Introducción a los modelos climáticos simples utilizados en el segundo informe de evaluación del IPCC.
- IPCC., 2007. Cambio climático 2007: Informe de síntesis. Contribución de los Grupos de trabajo I, II y III al Cuarto Informe de evaluación del Grupo Intergubernamental de Expertos sobre el Cambio Climático. *Nature* (Vol. 446).
- Jacoby, G. C., D'Arrigo, R. D., 1989. Reconstructed Northern Hemisphere annual temperature since 1671 based on high- latitude tree-ring data from North America. *Climate Change*, 14, 39–59.
- Johnson, N. C., Amaya, D. J., Ding, Q., Kosaka, Y., Tokinaga, H., Xie, S. P., 2020. Multidecadal modulations of key metrics of global climate change. *Global and Planetary Change*, 188(March), 103149. <https://doi.org/10.1016/j.gloplacha.2020.103149>
- Jones, P. D., Mann, M. E., 2004. Climate over past millennia. *Reviews of Geophysics*, 42 p.

- Kalteh, A. M., Hjorth, P., Berndtsson, R., 2008. Review of the self-organizing map (SOM) approach in water resources: Analysis, modelling and application. *Environmental Modelling and Software*, 23(7), 835–845. <https://doi.org/10.1016/j.envsoft.2007.10.001>
- Kantz, H., 1994. A robust method to estimate the maximal Lyapunov exponent of a time series. *Physics Letters A*, 185(1), 77–87. [https://doi.org/10.1016/0375-9601\(94\)90991-1](https://doi.org/10.1016/0375-9601(94)90991-1)
- Kantz, H., Schreiber, T., 1997. Nonlinear time series analysis. Cambridge University Press. Cambridge.
- Kara, A. B., Rochford, P., Hurlburt, H., 2000. An optimal definition for ocean mixed layer depth. *Journal geophysical Research*, 105 (C7), 16803–16821.
- Kemp, A., Pike, J., Pearce, R., Lange, C., 2000. The “Fall dump” a new perspective on the role of a “shade flora” in the annual cycle of diatom production and export flux. *Deep sea Research II*, 47, 2129–2154.
- Kennington, K., Haslett, S., Funnell, B., 1999. Offshore transport of neritic diatoms an indicator of surface current and trade wind strength in the Plio-Pleistocene eastern equatorial Pacific. *Palaeogeography Palaeoclimatology Palaeoecology*, 149, 171–181.
- Kohonen, T., 2001. Self-Organizing Maps, 3er. Extended Edition. Springer-verlag, Berlin Heidelberg.
- Krishnamurthy, V., 2018. Predictability of CFSv2 in the tropical Indo-Pacific region, at daily and subseasonal time scales. *Climate Dynamics*, 50(11–12), 3931–3948. <https://doi.org/10.1007/s00382-017-3855-y>
- Krishnamurthy, V., 2019. Predictability of Weather and Climate. *Earth and Space Science*, 6(7), 1043–1056. <https://doi.org/10.1029/2019EA000586>
- Kucera, M., Weinelt, M., Kiefer, T., Pflaumann, U., Hayes, A., Weinelt, M., Waelbroeck, C., 2005. Reconstruction of sea-surface temperatures from assemblages of planktonic foraminifera: Multi-technique approach based on geographically constrained calibration data sets and its application to glacial Atlantic and Pacific Oceans. *Quaternary Science Reviews*, 24, 951–998.
- Lechleitner, F.A., Breitenbach, S., Rehfeld, K., Ridley, H., Asmerom, Y., Pruger, K. M., Marwan, N., Goswami, B., Kennett, D. J., Aquino, V. V., Polyak, V., Haug, G. H., Eglinton, T. I., Baldini, J., 2017. Tropical rainfall over the last two millennia: evidence for a low-latitude hydrologic seesaw. *Scientific Reports*, 7, 45809. <https://doi.org/10.1038/srep45809>
- Lee, J. H. W., Huang, Y., Dickman, M., Jayawardena, A. W., 2003. Neural network modelling of coastal algal blooms. *Ecological Modelling journal*, 159, 179–201.
- Lei, M., Meng, G., 2011. Symplectic principal component analysis: A new method for time series analysis. *Mathematical Problems in Engineering*, 2011, 1–14. <https://doi.org/10.1155/2011/793429>

- Li, T., Sun, G., Yang, C., Liang, K., Ma, S., Huang, L., 2018. Using self-organizing map for coastal water quality classification: Towards a better understanding of patterns and processes. *Science of the Total Environment*, 628–629, 1446–1459. <https://doi.org/10.1016/j.scitotenv.2018.02.163>
- Li, X., Li, Z., Huang, W., Zhou, P., 2020. Performance of statistical and machine learning ensembles for daily temperature downscaling. *Theoretical and Applied Climatology*, 140(1–2), 571–588. <https://doi.org/10.1007/s00704-020-03098-3>
- Liu, Z., Peng, C., Xiang, W., Tian, D., Deng, X., Zhao, M., 2010. Application of artificial neural networks in global climate change and ecological research: An overview. *Chinese Science Bulletin*, 55, 3853–3863. <https://doi.org/10.1007/s11434-010-4183-3>
- Lopes, C., Mix, A. C., Abrantes, F., 2006. Diatoms in northeast Pacific surface sediments as paleoceanographic proxies. *Marine Micropaleontology*, 60, 45–65. <https://doi.org/10.1016/j.marmicro.2006.02.010>
- Luckman, B. H., Briffa, K. R., Jones, P. D., Schweingruber, F. H., 1997. “Tree-ring based reconstruction of summer temperatures at the Columbia Icefield, Alberta, Canada, AD 1073–1983”. *The Holocene*, 7(4), 375–389.
- Malmgren, A., Kucera, M., Nyberg, J., Waelbroeck, C., 2001. Comparison of statistical and artificial neural network techniques for estimating past sea surface temperatures from planktonic foraminifer census data: *Paleoceanography and Paleoclimatology*, 16(0), 1–11.
- Malmgren, B. A., Witon, E., Schrader, H., Huang, Y., 2002. An artificial neural network application based on diatom census data for estimation of north Atlantic marine quaternary paleotemperatures, In: European Geophysical Union, Nice. *Abstract EGS02-A-02116*.
- Maneta, M., Schnabel, S., 2003. Aplicación de redes neuronales artificiales para determinar la distribución espacial de la humedad del suelo en una pequeña cuenca de drenaje. Estudios preliminares. Estudios de la zona no saturada del suelo. Vol. VI.
- Mann, M. E., Bradley, R. S., Hughes, M. K., 1999. Northern Hemisphere temperatures during the past millennium: Inferences, uncertainties and limitations. *Geophysical Research Letters*, 26, 759–762.
- Mann, M. E., Bradley, R. S., Hughes, M. K., 1998. Global-scale temperature patterns and climate forcing over the past six centuries. *Nature*, 392(6678), 779–787.
- Mann, M. E., Cane, M. A., Zebiak, S. E., Clement, A., 2005. Volcanic and solar forcing of the tropical Pacific over the past 1000 years. *Journal of Climate*, 18(3), 447–456. <https://doi.org/10.1175/JCLI-3276.1>
- Mann, M. E., Jones, P. D., 2003. Global surface temperatures over the past two millennia, *Geophysical Research Letters*, 30(15), 1820. doi:10.1029/2003GL017814.

- Mann, M. E., Zhang, Z., Hughes, M. K., Bradley, R. S., Miller, S. K., Rutherford, S., Ni, F., 2008. Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia. *Proceedings of the National Academy of Sciences*, 105(36), 13252–13257.
- Martínez-López, A., Álvarez-Gómez, I. G., Pérez-Cruz, L., Verdugo-Díaz, G., Villegas-Aguilera, M., 2016. Production, exportation and preservation of silicoflagellates in Alfonso Basin, Gulf of California. *Journal of Sea Research*, 109, 52–62. <https://doi.org/10.1016/j.seares.2016.01.005>
- Martínez-López, A., Flores-Castillo, O., Saldívar-Lucio, R., Escobedo-Urías, D., Verdugo-Díaz, G., Pérez-Cruz, L., Albañez-Lucero, M., Acevedo-Acosta, J. D., 2019. Paleoclimate of the Gulf of California northwestern Mexico during the last 2000 years. In: Islebe, G., Torrescano-Valle, N., (eds.) *The Holocene and Anthropocene environmental and climate history of Mexico*. Springer Nature Switzerland, pp 1–37.
- Marwan, N., Trauth, M. H., Vuille, M., Kurths, J., 2003. Comparing modern and Pleistocene ENSO-like influences in NW Argentina using nonlinear time series analysis methods. *Climate Dynamics*, 21(3–4), 317–326. <https://doi.org/10.1007/s00382-003-0335-3>
- Matthews, J. A., Briffa, K. R., 2005. The “Little Ice Age”: Re-evaluation of an evolving concept. *Geografiska Annaler, Series A: Physical Geography*, 87(1), 17–36.
- Mccabe-Glynn, S., Johnson, K. R., Strong, C., Berkelhammer, M., Sinha, A., Cheng, H., Edwards, R. L., 2013. Variable North Pacific influence on drought in southwestern North America since AD 854. *Nature Geoscience*, 6(8), 617–621. <https://doi.org/10.1038/ngeo1862>
- Meehl, G., Covey, C., Delworth, T., Latif, M., McAvaney, B., Mitchell, J., Taylor, K., 2007. The WCRP CMIP3 multimodel dataset: A new era in climatic change research. *Bulletin of the American Meteorological Society*, 88(9), 1383–1394.
- Millán, H., Ghanbarian-Alavijeh, B., García-Fornaris, I., 2010. Nonlinear dynamics of mean daily temperature and dew point time series at Babolsar, Iran, 1961–2005. *Atmospheric Research*, 98(1), 89–101. <https://doi.org/10.1016/j.atmosres.2010.06.001>
- Miller, G. H., Geirsdóttir, Á., Zhong, Y., Larsen, D. J., Otto-Bliesner, B. L., Holland, M. M., Thordarson, T., 2012. Abrupt onset of the Little Ice Age triggered by volcanism and sustained by sea-ice/ocean feedbacks. *Geophysical Research Letters*, 39(2), 1–5. <https://doi.org/10.1029/2011GL050168>
- Monerie, P. A., Robson, J., Dong, B., Hodson, D., 2021. Role of the Atlantic multidecadal variability in modulating East Asian climate. *Climate Dynamics*, 56(1–2), 381–398. <https://doi.org/10.1007/s00382-020-05477-y>
- Monreal-Gómez, M., Molina-Cruz, A., Salas de León, D., 2001. Water masses and cyclonic circulation in bay of La Paz, Gulf of California, during June 1998. *Journal of Marine Systems*, 30(3–4), 305–315. doi:10.1016/S0924-7963(02)00075-1

- Nava-Sánchez, E. H., Gorsline, D., Molina-Cruz, A., 2001. The Baja California peninsula borderland: structural and sedimentological characteristics. *Sedimentary Geology*, 144, 63–82. doi:10.1016/S0037-0738(01)00135-X
- Neukom, R., Schurer, A. P., Steiger, N. J., Hegerl, G. C., 2018. Possible causes of data model discrepancy in the temperature history of the last Millennium. *Scientific Reports*, 8(1), 1–15. <https://doi.org/10.1038/s41598-018-25862-2>
- Neukom, R., Steiger, N., Gómez-Navarro, J. J., Wang, J., Werner, J. P., 2019. No evidence for globally coherent warm and cold periods over the preindustrial Common Era. *Nature*, 571(7766), 550–554.
- Obeso-Nieblas, M., Shirasago, B., Gaviño-Rodríguez, J., Pérez-Lezama, E., Obeso-Huerta, H., Jiménez-Illescas, A., 2008. Variabilidad hidrográfica en Bahía de La Paz, Golfo de California, México (1995-2005). *Revista de Biología Marina y Oceanografía*, 43(3), 559-567.
- Olawoyin, R., Nieto, A., Grayson, R. L., Hardisty, F., Oyewole, S., 2013. Application of artificial neural network (ANN)-self-organizing map (SOM) for the categorization of water, soil and sediment quality in petrochemical regions. *Expert Systems with Applications*, 40(9), 3634–3648. <https://doi.org/10.1016/j.eswa.2012.12.069>
- Olli, K., Wexels, C., Wassmann, P., Ratkova, T., Arashkevich, E., Pasternak, A., 2001. Vertical flux of biogenic matter during a Lagrangian study off the NW Spanish continental margin. *Progress in Oceanography*, 51(2), 443–466. DOI:10.1016/S0079-6611(01)00079-9
- Ortega, P., Robson, J., Moffa-Sanchez, P., Thornalley, D., Swingedouw, D., 2017. A last millennium perspective on North Atlantic variability: exploiting synergies between models and proxy data. *Past Global Changes Magazine*, 25(1), 61–67.
- PAGES2k Consortium., 2017. A global multiproxy database for temperature reconstructions of the Common Era. *Scientific Data*, 4, 1–33.
- Park, Y. S., Céréghino, R., Compin, A., Lek, S., 2003. Applications of artificial neural networks for patterning and predicting aquatic insect species richness in running waters. *Ecological Modelling*, 160(3), 265–280. [https://doi.org/10.1016/S0304-3800\(02\)00258-2](https://doi.org/10.1016/S0304-3800(02)00258-2)
- Park, Y. S., Gevrey, M., Lek, S., Giraudel, J. L., 2005. Evaluation of relevant species in communities: development of structuring indices for the classification of communities using a self organizing map. In: Lek, S., Scardi, M., Verdonschot, P., Descy, P., Park, Y., (eds.) *Modelling Community Structure in Freshwater Ecosystems*. Springer, Berlin, pp 369–380.
- Park, Y. S., Tison, J., Lek, S., Giraudel, J., Coste, M., Delmas, F., 2006. Application of a self-organizing map to select representative species in multivariate analysis: A case study determining diatom distribution patterns across France. *Ecological informatics*, 1(3), 247–257. <https://doi.org/10.1016/j.ecoinf.2006.03.005>

- Pasini, A., Lorè, M., Ameli, F., 2006. Neural network modelling for the analysis of forcings temperatures relationships at different scales in the climate system. *Ecological Modelling*, 191, 58–67.
- Peyron, O., De-Vernal, A., 2001. Application of artificial neural networks ANN to high-latitude dinocyst assemblages for the reconstruction of the past sea surface conditions in Arctic and sub-Arctic Seas. *Journal of Quaternary Science*, 16(7), 699–709. <https://doi.org/10.1002/jqs.651>
- Peyron, O., Guiot, J., Cheddadi, R., Tarasov, P., Reille, M., De Beaulieu, J.L., Bottema, S., Andrieu, V., 1998. Climatic reconstruction in Europe for 18,000 YR B.P. from pollen data. *Quat Res*, 49(2), 183–196. <https://doi.org/10.1006/qres.1997.1961>
- Racca, J., Racca, R., Pienitz, R., Prairie, Y., 2007. PaleoNet: new software for building, evaluating and applying neural network based transfer functions in paleoecology, 467–472.
- Ren, J., Gersonde, R., Esper, O., Sancetta, C., 2014. Diatom distributions in northern North Pacific surface sediments and their relationship to modern environmental variables. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 402, 81–103. <https://doi.org/10.1016/j.palaeo.2014.03.008>
- Schrader, H., Sorknes, R., 1991. Peruvian coastal upwelling: Late Quaternary productivity changes revealed by diatoms. *Marine Geology*, 97(3 – 4), 233–249.
- Schrader, H.J., Gersonde, R., 1978. Diatoms and silicoflagellates. *Utrecht Micro-Paleontological Bulletins*, 17, 67–81
- Schreiber, T., 1993. Extremely simple nonlinear noise-reduction method. *Physical Review E*, 47(4), 2401–2404. doi:10.1103/PhysRevE.47.2401
- Schreiber, T., Hegger, R., Schaffrath, C., Grassberger, P., Kantz, H., 2002. On noise reduction methods for chaotic data. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 3(2), 127–141. <https://doi.org/10.1063/1.165979>
- Shimada, C., Sato, T., Toyoshima, S., Yamasaki, M., Tanimura, Y., 2008. Paleocological significance of laminated diatomaceous oozes during the middle-to-late Pleistocene, North Atlantic Ocean IODP Site U1304. *Mar Micropaleontology*, 69(2), 139–150. <https://doi.org/10.1016/j.marmicro.2008.07.004>
- Sigl, M., McConnell, J., Toohey, M., Plunkett, G., Ludlow, F., Winstrup, M., Motizuki, Y., 2015. The history of volcanic eruptions since Roman times. *Past Global Change Magazine*, 23(2), 48–49. <https://doi.org/10.22498/pages.23.2.48>
- Sivakumar, B., 2000. Chaos theory in hydrology: Important issues and interpretations. *Journal of Hydrology*, 227(1–4), 1–20. [https://doi.org/10.1016/S0022-1694\(99\)00186-9](https://doi.org/10.1016/S0022-1694(99)00186-9)
- Sivakumar, B., 2004. Chaos theory in geophysics past, present and future. *Chaos, Solitons and Fractals*, 19(2), 441–462. doi:10.1016/S0960-0779(03)00055-9.

- Sivakumar, B., Phoon, K., Liong, S., Liaw, C., 1999. A systematic approach to noise reduction in chaotic hydrological time series. *Journal of Hydrology*, 219(3–4), 103–135. [https://doi.org/10.1016/S0022-1694\(99\)00051-7](https://doi.org/10.1016/S0022-1694(99)00051-7)
- Sun, J., Wang, H., 2006. Relationship between Arctic Oscillation and Pacific Decadal Oscillation on decadal timescales. *Chin. Sci. Bull.*, 51, 75–79.
- Suzuki, K., 2011. Artificial neural networks-methodological Advances and Biomedical Applications. Intech, Rijeka, Croatia. 374 p.
- Takens, F., 1985. Dynamical systems and bifurcations. In: Braaksma, B., Broer, H., Takens, F., (eds.) Lecture notes in mathematics. Springer-Verlag, Heidelberg.
- Tardif, R., Hakim, G. J., Perkins, W. A., Horlick, K. A., Erb, M. P., Emile-Geay, J., Noone, D., 2019. Last Millennium Reanalysis with an expanded proxy database and seasonal proxy modeling. *Climate of the Past*, 15(4), 1251–1273. <https://doi.org/10.5194/cp-15-1251-2019>
- Taylor, K. E., Stouffer, R. J., Meehl, G. A., 2012. An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, 93(4), 485–498. <https://doi.org/10.1175/BAMS-D-11-00094.1>
- Teng, H. Y., Branstator, G., Meehl, G. A., 2011. Predictability of the Atlantic Overturning Circulation and associated surface patterns in two CCSM3 climate change ensemble experiments. *J. Clim.*, 24, 6054–6076
- Tierney, J. E., Abram, N. J., Anchukaitis, K. J., Evans, M. N., Giry, C., Kilbourne, K. H., Zinke, J., 2015. Tropical sea surface temperatures for the past four centuries reconstructed from coral archives. *Paleoceanography*, 30(3), 226–252. <https://doi.org/10.1002/2014PA002717>
- Tran, T. T. K., Bateni, S. M., Ki, S. J., Vosoughifar, H., 2021. A review of neural networks for air temperature forecasting. *Water*, 13(9), 1–15. <https://doi.org/10.3390/w13091294>
- Trigo, R. M., Palutikof, J. P., 1999. Simulation of daily temperatures for climate change scenarios over Portugal: A neural network model approach. *Climate Research*, 13(1), 45–59. <https://doi.org/10.3354/cr013045>
- Tsai, W. P., Huang, S. P., Cheng, S. T., Shao, K. T., Chang, F. J., 2016. A data-mining framework for exploring the multi-relation between fish species and water quality through self-organizing map. *Science of the Total Environment*, 579(1), 474–483. <https://doi.org/10.1016/j.scitotenv.2016.11.071>
- Tseng, Y. H., Ding, R., Zhao, S., Kuo, Y. C., Liang, Y. C., 2020. Could the North Pacific Oscillation be modified by the initiation of the East Asian winter monsoon? *Journal of Climate*, 33(6), 2389–2406. <https://doi.org/10.1175/JCLI-D-19-0112.1>
- Tung, W., Gao, J., Hu, J., Yang, L., 2011. Detecting chaos in heavy-noise environments. *Physical Review E: Statistical, Nonlinear and Soft Matter Physics*, 83(4), 1–9. <https://doi.org/10.1103/PhysRevE.83.046210>

- Vallina, S., Follows, M., Dutkiewicz, M., Montoya, J., Cermeno, P., Loreau, M., 2014. Global relationship between phytoplankton diversity and productivity in the ocean. *Nature Communications*, 5, 4299. <https://doi.org/10.1038/ncomms5299>
- Verdugo–Díaz, G., 2004. Respuesta eco fisiológica del fitoplancton ante la variabilidad ambiental en una bahía subtropical de Baja California Sur, México. Tesis doctoral. Instituto Politécnico Nacional; centro interdisciplinario de ciencias marinas (CICIMAR). México. 138 p.
- Verdugo–Díaz, G., Martínez–López, A., Hakspiel–Segura, C., 2014. Producción primaria y eficiencia fotosintética en Cuenca Alfonso, Bahía de La Paz, Golfo de California, México. *Revista de biología marina y oceanografía*, 49(3), 527–536. <http://dx.doi.org/10.4067/S0718-19572014000300009>
- Verity, P. G., Wassmann, P., Frischer, M. E., Howard-Jones, M. H., Allen, A. E., 2002. Grazing of phytoplankton by microzooplankton in the Barents Sea during early summer. *Journal of Marine Systems*, 38(1–2), 109–123.
- Vesanto, J., 2000. Neural network tool for data mining: SOM Toolbox. SOM toolbox team, Helsinki University of technology, Finland
- Villegas-Aguilera, M., 2009. Fitoplancton silíceo de la zona eufótica, como señal de la productividad primaria en Cuenca Alfonso, Golfo de California. Tesis de maestría, Instituto Politécnico Nacional; centro interdisciplinario de ciencias marinas. México. 112p
- Wilson, C., Villareal, T. A., Maximenko, N., Bograd, S. J., Montoya, J. P., Schoenbaechler, C. A., 2006, Biological and physical forcings of late summer chlorophyll blooms at 30° N in the oligotrophic Pacific. *Journal of Marine Systems*, 69(3), 164 – 176. [10.1016/j.jmarsys.2005.09.018](https://doi.org/10.1016/j.jmarsys.2005.09.018)
- Wilson, R., Anchukaitis, K., Briffa, K. R., Büntgen, U., Cook, E., D'Arrigo, R., Zorita, E., 2016. Last millennium northern hemisphere summer temperatures from tree rings: Part I: The long term context. *Quaternary Science Reviews*, 134, 1–18. <https://doi.org/10.1016/j.quascirev.2015.12.005>
- Wu, C., Chen, Y., Peng, C., Li, Z., Hong, X., 2019. Modeling and estimating aboveground biomass of *Dacrydium pierrei* in China using machine learning with climate change. *Journal of Environmental Management*, 234, 167–179. <https://doi.org/10.1016/j.jenvman.2018.12.090>
- Yingyi, C., 1993. Predictability of the 500 hPa height field. *Advances in Atmospheric Sciences*, 10(4), 497–503. <https://doi.org/10.1007/BF02656975>
- Zanchettin, D., Timmreck, C., Graf, H. F., Rubino, A., Lorenz, S., Lohmann, K., Jungclaus, J. H., 2012. Bi-decadal variability excited in the coupled ocean-atmosphere system by strong tropical volcanic eruptions. *Climate Dynamics*, 39(1–2), 419–444. <https://doi.org/10.1007/s00382-011-1167-1>

- Zhang, J., Lam, K. C., Yan, W. J., Gao, H., Li, Y., 2004. Time series prediction using Lyapunov exponents in embedding phase space. *Computers and Electrical Engineering*, 30(1), 1–15. [https://doi.org/10.1016/S0045-7906\(03\)00015-6](https://doi.org/10.1016/S0045-7906(03)00015-6)
- Zhao, M., Eglinton, G., Read, G., Schimmelmann, A., 2000. An alkenone (U37/(K')) quasi-annual sea surface temperature record (A.D. 1440 to 1940) using varved sediments from the Santa Barbara Basin. *Organic Geochemistry*, 31(9), 903–917. [https://doi.org/10.1016/S0146-6380\(00\)00034-6](https://doi.org/10.1016/S0146-6380(00)00034-6)