



TESIS DOCTORAL

Visualización, predicción y análisis de variables climáticas del océano Atlántico mediante técnicas de aprendizaje automático

Autor

Juan José Carrasco Fernández

Directores

Dr. Antonio Geraldo Ferreira

Dr. Juan Gómez Sanchis

Dr. Joan Vila Francés

Tutor

Dr. Emilio Soria Olivas

Programa de Doctorado en Ingeniería Electrónica
Departament d'Enginyeria Electrònica
UNIVERSITAT DE VALÈNCIA – ESTUDI GENERAL
Valencia – Octubre, 2019

Visualización, predicción y análisis de variables
climáticas del océano Atlántico mediante técnicas de
aprendizaje automático

Juan José Carrasco Fernández, octubre de 2019

Departament d'Enginyeria Electrònica
Escola Tècnica Superior d'Enginyeria

D. Antonio Geraldo Ferreira, Doctor en Física del Medio Ambiente por la Universitat de València, Profesor Adjunto Nivel IV del Instituto de Ciencias do Mar - LABOMAR de la Universidade Federal do Ceará-Brasil,

D. Juan Gómez Sanchis, Doctor por la Universitat de València, Profesor Titular del Departamento de Ingeniería Electrónica de la Escola Tècnica Superior d'Enginyeria de la Universitat de València,

D. Joan Vila Francés, Doctor en Ingeniería Electrónica por la Universitat de València, Profesor Titular del Departamento de Ingeniería Electrónica de la Escola Tècnica Superior d'Enginyeria de la Universitat de València, y

D. Emilio Soria Olivas, Doctor en Ingeniería Electrónica por la Universitat de València, Profesor Titular del Departamento de Ingeniería Electrónica de la Escola Tècnica Superior d'Enginyeria de la Universitat de València

HACEN CONSTAR QUE:

El Graduado en Ingeniería Electrónica de Telecomunicación D. Juan José Carrasco Fernández ha realizado bajo nuestra dirección el trabajo titulado “Visualización, predicción y análisis de variables climáticas del océano Atlántico mediante técnicas de aprendizaje automático”, que se presenta en esta memoria para optar al grado de Doctor.

Y para que así conste a los efectos oportunos, firmamos el presente certificado, en Valencia, a _____ .

D. Antonio Geraldo Ferreira

D. Juan Gómez Sanchis

D. Joan Vila Francés

D. Emilio Soria Olivas

Enrique J. Sanchis Peris
Dir. del Departamento

Tesis Doctoral: VISUALIZACIÓN, PREDICCIÓN Y ANÁLISIS
DE VARIABLES CLIMÁTICAS DEL OCÉANO ATLÁNTICO
MEDIANTE TÉCNICAS DE APRENDIZAJE AUTOMÁTICO

Autor: JUAN JOSÉ CARRASCO FERNÁNDEZ

Directores: Dr. Antonio Geraldo Ferreira
Dr. Juan Gómez Sanchis
Dr. Joan Vila Francés

Tutor: Dr. Emilio Soria Olivas

El tribunal nombrado para juzgar la Tesis Doctoral arriba citada, compuesto por los doctores:

Presidente: _____

Vocal: _____

Secretario: _____

Acuerda otorgarle la calificación de _____

Y para que así conste a los efectos oportunos, firmamos el presente certificado.

Valencia, a

Agradecimientos

Además de esfuerzo y dedicación, la elaboración de una tesis doctoral requiere del apoyo de muchas personas. Por eso, me gustaría dar las gracias a todos los compañeros, familiares y amigos que, de una manera u otra, me han ayudado a finalizar este trabajo.

En primer lugar, quiero agradecer a mis directores Antonio Geraldo Ferreira, Juan Gómez, Joan Vila y a mi tutor Emilio Soria, el tiempo y dedicación invertidos. A Geraldo por aportarme sus conocimientos sobre oceanografía y meteorología, a Juan y a Joan por su experiencia en la aplicación de los algoritmos, y a Emilio por introducirme en el mundo de la ciencia de datos y proponerme la realización de este trabajo. Vuestra ayuda, ideas y consejos han sido fundamentales.

Por supuesto, agradezco también el apoyo recibido por parte del resto de compañeros del grupo IDAL (*Intelligent Data Analysis Laboratory*): Marcelino, Antonio, Rafa, José David, José María, Fran, Pablo, Delia, Fernando, Jose y Juanjo. Pese a que en los últimos años no hemos compartido lugar de trabajo, siempre es un placer disfrutar de vuestra compañía alrededor de una buena mesa.

También quiero dar las gracias por el apoyo y palabras de ánimo a todos los compañeros del Departamento de Fisioterapia, en especial a Sofía Pérez y Felipe Querol. Trabajar con vosotros en temas tan alejados de mi ámbito académico, está suponiendo una gran experiencia investigadora y personal.

Quiero dar las gracias también a Ana Paula de LABOMAR (*Instituto de Ciencias do MAR*) por contribuir con sus conocimientos sobre oceanografía física.

Fuera del entorno académico, mis amigos y amigas también merecen mi más sentido agradecimiento. El interés mostrado, las frecuentes muestras de apoyo y los buenos momentos juntos, han supuesto un impulso imprescindible. Por último, quiero dar las gracias a mis padres y a mi hermano por confiar siempre en mí y apoyarme de manera incondicional.

Juan José Carrasco Fernández
Valencia, 2019

*A Juan, Celia
y Alberto*

Índice general

Índice de figuras	VII
Índice de tablas	XV
Resumen	XXI
Abstract	XXIII
Acrónimos	XXV
1. Introducción	1
1.1. Variabilidad climática en el océano Atlántico tropical	1
1.2. Proyecto PIRATA	2
1.3. Aprendizaje automático	5
1.4. Organización de la Tesis Doctoral	9
1.5. Motivación y objetivos de la Tesis Doctoral	10
2. Descripción del área de estudio y las boyas PIRATA	13
2.1. Área de estudio	13
2.1.1. Zona de convergencia intertropical (ITCZ)	13
2.1.2. Caracterización climática	15
2.2. Boyas PIRATA	20
2.2.1. ATLAS <i>Hardware</i>	20
2.2.2. Variables, frecuencia de muestreo y control de calidad de los datos	23

2.2.3. Telemetría y distribución de datos	24
3. Estudio de la relación entre el dipolo PIRATA del Atlántico y las anomalías de lluvias en el Nordeste de Brasil	27
3.1. Introducción	27
3.2. Objetivos	28
3.3. Materiales y métodos	28
3.3.1. Área de estudio y caracterización climática	28
3.3.2. Datos	32
3.3.3. Análisis de correlación	34
3.3.3.1. Coeficiente de correlación de <i>Pearson</i>	34
3.3.3.2. Coeficiente de correlación de rangos de <i>Spearman</i>	35
3.3.3.3. Coeficiente de correlación de rangos de <i>Kendall</i>	35
3.4. Resultados	35
3.5. Conclusiones	42
4. Identificación de patrones en los datos PIRATA mediante mapas auto-organizados	43
4.1. Introducción	43
4.2. Objetivos	45
4.3. Materiales y métodos	46
4.3.1. Datos	46
4.3.2. Mapas auto-organizados	48
4.3.3. Ajustes de los parámetros de los mapas	50
4.4. Resultados	52
4.4.1. Análisis de la boya 0°N - 35°W	52
4.4.1.1. Análisis de la SST y de la temperatura en subsuperficie	52
4.4.1.2. Análisis de la salinidad	57
4.4.1.3. Análisis de la densidad	59
4.4.1.4. Análisis de la profundidad de la isoterma a 20 °C	62
4.4.1.5. Análisis de la relación entre velocidad del viento, humedad relativa y temperatura del aire	63
4.4.2. Análisis de la boya 0°N - 23°W	63

4.4.2.1.	Análisis de la SST y temperatura en subsuperficie . . .	64
4.4.2.2.	Relaciones entre densidad, temperatura y salinidad . . .	67
4.4.2.3.	Análisis de la profundidad de la isoterma a 20 °C . . .	68
4.4.2.4.	Análisis de la relación entre velocidad del viento, humedad relativa y temperatura del aire	70
4.4.3.	Análisis de las boyas 8°N – 38°W y 8°S – 30°W	71
4.4.3.1.	Análisis de la SST y temperatura en subsuperficie . . .	73
4.4.3.2.	Relación entre densidad, temperatura y salinidad . . .	76
4.4.3.3.	Análisis de la profundidad de la isoterma a 20 °C . . .	79
4.4.3.4.	Análisis de la relación entre velocidad del viento, humedad relativa y temperatura del aire	80
4.5.	Conclusiones	82
5. Predicción de la SST mediante técnicas de aprendizaje automático		85
5.1.	Introducción	85
5.2.	Objetivos	87
5.3.	Materiales y métodos	90
5.3.1.	Datos utilizados y procesado	90
5.3.2.	Metodología de los experimentos preliminares	92
5.3.3.	Reducción de la dimensionalidad. Selección y extracción de características	94
5.3.3.1.	Análisis de componentes principales	95
5.3.3.2.	Algoritmo <i>RReliefF</i>	95
5.3.3.3.	Análisis de correlación	96
5.3.4.	Metodología y definición de los modelos de predicción	97
5.3.4.1.	Modelo trivial	99
5.3.4.2.	Modelo climatología	100
5.3.4.3.	Regresión lineal múltiple robusta	100
5.3.4.4.	Perceptrón multicapa	101
5.3.4.5.	Máquinas de aprendizaje extremo	105
5.3.4.6.	Redes neuronales de regresión generalizada	108
5.3.4.7.	Árboles de regresión aleatorios	110

5.3.4.8.	Máquinas de vectores soporte	113
5.3.4.9.	Redes neuronales recurrentes LSTM	116
5.3.4.10.	Sistema de pronóstico climático NCEP-CFSv2	119
5.4.	Resultados	120
5.4.1.	Experimentos preliminares	120
5.4.2.	Modelo trivial y modelo climatología	124
5.4.3.	Modelos diarios	125
5.4.3.1.	Fecha codificada y SST como variables de entrada	125
5.4.3.2.	Modelos con todas las variables de entrada	129
5.4.3.3.	Modelos aplicando PCA	129
5.4.3.4.	Modelos aplicando <i>RReliefF</i>	131
5.4.3.5.	Modelos aplicando AC y <i>RReliefF</i>	132
5.4.4.	Modelos mensuales	137
5.4.5.	Modelos diarios y mensuales con LSTM	142
5.4.6.	Análisis estadístico de los resultados	143
5.4.6.1.	Análisis estadístico modelos diarios	144
5.4.6.2.	Análisis estadístico modelos mensuales	145
5.4.6.3.	Análisis estadístico modelos LSTM	148
5.4.6.4.	Coste computacional	148
5.5.	Conclusiones	149
6.	Visualización y predicción de datos PIRATA con Processing	151
6.1.	Introducción	151
6.2.	Objetivos	153
6.3.	Materiales y métodos	154
6.3.1.	Entorno de programación	154
6.3.2.	Descripción de la aplicación	155
6.3.3.	Funcionamiento interno	160
6.3.3.1.	Carga de fichero de datos	160
6.3.3.2.	Procesado de datos	162
6.3.3.3.	Cálculo de las predicciones	162
6.3.3.4.	Visualización	164

6.4. Resultados	165
6.4.1. Exploración de la profundidad de la isoterma a 20 °C y de la temperatura del mar	165
6.4.2. Dipolo negativo del año 2009	168
6.4.3. Dipolo positivo del año 2010	171
6.4.4. Predicciones para el periodo 2019 - 2020	172
6.5. Conclusiones	175
7. Conclusiones y proyección futura	177
7.1. Resumen y conclusiones generales	177
7.2. Conclusiones específicas	178
7.2.1. Estudio de la relación entre el dipolo PIRATA del Atlántico y las anomalías de lluvias en el Nordeste de Brasil	179
7.2.2. Identificación de patrones en los datos PIRATA mediante mapas auto-organizados	179
7.2.3. Predicción de la SST mediante técnicas de aprendizaje automático	181
7.2.4. Visualización y predicción de datos PIRATA con Processing . .	182
7.3. Proyección futura	183
Bibliografía	185

Índice de figuras

1.1. Red de boyas principal (cuadrados rojos) y expansiones Sudoeste (cuadrados amarillos), Sudeste (estrella marrón) y Nordeste (estrellas azules) del proyecto PIRATA en el océano Atlántico tropical. Estados Unidos, Brasil y Francia realizan el mantenimiento de las boyas (NOAA / PMEL).	4
1.2. Ejemplo de problema de clasificación (a) y de regresión (b) resueltos con máquinas de vectores soporte.	7
1.3. Ejemplo de problema de <i>clustering</i> con el algoritmo <i>k-means</i> y ejemplo de aplicación de un mapa auto-organizado.	7
1.4. Elementos que componen el problema de aprendizaje por refuerzo.	8
1.5. Etapas del aprendizaje automático.	9
2.1. Imágenes de la posición de la ITCZ tomadas con los satélites Meteosat-10 (2016) y Meteosat-11 (2018) en el canal 7 IR (adaptada de FUNCEME).	14
2.2. Climatologías mensuales (enero-junio) de la temperatura en la superficie del mar y del <i>pseudo-stress</i> del viento (datos promedio de 1974 a 2013). (Fuente de los datos: Servain/FUNCEME).	16
2.3. Climatologías mensuales (julio-diciembre) de la temperatura en la superficie del mar y del <i>pseudo-stress</i> del viento (datos promedio de 1974 a 2013). (Fuente de los datos: Servain/FUNCEME).	17
2.4. Inundaciones producidas en el estado de Ceará (región metropolitana de Fortaleza) debidas a las anomalías climáticas del año 2009 (Servain/FUNCEME).	18
2.5. Precipitación acumulada durante 2009 con respecto a la media histórica en el estado de Ceará. (Fuente de los datos: FUNCEME).	19

2.6. Anomalías de la SST ocurridas en el año 2009 en el Atlántico tropical. (Fuente de los datos: Servain/FUNCEME).	19
2.7. Boya ATLAS de superficie y anclaje (Servain/FUNCEME).	22
2.8. Esquema de la boya ATLAS (GTMBA, 2016).	22
2.9. Esquema del mantenimiento y difusión de datos del sistema ATLAS (GTMBA, 2016).	25
3.1. Localización de la región Nordeste de Brasil	29
3.2. Estados que conforman el Nordeste de Brasil.	29
3.3. Precipitaciones en los diferentes estados de la región Nordeste de Brasil (CPTEC/INPE, 2018).	30
3.4. Precipitaciones en los diferentes estados de la región Nordeste de Brasil (continuación).	31
3.5. Metodología para la obtención de los dipolos.	32
3.6. Metodología para la obtención de los dipolos (continuación).	33
3.7. Dipolo obtenido calculando las diferencias en las anomalías de SST entre las boyas 15°N - 38°W y 10°S - 10°W	33
3.8. Ejemplos de series temporales mensuales de anomalías de lluvia.	34
3.9. Comparación entre los dipolos PIRATA y el dipolo de Servain.	40
3.10. Dipolos mensuales (líneas azules) y medias de anomalías de precipitaciones (barras rojas) durante la temporada de lluvias (marzo a mayo) en el Nordeste de Brasil y el estado de Ceará (2006-2016). Se muestra el dipolo “clásico” 15°N / 10°S, el dipolo de Servain y los dipolos PIRATA con mayor número de correlaciones significativas con el NEB y Ceará (8°N / 14°S y 12°N / 14°S).	41
4.1. Perfil de temperatura promedio en función de la profundidad en la boya 12°N - 38°W.	44
4.2. Esquema del procesado realizado para obtener las matrices de datos.	46
4.3. Esquema de un SOM con una estructura rectangular de dos dimensiones.	48
4.4. Actualización de la BMU y sus vecinos hacia la muestra de entrada \mathbf{x} . Las líneas continuas corresponden a la situación anterior a la actualización mientras que las discontinuas corresponden a la situación posterior.	49
4.5. Ejemplo de “Hits map” para la boya 0°N - 0°E donde cada neurona se representa mediante un hexágono.	51

4.6. SOM de las variables meteo-oceanográficas de la boya PIRATA localizada en 0°N – 35°W. El tamaño del mapa es 17 x 7 neuronas.	52
4.7. Mapas de las variables SST, <i>Sea Temp 20m</i> y <i>Sea Temp 40m</i> de la boya PIRATA en 0°N – 35°W.	53
4.8. Temperatura diaria del océano Atlántico a 1 m (SST), 20 m y 40 m en 0°N - 35°W.	54
4.9. Mapas de la SST y temperaturas en subsuperficie de 60 a 120 m para la boya en 0°N – 35°W.	54
4.10. Temperatura diaria del océano Atlántico a 1 m (SST) y a varias profundidades en 0°N - 35°W.	55
4.11. Temperatura diaria de la SST y a 100 m en el océano Atlántico (eje derecho) y serie temporal de la profundidad de la isoterma a 20 °C (eje izquierdo) en 0°N - 35°W.	56
4.12. Perfiles climatológicos (WOA 2013) mostrando los extremos del espesor de la capa de mezcla (a) y la profundidad máxima de salinidad (b) (mínimo: líneas discontinuas; máximo: líneas continuas) para las boyas 0°N – 35°W, 0°N – 23°W, 8°S – 30°W y 8°N – 38°W.	57
4.13. Mapas de las variables SST, temperaturas a 120 m, 300 m y 500 m e isoterma a 20 °C en la boya PIRATA 0°N – 35°W.	58
4.14. Mapas SOM de la SST y salinidad en la boya PIRATA en 0°N – 35°W.	59
4.15. Series temporales diarias de salinidad y temperatura en subsuperficie en el océano Atlántico (profundidad 120 m). Boya PIRATA en 0°N – 35°W.	60
4.16. Mapas SOM de temperatura, densidad, salinidad y profundidad de la isoterma a 20 °C en la boya PIRATA 0°N – 35°W.	61
4.17. Densidad diaria del océano Atlántico (eje derecho) en 0°N – 35°W (profundidades: 1 m y 120 m), y serie temporal de la profundidad de la isoterma a 20 °C (eje izquierdo).	62
4.18. Mapas SOM de la temperatura del aire, SST, velocidad del viento y RH para la boya PIRATA en 0°N – 35°W.	63
4.19. SOM de las variables meteo-oceanográficas de la boya PIRATA en 0°N – 23°W. El tamaño del mapa es 15 x 12 neuronas.	64
4.20. Mapas SOM de la temperatura diaria de SST, a 20 m, 40 m y 60 m en 0°N - 23°W.	65
4.21. Temperatura diaria de la SST y a 40 m, 80 m y 140 m en la boya 0°N - 23°W.	66

4.22. Temperatura diaria de la SST (eje derecho) y serie temporal de la profundidad de la isoterma a 20 °C (eje izquierdo) en 0°N - 23°W . . .	66
4.23. Mapas SOM de la temperatura diaria de SST y profundidad de la isoterma a 20 °C en 0°N - 23°W.	67
4.24. Mapas SOM de temperatura y densidad hasta 40 m para la boya en 0°N - 23°W.	68
4.25. Mapas SOM de densidad y salinidad a 1 m, 20 m, 40 m y 120 m, así como temperatura a diferentes profundidades para la boya 0°N - 23°W.	69
4.26. Mapas SOM de la isoterma a 20 °C, de la temperatura, densidad, salinidad y velocidad del viento para la boya en 0°N - 23°W.	70
4.27. Series temporales diarias de la temperatura del aire y humedad relativa en 0°N - 23°W.	71
4.28. Mapas SOM de la SST, RH, temperatura del aire y velocidad del viento en la boya 0°N - 23°W.	71
4.29. Mapas SOM de las variables meteo-oceanográficas para la boya PIRATA en 8°N - 38°W. El tamaño del mapa es 13 x 9 neuronas.	72
4.30. Mapas SOM de las variables meteo-oceanográficas para la boya PIRATA en 8°S - 30°W. El tamaño del mapa es 11 x 9 neuronas.	72
4.31. Temperatura diaria del océano Atlántico (eje izquierdo) a 1 m (SST), 40 m, 80 m, 140 m, 180 m y serie temporal de la profundidad de la isoterma a 20 °C (eje derecho) en 8°N - 38°W.	73
4.32. Temperatura diaria del océano Atlántico (eje izquierdo) a 1 m (SST), 80 m, 100 m, 140 m y serie temporal de la profundidad de la isoterma a 20 °C (eje derecho) en 8°S - 30°W.	74
4.33. Mapas SOM de las boyas 8°N - 38°W y 8°S - 30°W para las variables (SST) y temperatura en la profundidad 20, 40, y 60 m.	74
4.34. Mapas SOM para las variables (SST) y temperaturas en las profundidades desde 60 a 180 m en la boya 8°N - 38°W.	75
4.35. Mapas SOM para las variables (SST) y temperaturas en las profundidades desde 100 a 180 m en la boya 8°S - 30°W.	76
4.36. Mapas SOM de salinidad para las boyas en 8°N - 38°W y 8°S - 30°W.	77
4.37. Mapas SOM de temperatura, salinidad, densidad y 20° C IsoD en la boya 8°N - 38°W.	78
4.38. Mapas SOM de temperatura, salinidad, densidad y 20° C IsoD en la boya 8°S - 30°W.	79
4.39. Mapas SOM para las variables <i>Wind Speed</i> , 20° C IsoD, SST, <i>Air Temp</i> y RH en la boya 8°N - 38°W.	81

4.40. Mapas SOM para las variables <i>Wind Speed</i> , <i>20° C ISOD</i> , <i>SST</i> , <i>Air Temp</i> y RH en la boya 8°S - 30°W.	82
5.1. Diagrama de la presentación de los resultados obtenidos.	89
5.2. Esquema del procesado de datos.	91
5.3. Representación esquemática de una neurona artificial.	101
5.4. Arquitectura del perceptrón multicapa con dos capas ocultas.	103
5.5. Índices MAE (°C) de entrenamiento, validación y test en función del número de neuronas del MLP utilizando como entradas SST y fecha codificada.	104
5.6. Representación esquemática de una ELM con N datos de entrada y L neuronas en la capa oculta.	106
5.7. Índice MAE (°C) en función del número de neuronas de la ELM para la boya 0°N - 0°E.	108
5.8. Función de base radial (RBF) para varios valores del factor de ensanchamiento.	109
5.9. Índices MAE (°C) en función del <i>spread factor</i> de la GRNN.	110
5.10. Ejemplo de árbol de decisión (boya 6°S - 8°E).	111
5.11. Error medio cuadrático (MSE) en función del número de árboles para la boya 0°N - 0°E.	113
5.12. Esquema de la SVM para regresión. La curva azul representa la curva de regresión. Las curvas rojas marcan los límites del “tubo”, los puntos dentro del tubo tienen unos valores de $\xi_i, \hat{\xi}_i = 0$	114
5.13. Esquema de una celda LSTM y propagación de la información a través de la misma. X_t es el vector de entrada, h_t el vector de salidas y C_t el estado o memoria de la celda. f_t, i_t, C'_t y O_t representan a la puerta olvidar, puerta de entrada, valor estado candidato y puerta de salida, respectivamente.	117
5.14. Autocorrelación de la variable SST en la boya 21°N - 23°W.	120
5.15. ACF y PACF de la SST en la boya 21°N - 23°W.	121
5.16. ACF y autocorrelación de la SST tras diferenciar (boya 21°N - 23°W).	121
5.17. Climatología diaria y serie histórica de la SST para la boya 20°N - 38°W.	122
5.18. Climatología y serie histórica de la SST para la boya 0°N - 10°W.	122
5.19. Climatología diaria y serie histórica de la SST para la boya 19°S - 34°W.	123

5.20. Ejemplos de series temporales diarias deseadas y estimadas de test para la SST en el horizonte $t + 7$ días. Resultados con la SST y fecha codificada como variables de entrada a los modelos.	127
5.21. Ejemplos de series temporales diarias deseadas y estimadas de test para la SST en el horizonte $t + 15$ días. Resultados con la SST y fecha codificada como variables de entrada a los modelos.	128
5.22. Ejemplos de series temporales diarias deseadas y estimadas de test para la SST en el horizonte $t + 30$ días. Resultados con la SST y fecha codificada como variables de entrada a los modelos.	128
5.23. Diagrama de Pareto de la varianza explicada por cada componete de la PCA para la boya $0^{\circ}\text{N} - 0^{\circ}\text{E}$	130
5.24. Gráfica de la importancia de las variables para la predicción según <i>RReliefF</i> (boya $0^{\circ}\text{N} - 0^{\circ}\text{E}$).	132
5.25. Gráfica de la importancia de las variables para la predicción según <i>RReliefF</i> (boya $0^{\circ}\text{N} - 35^{\circ}\text{W}$).	133
5.26. Visualización de la matriz de correlación de los predictores de la boya $0^{\circ}\text{N} - 0^{\circ}\text{E}$; se puede observar que existe un gran número de variables con altos valores de correlación.	135
5.27. Visualización de la matriz de correlación de los predictores de la boya $0^{\circ}\text{N} - 0^{\circ}\text{E}$ tras aplicar el análisis de correlación.	136
5.28. Ejemplos de series temporales mensuales deseadas y estimadas de test para la SST en el horizonte $t + 1$ mes. Resultados con la SST y fecha codificada como variables de entrada a los modelos.	140
5.29. Ejemplos de series temporales mensuales deseadas y estimadas de test para la SST en el horizonte $t + 2$ meses. Resultados con la SST y fecha codificada como variables de entrada a los modelos.	140
5.30. Ejemplos de series temporales deseadas y estimadas de test para la SST en el horizonte $t + 3$ meses. Resultados con la SST y fecha codificada como variables de entrada a los modelos.	141
6.1. Aplicación web de la <i>Global Tropical Moored Buoy Array Project Office</i> para la descarga y visualización de datos de los proyectos TAO/TRITON, PIRATA y RAMA (GT MBA, 2016).	152
6.2. Ejemplo de gráfica de las series mensuales de SST y climatología generada con la aplicación de la GT MBA para la boya $0^{\circ}\text{N} - 0^{\circ}\text{E}$	152
6.3. Vista principal de la aplicación.	156
6.4. Visualización de datos mensuales.	157
6.5. Visualización del modo dipolo.	158

6.6. Ejemplo de datos (“modo dipolo”) exportados a una hoja de cálculo. . .	159
6.7. Menú de selección de variables del viento.	159
6.8. Menú de selección de profundidades.	160
6.9. Esquema de funcionamiento y módulos principales de la aplicación. . .	161
6.10. Serie temporal diaria de la profundidad de la isoterma a 20 °C en la boya 0°N - 35°W.	166
6.11. Serie temporal diaria de la profundidad de la isoterma a 20 °C en la boya 0°N - 0°E.	166
6.12. Serie temporal diaria de la temperatura a 100 m de profundidad en la boya 0°N - 35°W.	167
6.13. Serie temporal diaria de la SST en la boya 0°N - 35°W.	167
6.14. Serie temporal diaria de de la temperatura a 40 m de profundidad en la boya 0°N - 0°E.	168
6.15. Serie temporal diaria de la SST en la boya 0°N - 0°E.	168
6.16. Series temporales de los dipolos PIRATA obtenidos mediante la aplica- ción desarrollada para analizar los eventos meteorológicos de los años 2009 y 2010.	170
6.17. Precipitación acumulada durante 2010 con respecto a la media histórica en el estado de Ceará.	171
6.18. Serie temporal de la SST, climatología y predicciones para la boya 15°N - 38°W.	173
6.19. Serie temporal de la SST, climatología y predicciones para la boya 10°S - 10°W.	173
6.20. Dipolo formado por las boyas 15°N - 38°W y 10°S - 10°W.	173
6.21. Serie temporal de la SST, climatología y predicciones para la boya 12°N - 38°W.	174
6.22. Serie temporal de la SST, climatología y predicciones para la boya 14°S - 32°W.	174
6.23. Dipolo formado por las boyas 12°N - 38°W y 14°S - 32°W.	174

Índice de tablas

2.1. Alturas estándar de los sensores meteorológicos de superficie ATLAS.	20
2.2. Profundidades estándar de los sensores de subsuperficie ATLAS.	21
2.3. Características de los sensores incorporados en las boyas PIRATA.	21
2.4. Definición de las variables registradas en las boyas PIRATA.	24
3.1. Resultados obtenidos con <i>Pearson</i> . Los valores indican el número de correlaciones significativas en cada caso y el total por dipolo, número de datos utilizado, número de correlaciones positivas, negativas y correlación media. Los casos con mayor número de correlaciones significativas se indican en negrita.	36
3.2. Resultados obtenidos con <i>Spearman</i> . Los valores indican el número de correlaciones significativas en cada caso y el total por dipolo, número de datos utilizado, número de correlaciones positivas, negativas y correlación media. Los casos con mayor número de correlaciones significativas se indican en negrita.	37
3.3. Resultados obtenidos con <i>Kendall</i> . Los valores indican el número de correlaciones significativas en cada caso y el total por dipolo, número de datos utilizado, número de correlaciones positivas, negativas y correlación media. Los casos con mayor número de correlaciones significativas se indican en negrita.	38
3.4. Resultados de las correlaciones mensuales entre los valores de temperatura del dipolo 12°N / 14°S (filas) y las anomalías de lluvia en el estado de Ceará (columnas) utilizando el coeficiente de <i>Kendall</i> . Las correlaciones estadísticamente significativas se muestran en negrita.	39
3.5. Resultados de las correlaciones mensuales entre los valores de temperatura del dipolo 8°N / 14°S (filas) y las anomalías de lluvia en el estado de Piauí (columnas) utilizando el coeficiente de <i>Kendall</i> . Las correlaciones estadísticamente significativas se muestran en negrita.	39

4.1. Descripción de las variables presentes en los conjuntos de datos.	47
4.2. Estadísticos básicos ($^{\circ}\text{C}$) y correlación de <i>Pearson</i> entre la SST diaria y la temperatura en subsuperficie a diferentes profundidades para la boya PIRATA $0^{\circ}\text{N} - 35^{\circ}\text{W}$; DE: Desviación Estándar.	53
4.3. Estadísticos básicos de la salinidad diaria a diferentes profundidades para la boya PIRATA en $0^{\circ}\text{N} - 35^{\circ}\text{W}$; DE: Desviación Estándar.	58
4.4. Estadísticos básicos: densidad diaria (kg/m^3) a diferentes profundidades para la boya PIRATA en $0^{\circ}\text{N} - 35^{\circ}\text{W}$; DE: Desviación Estándar.	60
4.5. Estadísticos básicos ($^{\circ}\text{C}$) y correlación <i>Pearson</i> entre la SST diaria y temperatura en subsuperficie a diferentes profundidades para la boya PIRATA en $0^{\circ}\text{N} - 23^{\circ}\text{W}$; DE: Desviación Estándar.	64
4.6. Estadísticos básicos, Salinidad diaria / Densidad (kg/m^3) a diferentes profundidades para la boya PIRATA en $0^{\circ}\text{N} - 23^{\circ}\text{W}$; DE: Desviación Estándar.	68
4.7. Estadísticos básicos ($^{\circ}\text{C}$) y correlación de <i>Perason</i> entre la SST y la temperatura en subsuperficie diaria a diferentes profundidades para las boyas en $8^{\circ}\text{N} - 38^{\circ}\text{W}$ y $8^{\circ}\text{S} - 30^{\circ}\text{W}$; DE: Desviación Estándar.	75
4.8. Estadísticos básicos, Salinidad / Densidad (kg/m^3) diaria a diferentes profundidades para la boya PIRATA en $8^{\circ}\text{N} - 38^{\circ}\text{W}$ y $8^{\circ}\text{S} - 30^{\circ}\text{W}$; DE: Desviación Estándar.	77
5.1. Número de patrones que disponen de la SST, patrones con todas las variables y fechas de los datos de cada boya.	91
5.2. Variables disponibles en cada boya. W, <i>Wind</i> ; Sw, <i>Short wave</i> ; RH, <i>Relative Humidity</i> ; DYN, <i>Dynamic Height</i> ; 20°C IsoD, <i>20^{\circ}\text{C isotherm depth</i> ; S, <i>Salinity</i> ; D, <i>Density</i> ; SST, <i>Sea Surface Temperature</i> , ST, <i>Sea Temperature</i> . Las X indican datos ausentes.	92
5.3. Mejor arquitectura del MLP para cada boya utilizando como entradas la SST y fecha, así como utilizando todas las variables de entrada. Los resultados muestran el número de neuronas en una capa.	104
5.4. Mejor arquitectura (número de neuronas) de la ELM en cada boya para datos diarios y mensuales utilizando como entradas SST y fecha codificada.	108
5.5. Factor de ensanchamiento de la GRNN con menor error de validación para SST y fecha como variables de entrada.	110
5.6. Valor mediano de los hiperparámetros (γ y σ^2) de la LS-SVM para las predicciones mensuales.	115

5.7. Número óptimo de unidades LSTM para cada boya en los conjuntos mensuales y diarios de SST entrenando la red en modo directo y modo recurrente.	118
5.8. Índices MAE (°C) obtenidos en datos diarios de test en las pruebas preliminares con la ELM (mediana de las 17 boyas).	123
5.9. Modelo climatología y modelo trivial diario en el conjunto de test. . .	124
5.10. Índices MAE (°C) de test obtenidos en las predicciones $t+7$ para cada boya utilizando la fecha codificada y la SST como variables de entrada. Los mejores resultados aparecen resaltados en negrita.	125
5.11. Índices MAE (°C) de test obtenidos en las predicciones $t+15$ para cada boya utilizando la fecha codificada y la SST como variables de entrada. Los mejores resultados aparecen resaltados en negrita.	126
5.12. Índices MAE (°C) de test obtenidos en las predicciones $t+30$ para cada boya utilizando la fecha codificada y la SST como variables de entrada. Los mejores resultados aparecen resaltados en negrita.	127
5.13. Número de variables en cada boya.	129
5.14. Índices MAE (°C) de test (mediana de las 17 boyas) obtenidos en los tres horizontes de predicción utilizando todas las variables de entrada. Los mejores resultados aparecen resaltados en negrita.	129
5.15. Coeficientes de las 9 primeras componentes principales de la PCA para la boya 0°N - 0°E. A mayor valor del peso (en valor absoluto), mayor es la aportación del predictor en la determinación del componente. . .	131
5.16. Índices MAE (°C) de test (promedio de las 17 boyas) obtenidos en los tres horizontes de predicción aplicando PCA a las variables de entrada. Los mejores resultados aparecen resaltados en negrita.	131
5.17. Variables ordenadas por nivel de importancia según el algoritmo <i>RReliefF</i> , las variables eliminadas aparecen resaltadas en negrita.	133
5.18. Índices MAE (°C) de test (mediana de las 17 boyas) obtenidos en los tres horizontes de predicción aplicando <i>RReliefF</i> a las variables de entrada. Los mejores resultados aparecen resaltados en negrita.	134
5.19. Número de variables de cada boya aplicando análisis de correlación y <i>RReliefF</i> a las entradas.	135
5.20. Variables seleccionadas por el AC y ordenadas según importancia mediante <i>RReliefF</i> , en negrita se muestran las variables eliminadas. . . .	136
5.21. Índices MAE (°C) de test (mediana de las 17 boyas) obtenidos en los tres horizontes de predicción aplicando análisis de correlación y <i>RReliefF</i> a las variables de entrada. Los mejores resultados aparecen resaltados en negrita.	137

5.22. Índices MAE (°C) de test obtenidos en las predicciones mensuales en $t + 1$ para cada boya utilizando la fecha codificada y la SST como variables de entrada. Los mejores resultados aparecen resaltados en negrita. *No se muestra el resultado por disponer solo de nueve datos.	138
5.23. Índices MAE (°C) de test obtenidos en las predicciones mensuales en $t + 2$ para cada boya utilizando la fecha codificada y la SST como variables de entrada. Los mejores resultados aparecen resaltados en negrita. *No se muestra el resultado por disponer solo de nueve datos.	138
5.24. Índices MAE (°C) de test obtenidos en las predicciones mensuales en $t + 3$ para cada boya utilizando la fecha codificada y la SST como variables de entrada. Los mejores resultados aparecen resaltados en negrita. *No se muestra el resultado por disponer solo de nueve datos.	139
5.25. Índices MAE (°C) de test (mediana de las 17 boyas) obtenidos en los datos mensuales utilizando los diferentes tipos de variables de entrada.	142
5.26. Índices MAE (°C) de test obtenidos en las predicciones diarias con la LSTM utilizando la SST y fecha como entradas.	143
5.27. Índices MAE (°C) de test obtenidos en las predicciones mensuales con LSTM utilizando la SST y fecha como entradas.	144
5.28. Índices MAE promedio y desviación estándar según el método de selección de variables para cada modelo en los datos diarios (medias de los tres horizontes y las 17 boyas).	145
5.29. Análisis estadístico de las diferencias entre pruebas con datos diarios. Únicamente se muestran los modelos y pruebas con diferencias significativas. C1: Conjunto 1; C2: Conjunto 2; IC: Intervalo de Confianza; LI: Límite Inferior; LS: Límite Superior.	146
5.30. Análisis estadístico de las diferencias entre modelos para el método SST más fecha codificada en los conjuntos diarios. Únicamente se muestran los resultados con diferencias significativas. IC: Intervalo de Confianza; LI: Límite Inferior; LS: Límite Superior.	146
5.31. Índices MAE promedio y desviación estándar (DE) de cada modelo en los datos mensuales utilizando SST y fecha como entradas (medias de los tres horizontes y las 17 boyas).	147
5.32. Análisis estadístico de las diferencias entre modelos mensuales para el método SST y fecha codificada. Únicamente se muestran los resultados con diferencias significativas. IC: Intervalo de Confianza; LI: Límite Inferior; LS: Límite Superior.	147
5.33. Índices MAE promedio y desviación estándar para los modelos diarios ELM y LSTM (medias de los tres horizontes y las 17 boyas).	148

5.34. Índices MAE promedio y desviación estándar para los modelos mensuales LS-SVM y LSTM (medias de los tres horizontes y las 17 boyas).	148
5.35. Coste computacional en segundos (media y desviación estándar de las 17 boyas) de cada modelo utilizando como entradas la SST y la fecha codificada.	149
6.1. Valores de enero a mayo de 2009 de anomalías de temperatura para las boyas seleccionadas.	169
6.2. Valores de enero a mayo de 2009 de los dipolos estudiados (°C).	170
6.3. Valores de enero a mayo de 2010 de anomalías de temperatura para las boyas seleccionadas.	171
6.4. Valores de enero a mayo de 2010 de los dipolos estudiados (°C).	172
6.5. Valores de las predicciones (en negrita) de la SST (°C), climatología, anomalías y dipolo de las boyas 12°N - 38°W y 14°S - 32°W.	175

Resumen

El clima del océano Atlántico tropical está caracterizado por un gran ciclo estacional, que presenta una significativa variabilidad océano-atmosférica a escalas interanuales y decadales. Existen dos fenómenos principales que explican estas variaciones climáticas, el denominado modo ecuatorial y el modo meridional o dipolo. El primero se caracteriza por fases cálidas, con vientos alisios débiles y temperaturas en la superficie del mar anormalmente altas, así como por fases frías con vientos fuertes y temperaturas del mar inusualmente bajas. El segundo de estos modos se define como la oscilación en el gradiente de temperatura del mar entre el norte y el sur (dipolo), y está relacionado con los desplazamientos de la zona de convergencia intertropical. Ambos modos de variabilidad afectan de manera importante al régimen de precipitaciones de las regiones tropicales. El alto coste socioeconómico que implica estas fluctuaciones, justifica el interés de la comunidad científica en estudiar dichos procesos climáticos. La red de boyas fijas en el Atlántico tropical denominada PIRATA (*Prediction and Research Moored Array in the Tropical Atlantic*), adquiere variables atmosféricas y oceánicas en diferentes localizaciones del Atlántico. La finalidad de esta monitorización del océano es proporcionar un conjunto de datos que permita mejorar el conocimiento sobre los diferentes procesos que afectan a la variabilidad climática y, por tanto, posibiliten la mejora de los modelos de predicción. En esta Tesis Doctoral, con la finalidad de mejorar la comprensión de estos fenómenos climáticos, se plantea el análisis de los datos PIRATA mediante diferentes técnicas estadísticas y algoritmos de aprendizaje automático. En concreto, se propone el análisis de la relación entre el dipolo y las anomalías de lluvia, la identificación de patrones en los conjuntos de datos, así como la obtención de predicciones diarias y mensuales de la temperatura en la superficie del mar. Además, se presenta una herramienta diseñada para descargar y visualizar los datos PIRATA, que permite también la obtención de predicciones de la temperatura en la superficie del mar y la observación del modo dipolo.

Abstract

The climate of the tropical Atlantic ocean is characterized by a large seasonal cycle, with a significant ocean-atmospheric variability at interannual and decadal timescales. There are two main phenomena that explain these climatic variations, the so-called equatorial mode and the meridional or dipole mode. The first is characterized by warm phases, with weak trade winds and unusually high sea surface temperatures, as well as cold phases with strong trade winds and unusually low temperatures. The second of these modes is defined as the gradient oscillation of the sea surface temperature between the north and the south (dipole), and is related to the displacements of the intertropical convergence zone. Both modes of variability significantly affect the precipitation regime of the tropical regions. The high socioeconomic cost that these fluctuations imply, justifies the interest of the scientific community in studying these climatic processes. The Prediction and Research Moored Array in the Tropical Atlantic (PIRATA) acquires atmospheric and oceanic variables in different Atlantic locations. The purpose of this ocean monitoring is to provide a dataset that allows to improve the knowledge on the different processes that affect the climatic variability and, therefore, make possible the improvement of the forecast models. This Doctoral Thesis proposes the application of different statistical techniques and machine learning algorithms in the PIRATA data to improve the understanding of these climatic events. Specifically, it is proposed the analysis of the relationship between the dipole and rainfall anomalies, the identification of patterns in the datasets, as well as obtaining daily and monthly forecasts of the sea surface temperature. In addition, a software tool designed to download and display PIRATA data is presented, which also allows to obtain sea surface temperature forecasts and observe the dipole mode.

Acrónimos

- AC** Análisis de Correlación
- ACF** *AutoCorrelation Function*
- AIC** criterio de información de Akaike
- ANN** *Artificial Neural Network*
- ANOVA** análisis de varianza
- ARIMA** *AutoRegressive Moving Average*
- ARMAX** *AutoRegressive Moving Average Exogenous Inputs*
- ARX** *AutoRegressive Exogenous Inputs*
- ATLAS** *Autonomous Temperature Line Acquisition System*
- ATRH** *Air, Temperature and Relative Humidity sensor*
- BMU** *Best-Matching Unit*
- BP** *Barometric Pressure sensor*
- BP** *Backpropagation Algorithm*
- CLIVAR** *International Climate Variability and Predictability Program*
- MIC** coeficiente de información máxima
- CFS** *Correlation-based Feature Subset Selection*
- CPTEC** *Centro de Previsão de Tempo e Estudos Climáticos*
- CPs** Componentes Principales
- DOE** *U.S. Department of Energy*

- EDD** *PMEL's Engineering Development Division*
- ELM** *Extreme Learning Machine*
- EOF** *Empirical Orthogonal Function*
- ENSO** *El Niño–Southern Oscillation*
- FUNCEME** *Fundación Cearense de Meteorología y Recursos Hídricos*
- FTP** *File Transfer Protocol*
- GHSOM** *Growing Hierarchical Self-Organizing Map*
- GRNN** *Generalized Regression Neural Networks*
- GTMBAA** *Global Tropical Moored Buoy Array Project Office*
- GTS** *Global Telecommunications System*
- INPE** *Instituto Nacional de Pesquisas Espaciais*
- IRI** *International Research Institute for Climate and Society*
- ITCZ** *Intertropical Convergence Zone*
- LM** *Levenberg-Marquardt*
- LS-SVM** *Least Squares - Support Vector Machines*
- LSTM** *Long Short-Term Memory*
- LWR** *Long Wave Radiation sensor*
- MAE** *Mean Absolute Error*
- MIMO** *Multi-Input Multi-Output*
- MLP** *Multilayer Perceptron*
- MLR** *Multiple Linear Regression*
- MRF** *Weather prediction Medium Range Forecast*
- MSW** *Maximum Salinity Water*
- NaN** *Not a Number*
- NASA** *National Aeronautics and Space Administration*

NCEP *National Center for Environmental Prediction*

NEB *región Nordeste de Brasil*

NSF *National Science Foundation*

NMME *North American Multi-Model Ensemble*

NOAA *US National Oceanic and Atmospheric Administration*

OOPC *Ocean Observations Panel for Climate*

PACF *Partial AutoCorrelation Function*

PCA *Principal Components Analysis*

PIRATA *Pilot Research Moored Array in the Tropical Atlantic*

PIRATA *Prediction and Research Moored Array in the Tropical Atlantic*

PIRATA-SWE *extensión PIRATA Sudoeste*

PIRATA-SEE *extensión PIRATA Sudeste*

PIRATA-NE *extensión PIRATA Nordeste*

PSO *Particle Swarm Optimization*

QP *Convex Quadratic Programming*

RBFNN *Radial Basis Function Network*

RFBT *Random Forest Bagging Trees*

RNN *Recurrent Neuroal Network*

SFS *Sequential Forward Selection*

SLFN *Single Layer Feedforward Network*

SOM *Self-Organizing Map*

SLR *Sparse Logistic Regression*

SSC *Sea Surface Conductivity*

SST *Sea Surface Temperature*

SWR *Short Wave Radiation sensor*

TAO *Tropical Atmosphere-Ocean*

T-FLEX *Tropical Flex*

TOGA *Tropical Ocean Global Atmosphere*

VOScim *Voluntary Observing Ship Climate*

WOA *World Ocean Atlas*

Capítulo 1

Introducción

El océano Atlántico, cuyo nombre deriva del titán Atlas, se encuentra al este de las Américas del Norte y del Sur y al oeste de Europa y África. Por el norte, el Atlántico se conecta al océano Ártico y, por el sur, al océano Antártico. En el Atlántico Norte, donde las aguas calientes de la corriente del Golfo se hunden hacia el fondo del océano tras ser enfriadas por las temperaturas árticas, tiene lugar el inicio del transporte oceánico global (*global ocean conveyor*), un patrón de circulación que ayuda a regular el clima de la Tierra. Según Castello y Krug (2015), el océano Atlántico comprende un área total de unos 74 millones de Km^2 y una profundidad media de 3.600 m, encontrándose las mayores profundidades junto a las fosas de Puerto Rico (9.220 m) y cerca de las islas Sandwich del Sur (8.264 m).

1.1. Variabilidad climática en el océano Atlántico tropical

La climatología del océano Atlántico tropical se caracteriza por un gran ciclo estacional en torno al cual se producen importantes variaciones interanuales y multidecadales. Estas variaciones océano-atmosféricas tienen una gran influencia en el clima regional, afectando a las economías de los países tropicales de América y África. Debido a este hecho, resulta de especial interés estudiar y predecir los diferentes mecanismos que afectan al ciclo estacional (Servain et al., 1998). Superpuestos a este ciclo estacional, existen dos modos principales de variabilidad climática en el océano Atlántico tropical:

- El primero de ellos, denominado modo ecuatorial, es similar al fenómeno El Niño-Oscilación del Sur (*El Niño-Southern Oscillation*, ENSO), que se produce en el océano Pacífico tropical. Sin embargo este fenómeno es mucho más fuerte

en el Pacífico que en el Atlántico. Durante la fase cálida del modo ecuatorial, los vientos alisios del Atlántico tropical occidental son débiles y la temperatura en la superficie del mar (*Sea Surface Temperature*, SST) en la zona ecuatorial es anormalmente alta, especialmente en el golfo de Guinea. Durante la fase fría, los vientos alisios son fuertes y la SST cerca del ecuador terrestre es anormalmente baja (Servain et al., 1998). Las fases calientes, o frías, pueden darse en periodos cortos de tiempo (semanas o meses) y suelen manifestarse cada dos o cuatro años afectando a la temperatura oceánica en superficie y en subsuperficie. Por ejemplo, durante el evento cálido de 1984 se produjeron anomalías positivas de SST mensuales de entre 3° y 4 °C cerca de las costas africanas. Estos eventos tienen un alto impacto económico y social ya que afectan a la pesca y producen variaciones en la distribución, e intensidad, de las lluvias. Otro ejemplo de los efectos de un evento cálido tuvo lugar en el año 1968, produciéndose fuertes lluvias que provocaron inundaciones en varios países del golfo de Guinea (Servain et al., 1999, 2000, 2003).

- El segundo de los modos principales de variabilidad climática es el denominado modo meridional, o dipolo del Atlántico, y no tiene equivalente en el Pacífico. Este modo se caracteriza por una oscilación del gradiente interhemisférico nortesur de la SST que está relacionada con las fluctuaciones en la posición media de la zona de convergencia intertropical (*Intertropical Convergence Zone*, ITCZ) y se puede observar en diferentes escalas de tiempo (estacional, interanual y decadal) (Moura y Shukla, 1981; Servain, 1991). Este modo de variabilidad está fuertemente relacionado con el régimen de lluvias en el Nordeste de Brasil (NEB), en la región Oeste de África y en el golfo de Guinea. Los desplazamientos anómalos de la ITCZ hacia el norte están relacionados con los periodos de sequías en el NEB, y suelen corresponderse con anomalías cálidas de SST al norte (entre 0° - 20°N) y anomalías frías de SST al sur (entre 0° - 20°S). De manera contraria, cuando los desplazamientos de la ITCZ son hacia el sur, se producen periodos de incremento de lluvias así como anomalías de SST frías en el norte y cálidas en el sur (Servain et al., 2000).

1.2. Proyecto PIRATA

Con la finalidad de mejorar el conocimiento y predicción de los dos modos principales de la variabilidad océano-atmosférica que se producen en las regiones tropicales del Atlántico, se desarrolló la red de investigación piloto de boyas fijas en el Atlántico tropical (*Pilot Research Moored Array in the Tropical Atlantic*, PIRATA). Los objetivos del proyecto PIRATA son (Servain et al., 1998; Bourlès et al., 2008):

- Proporcionar una mejor descripción de la variabilidad estacional e interanual que se produce en las capas límite entre la atmósfera y el océano en el Atlántico tropical.

- Mejorar el conocimiento de las contribuciones relativas de los diferentes componentes de los flujos aire-mar y de la dinámica del océano en la variabilidad estacional, e interanual, de la SST.
- Proporcionar un conjunto de datos que permita desarrollar y mejorar los modelos de predicción del sistema climático del Atlántico.
- Diseñar, implementar y mantener una red de boyas fijas que permita adquirir y transmitir datos atmosféricos y oceánicos vía satélite en tiempo real.

La red de boyas oceánicas PIRATA es similar a la red TAO (*Tropical Atmosphere-Ocean*) implantada en el proyecto TOGA (*Tropical Ocean Global Atmosphere*), que es utilizada para estudiar la variabilidad del fenómeno ENSO en el Pacífico ecuatorial. Al igual que en el proyecto TAO, las boyas desplegadas en el proyecto PIRATA son las denominadas ATLAS (*Autonomous Temperature Line Acquisition System*) (Hayes et al., 1991; McPhaden et al., 1998). Las características de las boyas ATLAS (construcción, tipos de sensores, protocolos de datos, etc.) se describirán en detalle en el capítulo 2 de la tesis.

La fase piloto del proyecto multinacional PIRATA se inició en 1997 con la participación de Brasil, Francia y Estados Unidos. Finalizada esta primera fase, que tuvo una duración de tres años, quedó implementada una red formada por 10 boyas. Cuatro de estas boyas están situadas sobre el ecuador (desde 35°W a 0°E), y el resto abarcan dos líneas meridionales (cuatro en 38°W desde 4°N a 15°N y dos en la línea 10°W, en 6°N y en 10°S) (Bourlès et al., 2008). La posición elegida para las boyas en la red principal permite observar los dos modos de variabilidad climática del Atlántico tropical (figura 1.1): las boyas situadas sobre la zona occidental del ecuador permiten estudiar las regiones donde la fuerza del viento es fuerte, mientras que las boyas situadas en la zona central y oriental permite estudiar la variabilidad de la SST. Por otra parte, las boyas situadas en las dos líneas meridionales cubren las regiones en las que se produce una alta variabilidad de la SST asociada al modo dipolo (Servain et al., 1998).

En 2001, los países involucrados en el proyecto decidieron extender en cinco años el mantenimiento de la red, entrando el proyecto en la fase de consolidación. Tras esta fase, las organizaciones CLIVAR (*International Climate Variability and Predictability Program*) y OOPC (*Ocean Observations Panel for Climate*) realizaron una evaluación del proyecto. El resultado de la evaluación respaldó la continuidad del proyecto PIRATA como parte del Sistema Mundial de Observación de los Océanos y del Sistema Mundial de Observación del Clima (Bourlès et al., 2008). Puesto que el proyecto superó la fase piloto, se actualizó el significado del acrónimo PIRATA a red de predicción e investigación de boyas fijas en el Atlántico tropical (*Prediction and Research Moored Array in the Tropical Atlantic*). Desde 2006 el proyecto se encuentra en la fase sostenida, y se ha ampliado la red principal con diversas extensiones, aumentando el número de boyas así como mejorando la capacidad de medición de las mismas (Bourlès et al., 2008).

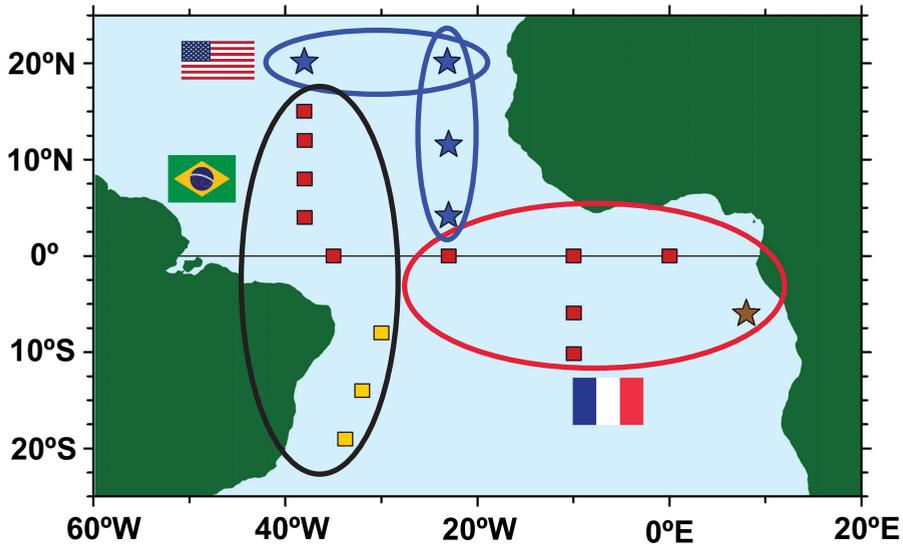


Figura 1.1

Red de boyas principal (cuadrados rojos) y expansiones Sudoeste (cuadrados amarillos), Sudeste (estrella marrón) y Nordeste (estrellas azules) del proyecto PIRATA en el océano Atlántico tropical. Estados Unidos, Brasil y Francia realizan el mantenimiento de las boyas (NOAA / PMEL).

Las tres extensiones realizadas a la red principal se describen a continuación:

- **Extensión Sudoeste (PIRATA-SWE).** En agosto de 2005 se añadieron tres boyas ATLAS en las posiciones $8^{\circ}\text{S} - 30^{\circ}\text{W}$, $14^{\circ}\text{S} - 32^{\circ}\text{W}$ y $19^{\circ}\text{S} - 34^{\circ}\text{W}$. Esta extensión fue propuesta por Brasil con la finalidad de estudiar la interacción entre la SST del Atlántico Sudoeste y la zona de convergencia del Atlántico Sur, investigar la formación de la ITCZ meridional, así como monitorizar la bifurcación de la corriente Ecuatorial del Sur que se produce en la región del sistema de corrientes fronterizas occidentales. Este sistema está formado por la corriente de Brasil, la corriente Norte de Brasil y la subcorriente Ecuatorial. Los dos primeros fenómenos tienen un efecto directo en la variabilidad climática de Brasil, mientras que el tercero además, está relacionado con la climatología global (Nobre, 2006).
- **Extensión Sudeste (PIRATA-SEE).** Sudáfrica, Angola y Namibia propusieron esta extensión, añadiéndose una boya en la posición $6^{\circ}\text{S} - 8^{\circ}\text{E}$ en junio de 2006 (Rouault, 2004). Francia fue la encargada del despliegue y posterior recogida de la boya durante la fase de pruebas (un año). Tras esta fase se decidió mantener la extensión de manera permanente con la finalidad de estudiar

la física del ciclo estacional de la SST, la variabilidad climática y los eventos cálidos que afectan a la región, la interacción entre el medio ambiente y la pesca, así como mejorar las predicciones regionales (Bourlès et al., 2008).

- **Extensión Nordeste (PIRATA-NE).** Estados Unidos propuso incluir cuatro boyas ATLAS en las coordenadas $4^{\circ}\text{N} - 23^{\circ}\text{W}$, $11,5^{\circ}\text{N} - 23^{\circ}\text{W}$, $20^{\circ}\text{N} - 23^{\circ}\text{W}$ y $20^{\circ}\text{N} - 38^{\circ}\text{W}$, realizándose el despliegue de estas boyas entre 2006 y 2007. Estas localizaciones se escogieron para poder estudiar los diferentes procesos océano-atmosféricos que afectan a la posición de la ITCZ, los vientos del Monzón Africano y las anomalías de SST en la región del Atlántico Norte tropical. Una mayor comprensión de estos procesos, que se producen en diferentes escalas temporales, puede ayudar a mejorar las predicciones de precipitación y formación de ciclones tropicales (Rick et al., 2006).

1.3. Aprendizaje automático

El aprendizaje automático es una rama de la inteligencia artificial interesada en reproducir las capacidades de aprendizaje de los seres humanos mediante algoritmos informáticos. Para que el sistema sea considerado inteligente, debe aprender y adaptarse a los cambios que se producen en el entorno. Si el sistema es capaz de adaptarse a nuevas situaciones de manera automática, no será necesario programar el algoritmo con todas las posibles soluciones. El proceso de aprendizaje consiste en optimizar los parámetros del modelo minimizando un criterio de error, utilizando para ello una serie de ejemplos o datos de entrenamiento (Alpaydin, 2010). El aprendizaje automático se aplica en campos como la robótica, reconocimiento de patrones, aplicaciones de internet y aplicaciones médicas (W. Hua y Lijuan, 2009). Debido a que el aprendizaje automático se utiliza en distintas disciplinas, existen múltiples sinónimos para denominar a algunos de los componentes que forman el proceso de aprendizaje (Kuhn y Johnson, 2013):

- Se denomina muestra, observación, ejemplo o patrón a cada uno de los casos individuales e independientes que forman un conjunto de datos.
- Al subconjunto de datos que se utiliza para desarrollar los modelos se le denomina conjunto de entrenamiento, mientras que a los subconjuntos empleados para evaluar el rendimiento de los modelos se les denomina conjuntos de test o validación.
- A las variables utilizadas como entradas a los modelos se les denomina predictores, atributos, descriptores, variables de entrada o variables independientes.
- A los resultados de los modelos se les denomina predicciones, objetivos, variables de salida o variables dependientes.

Dependiendo del tipo de problema a resolver, los algoritmos de aprendizaje automático se pueden clasificar en diferentes categorías (Haykin, 2009; Alpaydin, 2010; Brett, 2015; Das et al., 2015):

- **Aprendizaje supervisado.** Dado un número de ejemplos de entrenamiento asociados a unas salidas deseadas, los algoritmos de aprendizaje supervisado infieren una función que relaciona las entradas con las salidas. El aprendizaje se realiza minimizando el error entre las salidas obtenidas y las esperadas. El modelo o función resultante puede ser utilizado posteriormente para obtener predicciones con nuevos datos no conocidos por el modelo. Este tipo de algoritmos se utiliza en la detección de *spam*, procesamiento del lenguaje natural, reconocimiento de escritura, detección de rostros, reconocimiento de voz, visión por computador, etc (Das et al., 2015). Los problemas más comunes que resuelven los modelos de aprendizaje supervisado son:
 - **Clasificación.** En este tipo de modelos, el problema consiste en predecir a que categoría pertenece una observación determinada. La salida a predecir es de tipo categórica y se denomina clase, pudiéndose dividir en dos o más categorías denominadas niveles. Algunos de los algoritmos más utilizados en clasificación son: k-vecinos cercanos, máquinas de vectores soporte (*Support Vector Machines*, SVM), redes neuronales y árboles de decisión (Brett, 2015). En la figura 1.2.a se muestra un problema de clasificación multiclase resuelto con SVM, donde la curva roja representa el clasificador.
 - **Regresión.** Por otra parte, cuando se quiere obtener una predicción numérica (variable continua) se utilizan los modelos de regresión. Estos modelos pueden cuantificar de manera exacta la magnitud e incertidumbre de la relación entre las entradas y las salidas. Además de la regresión lineal y de los modelos autorregresivos, también son muy utilizados los modelos SVM para regresión, redes neuronales y árboles de regresión (Haykin, 2009). En la figura 1.2.b se muestra un problema de regresión simple en el que se puede observar el ajuste realizado por el modelo (curva roja).
- **Aprendizaje no supervisado.** Los modelos descriptivos se utilizan para descubrir estructuras o características ocultas en conjuntos de datos no etiquetados. Es decir, los datos de entrada no tienen asociados unas salidas deseadas. Los métodos no supervisados más utilizados son los algoritmos de agrupamiento (*clustering*) y se utilizan en tareas de clasificación de ADN, análisis de redes sociales, estudios de mercado, bioinformática, medicina, etc. Algunos algoritmos de este tipo son el *k-means* (figura 1.3.a), *fuzzy c-means* y DBSCAN. Además de estos algoritmos, otras técnicas de aprendizaje no supervisado ampliamente utilizadas son el análisis de componentes principales (*Principal Components Analysis*, PCA) y los mapas auto-organizados (*Self-Organizing Map*, SOM). La técnica PCA se emplea principalmente en tareas de reducción de variables, mientras que los SOM se utilizan en problemas de visualización de datos de alta

dimensionalidad (Haykin, 2009; Abdi y Williams, 2010). En la figura 1.3.b se puede ver un ejemplo de un SOM.

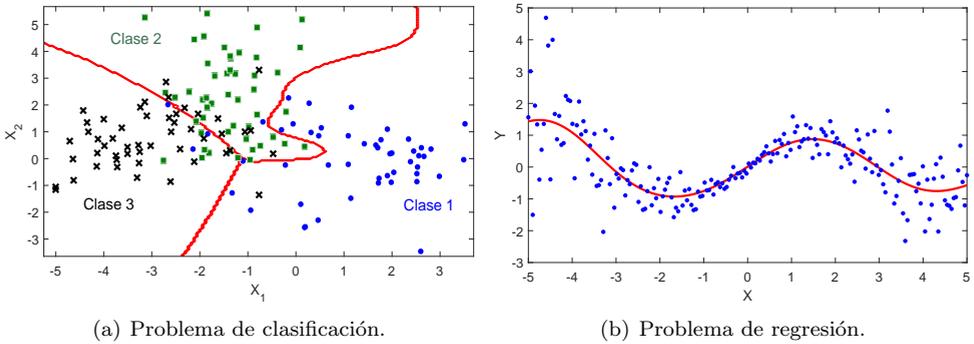


Figura 1.2

Ejemplo de problema de clasificación (a) y de regresión (b) resueltos con máquinas de vectores soporte.

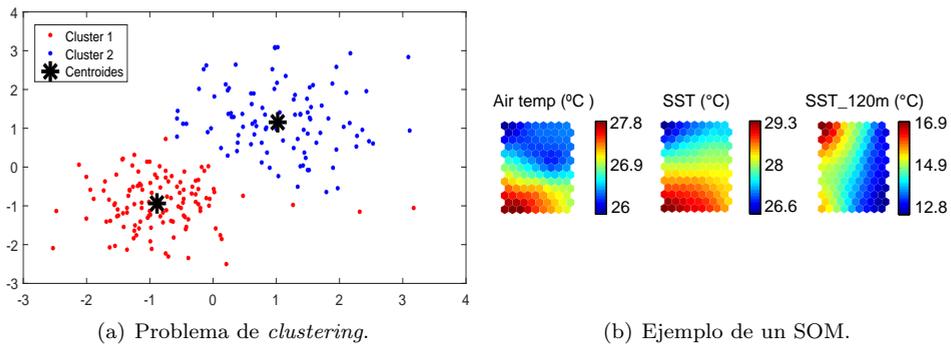


Figura 1.3

Ejemplo de problema de *clustering* con el algoritmo *k-means* y ejemplo de aplicación de un mapa auto-organizado.

- **Aprendizaje por refuerzo.** Los algoritmos de aprendizaje por refuerzo están orientados a la resolución de problemas de decisión secuenciales. El denominado *agente* interactúa con un entorno mediante una señal de estado, una señal de acción y una señal de recompensa (figura 1.4). El agente debe seleccionar las acciones que le permitan maximizar la función recompensa acumulada a largo plazo. Estos algoritmos se utilizan en problemas de predicción del tráfico, videojuegos, análisis de la bolsa de valores, etc. Algunos algoritmos conocidos son el método de Monte Carlo, *Temporal-Difference Learning* y *Q-Learning* (Haykin, 2009).



Figura 1.4

Elementos que componen el problema de aprendizaje por refuerzo.

El proceso de resolución de problemas mediante técnicas de aprendizaje automático puede llevarse a cabo en los cinco pasos representados en la figura 1.5. A continuación se describen cada una de esas etapas (Brett, 2015):

1. **Recogida de datos.** La primera etapa consiste en obtener el conjunto de datos en el formato adecuado para su análisis. Estos datos pueden proceder de una URL (internet), de una máquina local o de una base de datos (SQL, *Oracle*, *Acces*, etc.) en diferentes formatos (*Excel*, CSV, ASCII, etc.).
2. **Preprocesado.** El resultado de los modelos de aprendizaje dependerá en gran medida de la calidad de los datos, por lo que es necesario realizar una etapa de limpieza y procesado de los mismos. En esta etapa se realizan tareas de recodificación de variables, selección de predictores, normalización de los datos, así como procesado de datos ausentes y datos extremos.
3. **Aprendizaje.** En la siguiente etapa se debe seleccionar el algoritmo que mejor se ajuste al tipo de problema a resolver (clasificación, regresión, agrupamiento, etc.). Una vez entrenado el algoritmo escogido, se obtiene una representación de los datos en forma de modelo. Este modelo permitirá tomar decisiones sobre nuevos datos no vistos en el entrenamiento.
4. **Evaluación del modelo.** Una vez entrenado el modelo es necesario evaluar su rendimiento ante nuevos datos (capacidad de aprendizaje). Una de las técnicas más utilizadas para garantizar que el resultado del modelo es independiente de la partición de los datos (entrenamiento/validación), es la denominada validación cruzada (Kuhn y Johnson, 2013). Mediante un índice de error se evalúa la precisión del modelo en diferentes particiones del conjunto de datos. En especial, es importante controlar que no se produzca el denominado sobreentrenamiento del modelo (pérdida de generalización).
5. **Interpretación de resultados.** En la última etapa se obtienen los resultados del problema planteado. En el caso de no obtenerse el rendimiento deseado, se puede optar por mejorar la etapa de preprocesado (interpolación de datos ausentes o mejorar la selección de características), escoger modelos más avanzados, o una

combinación de ambas. Además, los resultados del problema pueden llevar a dos acciones: utilizar el modelo para obtener predicciones en el futuro (nuevos datos), o utilizarlo para hacer simulaciones del comportamiento de un sistema para comprender mejor su funcionamiento.



Figura 1.5
Etapas del aprendizaje automático.

1.4. Organización de la Tesis Doctoral

La presente Tesis Doctoral se ha estructurado en siete capítulos. En el capítulo 1 se realiza una introducción general al problema planteado. En la primera parte del capítulo 2 se describen las principales características de la climatología en el océano Atlántico intertropical. A continuación, en la segunda parte del capítulo se detallan las particularidades de la construcción, instrumentación meteorológica y oceanográfica, así como del control de calidad y difusión de los datos registrados por las boyas ATLAS que componen la red PIRATA.

En el capítulo 3 se detallan las características del denominado modo dipolo del Atlántico y se propone un análisis de correlación entre las precipitaciones en el NEB y los dipolos de SST. En el capítulo 4 se presenta un análisis cualitativo de los datos meteo-oceanográficos del proyecto PIRATA mediante el uso de redes neuronales SOM. El capítulo 5 aborda el problema de las predicciones de la SST comparando una amplia variedad de modelos de predicción, así como utilizando diferentes técnicas de extracción y selección de características. En el capítulo 6 se presenta una herramienta que posibilita la descarga y visualización interactiva de las diferentes variables registradas por las boyas PIRATA, permitiendo además la obtención de predicciones mensuales de SST y el estudio del dipolo del Atlántico. Por último, en el capítulo 7 se exponen las conclusiones que se derivan de los resultados obtenidos en los diferentes experimentos. Además se detallan las diferentes contribuciones realizadas y las posibles líneas de investigación futuras.

A excepción de los capítulos introductorios (capítulos 1 y 2) y del capítulo de conclusiones (capítulo 7), cada uno de los capítulos principales de esta Tesis Doctoral (capítulos 3 a 6) está organizado siguiendo la estructura habitual de un artículo científico (introducción, objetivos, materiales y métodos, resultados y conclusiones) para facilitar su lectura. Esta estructura permite que los capítulos sean autocontenidos y puedan leerse de forma independiente aunque, no obstante, se recomienda

leerlos en el orden presentado. En cada uno de los capítulos se detallan los experimentos realizados para abordar el complejo problema que supone la comprensión de los diversos procesos de interacción océano-atmósfera que tienen lugar en el océano Atlántico. El análisis global de estos capítulos permite avanzar en el conocimiento de dichos procesos, en especial, en los relacionados con la SST.

1.5. Motivación y objetivos de la Tesis Doctoral

Las variaciones en las anomalías de la SST en el Atlántico tienen una influencia directa en las condiciones climáticas, en particular en las zonas costeras bañadas por este océano (Ropelewski y Halpert, 1987; Moura y Hastenrath, 2004). El impacto económico y social que suponen las fluctuaciones climáticas en estas regiones, justifica el interés de la comunidad científica en estudiar los diferentes procesos involucrados en la interacción océano-atmósfera que se produce en el Atlántico. Para este fin es necesario disponer de datos meteo-oceanográficos precisos como los que proporcionan las boyas del proyecto PIRATA.

En consecuencia, la principal motivación de esta Tesis Doctoral es analizar los datos PIRATA para mejorar la comprensión de las interacciones océano-atmosféricas del Atlántico, empleando para este fin diferentes técnicas estadísticas y algoritmos de aprendizaje automático. De manera detallada, los objetivos que se persiguen en esta Tesis Doctoral son:

1. **Obtener la combinación de boyas (dipolo) que presente un mayor potencial para el pronóstico de la temporada de lluvias en el NEB.** Se obtendrán los dipolos de SST a partir de las diferencias de temperatura resultantes de combinar las boyas PIRATA fondeadas en el Atlántico Norte y en el Atlántico Sur. Posteriormente se estudiará la posible relación lineal existente entre estos dipolos y las anomalías de precipitación del NEB mediante un análisis de correlación.
2. **Identificar patrones y relaciones entre las diferentes variables de los datos PIRATA.** Puesto que en los procesos que modulan la variabilidad de la SST intervienen multitud de variables, se emplearán los SOMs para realizar un análisis visual, cualitativo y multivariante de los datos meteo-oceanográficos adquiridos por las boyas PIRATA. Se pretende, por tanto, identificar patrones generales entre las variables y extraer información útil para mejorar el conocimiento sobre los diversos procesos que tienen lugar en el océano Atlántico.
3. **Obtener predicciones de la SST mediante técnicas de aprendizaje automático.** Debido a que la SST tiene un papel fundamental en la comprensión de la variabilidad climática del Atlántico, uno de los objetivos prioritarios de esta Tesis Doctoral es la obtención de predicciones diarias y mensuales de esta variable. Para este fin se plantea el uso de diferentes técnicas de selección

y extracción de características. Asimismo se realizará una amplia revisión del estado del arte en cuanto a modelos de predicción basados en aprendizaje automático. Utilizando los diferentes subconjuntos obtenidos, se compararán las predicciones logradas por los modelos de aprendizaje automático propuestos. La finalidad de esta comparación será descubrir la combinación de variables de entrada y modelo de predicción que permite obtener las mejores predicciones de la SST.

4. **Desarrollar una herramienta para el estudio de los datos PIRATA.** Utilizando las conclusiones más importantes de los capítulos anteriores, se implementará una interfaz gráfica de usuario para el análisis, predicción y visualización de datos PIRATA. Para ello se diseñará una herramienta que permita la descarga de datos, el análisis y visualización interactiva de las diferentes variables, el estudio del modo dipolo, la obtención de predicciones de SST y la exportación de resultados. Dichas predicciones se obtendrán con las variables de entrada y el algoritmo de aprendizaje automático que presenten un mejor rendimiento.

Capítulo 2

Descripción del área de estudio y las boyas PIRATA

2.1. Área de estudio

El área de estudio de esta tesis se centra en el océano Atlántico intertropical, comprendido entre las latitudes 30°N y 20°S y las longitudes 60°W y 16°E (ver figura 1.1). Con la finalidad de caracterizar el comportamiento del clima en esta zona de la Tierra, a continuación se describirán las características de la ITCZ y su influencia en las fluctuaciones climáticas. Además, se detallará la evolución de la SST y del viento en los diferentes periodos del año.

2.1.1. Zona de convergencia intertropical (ITCZ)

En la zona del ecuador terrestre, el aire cálido asciende y fluye hacia los polos, generando zonas de bajas y altas presiones. A medida, por ejemplo, que la masa de aire se aleja del ecuador, ésta se enfría y desciende en las zonas subtropicales (30° - 35° de latitud), generando una zona de altas presiones. Además, el movimiento de rotación de la Tierra desvía la masa de aire hacia occidente (efecto de Coriolis). En el hemisferio norte, el aire descendiente sopla desde el nordeste al sudoeste, volviendo hacia el ecuador. De manera similar, en el hemisferio sur, el aire sopla desde el sudeste hacia el noroeste. A estos vientos que soplan desde las zonas de altas presiones subtropicales hacia la zona de baja presión ecuatorial se les denominan vientos alisios (Ross, 1995).

La región en la que convergen los vientos alisios procedentes del hemisferio norte con los procedentes del hemisferio sur, se le denomina zona de convergencia intertropical. En esta zona, la convergencia de los vientos fuerza al aire a ascender hacia la atmósfera formando la ITCZ, que aparece como un ancho cinturón de nubes cerca

del ecuador que puede extenderse cientos de kilómetros (figura 2.1) (Waliser y Jiang, 2015). Siguiendo la posición del sol, la ITCZ se desplaza hacia el norte durante el verano boreal (junio-agosto) y hacia el sur durante la primavera boreal (marzo-abril). Los periodos de aumento de precipitaciones en el Nordeste de Brasil están asociados con desplazamientos de la ITCZ hacia el sur y, los periodos de sequías, están asociados a desplazamientos de la ITCZ hacia el norte. Por tanto, la variación del desplazamiento anual de la ITCZ interviene directamente en las fluctuaciones climáticas en el Atlántico tropical (Servain et al., 2000; Chang et al., 2006). En las figuras 2.1.a y 2.1.b se puede observar que, durante los meses de enero y marzo, la ITCZ se sitúa sobre el NEB (hemisferio sur). Sin embargo, en las figuras 2.1.c y 2.1.d se muestra que la ITCZ se posiciona por encima del ecuador (hemisferio norte) en el segundo semestre del año.

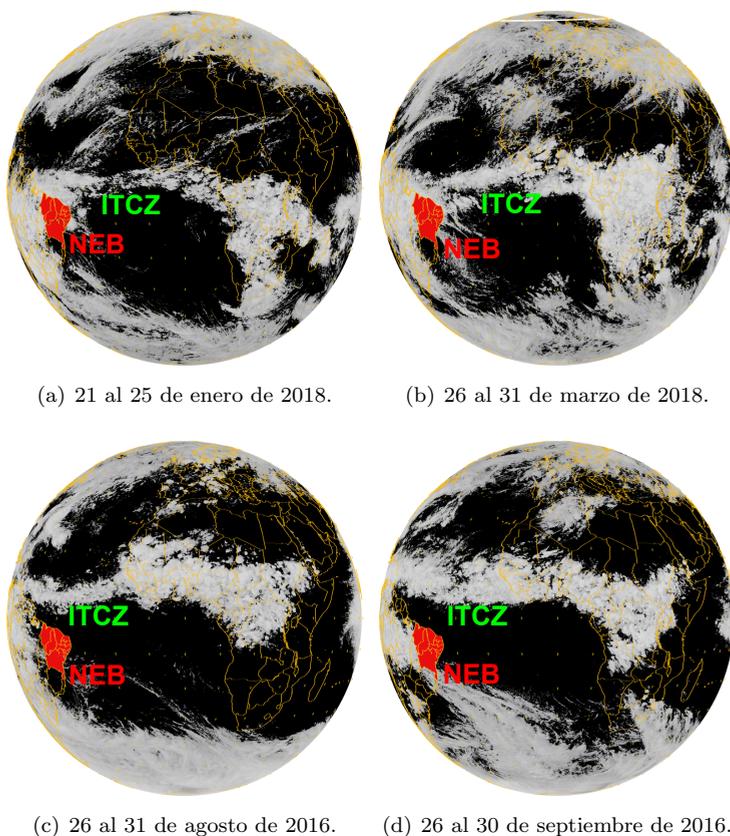


Figura 2.1

Imágenes de la posición de la ITCZ tomadas con los satélites Meteosat-10 (2016) y Meteosat-11 (2018) en el canal 7 IR (adaptada de FUNCEME).

2.1.2. Caracterización climática

Las figuras 2.2 y 2.3 muestran la climatología mensual de la SST y del *pseudo-stress* del viento en el océano Atlántico intertropical. El *pseudo-stress* del viento es el producto entre el vector viento y la velocidad (m^2s^{-2}). Los datos utilizados para generar las gráficas fueron recogidos por navíos mercantes y de investigación desde el año 1974 al 2013, con una resolución espacial de 2° de longitud por 2° de latitud. El dominio espacial varía de $30^\circ N$ a $20^\circ S$ (25 puntos de latitud) y de $60^\circ W$ a $16^\circ E$ (38 puntos de longitud). Como se puede observar en la figura 2.2, en el período de enero a mayo (correspondiente al verano/otoño austral), el océano Atlántico presenta altas temperaturas en superficie. La temperatura se sitúa por encima de los $27^\circ C$ especialmente en enero (entre las latitudes $0^\circ N$ y $5^\circ N$), febrero (entre $5^\circ S$ y $5^\circ N$) y marzo (entre $10^\circ S$ y $5^\circ N$). Respecto al mes de abril (figura 2.2.d), las aguas superficiales presentan temperaturas superiores a $28^\circ C$ entre las latitudes $10^\circ S$ y $5^\circ N$.

Identificando la zona de confluencia entre los vientos alisios del nordeste y del sudeste, en estas figuras también se puede observar la posición que ocupa la ITCZ a lo largo de los meses. La ITCZ se posiciona sobre las aguas más calientes del océano Atlántico y, en los años considerados normales, permanece sobre la región del NEB entre febrero y mayo de cada año, generando precipitaciones durante la estación de lluvias. En marzo y abril, los meses climatológicamente más lluviosos del NEB, la ITCZ se posiciona por debajo de la línea del Ecuador entre, aproximadamente, 2° y $2,5^\circ S$ (figuras 2.2.c y 2.2.d). A partir del mes de mayo (figura 2.2.e), la masa de agua caliente de esta zona comienza a ser “empujada” hacia el norte como consecuencia de la penetración de aguas con temperaturas más bajas. Estas aguas tienen una temperatura aproximada de $21^\circ C$ en el extremo sudeste, área comprendida entre las latitudes 18° y $20^\circ S$ y longitudes $22^\circ W$ y $5^\circ E$.

Este avance gradual de aguas frías del Atlántico Sur se produce hasta el mes de septiembre (figura 2.3.c), cuando las aguas superficiales se encuentran entre 17° y $22^\circ C$ de temperatura. A partir de octubre (figura 2.3.d), las aguas más frías (temperaturas inferiores a $24^\circ C$) comienzan a retroceder dando espacio a la penetración de aguas con temperaturas más elevadas (superiores a $26^\circ C$) que van, nuevamente, envolviendo toda la costa norte/nordeste de Brasil y el Atlántico ecuatorial ($5^\circ S$ - $5^\circ N$). Una característica importante de la distribución de la SST en el litoral del NEB, es que su gradiente térmico es perpendicular a la línea de la costa, disminuyendo desde la zona costera hacia la zona oceánica. Este hecho está directamente relacionado con la propia dirección de la línea de la costa, con el estrangulamiento de la cuenca Atlántica en la región tropical, así como con la hidrodinámica de la región en lo referente a las corrientes y al desplazamiento de las isoterma en las diferentes épocas del año.

Las figuras 2.2 y 2.3 también ilustran las medias mensuales del viento sobre el Atlántico intertropical. En ellas se puede observar que los vientos alisios de nordeste son más intensos que los vientos alisios de sudeste en las estaciones de invierno y primavera boreal para una amplia área centrada alrededor de $10^\circ N$ y $45^\circ W$ (figuras 2.2.a - 2.2.e).

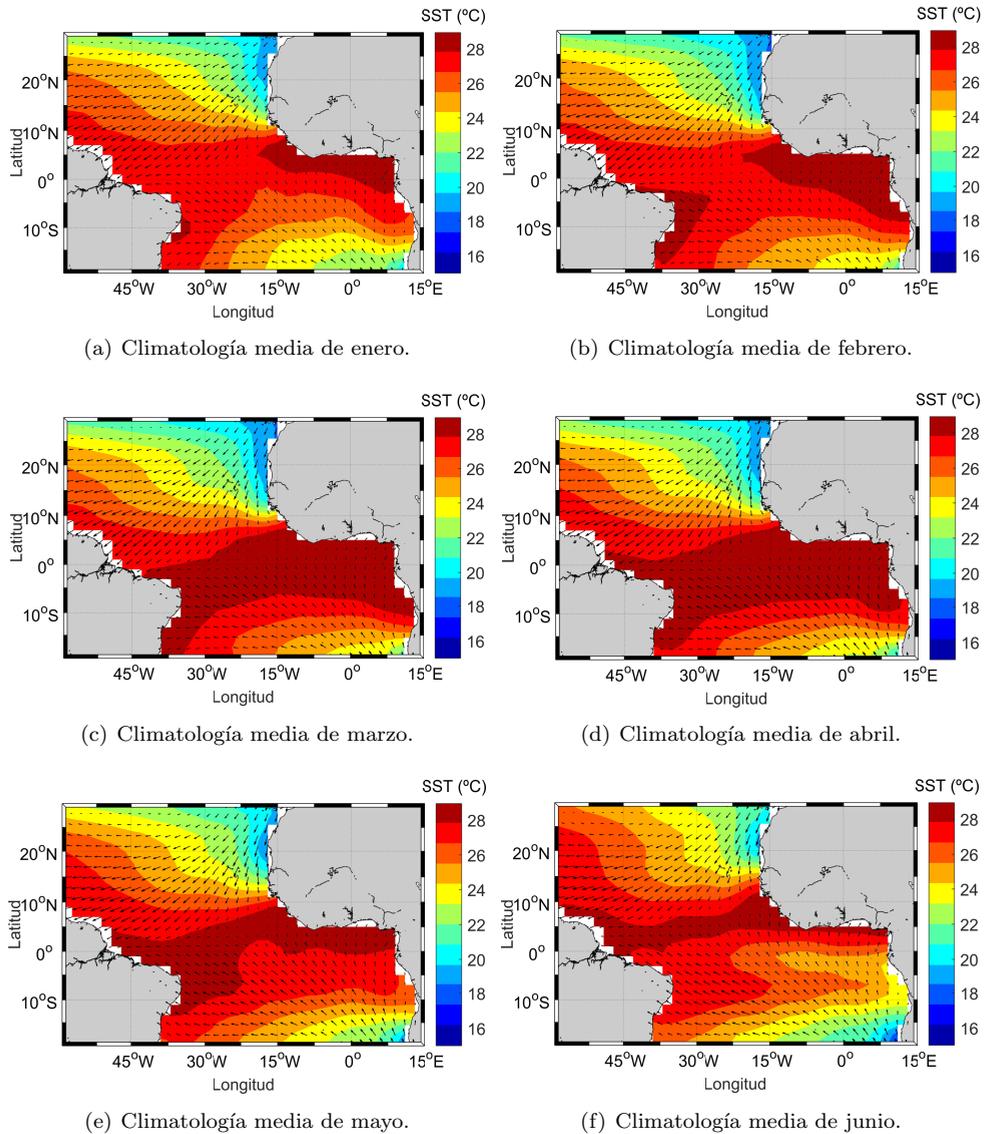
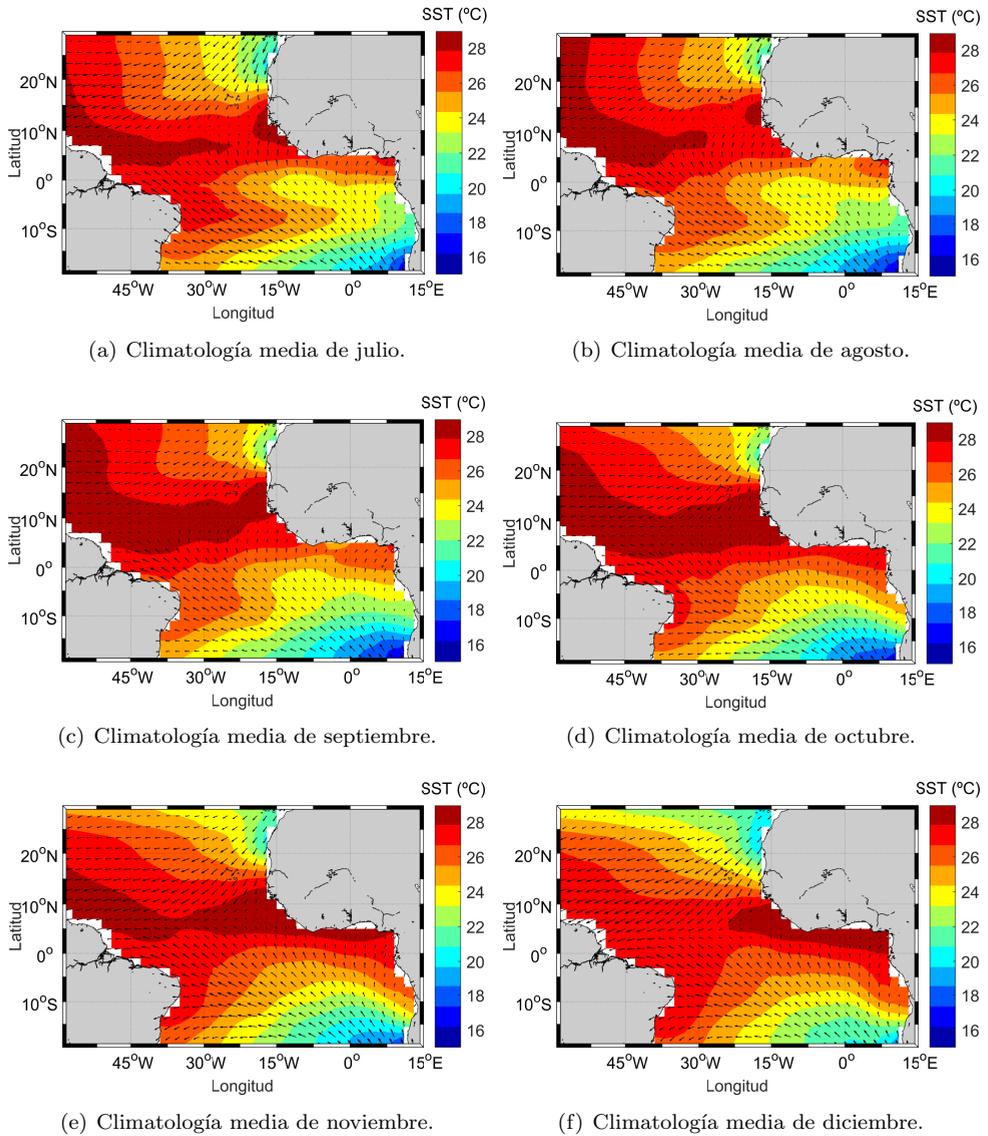


Figura 2.2

Climatologías mensuales (enero-junio) de la temperatura en la superficie del mar y del *pseudo-stress* del viento (datos promedio de 1974 a 2013). (Fuente de los datos: Servain/FUNCEME).

**Figura 2.3**

Climatologías mensuales (julio-diciembre) de la temperatura en la superficie del mar y del *pseudo-stress* del viento (datos promedio de 1974 a 2013). (Fuente de los datos: Servain/FUNCEME).

En los meses de junio a noviembre, los vientos alisios del sudeste son más intensos (área centrada en 10°S y 20°W) que los vientos alisios del nordeste (figuras 2.2.f - 2.3.e). Estos resultados coinciden con los encontrados por Servain y Legler (Servain y Legler, 1986). Además, se puede observar que a partir del mes de mayo, la confluencia de los vientos alisios del nordeste y del sudeste, indicativo de la presencia de la ITCZ siguiendo el movimiento aparente del sol, se desplaza hacia posiciones al norte de la línea del ecuador. En septiembre (figura 2.3.c), la ITCZ alcanza su posición máxima (entre 8° y 10°N), siendo a partir de octubre cuando la ITCZ comienza a migrar de nuevo hacia la línea del ecuador. Durante el invierno boreal y austral, los vientos se hacen más intensos debido al fortalecimiento del Sistema Semi-Permanente de Alta Presión del Atlántico Norte y del Atlántico Sur, respectivamente.

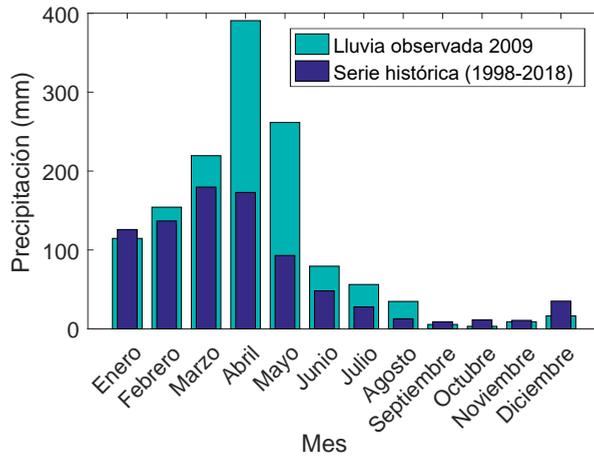
En general, la presencia de SSTs anormalmente altas (bajas) están asociadas con presiones más bajas (altas) que provocan debilitación (fortalecimiento) del Centro Semi-Permanente de Alta Presión climatológico. La intensidad de las altas presiones subtropicales, a su vez, influyen directamente en la dirección y magnitud de los vientos alisios de nordeste y sudeste sobre la superficie del océano (Souza, 1999). Estos vientos alisios están separados por un eje cinemático, que se extiende hacia el nordeste a partir del contorno de la costa de África. La posición más septentrional (entre 8°N y 12°N) y la más meridional (entre 2° y 4°S) del ciclo anual de la ITCZ, ocurre durante el verano de cada hemisferio (Servain y Legler, 1986).

En las figuras 2.4.a y 2.4.b se muestran dos ejemplos del impacto que produjo el dipolo negativo del Atlántico en las precipitaciones del estado de Ceará (región metropolitana de Fortaleza) en el período de marzo a mayo del año 2009. Durante estos meses, la precipitación acumulada en Ceará fue el doble de lo habitual, tal y como demuestra la figura 2.5. Las anomalías de la SST que tuvieron lugar en este período de tiempo se muestran en las figuras 2.6.a, 2.6.b y 2.6.c. Estas figuras evidencian que las temperaturas en el sur fueron anormalmente altas (en rojo), mientras que en el norte fueron anormalmente bajas (en azul), siendo los valores del dipolo de Servain en los meses de marzo, abril y mayo, $-1,51^{\circ}\text{C}$, $-1,41^{\circ}\text{C}$ y $-1,83^{\circ}\text{C}$, respectivamente.

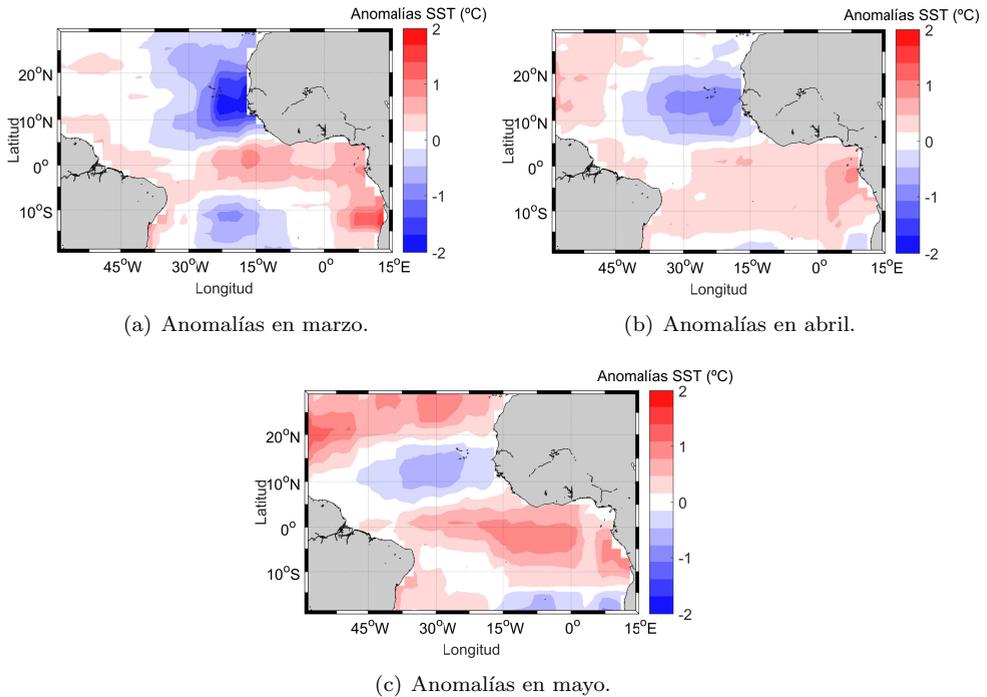


Figura 2.4

Inundaciones producidas en el estado de Ceará (región metropolitana de Fortaleza) debidas a las anomalías climáticas del año 2009 (Servain/FUNCEME).

**Figura 2.5**

Precipitación acumulada durante 2009 con respecto a la media histórica en el estado de Ceará. (Fuente de los datos: FUNCEME).

**Figura 2.6**

Anomalías de la SST ocurridas en el año 2009 en el Atlántico tropical. (Fuente de los datos: Servain/FUNCEME).

2.2. Boyas PIRATA

En el proyecto PIRATA se utilizan las boyas del tipo ATLAS, capaces de registrar variables meteorológicas de la atmósfera (dirección y velocidad del viento, temperatura y humedad relativa del aire, lluvia y radiación solar) y diferentes propiedades del océano en superficie, así como a diferentes profundidades (temperatura, salinidad y presión) (Bourlès et al., 2008). Las boyas PIRATA son tecnológicamente idénticas (*hardware*, tipo de sensores, procedimientos de calibración, resolución y muestreado, procesamiento de datos, protocolos de transmisión de datos, etc.) a las boyas utilizadas en el Pacífico ecuatorial de la red TAO (Hayes et al., 1991; McPhaden et al., 1998). El diseño de las boyas ATLAS fue iniciado por *PMEL's Engineering Development Division* (EDD) en 1984. Posteriormente el diseño inicial fue actualizado para mejorar la calidad de los datos (resolución, precisión y fiabilidad), añadir nuevos sensores, aumentar el número de variables registradas y simplificar los procedimientos de fabricación con el fin de reducir sus costes. A principios de 1996 comenzaron a utilizarse estas boyas ATLAS de nueva generación (Milburn et al., 1996). Siguiendo con el programa de actualizaciones de las boyas, de 2011 a 2015 se llevó a cabo la fase de pruebas del nuevo modelo denominado T-FLEX (*Tropical Flex*), que incorpora nuevas mejoras en la construcción, comunicación y sensores. En 2016 comenzaron los primeros reemplazos de boyas ATLAS por unidades T-FLEX, fase que se extenderá hasta el 2023 (GT MBA, 2016).

2.2.1. ATLAS *Hardware*

Las boyas ATLAS disponen de un anemómetro (viento), un sensor de radiación de onda larga (*Long Wave Radiation sensor*, LWR) y de onda corta (*Short Wave Radiation sensor*, SWR), un pluviómetro (lluvia), un sensor de humedad relativa (*Air, Temperature and Relative Humidity sensor*, ATRH) y un sensor de presión barométrica (*Barometric Pressure sensor*, BP). En la tabla 2.1 se muestran las alturas estándar de estos instrumentos meteorológicos.

Tabla 2.1

Alturas estándar de los sensores meteorológicos de superficie ATLAS.

Sensor Meteorológico	Viento	LWR	SWR	Lluvia	ATRH	BP
Altura (m)	4	3,5	3,5	3,5	3	3

Además de estos instrumentos, dispone de sensores que miden la temperatura y la conductividad en la superficie del mar (*Sea Surface Temperature and Conductivity sensor*, SST y SSC), así como sensores para la medición de temperatura (T1, T2, ..., T10), conductividad (C1, C2 y C6) y presión (P9 y P10) a diferentes profundidades de la cuenca oceánica. La tabla 2.2 indica las profundidades a las que se encuentran estos sensores, mientras que en la tabla 2.3 se muestran sus características más importantes (modelo, resolución, rango y precisión) (Servain et al., 1998; GT MBA, 2016).

Tabla 2.2

Profundidades estándar de los sensores de subsuperficie ATLAS.

Sensor	SST SSC	T1 C1	T2 C2	T3	T4	T5	T6 C6	T7	T8	T9 P9	T10 P10
Profundidad (m)	1	20	40	60	80	100	120	140	180	300	500

Tabla 2.3

Características de los sensores incorporados en las boyas PIRATA.

Medición	Tipo sensor	Fabricante: Modelo	Resolución	Rango	Precisión
Velocidad viento	Hélice	R. M. Young: 05103	0,2 m s ⁻¹	1-20 m s ⁻¹ (0,4 - 36 m s ⁻¹)	± 0,3 m s ⁻¹ o 3 %
Dirección viento	Veleta	R. M. Young: 05103	1,4°	0-355°	4°- 7,8°
	Brújula de flujo	E.G. and G. 63764 o KVH LP101-5	1,4°	0-359°	
Temperatura Aire	Pt-100 RTD	Rotronic Instrument	0,01°C	14-32°C (0-40°C)	±0,2°C
Humedad Relativa	Capacitancia	Corp.: MP-101	0,4 % RH tiempo real 0,02 % RH modo retardo	55-95 % RH (0-100 % RH)	±2,7 % RH
Lluvia	Capacitancia	R. M. Young: 50203-34	0,2 mm hr ⁻¹	0-50 mm	±0,4 mm hr ⁻¹
Radiación onda corta	Piranómetro	Eppley Laboratory: PSP-TAO, Delrin case	0,4 W m ⁻²	200-1000 W m ⁻² (0-1600 W ⁻²)	±2 %
Radiación onda larga	Pirgeómetro	Eppley Laboratory: PIR-TAO, Delrin case	0,1 W m ⁻² 0,03°C	200 W m ⁻² (sólo termopila)	±1 %
Presión Barométrica	Transductor de presión	Paroscientific: MET1-2	0,1 hPa	800-1100 hPa	±0,01 % de lectura
Temp. mar superficie y subsuperficie	Termistor	Sea Bird Electronics: SBE16, SBE37	0,001°C	1-31°C (-5-35°C)	±0,003 °C
Salinidad	Célula de conductividad de campo interno	Sea Bird Electronics: SBE16 (Seacat)	0,0001 S m ⁻¹	3-6 S m ⁻¹	±0,02
		Sea Bird cell with PMEL electronics	0,002 S m ⁻¹		
Presión	Transductor	Paine: 211-30-660-01	0,03 psi	400-800 psi (0-1000 psi)	±1,4psi

En cuanto a su construcción física, la boya de superficie ATLAS (figura 2.7) está formada por un toroide de fibra de vidrio y espuma de 2,3 m de diámetro así como por una torre de aluminio en la que se fijan los instrumentos meteorológicos. El sistema completo tiene un peso aproximado de 660 kg y una altura de 4,9 m. Se utiliza un cable reforzado de 0,92 cm de diámetro en los primeros 700 metros de profundidad para protegerlo de posibles daños. Los sensores de subsuperficie están fijados a este cable. El resto de la línea hasta llegar al fondo marino está compuesta por una cuerda de nylon trenzada de 1,9 cm de diámetro. Finalmente, se fija la boya al fondo marino mediante un anclaje formado por ruedas de ferrocarril con un peso de entre 1500 y 2000 kg. En la figura 2.8 se puede observar un esquema completo de la boya ATLAS (Milburn et al., 1996).

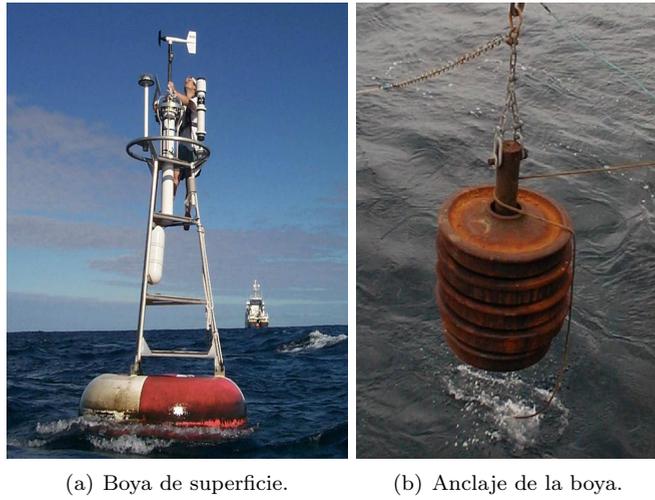


Figura 2.7
Boya ATLAS de superficie y anclaje (Servain/FUNCEME).

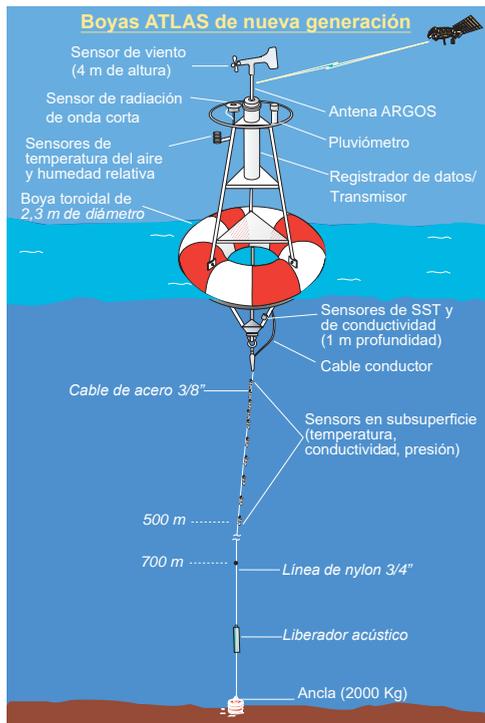


Figura 2.8
Esquema de la boya ATLAS (GT MBA, 2016).

2.2.2. Variables, frecuencia de muestreo y control de calidad de los datos

La definición de las diferentes variables que registran los instrumentos y sensores de las boyas PIRATA se muestra en la tabla 2.4. La medición de la presión barométrica se realiza con una frecuencia de muestreo de 1 hora, la radiación de onda corta y onda larga con una frecuencia de 2 minutos y la precipitación de 1 minuto. Las mediciones del resto de parámetros se realiza con una frecuencia de 10 minutos. Estos datos de alta resolución quedan registrados internamente y se recuperan durante las tareas de mantenimiento de las boyas (una vez al año aproximadamente). Además, también se generan las medias por hora de las variables medidas en superficie y las medias diarias de las variables de subsuperficie. Estos datos se transmiten casi en tiempo real a los centros de operaciones vía satélite (GT MBA, 2016).

A los datos en tiempo real se les realiza un control de calidad de manera diaria, semanal y mensual para asegurar que cumplen los estándares de precisión. El control diario consiste en un análisis automático que detecta datos fuera de unos rangos de error especificados que dependen del tipo de variable medida. Por ejemplo, se considera erróneo:

- Un dato de velocidad del viento superior a 5 ms^{-1} con respecto al día anterior.
- Un valor de SST si:
 - Es mayor a $33,0 \text{ }^\circ\text{C}$ o menor a $-9,0 \text{ }^\circ\text{C}$.
 - Si se produce un cambio mayor a $5 \text{ }^\circ\text{C}$ con respecto al día anterior.
 - Si la SST por hora se encuentra fuera del rango $20^\circ - 30 \text{ }^\circ\text{C}$ en las dos últimas semanas.

Los datos que han sido marcados como erróneos se comprueban posteriormente por un analista experto. En cuanto al control semanal, el *National Center for Environmental Prediction* (NCEP) compara los datos con un modelo de predicción meteorológica (*weather prediction Medium Range Forecast*, MRF) y genera alertas para los datos que están fuera del rango especificado para cada variable. Por último, también se comprueban los datos mensuales comparándolos con modelos climatológicos (GT MBA, 2016).

A los datos de alta resolución, recuperados de la memoria interna de la boya, también se les aplica un procesado para la detección de datos ausentes y datos fuera de los rangos definidos. En el siguiente paso se generan gráficas de las series temporales, diagramas espectrales e histogramas y se calculan estadísticos básicos para ser revisados por un analista experimentado. También se verifican los registros para comprobar errores en la carga de la batería, daños en los sensores, fallos en la sincronización, etc. En algunas variables (lluvia, presión y salinidad) se aplica un procesado adicional para garantizar la calidad de los datos (GT MBA, 2016).

Tabla 2.4

Definición de las variables registradas en las boyas PIRATA.

Variable	Definición
YYYYMMDD	Fecha (año, mes y día) de registro del dato.
AIRT (°C)	Temperatura del aire.
UWND (m/s)	Componente este-oeste del vector viento, también denominado U o componente zonal del viento. Valores positivos indican la dirección este.
VWND (m/s)	Componente norte-sur del vector viento, también denominado como V o componente meridional. Los valores positivos indican la dirección norte.
WSPD (m/s)	Velocidad escalar del viento.
WDIR (grados)	Dirección del viento en grados en sentido horario desde el norte.
LWRad (W/m^2)	Radiación solar de onda larga.
SWRad (W/m^2)	Radiación solar de onda corta.
RH (%)	Humedad relativa.
Prec (mm/hr)	Lluvia o precipitación acumulada.
SLP (hPa)	Presión barométrica al nivel del mar.
DYN (dyn-cm)	La altura dinámica es una medida de la elevación de la superficie del mar (sin olas) y se calcula integrando la anomalía específica de volumen del agua de mar entre la superficie y a 500 metros.
ISO (m)	La profundidad de la isoterma (zona de temperatura constante) es la profundidad del océano en la que el agua se encuentra a una temperatura de 20°C. Se calcula mediante la interpolación lineal de los perfiles de temperatura vs profundidad.
SSS	Salinidad en superficie. El cálculo deriva de las mediciones de conductividad y temperatura.
S(z)	Salinidad a diferentes profundidades.
SSD (kg/m^3)	Densidad en superficie (sigma-theta). Deriva de la temperatura y la salinidad.
D(z) (kg/m^3)	Densidad en subsuperficie (sigma-theta) a diferentes profundidades.
SST (°C)	Temperatura de la superficie del mar.
T(z) (°C)	Temperatura del mar a diferentes profundidades.

2.2.3. Telemetría y distribución de datos

Los datos de las boyas son enviados casi en tiempo real mediante el servicio Argos que utiliza los satélites de órbita polar de la NOAA (*US National Oceanic and Atmospheric Administration*). Una vez aplicados los controles de calidad y procesado a los datos, estos quedan disponibles tanto para la comunidad científica como para el público general. Desde las páginas web de los diferentes proyectos (TAO/TRITON y PIRATA) es posible acceder a las bases de datos (GTMBA, 2016).

Las boyas ATLAS transmiten las observaciones horarias más recientes y las medias diarias del día anterior. Desde el año 2005, la transmisión de datos se realiza durante 16 horas al día (de 00:00 a 04:00, de 06:00 a 10:00, de 12:00 a 16:00, y de 18:00 a 22:00), coincidiendo con el paso de los satélites (de 8 a 14 veces diarias). Estos datos se actualizan diariamente en la página web y además se transmiten mediante el sistema GTS (*Global Telecommunications System*) del servicio Argos a los diferentes centros de operaciones del mundo. En la figura 2.9 se muestra un esquema de la transmisión de los datos enviados por las boyas (Hayes et al., 1991; Servain et al., 1998).

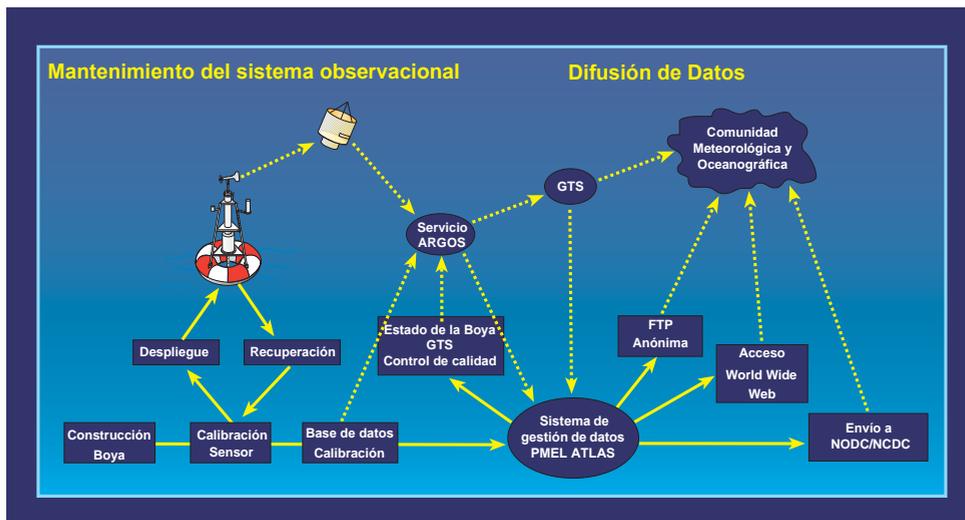


Figura 2.9

Esquema del mantenimiento y difusión de datos del sistema ATLAS (GT MBA, 2016).

Capítulo 3

Estudio de la relación entre el dipolo PIRATA del Atlántico y las anomalías de lluvias en el Nordeste de Brasil

3.1. Introducción

Tal y como se ha descrito en el capítulo 1, la red de boyas PIRATA se implementó inicialmente en 1995 con la finalidad de monitorizar los dos modos principales de la variabilidad climática que se producen en el Atlántico tropical. Estos modos, conocidos como modo ecuatorial y modo meridional (dipolo), están fuertemente relacionados con las anomalías que se producen en las precipitaciones estacionales del NEB y en la región Oeste de África (Moura y Shukla, 1981; Servain et al., 2000, 2003).

Desde el inicio del proyecto PIRATA, las boyas situadas en las líneas meridionales 38°W (4 boyas entre 4°N y 20°N) y 10°W (3 boyas entre 0° y 10°S) se han empleado en el estudio del modo dipolo. Por ejemplo, en la Fundación Cearense de Meteorología y Recursos Hídricos (FUNCEME), responsable de la previsión del tiempo y del clima en Ceará, se monitoriza habitualmente el gradiente inter-hemisférico o dipolo representado por las diferencias diarias en la SST que se dan entre las boyas 15°N - 38°W (noroeste) y 10°S - 10°W (sudeste) como aproximación al dipolo del Atlántico tropical. Los datos de este dipolo están disponibles en la página web de FUNCEME (2016). Además de esta aproximación, también se utiliza habitualmente el denominado dipolo SST de Servain, que se obtiene calculando las diferencias mensuales de anomalías de SST entre la región del Norte (de 28°N a 5°N y de 60°W a 15°E) y la

región del Sur (de 5°N a 20°S y de 60°W a 15°E). Los datos de SST para el cálculo de este dipolo son registrados por barcos participantes en el programa VOSCLim (*Voluntary Observing Ship Climate*).

En el NEB, la ITCZ es el principal sistema inductor de precipitaciones en la región. Su posicionamiento depende, de entre otras variables de interacción océano-atmósfera, de las temperaturas observadas sobre el océano Atlántico tropical. Por tanto, el estudio del gradiente inter-hemisférico de la SST proporciona información importante para la previsión del clima.

3.2. Objetivos

En este capítulo se va analizar la relación existente entre los dipolos de SST y las precipitaciones en los diferentes estados del NEB. Para ello, se utilizarán los datos de SST proporcionados por las boyas situadas en la línea meridional 38°W, la boya 10°S - 10°W y las boyas de la extensión sudoeste de la red PIRATA. Además, también se analizará el dipolo de Servain. Los objetivos del presente capítulo son:

- Obtener los dipolos mensuales a partir de todas las combinaciones posibles entre las boyas PIRATA fondeadas en el Atlántico Norte (línea meridional 38°W) y las fondeadas en el Atlántico Sur (boya 10°S - 10°W y boyas de la PIRATA-SWE).
- Correlacionar los valores de los dipolos obtenidos con las anomalías de lluvias que se producen en los estados que forman el NEB.
- Comparar los resultados de los dipolos PIRATA obtenidos con el dipolo “clásico” (15°N / 10°S) y el dipolo SST de Servain.
- Obtener el dipolo o dipolos con mayor potencial para el pronóstico de la temporada de lluvias en el NEB.

3.3. Materiales y métodos

3.3.1. Área de estudio y caracterización climática

La región del NEB ocupa un área de 1.554.257 Km², lo que equivale al 18,25% del territorio brasileño (Rocha et al., 2011). En la figura 3.1 se muestra un mapa de América del Sur y la posición geográfica que ocupa la región del NEB, mientras que en la figura 3.2 se muestra un mapa ampliado del NEB y de los nueve estados que lo conforman (Alagoas, Sergipe, Rio Grande do Norte, Piauí, Pernambuco, Paraíba, Maranhão, Ceará y Bahia).

Cerca del 70% de territorio del NEB presenta un clima semiárido. Este tipo de clima está caracterizado por unas temperaturas medias elevadas (entorno a 28 °C) y

por lluvias escasas e irregulares, que son fuertemente moduladas y modificadas por los patrones termodinámicos de las cuencas del océano Pacífico y océano Atlántico tropicales (Moura y Shukla, 1981; Carton y Zhou, 1997; Rocha et al., 2011).



Figura 3.1
Localización de la región Nordeste de Brasil



Figura 3.2
Estados que conforman el Nordeste de Brasil.

Los valores mensuales de temperatura y viento del Atlántico intertropical se pueden consultar en el capítulo 2. Además, en las figuras 3.3 y 3.4 están representadas las precipitaciones medias mensuales de la región del NEB y sus estados. Anualmente en el NEB se registra una precipitación media de 952 mm (79 mm mensuales), siendo el periodo comprendido entre enero y abril en el que se registra la mayor parte de estas lluvias (figura 3.3.a). Sin embargo, como se muestra en el resto de figuras, la época de lluvias y las precipitaciones acumuladas varían en cada estado. Por ejemplo, Pernambuco (figura 3.4.b) es el estado con la menor media de precipitación anual (639 mm), mientras que en Maranhão (figura 3.4.d) se registra la mayor media (1465 mm). De igual manera, la estación de lluvias también difiere en las diferentes regiones. En estados como Alagoas y Sergipe (figuras 3.3.b y 3.3.c), ésta tiene lugar entre los meses de abril y agosto, mientras que en Bahia (figura 3.4.f) tiene lugar entre noviembre y abril. Los detalles de los sistemas atmosféricos que causan estas lluvias están descritos en Ferreira y Mello. (2005).

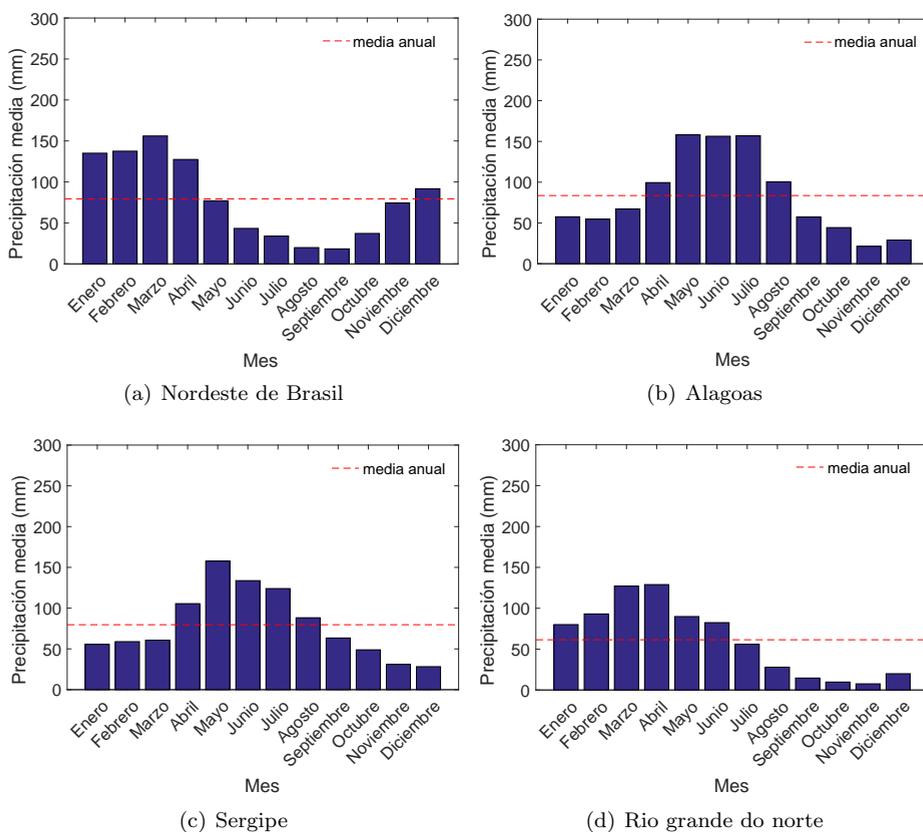
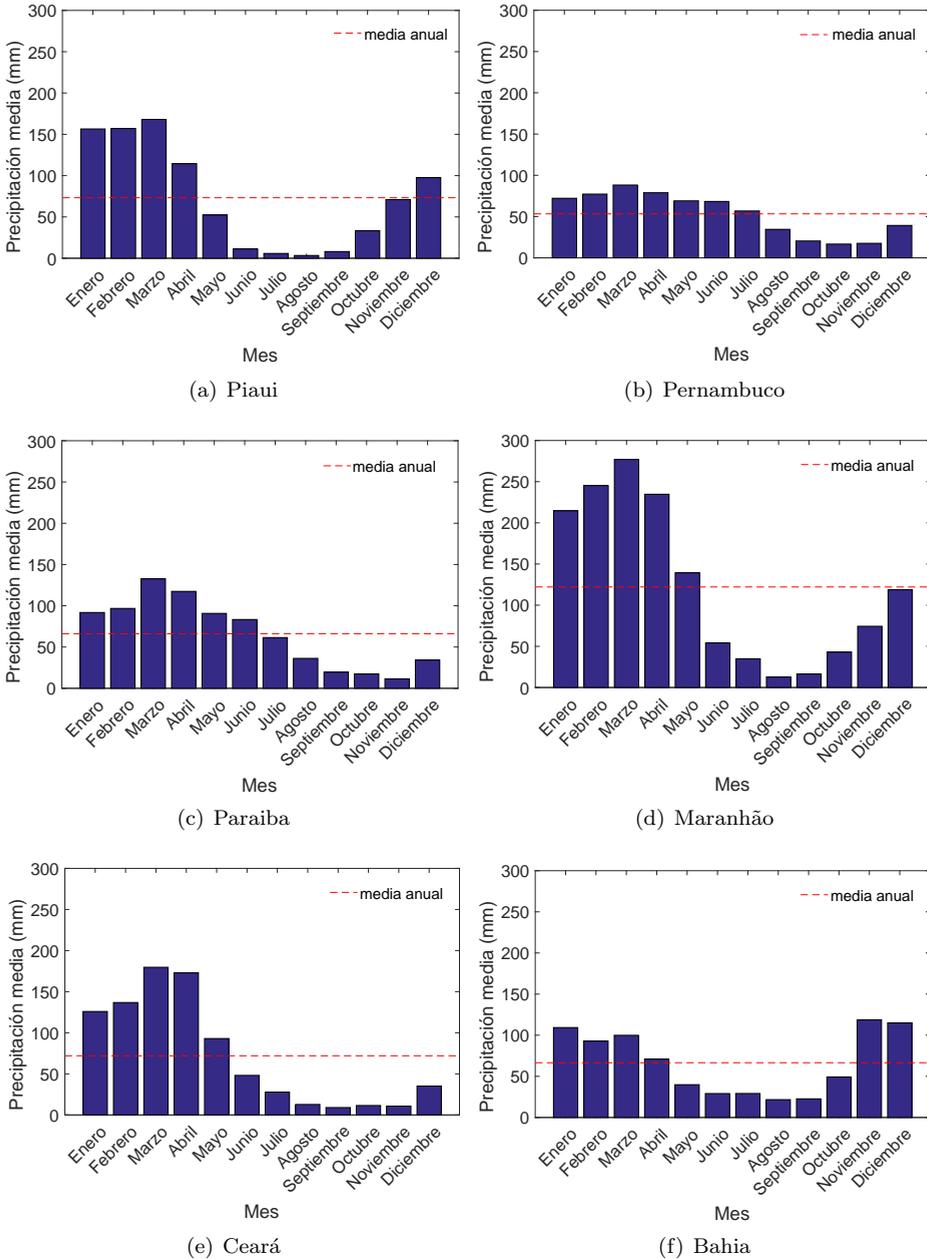


Figura 3.3 Precipitaciones en los diferentes estados de la región Nordeste de Brasil (CPTEC/INPE, 2018).

**Figura 3.4**

Precipitaciones en los diferentes estados de la región Nordeste de Brasil (continuación).

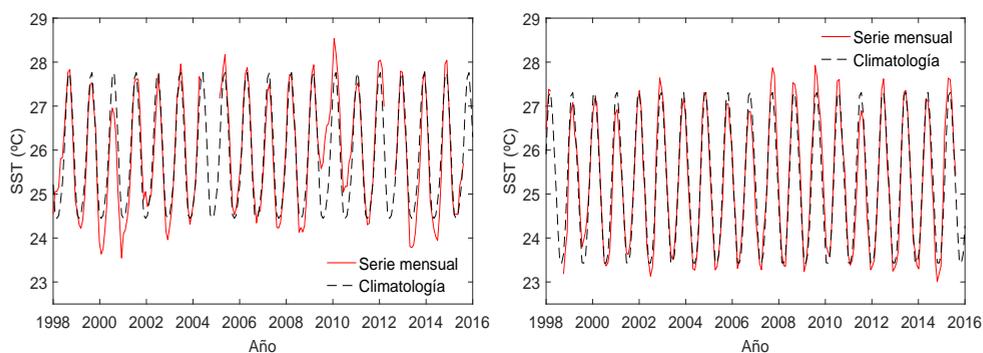
En cuanto a las boyas PIRATA, se han seleccionado las localizadas en la región Noroeste a lo largo de la línea 38°W (15°N , 12°N , 8°N y 4°N), las tres boyas de la PIRATA-SWE (8°S - 30°W , 14°S - 32°W y 19°S - 34°W) y la boya fondeada en 10°S - 10°W . La posición de estas boyas se muestra en la figura 1.1 del capítulo 1.

3.3.2. Datos

En primer lugar se obtuvieron los datos de SST desde el 1998 al 2017 de las boyas de interés en la página web de la *Global Tropical Moored Buoy Array Project Office* (GT MBA) perteneciente a la NOAA (GT MBA, 2016). A continuación, tras eliminar los datos ausentes, se siguieron los siguientes pasos para la obtención de los dipolos:

1. Obtener las series mensuales de SST observada promediando las series diarias.
2. Obtener la climatología mensual calculando el valor medio de SST en cada mes a lo largo de toda la serie histórica disponible.
3. Calcular las anomalías de SST restando a la SST observada la climatología.
4. Calcular el dipolo a partir de la diferencia entre las anomalías de SST en la boya norte y las anomalías de SST en la boya sur.

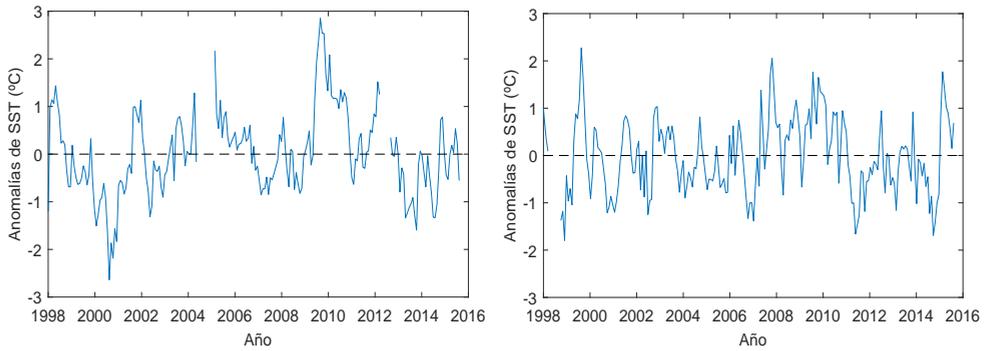
Como ejemplo, en las figuras 3.5 y 3.6 se muestran las series temporales que se obtienen al aplicar esta metodología en las boyas 15°N - 38°W y 10°S - 10°W . Puesto que siempre se calcula la diferencia entre la boya norte y la boya sur, los valores positivos (negativos) del dipolo indican una mayor (menor) temperatura en el hemisferio norte con respecto al hemisferio sur.



(a) SST mensual y climatología (15°N - 38°W). (b) SST mensual y climatología (10°S - 10°W).

Figura 3.5

Metodología para la obtención de los dipolos.



(a) Anomalías mensuales de SST ($15^{\circ}\text{N} - 38^{\circ}\text{W}$). (b) Anomalías mensuales de SST ($10^{\circ}\text{S} - 10^{\circ}\text{W}$).

Figura 3.6

Metodología para la obtención de los dipolos (continuación).

En la figura 3.7 se muestra el dipolo resultante, en el que se puede observar como el dipolo oscila entre valores positivos y negativos a lo largo de la serie histórica. Es decir, se observa la oscilación en el gradiente interhemiférico norte-sur de la SST que caracteriza al modo dipolo.

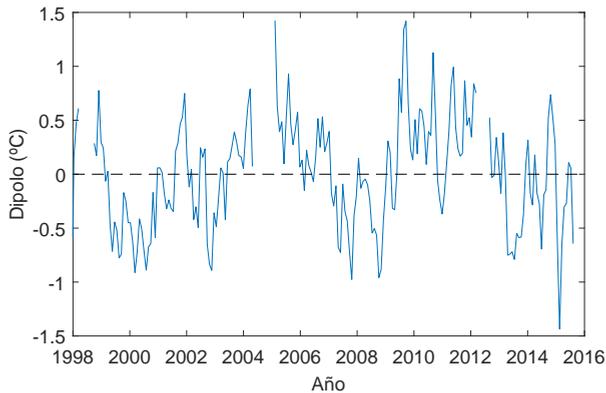


Figura 3.7

Dipolo obtenido calculando las diferencias en las anomalías de SST entre las boyas $15^{\circ}\text{N} - 38^{\circ}\text{W}$ y $10^{\circ}\text{S} - 10^{\circ}\text{W}$

En cuanto a los datos de precipitaciones, estos se obtuvieron del *Centro de Previsão de Tempo e Estudos Climáticos, Instituto Nacional de Pesquisas Espaciais* (CPTEC/INPE, 2018). Posteriormente, se calcularon las anomalías de lluvia mensuales restando a la lluvia mensual observada la media de precipitaciones mensuales de la serie histórica. En la figura 3.8 se muestran dos ejemplos de series de anomalías de lluvia.

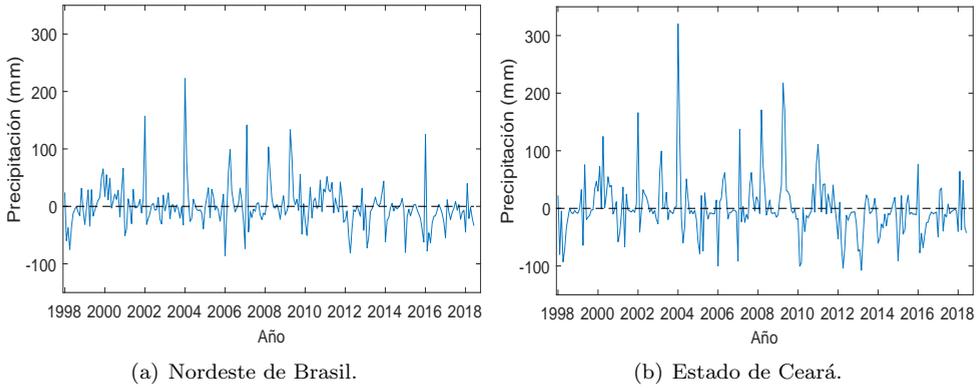


Figura 3.8

Ejemplos de series temporales mensuales de anomalías de lluvia.

3.3.3. Análisis de correlación

Una vez obtenidos los datos necesarios, se procedió con el cálculo de las correlaciones. Para determinar esta posible relación lineal, se utilizaron los coeficientes de correlación de *Pearson*, *Spearman* y *Kendall*.

3.3.3.1. Coeficiente de correlación de *Pearson*

La función de correlación más utilizada habitualmente es el coeficiente de correlación de *Pearson*, que fue desarrollado por el matemático Karl Pearson en el siglo XX. Formalmente, la correlación lineal r_{xy} entre dos variables aleatorias $\mathbf{x}(x_1, \dots, x_n)$ e $\mathbf{y}(y_1, \dots, y_n)$ se define como (Guyon y Elisseeff, 2006):

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (3.1)$$

donde \bar{x} y \bar{y} son los valores medios de \mathbf{x} e \mathbf{y} .

El coeficiente de correlación indica el grado de asociación lineal entre dos variables y puede variar entre -1 y 1. El signo + indica que existe una relación lineal creciente, mientras que el signo - indica una relación lineal decreciente (cuando una variable aumenta, la otra disminuye). Un coeficiente de 1 (o -1) representa una relación lineal perfecta, mientras que un valor de 0 indica que no existe relación lineal. Pese a ser la función más popular, la correlación de *Pearson* es sensible al número de datos y su distribución. Por tanto, su uso debe limitarse a los casos en los que la distribución de los datos sea normal (distribución gaussiana), el número de datos sea grande y no existan valores extremos (Hair et al., 2010).

3.3.3.2. Coeficiente de correlación de rangos de *Spearman*

Otra medición de la dependencia entre variables frecuentemente utilizada es la conocida como correlación de rangos de *Spearman*, nombrada así en honor del psicólogo C. Spearman. La correlación de *Spearman*, a diferencia de la de *Pearson*, es un test no paramétrico y, por tanto, es menos sensible a la distribución de los datos y valores extremos.

Siendo $\mathbf{R}(R_1, \dots, R_n)$ y $\mathbf{S}(S_1, \dots, S_n)$ los rangos de las observaciones originales $\mathbf{x}(x_1, \dots, x_n)$ e $\mathbf{y}(y_1, \dots, y_n)$, el coeficiente ρ se obtiene mediante la siguiente fórmula (Daniel, 1990):

$$\rho_{xy} = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)} \quad (3.2)$$

donde $d_i = R_i - S_i, i = 1, \dots, n$

De manera similar a la correlación de *Pearson*, el valor de ρ_{xy} varía entre -1 y 1, obteniéndose -1 cuando la secuencia de rangos es completamente opuesta y 1 cuando la secuencia coincide completamente.

3.3.3.3. Coeficiente de correlación de rangos de *Kendall*

Al igual que el coeficiente de *Spearman*, la correlación de *Kendall* (o τ de Kendall) es un test no paramétrico y está basado en el análisis de los rangos. De los índices utilizados, el de *Kendall* es el más robusto frente a valores extremos. Se dice que un par de observaciones (x_i, y_i) y (x_j, y_j) es concordante cuando los rangos de los elementos en ambos vectores coinciden, es decir, $x_i < x_j$ e $y_i < y_j$ o $x_i > x_j$ e $y_i > y_j$. En caso contrario, el par de observaciones es discordante. Siendo n_c el número de pares concordantes y n_d el número de pares discordantes, el coeficiente τ de *Kendall* se define como (Kendall y Gibbons, 1990):

$$\tau_{xy} = 2 \frac{n_c - n_d}{n(n - 1)} \quad (3.3)$$

Igual que en los casos anteriores, el valor de τ_{xy} varía entre -1 y 1.

3.4. Resultados

Como resultado de combinar los datos de las diferentes boyas, se obtuvieron 16 dipolos. Los valores mensuales de estos dipolos más el dipolo de Servain se correlacionaron con las anomalías de precipitación de cada mes en los nueve estados del NEB y con la media de toda la región del NEB, obteniéndose un total de 24480 correlaciones.

La normalidad de los datos se comprobó mediante el test de Shapiro-Wilk (Royston, 1992). De los 24480 subconjuntos de datos, no se cumplió la hipótesis de normalidad en 1560 subconjuntos de los dipolos y en 6491 series de anomalía de lluvia.

En las tablas 3.1, 3.2 y 3.3, se muestra un resumen de los resultados obtenidos utilizando las coeficientes de *Pearson*, *Spearman* y *Kendall* respectivamente. En estas tablas se indica el número de correlaciones estadísticamente significativas ($p < 0,05$) obtenidas en cada caso, el total de correlaciones por dipolo, el número de datos utilizado, el número de correlaciones positivas y negativas, así como la correlación media obtenida. Para el cálculo de esta correlación media se han utilizado los valores absolutos de las correlaciones significativas.

Utilizando la correlación de *Pearson* (tabla 3.1), los dipolos que presentan en general un mayor número de correlaciones significativas totales (71) y, además, con una media de correlación más alta ($r = 0,88$ y $r = 0,91$), son los dipolos $4^\circ\text{N} / 8^\circ\text{S}$ y $4^\circ\text{N} / 14^\circ\text{S}$. Sin embargo, se puede observar que el número de datos disponible en estas correlaciones es sólo de seis. Es decir, se dispone únicamente de los valores de seis años en las series de dipolo y lluvia. En casos particulares, como el de los estados de Paraíba y Maranhão, el mejor dipolo es el $15^\circ\text{N} / 19^\circ\text{S}$. Estos resultados son mejores que los obtenidos con el dipolo “clásico” ($15^\circ\text{N} / 10^\circ\text{S}$) y con el dipolo de Servain, ya que estos presentan un menor número de correlaciones significativas y una correlación media inferior.

Tabla 3.1

Resultados obtenidos con *Pearson*. Los valores indican el número de correlaciones significativas en cada caso y el total por dipolo, número de datos utilizado, número de correlaciones positivas, negativas y correlación media. Los casos con mayor número de correlaciones significativas se indican en negrita.

Dipolo	NEB	Alagoas	Sergipe	Rio Grande	Piauí	Pernambuco	Paraíba	Maranhão	Ceará	Bahia	Total	N datos	N Corr. Pos.	N Corr. Neg.	Corr. media
15N8S	1	5	5	6	3	4	6	4	7	4	45	9	38	7	0,74
15N10S	6	3	4	4	5	2	3	3	5	6	41	17	13	28	0,56
15N14S	1	4	0	3	1	1	2	0	4	5	21	10	4	17	0,69
15N19S	4	5	6	5	6	2	7	10	9	1	55	10	48	7	0,70
12N8S	9	2	4	7	6	4	4	3	4	5	48	8	27	21	0,78
12N10S	2	3	4	5	1	3	7	1	6	4	36	14	8	28	0,57
12N14S	4	6	2	6	5	2	6	6	3	4	44	9	5	39	0,78
12N19S	1	6	2	6	4	4	4	4	8	3	42	9	29	13	0,73
8N8S	5	5	1	9	8	7	4	5	6	4	54	8	17	37	0,80
8N10S	5	4	9	2	5	4	1	4	7	7	48	16	11	37	0,54
8N14S	6	2	0	5	3	4	4	3	8	7	42	9	14	28	0,76
8N19S	3	4	7	4	3	3	2	2	2	2	32	9	22	10	0,74
4N8S	10	5	3	9	9	7	4	5	10	9	71	6	12	59	0,88
4N10S	4	0	2	2	3	1	0	0	5	2	19	12	4	15	0,61
4N14S	10	7	2	12	6	5	4	8	12	5	71	6	8	63	0,91
4N19S	2	6	5	6	5	4	2	3	4	2	39	6	24	15	0,84
Servain	5	0	0	7	3	1	2	5	2	4	29	20	6	23	0,53

Como se ha comentado anteriormente, la correlación de *Pearson* es sensible a las características de los datos. En este caso, utilizando el número de datos y el número total de correlaciones significativas, se obtuvo una correlación de $r = -0,35$. Esta correlación puede indicar que a menor número de datos, se obtiene un mayor número de correlaciones significativas, es decir, se obtienen correlaciones más altas. Además, en más de un tercio de los casos no se cumplía el criterio de normalidad en las anomalías de lluvia, pudiendo indicar un alto número de valores extremos. Puesto que los resultados con *Pearson* pueden estar sesgados por estos motivos, se utilizaron también los tests no paramétricos de *Spearman* y *Kendall*.

En la tabla 3.2 se muestran los resultados del coeficiente de *Spearman*. En este caso, los dipolos con un mayor número de correlaciones totales son 12°N / 14°S y 8°N / 14°S, con 50 y 48 correlaciones significativas respectivamente, siendo la media de correlación de 0,78 en ambos casos. Con *Spearman*, los dipolos 4°N / 8°S y 4°N / 14°S presentan 36 y 44 correlaciones significativas respectivamente. Este número es inferior al obtenido mediante *Pearson*, sin embargo, la correlación media en estos dipolos sigue siendo alta ($\rho = 0,92$ y $\rho = 0,88$). En cuanto a la correlación entre el número de datos y el número de correlaciones significativas, el valor obtenido es de $r = -0,13$, inferior al obtenido con *Pearson*. Esto podría indicar una menor influencia del número de datos en los resultados.

Tabla 3.2

Resultados obtenidos con *Spearman*. Los valores indican el número de correlaciones significativas en cada caso y el total por dipolo, número de datos utilizado, número de correlaciones positivas, negativas y correlación media. Los casos con mayor número de correlaciones significativas se indican en negrita.

Dipolo	NEB	Alagoas	Sergipe	Rio Grande	Piauí	Pernambuco	Paraíba	Maranhão	Ceará	Bahía	Total	N datos	N Corr. Pos.	N Corr. Neg.	Corr. media
15N8S	1	8	4	3	2	1	4	2	3	5	33	9	18	15	0,77
15N10S	2	4	4	7	3	2	2	3	5	5	37	17	6	31	0,58
15N14S	2	5	3	2	1	1	5	1	4	3	27	10	5	22	0,70
15N19S	4	3	8	1	7	1	4	5	6	4	43	10	35	8	0,70
12N8S	7	2	3	5	4	3	3	4	6	9	46	8	32	14	0,79
12N10S	3	3	4	5	0	4	2	0	7	5	33	14	6	27	0,59
12N14S	3	7	0	5	6	3	8	4	12	2	50	9	7	43	0,78
12N19S	2	6	4	4	6	0	2	4	5	4	37	9	28	9	0,74
8N8S	6	5	0	5	6	4	4	4	4	5	43	8	19	24	0,84
8N10S	3	5	5	1	0	5	4	2	5	3	33	16	9	24	0,57
8N14S	8	2	1	4	8	3	6	4	7	5	48	9	18	30	0,78
8N19S	2	4	5	3	3	3	0	1	1	3	25	9	15	10	0,74
4N8S	6	2	0	7	4	3	3	2	7	2	36	6	8	28	0,93
4N10S	1	2	2	1	2	3	1	1	1	2	16	12	6	10	0,61
4N14S	7	5	1	5	4	1	5	8	6	2	44	6	6	38	0,92
4N19S	2	3	5	2	4	3	1	1	1	1	23	6	16	7	0,88
Servain	3	2	2	3	5	4	5	4	3	1	32	20	8	24	0,50

Por otra parte, los resultados con el test de *Kendall* se presentan en la tabla 3.3. En este caso, los dipolos con un mayor número total de correlaciones significativas

son 15°N / 19°S, 12°N / 14°S y 8°N / 14°S. Estos resultados son similares a los obtenidos mediante *Spearman*, aunque ahora las correlaciones medias son inferiores ($\tau = 0,54$, $\tau = 0,64$ y $\tau = 0,64$). En el caso de los dipolos 4°N / 8°S y 4°N / 14°S, siguen manteniendo unos valores medios de correlación altos ($\tau = 0,90$ y $\tau = 0,88$), pero presentan un menor número de correlaciones significativas (27 y 32) que en los anteriores casos. La correlación entre el número de datos y el número de correlaciones significativas es $r = 0,04$, siendo un valor inferior al de los otros métodos analizados.

Tabla 3.3

Resultados obtenidos con *Kendall*. Los valores indican el número de correlaciones significativas en cada caso y el total por dipolo, número de datos utilizado, número de correlaciones positivas, negativas y correlación media. Los casos con mayor número de correlaciones significativas se indican en negrita.

Dipolo	NEB	Alagoas	Sergipe	Rio Grande	Piauí	Pernambuco	Paraíba	Maranhão	Ceará	Bahia	Total	N datos	N Corr. Pos.	N Corr. Neg.	Corr. media
15N8S	1	5	4	3	3	0	5	1	3	4	29	9	14	15	0,62
15N10S	2	4	3	5	2	2	1	3	5	5	32	17	6	26	0,44
15N14S	0	5	4	4	1	1	3	1	4	4	27	10	7	20	0,55
15N19S	3	4	9	2	5	2	5	5	4	4	43	10	36	7	0,54
12N8S	5	0	1	4	3	3	2	3	3	4	28	8	20	8	0,68
12N10S	2	2	2	4	0	4	3	0	7	3	27	14	4	23	0,45
12N14S	2	5	0	4	5	4	6	6	10	1	43	9	5	38	0,64
12N19S	3	6	3	3	7	1	3	4	6	4	40	9	28	12	0,58
8N8S	4	4	0	5	5	4	2	2	4	3	33	8	15	18	0,73
8N10S	3	4	4	2	1	4	4	2	2	4	30	16	8	22	0,42
8N14S	8	3	1	6	8	3	4	4	5	5	47	9	19	28	0,64
8N19S	1	3	5	3	4	1	2	1	2	4	26	9	16	10	0,60
4N8S	5	2	0	5	3	1	2	2	5	2	27	6	3	24	0,90
4N10S	1	3	1	2	1	2	0	0	0	1	11	12	4	7	0,48
4N14S	5	4	1	4	4	1	3	6	4	0	32	6	2	30	0,88
4N19S	1	3	5	0	3	3	0	1	1	1	18	6	13	5	0,80
Servain	2	2	1	5	5	3	4	4	3	1	30	20	9	21	0,36

En la tabla 3.4 se muestra en detalle (mes a mes) los resultados obtenidos entre los valores del dipolo 12°N / 14°S y las anomalías de lluvia del estado de Ceará utilizando el coeficiente de *Kendall*. Se puede observar que, por ejemplo, el dipolo de abril presenta correlaciones significativas con las anomalías de lluvias en los meses de mayo, julio, agosto y noviembre. Estas correlaciones son negativas, indicando que las anomalías cálidas de SST al norte con respecto al sur (dipolos positivos), están relacionados con periodos de sequías. De igual manera, los dipolos negativos estarían relacionados con periodos en los que las precipitaciones acumuladas serían superiores a la media.

Otro ejemplo de los resultados mes a mes se muestra en la tabla 3.5. En este caso también se obtienen correlaciones negativas significativas en ciertos meses. Sin embargo, los datos del dipolo de enero con las lluvias de abril, junio, agosto y octubre proporcionan correlaciones positivas. Este hecho parece contradecir lo expuesto anteriormente para la región de Ceará. En el caso de la región de Piauí, los resultados

sugieren que el modo dipolo presenta una menor influencia en sus regímenes de lluvia, pudiendo tener una mayor influencia otras variables de interacción océano-atmosféricas.

Tabla 3.4

Resultados de las correlaciones mensuales entre los valores de temperatura del dipolo 12°N / 14°S (filas) y las anomalías de lluvia en el estado de Ceará (columnas) utilizando el coeficiente de *Kendall*. Las correlaciones estadísticamente significativas se muestran en negrita.

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
Enero	-0,06	0,50	0,28	0,28	0,28	0,67	0,06	0,17	-0,33	0,11	0,22	-0,28
Febrero		-0,17	-0,39	-0,28	-0,28	0,11	-0,28	-0,39	-0,22	0,22	-0,11	-0,06
Marzo			-0,17	0,06	0,06	0,22	0,17	0,17	0,00	0,44	0,00	-0,06
Abril				-0,44	-0,56	-0,39	-0,67	-0,56	-0,17	-0,17	-0,61	0,22
Mayo					-0,79	-0,71	-0,57	-0,57	-0,07	0,07	-0,50	0,14
Junio						-0,33	-0,44	-0,17	-0,17	-0,17	-0,72	0,28
Julio							-0,06	0,11	-0,33	0,22	-0,44	0,22
Agosto								-0,11	-0,56	-0,33	-0,78	0,28
Septiembre									-0,28	0,06	-0,39	0,11
Octubre										-0,11	-0,42	0,16
Noviembre											-0,13	0,09
Diciembre												0,16

Tabla 3.5

Resultados de las correlaciones mensuales entre los valores de temperatura del dipolo 8°N / 14°S (filas) y las anomalías de lluvia en el estado de Piauí (columnas) utilizando el coeficiente de *Kendall*. Las correlaciones estadísticamente significativas se muestran en negrita.

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
Enero	0,17	0,00	0,39	0,61	0,44	0,56	0,22	0,56	0,11	0,56	-0,11	0,50
Febrero		0,07	-0,21	-0,64	-0,50	0,14	-0,14	-0,43	-0,43	-0,21	0,21	-0,57
Marzo			-0,29	-0,29	-0,29	-0,50	-0,07	-0,07	0,21	-0,14	-0,29	-0,36
Abril				-0,50	-0,61	0,11	-0,50	-0,33	0,00	-0,33	0,33	-0,28
Mayo					-0,71	0,14	-0,64	-0,29	0,07	-0,43	0,36	-0,29
Junio						-0,07	-0,57	0,14	0,14	-0,57	0,36	-0,14
Julio							-0,50	0,36	0,36	-0,21	0,29	0,07
Agosto								-0,02	-0,02	-0,42	0,07	-0,20
Septiembre									0,00	0,06	-0,06	0,39
Octubre										-0,33	-0,02	-0,07
Noviembre											-0,06	0,17
Diciembre												0,17

Con el fin de complementar estos resultados, se obtuvieron además las correlaciones entre los dipolos de octubre a diciembre con las anomalías promedio de la temporada de lluvias (marzo a mayo) del año siguiente en el NEB y Ceará, así como las correlaciones entre los dipolos de enero y febrero con las anomalías de lluvia promedio del mismo año. Con el dipolo de Servain se obtuvieron correlaciones signi-

ficativas únicamente con los datos de febrero (NEB: $\rho = -0,54$ y Ceará: $\rho = -0,46$). Con el resto de dipolos se obtuvieron las siguientes correlaciones significativas:

- **NEB:** diciembre ($\rho = 0,83$) y enero ($\rho = 0,88$) del dipolo $8^\circ\text{N} / 8^\circ\text{S}$, enero ($\rho = 0,73$) del dipolo $8^\circ\text{N} / 14^\circ\text{S}$ y diciembre ($\rho = 0,58$) del dipolo $8^\circ\text{N} / 10^\circ\text{S}$
- **Ceará:** octubre ($\rho = 0,65$) del dipolo $8^\circ\text{N} / 14^\circ\text{S}$.

Estos resultados no consiguen mejorar a los obtenidos correlacionando mensualmente los dipolos con las anomalía de lluvia. Sin embargo, este nuevo análisis apunta a que el dipolo de Servain es mejor indicador que los dipolos PIRATA al estudiar el promedio de anomalías en la temporada de lluvias, mientras que los dipolos PIRATA son mejores en el análisis mes a mes. Esta diferencia puede deberse a que el dipolo de Servain utiliza valores de la SST de regiones extensas del océano, mientras que los dipolos PIRATA utilizan valores de la SST en puntos concretos.

En la figura 3.9 se muestra el dipolo “clásico”, el dipolo de Servain, así como los dipolos PIRATA $12^\circ\text{N} / 14^\circ\text{S}$ y $8^\circ\text{N} / 14^\circ\text{S}$ de manera superpuesta. Aunque dependiendo del periodo, se pueden observar mayores o menores similitudes en el comportamiento de las series temporales, en general, el dipolo de Servain muestra unos valores de anomalías mayores a los de los dipolos PIRATA. Finalmente, en la figura 3.10 se ilustran las series temporales del dipolo “clásico” $15^\circ\text{N} / 10^\circ\text{S}$, del dipolo de Servain y de los dipolos PIRATA con mayor número de correlaciones significativas con las anomalías de precipitaciones en el NEB ($8^\circ\text{N} / 14^\circ\text{S}$) y en el estado de Ceará ($12^\circ\text{N} / 14^\circ\text{S}$). Superpuestas a estas series temporales, se muestra la media de precipitaciones durante la temporada de lluvias de marzo a mayo.

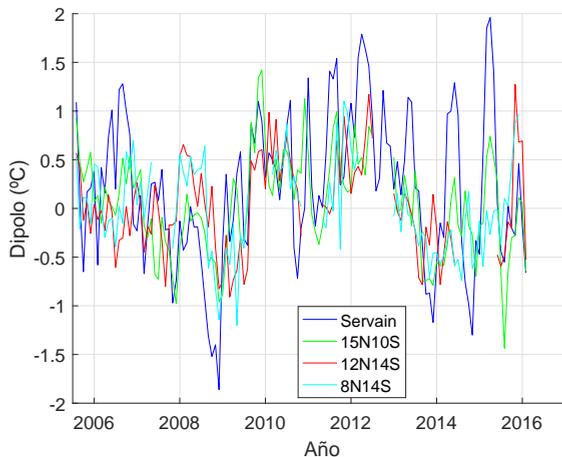
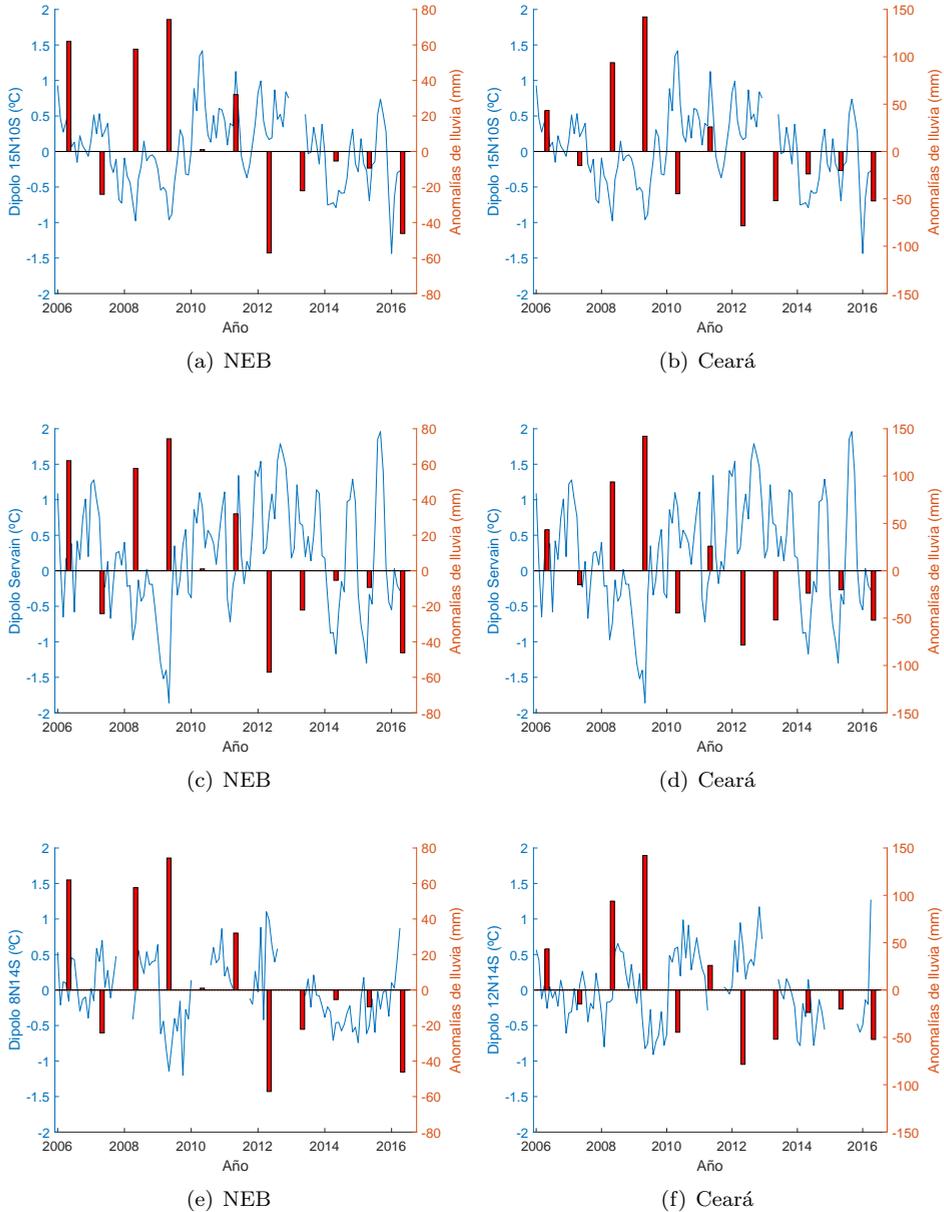


Figura 3.9
Comparación entre los dipolos PIRATA y el dipolo de Servain.

**Figura 3.10**

Dipolos mensuales (líneas azules) y medias de anomalías de precipitaciones (barras rojas) durante la temporada de lluvias (marzo a mayo) en el Nordeste de Brasil y el estado de Ceará (2006-2016). Se muestra el dipolo “clásico” $15^{\circ}\text{N} / 10^{\circ}\text{S}$, el dipolo de Servain y los dipolos PIRATA con mayor número de correlaciones significativas con el NEB y Ceará ($8^{\circ}\text{N} / 14^{\circ}\text{S}$ y $12^{\circ}\text{N} / 14^{\circ}\text{S}$).

3.5. Conclusiones

En este capítulo de la Tesis Doctoral se ha analizado la relación entre los valores de temperatura del dipolo del Atlántico y las anomalías de precipitación en los diferentes estados del NEB. En primer lugar, se han obtenido los datos de SST proporcionados por las boyas de la línea meridional 38°W y 10°W, la boya 10°S - 10°W y las boyas de la PIRATA-SWE. Utilizando estos datos, se han obtenido 16 dipolos mediante las combinaciones entre las boyas del norte y las boyas del sur. Posteriormente, utilizando los coeficientes de correlación de *Pearson*, *Spearman* y *Kendall*, se ha determinado la relación entre estos dipolos y las series históricas de lluvia del NEB. Las conclusiones obtenidas son:

- Aunque los resultados varían en función del método utilizado y de la región estudiada, los experimentos muestran que los nuevos dipolos analizados presentan, en general, un mayor número de correlaciones significativas y una correlación media mayor que el dipolo “clásico” (15°N / 10°S) y el dipolo de Servain.
- Puesto que los métodos *Spearman* y *Kendall* son más robustos frente a datos extremos que *Pearson*, los resultados obtenidos sugieren que los dipolos 12°N / 14°S y 8°N / 14°S son los mejores candidatos para representar el dipolo del Atlántico.
- Junto al dipolo “clásico” y al dipolo de Servain, sería importante la monitorización de estos nuevos dipolos con la finalidad de mejorar las previsiones de tiempo y clima en el NEB.

Capítulo 4

Identificación de patrones en los datos PIRATA mediante mapas auto-organizados

4.1. Introducción

En los anteriores capítulos se han detallado varios de los diversos procesos de interacción océano-atmósfera que están implicados en la variabilidad del clima del océano Atlántico. Para comprender estos procesos es necesario estudiar el efecto de las variables atmosféricas, especialmente las relativas al viento, en el comportamiento de las variables oceánicas. Por ejemplo, en la zona ecuatorial, conocer la migración de la ITCZ y su influencia en el comportamiento de los vientos es fundamental para comprender la variabilidad de la SST (Csanady, 2001; Ferreira y Mello., 2005).

Los vientos y las olas mezclan las aguas de la parte superior del océano, distribuyendo el calor de manera casi uniforme. Por debajo de la capa superior, denominada capa de mezcla, se sitúa la termoclina. Esta capa de transición, separa las aguas calientes de la superficie de las aguas frías de las profundidades. Mientras que en la capa de mezcla la temperatura permanece prácticamente constante, en la termoclina disminuye rápidamente al aumentar la profundidad. Una vez alcanzada la base de la termoclina, la temperatura de las aguas profundas continúa disminuyendo, pero mucho más lentamente (NOAA, 2018). La profundidad de las capas de mezcla y termoclina varían dependiendo de la latitud, longitud y época del año, afectando al comportamiento de la SST (Rugg et al., 2016). En la figura 4.1 se muestra un perfil de temperatura para la boya 12°N - 38°W, en la que se observa la rápida disminución de la temperatura en la capa termoclina. En esta localización, la capa de mezcla presenta una profundidad promedio anual de 43 m (Monterey y Levitus, 1997).

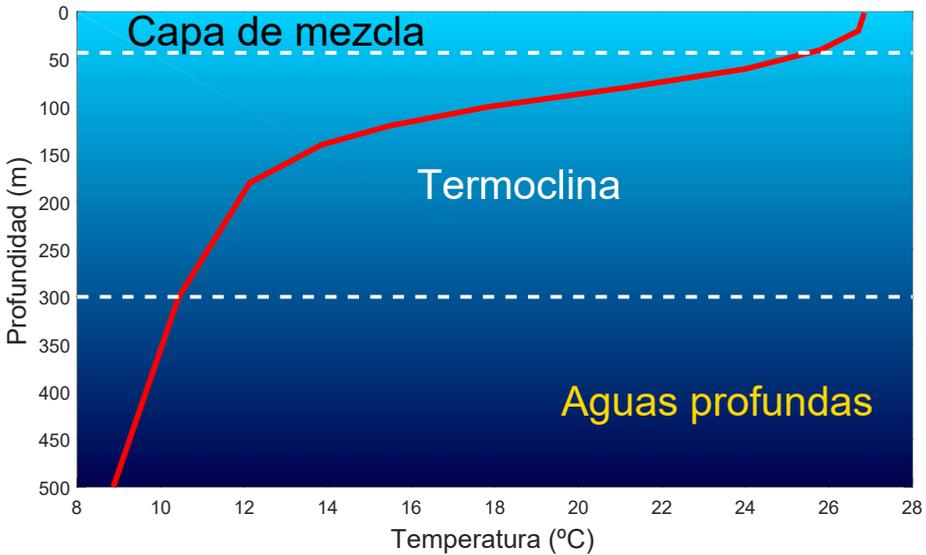


Figura 4.1

Perfil de temperatura promedio en función de la profundidad en la boya 12°N - 38°W.

Además de las variables relativas al viento y las propiedades de las capas de mezcla y termoclina, otras variables meteo-atmosféricas como la temperatura del aire, la humedad relativa, la salinidad o la densidad están implicadas en la variabilidad de la temperatura de los océanos. Por tanto, el análisis de todas estas relaciones plantea un problema multivariante. Una posible aproximación a la resolución de este problema consiste en realizar un análisis mediante redes SOM. Un SOM es un tipo de red neuronal artificial (*Artificial Neural Network*, ANN) que se utiliza habitualmente en tareas de visualización de datos de alta dimensionalidad y fue descrito por primera vez por Teuvo Kohonen en 1982 (Kohonen, 2001). Los SOM mapean las relaciones complejas que existen entre las variables de un conjunto de datos de alta dimensionalidad en una relación más sencilla de baja dimensionalidad, facilitando la visualización y extracción de información de los datos (Haykin, 2009).

Esta técnica ha sido utilizada anteriormente en física oceanográfica, principalmente para detectar patrones en la variabilidad de la temperatura y del clima. Por ejemplo, utilizando una variante del SOM, el GHSOM (*Growing Hierarchical Self-Organizing Map*), los autores Liu et al. (2006) analizaron series temporales diarias de SST con la finalidad de obtener patrones de variabilidad sobre la plataforma continental de Florida occidental. Los SOMs obtenidos permitieron visualizar y caracterizar los patrones que presenta el ciclo estacional (invierno/verano) de la SST. Además, los autores sostienen que el método SOM permitió obtener patrones que no fueron fácilmente detectables mediante el análisis con la función ortogonal empírica (EOF), método lineal ampliamente utilizado en oceanografía.

Por otra parte, Reusch et al. (2007) y Johnson et al. (2008) estudiaron la variabilidad climática en el Atlántico Norte y la Oscilación del Atlántico Norte. En el trabajo de Reusch et al. (2007), estos fenómenos se analizaron mediante los mapas SOM entrenados utilizando como entradas la presión media del nivel del mar durante la estación de invierno. Los autores concluyen que, comparado con el método PCA, el SOM posibilita el análisis de la variabilidad de manera más sencilla e intuitiva. En el trabajo de Johnson et al. (2008) también se utiliza la presión media del nivel del mar y concluyen que análisis SOM supone un método adicional para explorar las relaciones en las conexiones de diferentes zonas del Atlántico Norte, así como entre la Oscilación del Atlántico Norte y otras variables climáticas. Continuando con estudios en el Atlántico Norte, Wu et al. (2012) utilizó los mapas SOM para estimar las anomalías de temperatura en subsuperficie entrenando los modelos con series mensuales de anomalías de SST, altura del mar y salinidad, obteniendo buenos resultados a profundidades superiores a 700 m. Por último, entrenando los SOMs con patrones anómalos de la SST, Li et al. (2015) consiguió clasificar con éxito las fases de inicio, pico y fin de los eventos el Niño y la Niña.

A diferencia de los trabajos mencionados, en este capítulo de la Tesis Doctoral se plantea el uso de los SOMs para analizar y extraer patrones generales sobre las posibles relaciones existentes entre un mayor número de variables que caracterizan al océano Atlántico. Para este análisis se pretende utilizar todas las variables meteorológicas proporcionadas por las boyas PIRATA situadas en $0^{\circ}\text{N} - 35^{\circ}\text{W}$, $0^{\circ}\text{N} - 23^{\circ}\text{W}$, $8^{\circ}\text{N} - 38^{\circ}\text{W}$ y $8^{\circ}\text{S} - 30^{\circ}\text{W}$. La localización de estas boyas permite estudiar los sistemas de interacción océano-atmósfera en la zona ecuatorial, así como en las regiones norte y sur del Atlántico tropical.

4.2. Objetivos

El objetivo de este capítulo es utilizar el análisis SOM, que permite visualizar la estructura general de todo el conjunto de datos disponible en cada boya, con la finalidad de contribuir en la mejora del conocimiento sobre los procesos que modulan la variabilidad de la SST en el océano Atlántico tropical. Por tanto, los objetivos principales de este capítulo de la Tesis Doctoral son:

- Demostrar la utilidad de la técnica SOM en la identificación de patrones generales en los conjuntos de datos oceánicos y atmosféricos proporcionados por las boyas del proyecto PIRATA.
- Describir algunos de los patrones meteorológicos-oceanográficos en el océano Atlántico tropical y los procesos involucrados en su evolución temporal.
- Estudiar y establecer relaciones entre la capa de mezcla y la termoclina; entre densidad, salinidad, temperatura en superficie y subsuperficie, así como entre viento, humedad relativa y temperatura del aire.

Las series temporales diarias se obtuvieron de la página web de la NOAA (GTM-BA, 2016). Cada tipo de variable se proporciona en un archivo *.ascii* individual que presenta una estructura de datos por bloques. Por tanto, la primera etapa del procesamiento consistió en generar una única matriz de datos para cada boya a partir de las variables individuales, tomando como referencia la fecha de registro de cada dato. En la segunda etapa de procesamiento se realizó un análisis de los datos ausentes en cada variable, eliminando aquellas variables con un número elevado de valores ausentes respecto a la media del resto de variables. En la siguiente etapa se eliminaron los patrones con datos ausentes. Finalmente, todas las variables fueron estandarizadas (media cero y varianza unidad). Tras un primer análisis general de los SOMs obtenidos en cada boya, se decidió estudiar con mayor detalle cuatro boyas representativas, dos situadas sobre el ecuador, una en el hemisferio norte y otra en el hemisferio sur:

1. 0°N – 35°W (22/01/1988 a 31/12/2013).
2. 0°N – 23°W (03/03/1999 a 31/12/2013).
3. 8°N – 38°W (31/01/1998 a 31/12/2013).
4. 8°S – 30°W (21/08/2005 a 31/12/2013).

En la tabla 4.1 se muestra la definición de las variables presentes en las matrices de datos. En la sección 2.2.1 del capítulo 2 se puede consultar una descripción más detallada de estas variables. Con respecto a las originales, las variables han sido renombradas para facilitar su identificación en los mapas SOM.

Tabla 4.1

Descripción de las variables presentes en los conjuntos de datos.

Variable	Definición
<i>Air Temp</i> (°C)	Temperatura del aire.
<i>U Wind</i> (m/s)	Viento zonal (componente este-oeste).
<i>V Wind</i> (m/s)	Viento meridional (componente norte-sur).
<i>Wind Speed</i> (m/s)	Velocidad escalar del viento.
<i>Wind Dir</i> (deg)	Dirección del viento en grados.
<i>Max Sw rad</i> (W/m ²)	Radiación máxima de onda corta.
RH (%)	Humedad relativa.
<i>Rain</i> (mm/hr)	Lluvia o precipitación acumulada.
<i>Dyn Ht</i> (0-500 m)	Altura dinámica.
<i>20° C IsoD</i> (m)	Profundidad de la isoterma a 20 °C.
<i>Sea Salt</i> 1m, 20m, 40m y 120m	Salinidad del mar a diferentes profundidades.
<i>Sea Dens</i> (Kg/m ³) 1m, 20m, 40m y 120m	Densidad del mar a diferentes profundidades.
<i>Sea Temp</i> (°C) 20m, 40m, 60m, 80m, 100m, 120m, 140m, 180m, 300m y 500m	Temperatura del mar en subsuperficie a diferentes profundidades.
SST (°C)	Temperatura en la superficie del mar.

4.3.2. Mapas auto-organizados

Un SOM está formado por una serie de neuronas ordenadas en dos capas: una capa de entrada formada por N neuronas (N es igual al número de variables de entrada) y una capa de salida en la que se procesa la información. Las neuronas de la capa de salida se configuran en un espacio de baja dimensionalidad en forma de rejilla rectangular o hexagonal. Esta capa normalmente sigue una estructura de dos dimensiones ya que la extracción de información resulta más sencilla que en una estructura con un mayor número de dimensiones. En el mapa de dos dimensiones resultante o “*components plane*”, se mantienen las relaciones de vecindad existentes en el espacio original, de manera que los datos con características similares se encuentran próximos entre sí, mientras que los que presentan características distintas quedan alejados de los anteriores (Kiviluoto, 1996). Por lo tanto se produce una agrupación de los datos según sus características. En el problema analizado esto significa que patrones de entrada con similares características de temperatura, salinidad, viento, etc., se sitúan en áreas cercanas en el mapa resultante.

El número de neuronas necesarias para formar la red varía en función del número de patrones de entrada, pudiendo contener desde unas pocas decenas hasta unos cientos. Cada neurona está representada por un vector de pesos N -dimensional (vector prototipo) definido como $\mathbf{m}_{ij} = [m_{ij}^1, \dots, m_{ij}^N]$, donde N es el número de dimensiones del vector de entrada. Las neuronas que componen la red están conectadas a las neuronas adyacentes mediante una relación de vecindad. Esta relación fija la estructura de la red, pudiendo ser rectangular, hexagonal, cilíndrica o toroidal, siendo la rectangular la estructura más habitual.

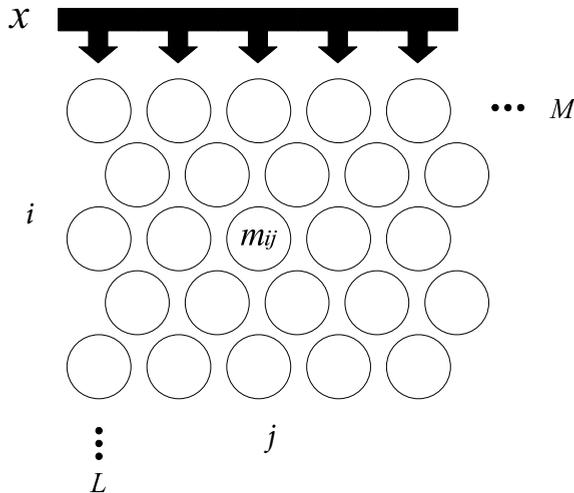


Figura 4.3

Esquema de un SOM con una estructura rectangular de dos dimensiones.

Antes de comenzar el entrenamiento es necesario inicializar los pesos del vector prototipo de las neuronas; esta inicialización usualmente es de tipo aleatoria. En el siguiente paso, el algoritmo de entrenamiento ajusta los valores de los pesos mediante un proceso iterativo. En cada iteración se selecciona un patrón \mathbf{x} de los datos de entrada y se calcula la distancia existente entre el patrón seleccionado y los vectores prototipos del mapa. La medida de distancia habitualmente utilizada es la distancia euclídea. La neurona cuyo vector de pesos es el más cercano al patrón de entrada \mathbf{x} se denomina *Best-Matching Unit* (BMU) denotado como c :

$$\|\mathbf{x} - \mathbf{m}_c\| = \min_{ij} \{\|\mathbf{x} - \mathbf{m}_{ij}\|\} \quad (4.1)$$

donde $\|\cdot\|$ es la distancia utilizada.

Este proceso de ajuste de los pesos es similar a los algoritmos basados en *Vector Quantization*, como el algoritmo de *clustering k-means* (Gray, 1984). La principal diferencia radica en que además de actualizar los vector de pesos de la red, también se actualizan las relaciones topológicas de vecindad, situando la BMU en una posición más cercana al vector de entrada. El proceso de ajuste se muestra de manera gráfica en la figura 4.4.

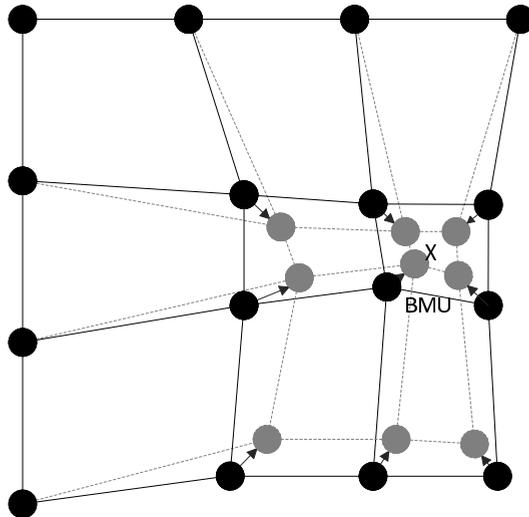


Figura 4.4

Actualización de la BMU y sus vecinos hacia la muestra de entrada \mathbf{x} . Las líneas continuas corresponden a la situación anterior a la actualización mientras que las discontinuas corresponden a la situación posterior.

La regla de actualización del vector de pesos de la unidad ij del SOM se define como:

$$\mathbf{m}_{ij}(t+1) = \mathbf{m}_{ij}(t) + \alpha(t)h_{cij}(t)[\mathbf{x}(t) - \mathbf{m}_{ij}(t)], \quad (4.2)$$

donde t indica el tiempo. El patrón de entrada $\mathbf{x}(t)$, se selecciona aleatoriamente del conjunto de datos de entrada en el tiempo t , $h_{cij}(t)$ es la función de vecindad alrededor de la unidad ganadora c y $\alpha(t)$ es la tasa de aprendizaje. La función de vecindad es una función decreciente (distancia con la neurona ganadora) de la neurona ij . Esta función de vecindad define la región de influencia que el patrón de entrada tiene en el SOM.

El entrenamiento se realiza normalmente en dos fases: en la primera, se asigna una tasa inicial de aprendizaje y un radio de vecindad relativamente grandes. En la segunda fase, se reduce el tamaño de ambos parámetros. Con este procedimiento, en la primera fase se inicializa el SOM acorde a los datos de entrada y en la segunda fase se ajustan los parámetros del mapa. Cuando finaliza el proceso de entrenamiento, es posible visualizar en el mapa de dos dimensiones resultante denominado “*components plane*”. Este mapa permite obtener información cualitativa de la relación que existe entre las variables de entrada del conjunto de datos utilizado en el entrenamiento. Para visualizar la respuesta de los datos de entrada en el SOM se emplea el “*Hits map*”. En este mapa se representa el número de veces que una neurona es el BMU para cada patrón de entrada. Como ejemplo, en la figura 4.5 se muestra el “*Hits map*” correspondiente a la boya 0°N - 0°E. En esta figura se puede observar la distribución de los BMU para el conjunto de datos de esta boya. El área coloreada de cada hexágono es proporcional al número de patrones de entrada que ha sido asignado a esa neurona.

4.3.3. Ajustes de los parámetros de los mapas

Para el entrenamiento, análisis y visualización de los mapas autoorganizados se ha utilizado la librería desarrollada para *Matlab* por el *Laboratory of Information and Computer Science* de la *Helsinki University of Technology* (Vesanto et al., 1999). Los SOMs se implementaron en la versión R2016a de *Matlab* en un servidor con procesador Intel® Xeon® CPU E5530 @2,4GHz x16, 70Gb de memoria RAM y sistema operativo Linux (Ubuntu).

El primer parámetro de ajuste es el número de neuronas de la red (tamaño del mapa). Este tamaño se define comúnmente como $5\sqrt{n}$, donde n es el número de muestras de entrada o patrones (Vesanto et al., 1999). Este tamaño es distinto en cada caso ya que el número de patrones disponible varía según la boya. En el caso de la boya 0°N - 0°E el tamaño del mapa es 14 x 11 neuronas ($5\sqrt{1040} = 161$). El resto de parámetros de entrenamiento se evalúa mediante la combinación de las diferentes posibilidades.

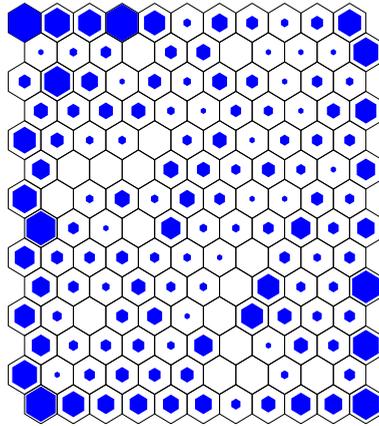


Figura 4.5

Ejemplo de “Hits map” para la boya 0°N - 0°E donde cada neurona se representa mediante un hexágono.

En total, se han entrenado 4008 mapas para cada boya combinando los diferentes tipos de inicialización de los vectores de pesos de las neuronas (*random* y *linear*), con los diferentes algoritmos de entrenamiento disponibles (*batch* y *sequential*) y con cuatro tipos de funciones de vecindad (*gaussian*, *cut gaussian*, *epanechnikov* y *bubble*). El resto de parámetros como el tiempo de entrenamiento, tasa de aprendizaje y el radio de vecindad, se han ajustado siguiendo las recomendaciones de la documentación de la librería del SOM (Vesanto et al., 1999). Una vez obtenidos los mapas entrenados con los diferentes parámetros, es necesario realizar la búsqueda de los mejores mapas o arquitecturas. Generalmente, se utilizan dos tipos de criterio para realizar esta selección (Kohonen, 2001):

- **Error de resolución o cuantización.** Este índice se calcula como la media de las distancias entre cada vector de datos y su BMU.
- **Error de conservación de la topología o error topográfico.** Este criterio indica si se conserva la relación de vecindad entre el espacio original y el espacio final. Se calcula a partir de la proporción de vectores de datos para los que su primer y segundo BMU son unidades no adyacentes.

Con la finalidad de obtener el mapa con el menor error combinando ambos criterios, de los mapas con error topográfico igual a 0, se ha escogido el mapa con menor error de cuantización.

4.4. Resultados

4.4.1. Análisis de la boya 0°N - 35°W

El plano de componentes obtenido para las variables de la boya 0°N - 35°W se muestra en la figura 4.6. Para facilitar la interpretación del análisis realizado, la presentación de los resultados se ha dividido en diferentes apartados, mostrando una ampliación de los mapas y zonas de interés en cada caso. Además, las descripciones de los SOMs se han complementado mediante tablas de estadísticos y gráficas de series temporales relacionadas con las variables analizadas.

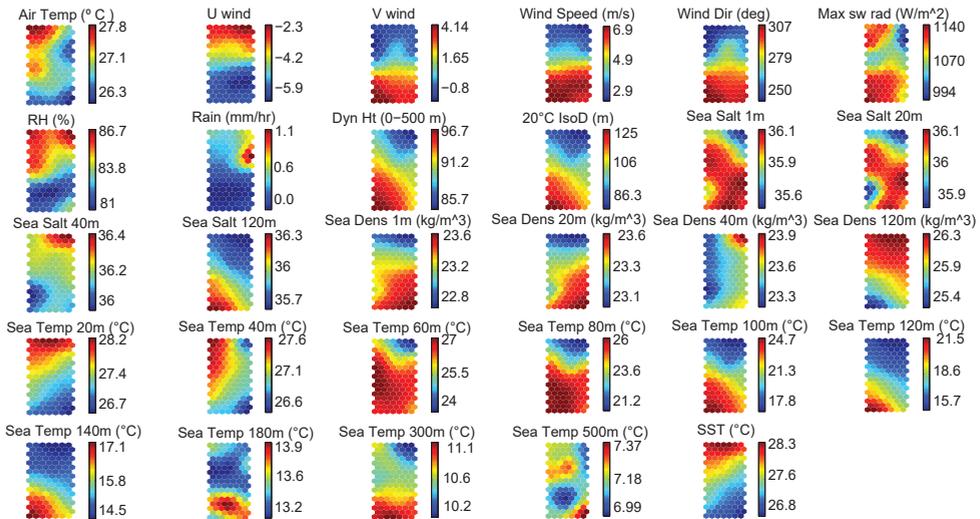
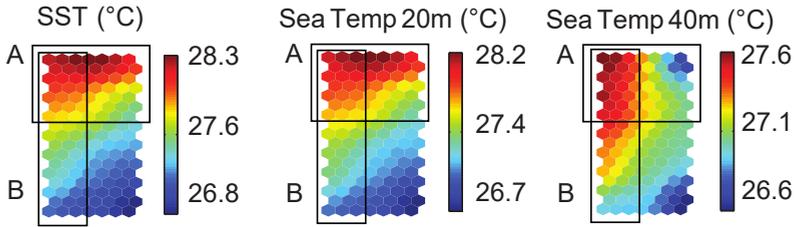


Figura 4.6

SOM de las variables meteo-oceanográficas de la boya PIRATA localizada en 0°N - 35°W. El tamaño del mapa es 17 x 7 neuronas.

4.4.1.1. Análisis de la SST y de la temperatura en subsuperficie

Durante el mes de abril la ITCZ se encuentra en su posición más al sur (Csanady, 2001), dando lugar a una menor velocidad del viento en la zona ecuatorial, hecho que favorece una menor mezcla vertical del viento en la columna de agua y un menor flujo de calor latente (evaporación del agua). Por otra parte, según Ferreira y Mello. (2005), en agosto la ITCZ se encuentra en su posición más al norte, ocasionando vientos alisios del sureste más fuertes a lo largo del ecuador, favoreciendo un flujo de calor latente más fuerte en los trópicos profundos. Estos procesos afectan a la temperatura en la parte superficial de la columna de agua (entre la superficie y 20 m), tal y como reflejan los resultados del SOM que se muestran en la figura 4.7.

**Figura 4.7**

Mapas de las variables SST, *Sea Temp 20m* y *Sea Temp 40m* de la boya PIRATA en $0^{\circ}\text{N} - 35^{\circ}\text{W}$.

Comparando los mapas correspondientes a las variables SST y *Sea Temp 20m*, se observan patrones similares de temperatura (las áreas rojas, amarillas, verdes y azules se encuentran en las mismas regiones de los mapas), evidenciando similitud entre la SST y la temperatura a 20 m de profundidad. Además, esta similitud indica que la capa de mezcla se extiende desde la superficie hasta, al menos, 20 m de profundidad. Cabe aclarar que los nodos rojos indican altas temperaturas, los nodos amarillos y verdes temperaturas intermedias y, los nodos azules, bajas temperaturas. La tabla 4.2 muestra los estadísticos básicos, así como el valor de la correlación entre la SST y la temperatura del mar a varias profundidades. El coeficiente de correlación es de 0,98 entre la SST y la temperatura a 20 m, respaldando lo observado en la figura 4.7.

Tabla 4.2

Estadísticos básicos ($^{\circ}\text{C}$) y correlación de *Pearson* entre la SST diaria y la temperatura en subsuperficie a diferentes profundidades para la boya PIRATA $0^{\circ}\text{N} - 35^{\circ}\text{W}$; DE: Desviación Estándar.

	SST	20 m	40 m	60 m	80 m	100 m	120 m	140 m	180 m	300 m	500 m
Mín.	25,7	25,7	23,9	20,9	16,1	14,7	14,0	12,9	11,3	9,0	6,2
Máx.	29,6	29,1	28,5	27,8	27,5	27,1	25,8	22,9	16,5	13,0	8,2
Media	27,5	27,4	27,1	26,3	24,3	21,5	18,1	15,6	13,5	10,8	7,2
DE	0,6	0,6	0,5	1,1	2,2	2,9	2,3	1,3	0,6	0,6	0,3
<i>Pearson</i>	-	0,98	0,67	-0,20	-0,50	-0,53	-0,45	-0,40	-0,19	-0,26	0,03

La serie temporal de la SST (figura 4.8) muestra una variabilidad anual con valores máximos aproximadamente en abril (cuando la ITCZ está en su posición más meridional) y valores mínimos en agosto (cuando la ITCZ está en su posición más septentrional). Los datos ausentes en esta figura y siguientes se deben a fallos de sensores o vandalismo asociado con actividades pesqueras. Comparando las series temporales de temperatura, se observa que a 40 m de profundidad la serie presenta un comportamiento diferente al de la porción superficial de la columna de agua. Consecuentemente, el mapa SOM de temperatura a 40 m es diferente a los mapas de la temperatura en superficie y en 20 m (figura 4.7). En los mapas de SST y *Sea Temp 20m* el color rojo predomina en la región superior del mapa (zona A), mientras que en esa misma zona del mapa *Sea Temp 40m* se observan nodos de todos los tipos de colores. Además, en la parte izquierda (zona B) de los mapas de SST y *Sea Temp 20m* las áreas de diferentes colores aparecen distribuidos de manera aproximadamente

equitativas, mientras que en la zona B del mapa *Sea Temp 40m*, el color rojo es el predominante. Coincidiendo con esta observación, el valor de correlación entre la SST y *Sea Temp 40m* ha disminuido (0,67) con respecto a *Sea Temp 20m* (0,98).

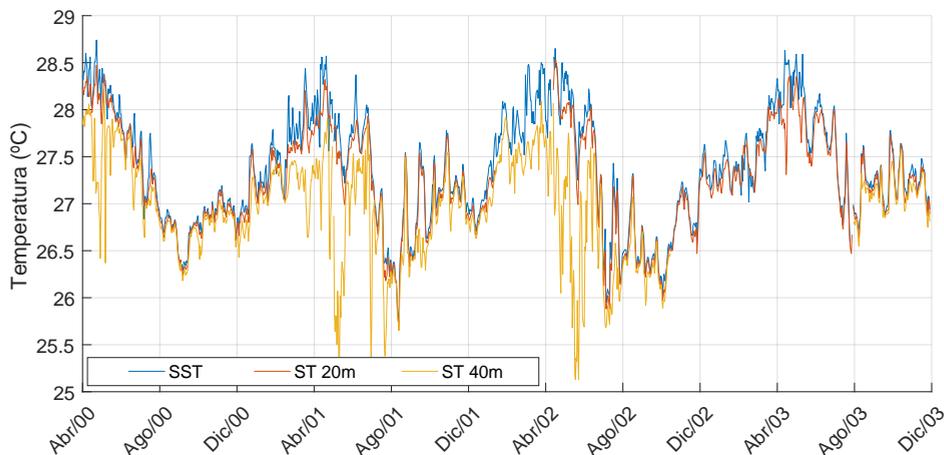


Figura 4.8
Temperatura diaria del océano Atlántico a 1 m (SST), 20 m y 40 m en 0°N - 35°W.

La diferencia entre temperaturas aumenta al profundizar en la columna de agua como muestran las figuras 4.9 y 4.10. A partir de 60 m se amplía la diferencia con respecto a las series temporales de temperatura en superficie y en subsuperficie a 20 y 40 m, con una tendencia a la oposición de fase. Es decir, los valores de temperatura máximos (mínimos) en subsuperficie tienden a coincidir con los valores mínimos (máximos) en superficie. Esto indica que, a partir de 60 m, la temperatura está fuertemente influenciada por la variabilidad del espesor de la capa de mezcla.

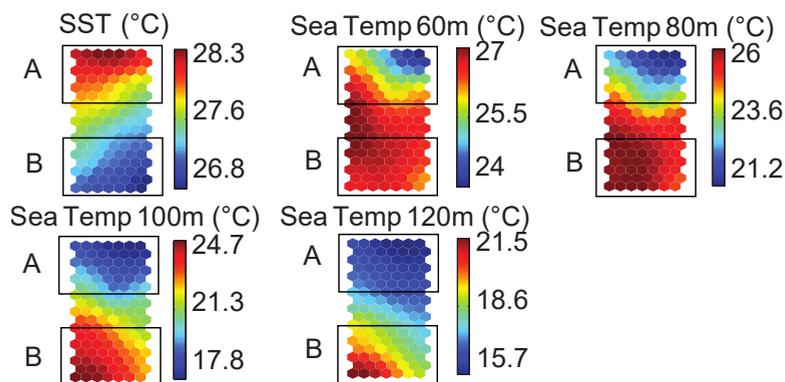


Figura 4.9
Mapas de la SST y temperaturas en subsuperficie de 60 a 120 m para la boya en 0°N - 35°W.

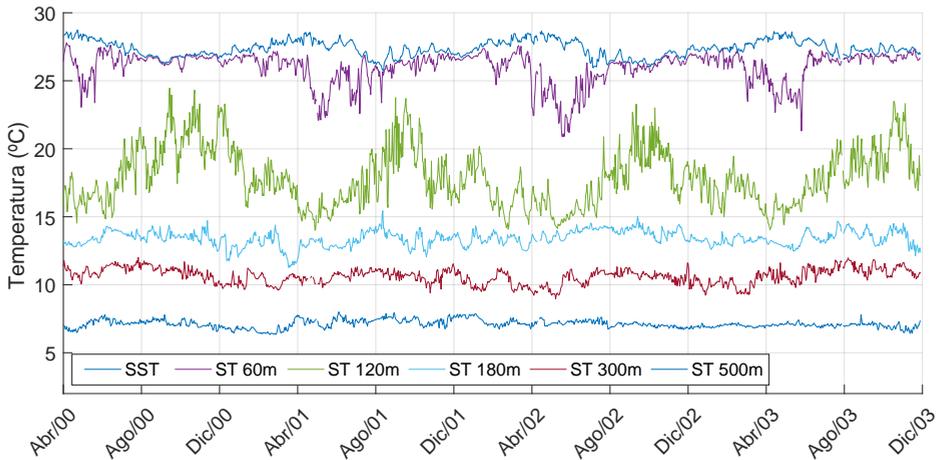


Figura 4.10

Temperatura diaria del océano Atlántico a 1 m (SST) y a varias profundidades en 0°N - 35°W .

A mayores profundidades, sigue aumentando la diferencia de temperatura con respecto a la superficie y, además, esta temperatura presenta un mayor rango, alcanzando los máximos valores de amplitud entre los 100 - 120 m de profundidad. En estas profundidades los valores de la serie temporal de temperatura y la serie de la SST están en oposición de fase según se observa en la figura 4.10 y en los valores de las correlaciones, que ahora son negativos (tabla 4.2). Nuevamente, esta relación se evidencia comparando los mapas SOM resultantes (figura 4.9). Por ejemplo, el área de color rojo (zona A) se encuentra en la región superior del mapa SST, mientras que el área de color azul (zona B) se encuentra en la región inferior. Sin embargo, en los mapas *Sea Temp 100m* y *Sea Temp 120m* estas áreas aparecen invertidas, el color rojo predomina en la región inferior (zona B) y el color azul predomina en la región superior (zona A).

El hecho de que la serie temporal de temperatura a esas profundidades (100 – 120 m) esté aproximadamente en oposición de fase con la SST y tenga una mayor variación de amplitud durante el año, sugiere que la base de la capa de mezcla se encuentra cerca de esas profundidades, pudiéndose encontrar la capa de mezcla o la termoclina dependiendo de la época del año. En los momentos en los que la temperatura a esas profundidades alcanza los valores más altos (septiembre), es probable que se produzca un espesamiento de la capa de mezcla y, a estas profundidades, se encuentre dicha capa. Además, durante esos periodos la SST alcanza los valores mínimos. Por otro lado, cuando la temperatura en esas profundidades es menor (mayo) la SST es alta y es esperable que la capa de mezcla se estreche, encontrándose la termoclina en esas profundidades. Estas observaciones quedan respaldadas por las variaciones anuales de la isoterma a 20°C , utilizada como referencia de la profundidad de la termoclina (figura 4.11), y los perfiles climatológicos de temperatura del Atlas Oceánico Mundial

(*World Ocean Atlas*, WOA; figura 4.12.a) (Locarnini et al., 2013). La profundidad de la isoterma a 20 °C muestra una termoclina menos profunda en mayo y, por tanto, una capa de mezcla más delgada, así como una termoclina más profunda en septiembre y, consecuentemente, una capa de mezcla más gruesa. El perfil de temperatura de la WOA también refleja este hecho, mostrando una capa de mezcla menos profunda en mayo (línea verde discontinua) y más profunda en septiembre (línea verde continua).

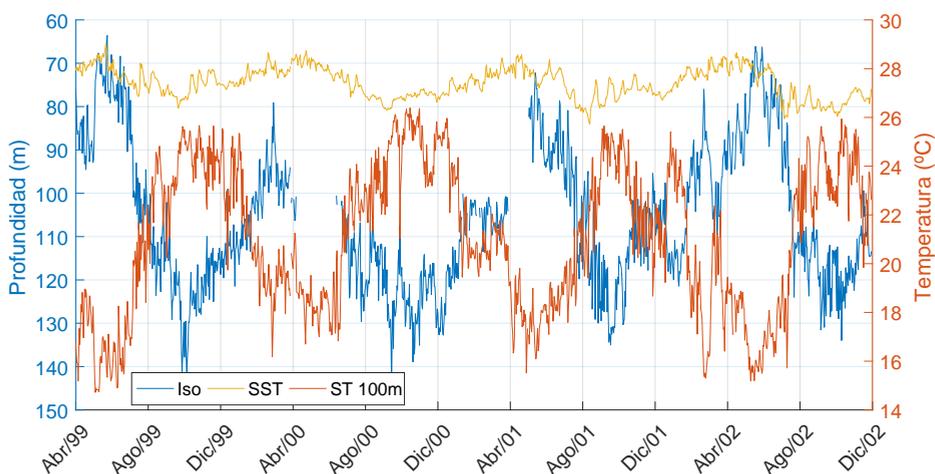
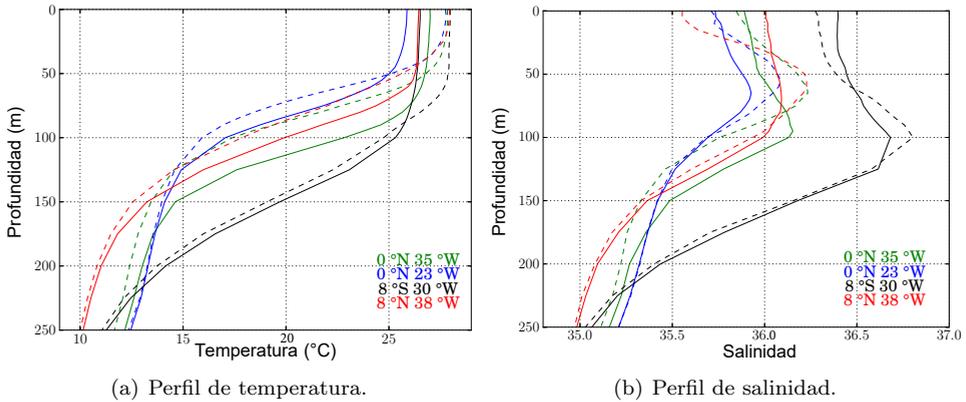


Figura 4.11

Temperatura diaria de la SST y a 100 m en el océano Atlántico (eje derecho) y serie temporal de la profundidad de la isoterma a 20 °C (eje izquierdo) en 0°N - 35°W.

La variabilidad estacional observada en el espesor de la capa de mezcla es consistente con el comportamiento indicado en estudios previos (Bourlès et al., 2008; Rugg et al., 2016). Además, el comportamiento de la temperatura en superficie y subsuperficie en esta localización sugiere un proceso similar al descrito por Yu et al. (2006): una mayor (menor) concentración de calor superficial neto en una capa de mezcla más delgada (ancha) produce una temperatura más alta (más baja) en la superficie del mar. Esto significa que el calentamiento de la superficie del mar no está asociado con el aumento del contenido de calor en subsuperficie relativo a la profundidad de la termoclina (Merle, 1980; Houghton, 1991; Reverdin et al., 1997), sino con el aumento del calor en la capa de mezcla resultante del incremento del flujo de calor neto (Foltz et al., 2013). Esto corrobora los resultados de Niiler y Kraus (2009), quien sugiere que, en ausencia de arrastre, el calentamiento superficial reduce el mezclado cerca de la superficie, generando una estratificación estable en la base de la capa de mezcla y evitando la mezcla con aguas más profundas, modulando la temperatura de esta capa y, consecuentemente la temperatura superficial del mar. En estas regiones gran parte del flujo de calor superficial neto se almacena en la capa de mezcla, desacoplándose de la termoclina. Por lo tanto, los niveles a 100 y 120 m parecen estar cerca de la base de la capa de mezcla, con temperaturas en oposición de fase a la SST.

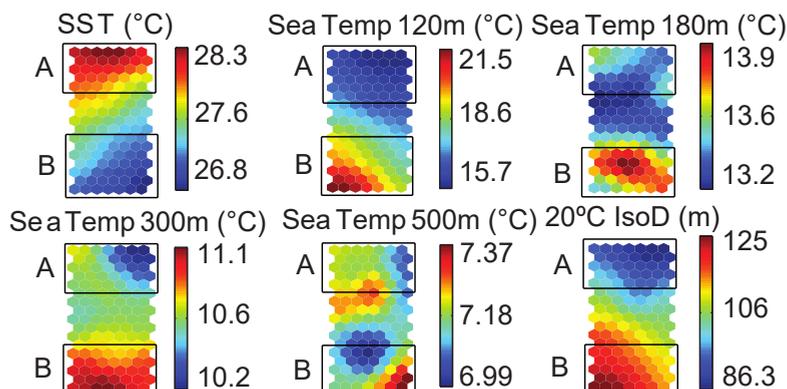
**Figura 4.12**

Perfiles climatológicos (WOA 2013) mostrando los extremos del espesor de la capa de mezcla (a) y la profundidad máxima de salinidad (b) (mínimo: líneas discontinuas; máximo: líneas continuas) para las boyas $0^{\circ}\text{N} - 35^{\circ}\text{W}$, $0^{\circ}\text{N} - 23^{\circ}\text{W}$, $8^{\circ}\text{S} - 30^{\circ}\text{W}$ y $8^{\circ}\text{N} - 38^{\circ}\text{W}$.

La temperatura a 180 m está ligeramente influenciada por la variabilidad del espesor de la capa de mezcla (figura 4.13). En la parte superior de los mapas SOM 20°C IsoD y *Sea Temp 180m* (zona A), predomina el color azul (representando una capa de mezcla más delgada y temperatura más baja), indicando la influencia de la variabilidad en la profundidad de la capa de mezcla en la temperatura a 180 m. Por debajo de 180 m, la amplitud de la variación de temperatura es menor de 1°C (figura 4.10, tabla 4.2), sugiriendo que la termoclina se encuentra por encima de esas profundidades (300 y 500 m) a lo largo del año con una baja influencia de la temperatura superficial. Analizando la figura 4.13, en la parte superior del mapa de SST el color rojo es el dominante (representando valores altos de SST), mientras que en la parte superior de las variables *Sea Temp 300m* y *Sea Temp 500m* predominan los colores azules y verdes (representando temperaturas intermedias y bajas). En la parte inferior (zona B) del mapa SST, predominan los colores azules y, en la parte inferior de *Sea Temp 300m* predominan los colores rojos y amarillos. En los perfiles climatológicos de temperatura de la WOA (Locarnini et al., 2013) (figura 4.12.a) se observa que la media anual de la termoclina se extiende desde los 60 a los 150 m de profundidad en esta localización, respaldando las conclusiones previas.

4.4.1.2. Análisis de la salinidad

En las profundidades muestreadas de la columna de agua, los valores de salinidad oscilan entre 32,5 y 36,9, con el máximo valor de salinidad en 120 m de profundidad (tabla 4.3).

**Figura 4.13**

Mapas de las variables SST, temperaturas a 120 m, 300 m y 500 m e isoterma a 20 °C en la boya PIRATA 0°N – 35°W.

Tabla 4.3

Estadísticos básicos de la salinidad diaria a diferentes profundidades para la boya PIRATA en 0°N – 35°W; DE: Desviación Estándar.

Estadísticos	1 m	20 m	40 m	80 m	120 m
Mín.	32,5	35,3	35,5	35,7	35,4
Máx.	36,4	36,4	36,8	36,5	36,9
Media	35,9	36,0	36,1	36,2	35,9
DE	0,3	0,2	0,2	0,1	0,3

Los resultados SOM de la salinidad cerca de la superficie de la columna de agua (1 m, 20 m) presentan patrones muy similares (figura 4.14: *Sea Salt 1m*, *Sea Salt 20m*), es decir, la distribución de color es similar entre ambos mapas. En la parte superior derecha de ambos mapas predominan los colores azules y en la parte inferior derecha predominan los colores rojos, con mayor variación en la superficie. Al profundizar en la columna de agua aumentan las diferencias con la salinidad en superficie (figura 4.14: *Sea Salt 40m*). La parte superior derecha en los mapas *Sea Salt 1m* y *Sea Salt 20m* presenta colores azules (indicando baja salinidad), mientras que en esa zona del mapa *Sea Salt 40m* predomina el color rojo (indicando alta salinidad). Por otra parte, la zona inferior derecha de los mapas *Sea Salt 1m* y *Sea Salt 20m* presentan colores rojos mientras que *Sea Salt 40m* presenta colores azules. La profundidad con mayor rango de variación y con una notable periodicidad es 120 m (figura 4.15), presentando sus valores más altos aproximadamente en septiembre y los más bajos en mayo. En este caso, los mapas SOM no han permitido establecer una relación clara entre la salinidad en superficie y a 120 m (figura 4.14: *Sea Salt 1m*, *Sea Salt 120m*).

La serie temporal de salinidad a 120 m (figura 4.15) muestra que, a esta profundidad, los valores más altos de salinidad coinciden con los periodos de tiempo en los que la capa de mezcla es más gruesa (agosto). Esto ocurre debido a que la masa de agua

de salinidad máxima (*Maximum Salinity Water*, MSW), una masa de agua vertical situada cerca de la base de la capa de mezcla, puede alcanzar esta profundidad en esta época del año (Stramma et al., 2005; Silva et al., 2009). Los valores más bajos de salinidad a 120 m de profundidad se dan en mayo, cuando la capa de mezcla es más estrecha y, por tanto, la MSW se encuentra cerca de la superficie (entre 80 y 100 m). El periodo de variabilidad de la masa MSW se confirma con la ayuda de los perfiles climáticos de salinidad de la WOA 2013 (Zweng et al., 2013), que muestran que estos valores máximos de salinidad se encuentran cerca de la superficie durante los meses de marzo, abril y mayo (figura 4.12.b). De este modo, puesto que la salinidad a 120 m es mayor cuando la capa de mezcla es más gruesa, lo que ocurre cuando la SST es menor, las variables SST y salinidad a 120 m presentan patrones opuestos (figura 4.14: *Sea Salt 120m*, SST). La región inferior del mapa SST presenta colores azules indicando bajas temperaturas y el mapa *Sea Salt 120m* presenta colores rojos en esa misma zona, indicando alta salinidad.

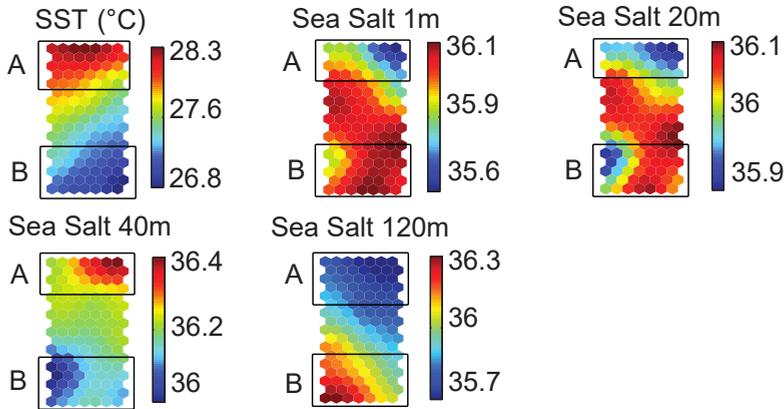


Figura 4.14
Mapas SOM de la SST y salinidad en la boya PIRATA en $0^{\circ}\text{N} - 35^{\circ}\text{W}$.

4.4.1.3. Análisis de la densidad

De manera similar a la salinidad y la temperatura, la densidad presenta patrones muy parecidos en la porción cercana a la superficie (figura 4.16: *Sea Dens 1m*, *Sea Dens 20m*), aumentando las diferencias con la superficie al profundizar en la columna de agua (figura 4.16: *Sea Dens 1m*, *Sea Dens 40m*, *Sea Dens 120m*). Como se esperaba, la densidad en superficie tiene valores más bajos que en subsuperficie tal y como muestran los estadísticos de la tabla 4.4.

En cuanto a la relación entre densidad, temperatura y salinidad, la densidad en los océanos está determinada por la ecuación de estado del agua de mar, que en su versión lineal puede expresarse como (Pedlosky, 1987; Talley, 2011):

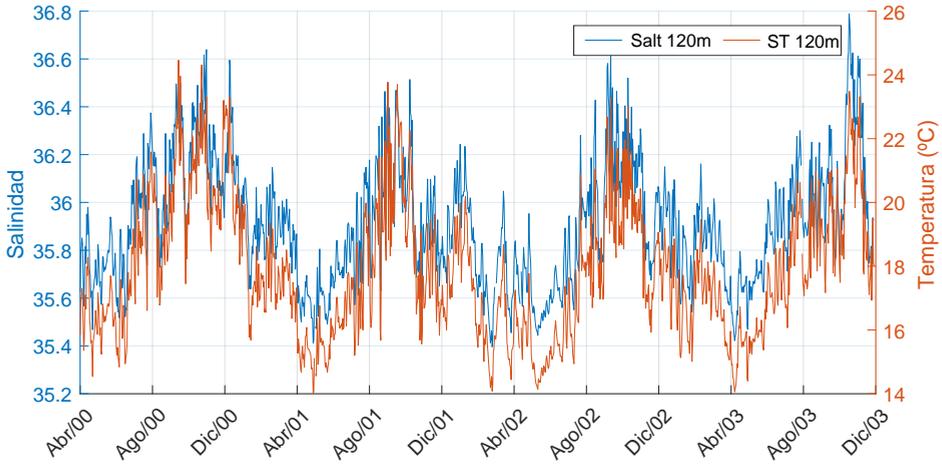


Figura 4.15
Series temporales diarias de salinidad y temperatura en subsuperficie en el océano Atlántico (profundidad 120 m). Boya PIRATA en 0°N - 35°W.

$$\rho = \rho_0 + \rho_0 [1 - \alpha\Delta T + \beta\Delta S] \tag{4.3}$$

donde ρ es la densidad, ρ_0 es una densidad de referencia, α es el coeficiente de expansión termal, ΔT es la variación de temperatura, β es el coeficiente de concentración y ΔS es la variación de salinidad. Por tanto, de acuerdo a la ecuación 4.3, la densidad y la temperatura son inversamente proporcionales, mientras que la densidad y la salinidad son directamente proporcionales. Sin embargo, en general, la temperatura en los océanos es más relevante que la salinidad en la determinación de la densidad, ya que el rango de variación de la temperatura es mayor que el de la salinidad.

Tabla 4.4
Estadísticos básicos: densidad diaria (kg/m^3) a diferentes profundidades para la boya PIRATA en 0°N - 35°W; DE: Desviación Estándar.

Estadísticos	1 m	20 m	40 m	80 m	120 m
Mín.	21,0	22,6	22,8	23,6	24,2
Máx.	23,9	24,1	24,8	26,3	26,5
Media	23,2	23,4	23,5	24,6	26,0
DE	0,3	0,2	0,2	0,7	0,4

Con esto en mente, cerca de la superficie (1, 20 y 40 m) el comportamiento de la densidad es opuesto al de la temperatura (figura 4.16: comparar los patrones SOM de *Sea Dens 1m* con *SST*, *Sea Dens 20m* con *Sea Temp 20m* y *Sea Dens 40m* con *Sea Temp 40m*) y presenta patrones algo similares a los de salinidad en cada profundidad

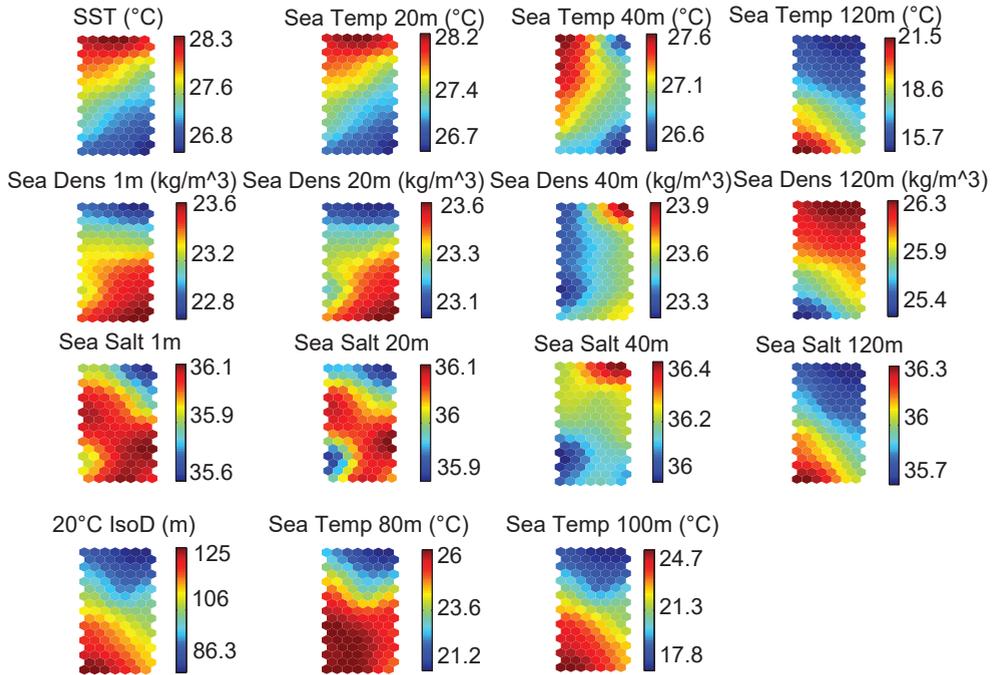


Figura 4.16

Mapas SOM de temperatura, densidad, salinidad y profundidad de la isoterma a 20 °C en la boya PIRATA 0°N – 35°W.

(figura 4.16: comparar los mapas SOM de *Sea Dens 1m* con *Sea Salt 1m*, *Sea Dens 20m* con *Sea Salt 20m* y *Sea Dens 40m* con *Sea Salt 40m*). La fuerte relación con la temperatura está relacionada con el gran rango de variación de temperatura que se produce en esta localización durante el año y, consecuentemente, la densidad tiene una mayor dependencia de la temperatura que la salinidad.

A 120 m de profundidad el rango de temperatura aumenta, alcanzando unos 10 °C, mientras que el rango de variación de la salinidad es de aproximadamente una unidad (figura 4.15), provocando que la temperatura sea todavía más dominante en la variación de la densidad. Por tanto, a 120 m, la densidad es inversamente proporcional a la temperatura y también inversamente proporcional a la salinidad, debido a que el rango de variabilidad de este último es relativamente pequeño (figura 4.16: *Sea Dens 120m*, *Sea Temp 120m*, *Sea Salt 120m*). La distribución de colores en los mapas *Sea Temp 120m* y *Sea Salt 120m* coinciden, existiendo una relación directa, mientras que en el mapa *Sea Dens 120m* esta distribución es opuesta. Cabe recordar que la gran variación de temperatura que se produce a esta profundidad indica la proximidad de la base de la capa de mezcla.

4.4.1.4. Análisis de la profundidad de la isoterma a 20 °C

La profundidad de la isoterma a 20 °C presenta una variación anual de entre 63,6 y 155,3 m (figura 4.17). Los valores más bajos se dan aproximadamente en mayo, mes en el que el viento presenta la menor velocidad y, consecuentemente, la menor capacidad de mezclado. Además, la capa de mezcla presenta su menor espesor y la SST los valores más altos (Yu et al., 2006; Rugg et al., 2016). Así, la capa de mezcla más delgada (más gruesa) está relacionada con el mayor (menor) valor de SST y, como consecuencia, una menor (mayor) densidad superficial. A su vez, los niveles próximos a la base de la capa de mezcla presentan un menor (mayor) valor de temperatura y un mayor (menor) valor de densidad. Comparando los mapas de 20° C IsoD y de Sea Dens 1m, la localización de las áreas rojas y azules coinciden aproximadamente, presentando una posición opuesta en el mapa de la SST (figura 4.16). En 120 m, la posición de los nodos rojos y azules coinciden en los mapas de 20° C IsoD y Sea Temp 120m, siendo opuestos en el mapa Sea Dens 120.

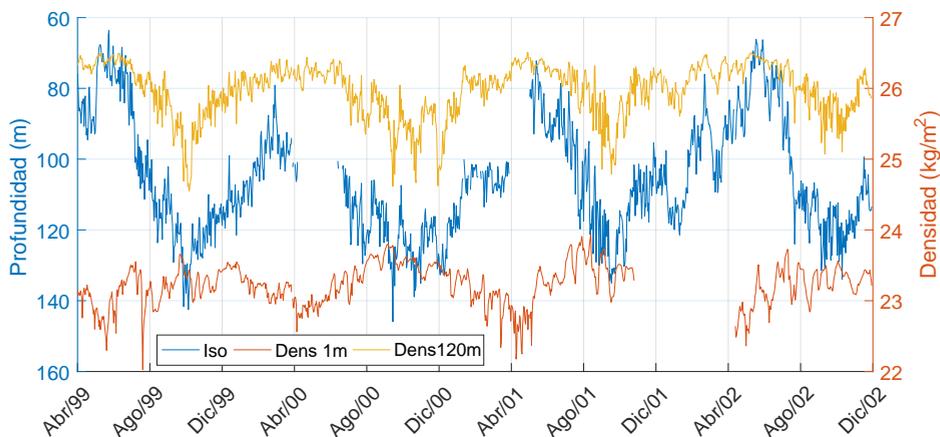


Figura 4.17

Densidad diaria del océano Atlántico (eje derecho) en 0°N – 35°W (profundidades: 1 m y 120 m), y serie temporal de la profundidad de la isoterma a 20 °C (eje izquierdo).

Por tanto, el comportamiento de la profundidad de la isoterma a 20 °C es contrario al de la densidad a 120 m (figura 4.17), lo que confirma que cuando la capa de mezcla es más gruesa se reduce la densidad a 120 m de profundidad. Esto se debe a que cuando la capa de mezcla es más gruesa (aproximadamente en septiembre), la profundidad de 120 m está bajo una mayor influencia de la superficie, aumentando la temperatura a esta profundidad. Teniendo en cuenta que a 120 m la temperatura es el principal determinante de la densidad, esta se ve reducida. Un comportamiento similar se observa en las profundidades 80 y 100 m (figura 4.16).

Con el engrosamiento de la capa de mezcla, la porción de agua cercana a la superficie (1 – 20 m) presenta una menor temperatura y, consecuentemente, una mayor

densidad (figura 4.16: 20°C IsoD , SST, *Sea Dens 1m*). A 40 m de profundidad existe un comportamiento intermedio entre los regímenes de superficie y subsuperficie (figura 4.16: *Sea Temp 40m*, *Sea Dens 40m*, 20°C IsoD). En cuanto a la salinidad, la relación más fuerte con la profundidad de la isoterma a 20°C se encuentra a 120 m, donde el aumento de salinidad está relacionado con un engrosamiento de la capa de mezcla y, por tanto, con una mayor profundidad del núcleo del MSW (figura 4.16: 20°C IsoD , *Sea Salt 120m*).

4.4.1.5. Análisis de la relación entre velocidad del viento, humedad relativa y temperatura del aire

La temperatura del aire, como se esperaba, muestra un comportamiento similar al de la SST y contrario a la velocidad del viento (figura 4.18). Como se discutió anteriormente, los vientos fuertes están relacionados con un flujo de calor latente y mezcla más fuerte y, por tanto, con una disminución de la SST así como de la temperatura del aire. La humedad relativa y velocidad del viento tienen un comportamiento aproximadamente opuesto debido a que el transporte de humedad por el viento reduce la presión de vapor, reduciéndose así la humedad relativa a pesar de la reducción de las temperaturas del aire y del mar (figura 4.18). Por lo tanto, velocidades de viento altas (bajas) y temperaturas bajas (altas) de SST y aire están asociadas con una humedad relativa baja (alta).

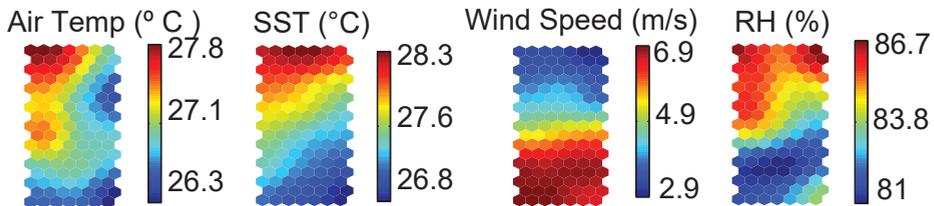


Figura 4.18

Mapas SOM de la temperatura del aire, SST, velocidad del viento y RH para la boya PIRATA en $0^{\circ}\text{N} - 35^{\circ}\text{W}$.

4.4.2. Análisis de la boya $0^{\circ}\text{N} - 23^{\circ}\text{W}$

En esta sección se analizan los resultados del SOM en la boya $0^{\circ}\text{N} - 23^{\circ}\text{W}$ (figura 4.19). Ya que se encuentra en la misma latitud que la boya $0^{\circ}\text{N} - 35^{\circ}\text{W}$, los principales procesos físicos son similares a los anteriormente descritos. Los resultados se centrarán por tanto, en describir las diferencias en los procesos entre ambas localizaciones.

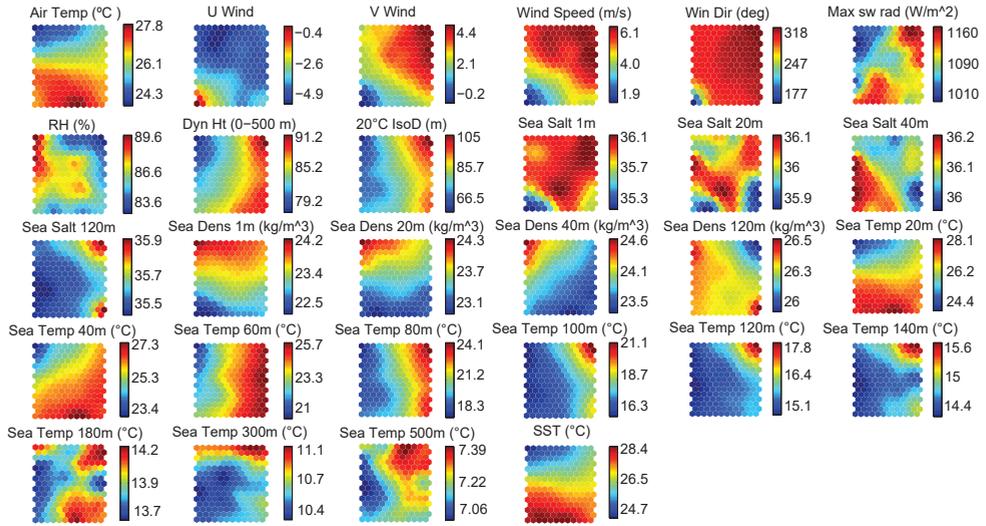


Figura 4.19

SOM de las variables meteo-oceanográficas de la boya PIRATA en $0^{\circ}\text{N} - 23^{\circ}\text{W}$. El tamaño del mapa es 15 x 12 neuronas.

4.4.2.1. Análisis de la SST y temperatura en subsuperficie

En general, en la localización $0^{\circ}\text{N} - 23^{\circ}\text{W}$, el comportamiento de la temperatura es muy similar entre los niveles cercanos a la superficie (SST 1 m - 20 m) como muestran los estadísticos de la tabla 4.5 y los mapas SOM de la figura 4.20. Los valores mínimos se producen aproximadamente en julio, cuando el ascenso de vientos ecuatoriales es más intenso (Servain et al., 1987; Blanke y Delecluse, 1993). Por tanto, existe un desfase entre la temperatura en $0^{\circ}\text{N} - 23^{\circ}\text{W}$, con mínimos valores en julio, y la temperatura en $0^{\circ}\text{N} - 35^{\circ}\text{W}$, donde los mínimos valores de SST ocurren entre agosto y septiembre.

Tabla 4.5

Estadísticos básicos ($^{\circ}\text{C}$) y correlación *Pearson* entre la SST diaria y temperatura en subsuperficie a diferentes profundidades para la boya PIRATA en $0^{\circ}\text{N} - 23^{\circ}\text{W}$; DE: Desviación Estándar.

	SST	20 m	40 m	60 m	80 m	100 m	120 m	140 m	180 m	300 m	500 m
Mín.	23,2	22,6	18,6	16,3	15,7	14,3	13,9	13,0	12,5	9,3	6,4
Máx.	29,8	29,3	28,5	27,2	26,3	24,7	21,1	18,3	15,3	12,5	8,2
Media	26,7	26,6	25,7	23,6	20,6	17,7	15,8	14,8	13,9	10,7	7,2
DE	1,2	1,1	1,1	1,8	2,0	1,5	0,9	0,5	0,3	0,5	0,3
<i>Pearson</i>	-	0,97	0,72	0,10	-0,12	-0,19	-0,16	-0,10	0,08	-0,29	-0,13

A 40 m de profundidad y por debajo, existe un desfase creciente con la serie temporal de SST y, a 60 - 80 m de profundidad, las temperaturas presentan un desfase de tres meses respecto a la SST (figura 4.21). Este desfase es el mayor observado en

esta localización entre las temperaturas en superficie y subsuperficie. Como se puede observar en los mapas SOM de la figura 4.20, este desfase entre las series temporales de temperatura entre diferentes niveles de la columna de agua no representa una relación directa o inversa, implica una rotación en los patrones de los mapas. Las zonas de nodos A y B de la temperatura a 60 m aparecen rotadas 90 grados con respecto a las mismas zonas del mapa de la SST.

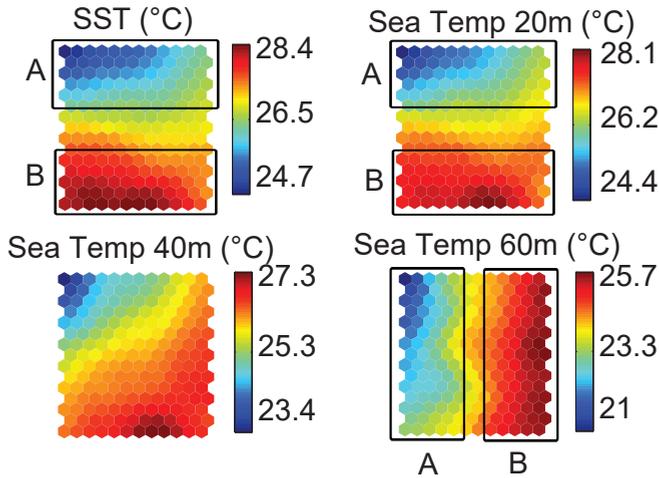


Figura 4.20

Mapas SOM de la temperatura diaria de SST, a 20 m, 40 m y 60 m en $0^{\circ}\text{N} - 23^{\circ}\text{W}$.

En las boyas $0^{\circ}\text{N} - 35^{\circ}\text{W}$ y $0^{\circ}\text{N} - 23^{\circ}\text{W}$ el máximo desfase entre la SST y la temperatura en subsuperficie se produce entre los 100 - 120 m y entre los 60 - 80 m respectivamente (desfase de seis meses en $0^{\circ}\text{N} - 35^{\circ}\text{W}$ y de tres meses en $0^{\circ}\text{N} - 23^{\circ}\text{W}$). Esto indica que la base de la capa de mezcla se encuentra cerca de estas profundidades en estas localizaciones y, además, se produce una disminución del espesor de la capa de mezcla hacia el este a lo largo del ecuador. Esta disminución se menciona en la literatura (Rugg et al., 2016) y se observa en los perfiles climáticos de temperatura, con la base de la capa de mezcla situándose de media a 50 m en $0^{\circ}\text{N} - 35^{\circ}\text{W}$ y a 40 m en $0^{\circ}\text{N} - 23^{\circ}\text{W}$, (figura 4.12.a).

La boya $0^{\circ}\text{N} - 23^{\circ}\text{W}$ presenta un desfase temporal de tres meses entre la SST y la profundidad de la isoterma a 20°C (figura 4.22), que se muestra como una rotación en los mapas SOM (figura 4.23). Las zonas de nodos A y B de la isoterma aparecen rotadas 90 grados con respecto a las mismas zonas de la SST. Esta rotación, considerando que la variabilidad principal presenta un periodo anual, representa un desfase de tres meses. Este desfase es diferente al observado en las series temporales (figura 4.11) y en los mapas SOM (figura 4.13: SST y 20°C IsoD) de la boya $0^{\circ}\text{N} - 35^{\circ}\text{W}$. Las áreas rojas y azules presentan patrones opuestos, indicando una rotación de 180 grados entre ellas, representando un desfase de seis meses.

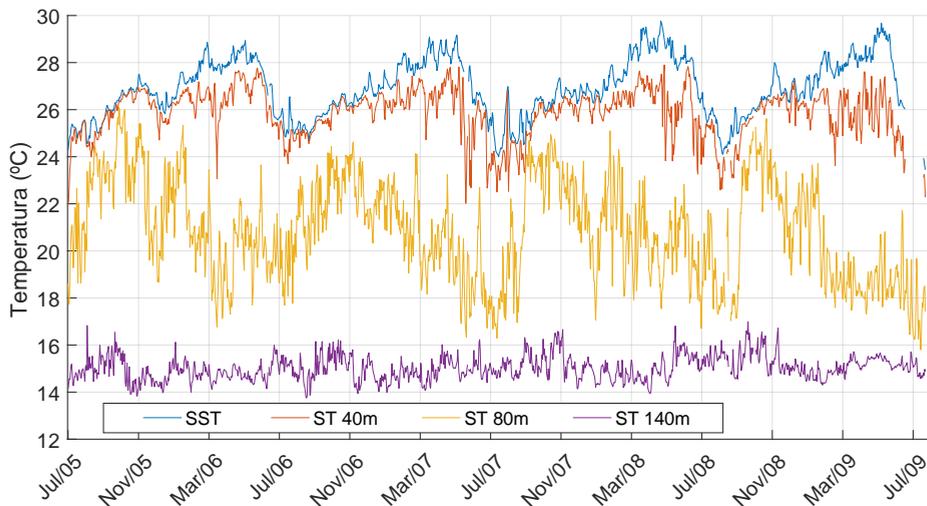


Figura 4.21
Temperatura diaria de la SST y a 40 m, 80 m y 140 m en la boya 0°N - 23°W.

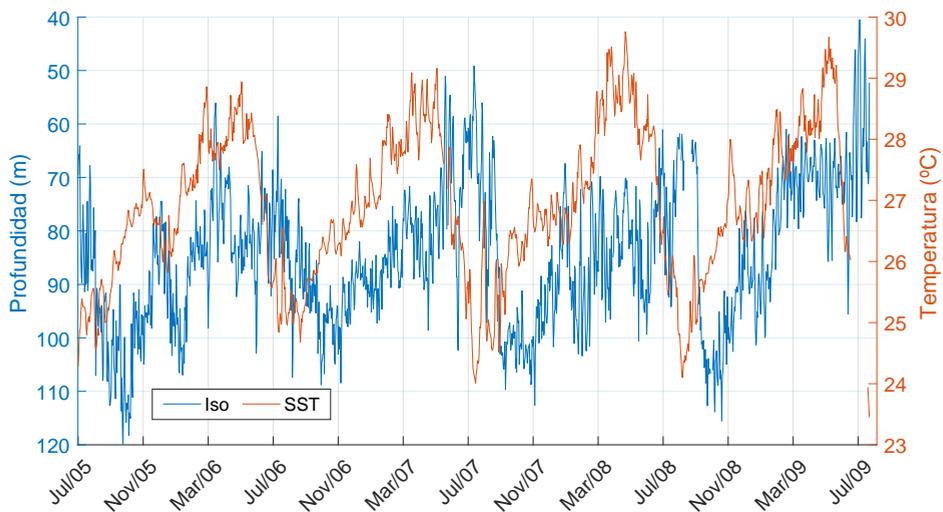


Figura 4.22
Temperatura diaria de la SST (eje derecho) y serie temporal de la profundidad de la isoterma a 20 °C (eje izquierdo) en 0°N - 23°W

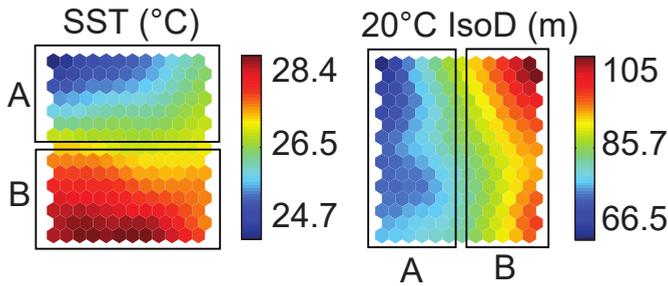


Figura 4.23

Mapas SOM de la temperatura diaria de SST y profundidad de la isoterma a 20 °C en 0°N - 23°W.

Este desfase puede indicar que, en la posición 23°W, la variabilidad de la SST y de la profundidad de la capa de mezcla responde a mecanismos diferentes que en posiciones más al oeste, como en 35°W. En la zona ecuatorial más occidental, los flujos de calor en superficie son más importantes en la determinación de la temperatura superficial y de la profundidad de la capa de mezcla. Con una mayor estratificación en su base, se desacopla de la termoclina, por lo que la SST es mayor cuando la capa de mezcla es más delgada (Niiler y Kraus, 2009). En la parte ecuatorial oriental, la temperatura en superficie se determina principalmente por la advección vertical, por lo que los mayores valores de SST se dan cuando la capa de mezcla es más gruesa (Weingartner y Weisberg, 1991; Yu et al., 2006). La posición 0°N - 23°W es una región intermedia entre esos regímenes, por lo que para la determinación de la temperatura superficial y el espesor de la capa de mezcla, actúan conjuntamente los procesos de interacción aire-mar (flujos térmicos superficiales), así como los procesos oceánicos (ascensión de vientos ecuatorial).

Algunas de las principales diferencias entre las series temporales de las variables en 0°N - 23°W y en 0°N - 35°W están relacionadas con la profundidad de la capa de mezcla y con la mayor importancia de la ascensión de vientos ecuatoriales que tienen lugar hacia el este del ecuador. En las siguientes secciones se comentarán las principales consecuencias de la reducción del espesor de la capa de mezcla.

4.4.2.2. Relaciones entre densidad, temperatura y salinidad

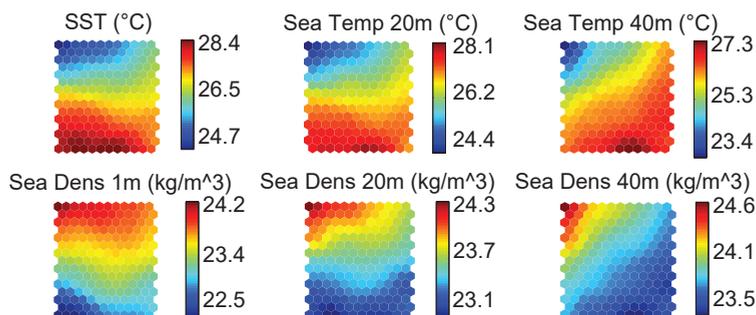
Entre la porción de la columna de agua cercana a la superficie y la base de la capa de mezcla (entre 60 - 80 m), la densidad en la posición 0°N - 23°W presenta un patrón aproximadamente opuesto al de la temperatura (figura 4.24: SST, *Sea Temp 20m*, *Sea Temp 40m*, *Sea Dens 1m*, *Sea Dens 20m*, *Sea Dens 40m*), de manera similar a lo encontrado para la boya 0°N - 35°W. Este comportamiento es el esperado, como se ha mencionado anteriormente, dado que la densidad es inversamente proporcional a la temperatura. Para la salinidad no se ha encontrado una relación clara con la densidad en ninguna profundidad (tabla 4.6, figura 4.25: *Sea Dens 1m*, *Sea Dens*

20m, Sea Dens 40m, Sea Dens 120m, Sea Salt 1m, Sea Salt 20m, Sea Salt 40m, Sea Salt 120m), sugiriendo que en esta localización existen más procesos que intervienen en la variabilidad de la profundidad de la capa de mezcla y, consecuentemente, en la profundidad de máxima salinidad. A 120 m (figura 4.25: *Sea Dens 120m, Sea Temp 120m, Sea Salt 120m*), un alto valor de densidad está asociado con una baja temperatura independientemente de los valores de salinidad y, una baja densidad, está relacionada con una alta temperatura y alta salinidad. Esto indica que para la posición $0^{\circ}\text{N} - 23^{\circ}\text{W}$ y a esta profundidad, dado que el rango de variación de temperatura ($\sim 3^{\circ}\text{C}$) es mayor que el de la salinidad ($\sim 0,5$), la temperatura es dominante en la determinación de la densidad.

Tabla 4.6

Estadísticos básicos, Salinidad diaria / Densidad (kg/m^3) a diferentes profundidades para la boya PIRATA en $0^{\circ}\text{N} - 23^{\circ}\text{W}$; DE: Desviación Estándar.

Estadísticos	1 m	20 m	40 m	60 m	80 m	120 m
Mín.	34,1 / 21,6	35,0 / 22,1	35,4 / 23,0	35,2 / 23,5	35,7 / 23,7	35,4 / 25,3
Máx.	36,5 / 24,7	36,5 / 24,8	36,6 / 25,8	36,8 / 26,2	36,9 / 26,3	36,5 / 27,0
Media	35,9 / 23,5	36,0 / 23,6	36,1 / 23,9	36,2 / 24,6	36,2 / 25,5	35,7 / 26,3
DE	0,3 / 0,5	0,2 / 0,4	0,2 / 0,4	0,2 / 0,5	0,2 / 0,5	0,1 / 0,1

**Figura 4.24**

Mapas SOM de temperatura y densidad hasta 40 m para la boya en $0^{\circ}\text{N} - 23^{\circ}\text{W}$.

4.4.2.3. Análisis de la profundidad de la isoterma a 20°C

La profundidad de la isoterma a 20°C se aproxima a la superficie a lo largo del ecuador en dirección este, variando aproximadamente entre 65 y 150 m en $0^{\circ}\text{N} - 35^{\circ}\text{W}$, así como entre 40 y 120 m en $0^{\circ}\text{N} - 23^{\circ}\text{W}$ (figura 4.11 y figura 4.22). En $0^{\circ}\text{N} - 23^{\circ}\text{W}$, el comportamiento de la profundidad de la isoterma a 20°C es similar al indicado por Blanke y Delecluse (1993), con mínimos valores aproximadamente en junio y máximos en octubre (figura 4.22). Cerca de la superficie, como se ha explicado anteriormente, la temperatura y la densidad presentan patrones opuestos.

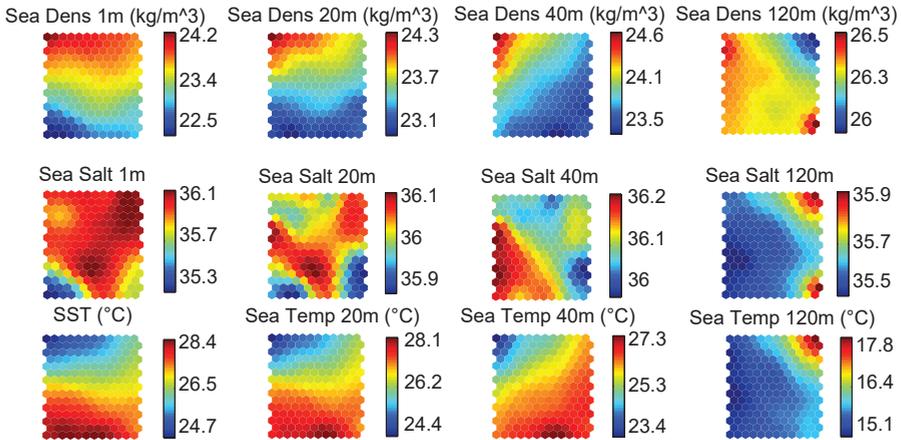


Figura 4.25

Mapas SOM de densidad y salinidad a 1 m, 20 m, 40 m y 120 m, así como temperatura a diferentes profundidades para la boya $0^{\circ}\text{N} - 23^{\circ}\text{W}$.

La variable 20°C IsoD presenta un desfase de tres meses con la temperatura y la densidad, es decir, la termoclina es más gruesa alrededor de tres meses después de que la temperatura (densidad) cerca de la superficie alcance su valor mínimo (máximo). (Figura 4.22, figura 4.26: 20°C IsoD , *Sea Temp 20m*, SST, *Sea Dens 1m*, *Sea Dens 20m*). De nuevo, este retardo sugiere que además de la mezcla vertical promovida por el viento de la región, pueden estar presentes otros factores como la advección vertical, la mezcla vertical turbulenta (Rugg et al., 2016) y el ascenso vertical de vientos (Yu et al., 2006).

En esta localización, la similitud entre los patrones de temperatura a 60 - 80 m y la profundidad de la isoterma a 20°C así como la ausencia de desfase, sugiere que la base de la capa de mezcla se encuentra aproximadamente en dichas profundidades (figura 4.26: 20°C IsoD , *Sea Temp 80m*). Esta relación se debilita al profundizar en la columna de agua y, en general, por debajo de los 140 m la conexión entre la temperatura y 20°C IsoD se reduce, ya que la influencia de la superficie también disminuye (figura 4.19: 20°C IsoD , *Sea Temp 100m*, *Sea Temp 120m*, *Sea Temp 140m*, *Sea Temp 180m*, *Sea Temp 300m*, *Sea Temp 500m*).

En ambas posiciones ($0^{\circ}\text{N} - 35^{\circ}\text{W}$ y $0^{\circ}\text{N} - 23^{\circ}\text{W}$) la serie de salinidad a 120 m presenta un patrón similar a la profundidad de la isoterma 20°C (figura 4.16: 20°C IsoD , *Sea Salt 120m* y figura 4.26: 20°C IsoD , *Sea Salt 120m*). Esto indica que en $0^{\circ}\text{N} - 23^{\circ}\text{W}$, cuando la capa de mezcla es más delgada, el máximo de salinidad se encuentra a profundidades cercanas a 120 m y, cuando la capa de mezcla se ensancha (la profundidad de la isoterma 20°C disminuye), este máximo de salinidad se desplaza hacia mayores profundidades. Estas conclusiones quedan respaldadas por los perfiles climáticos de salinidad de esta localización (figura 4.12.b). De nuevo, la reducción

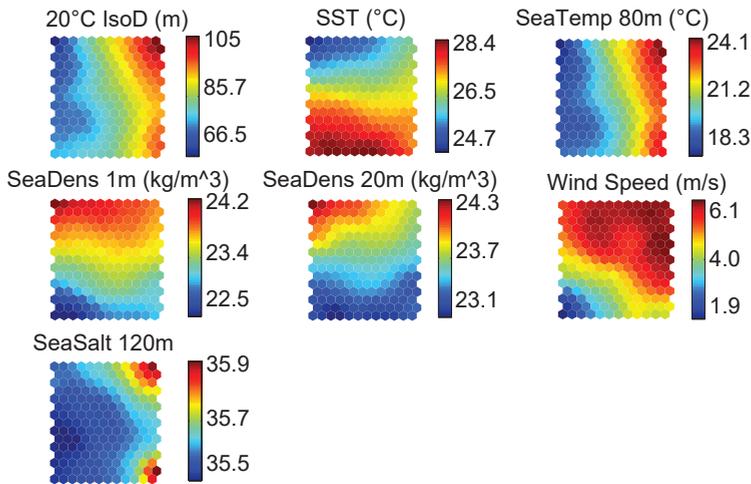


Figura 4.26

Mapas SOM de la isoterma a 20°C , de la temperatura, densidad, salinidad y velocidad del viento para la boya en $0^{\circ}\text{N} - 23^{\circ}\text{W}$.

de la profundidad del valor máximo de salinidad entre las diferentes localizaciones es otra evidencia de la reducción del espesor de la capa de mezcla en la zona este de la región ecuatorial. Para las profundidades cercanas a la superficie no se han podido determinar patrones claros de la variabilidad entre salinidad y 20°C IsoD (figura 4.19: *Sea Salt 1m*, *Sea Salt 20m*, *Sea Salt 40m*, 20°C IsoD).

En la posición $0^{\circ}\text{N} - 35^{\circ}\text{W}$, el comportamiento de la profundidad de la isoterma de 20°C es similar al de la velocidad del viento: cuando la velocidad del viento aumenta, la capa de mezcla se vuelve más gruesa. Sin embargo, para la localización $0^{\circ}\text{N} - 23^{\circ}\text{W}$, la velocidad del viento parece presentar un desfase de tres meses (no se muestra la serie temporal) con respecto a 20°C IsoD . Este desfase se muestra como una rotación relativa entre los mapas SOM de la velocidad del viento y 20°C IsoD (figura 4.26: *Wind Speed*, 20°C IsoD). De nuevo, este desfase sugiere que en esta localización intervienen otros factores en la determinación del espesor de la capa de mezcla además de la variación del flujo neto de calor en superficie y la fuerza del viento.

4.4.2.4. Análisis de la relación entre velocidad del viento, humedad relativa y temperatura del aire

Para la boya localizada en $0^{\circ}\text{N} - 23^{\circ}\text{W}$, a diferencia de la boya en $0^{\circ}\text{N} - 35^{\circ}\text{W}$, en el SOM no se aprecia una relación directa entre la humedad relativa y la velocidad del viento (figura 4.28). En cuanto a la humedad relativa y la temperatura del aire, en $0^{\circ}\text{N} - 35^{\circ}\text{W}$ los resultados del SOM presentaban patrones opuestos (figura 4.18), indicando

la importancia del viento en la reducción de la presión de vapor. Sin embargo, en $0^{\circ}\text{N} - 23^{\circ}\text{W}$ tampoco se ha podido obtener una relación clara entre la humedad relativa y la temperatura del aire, estando en fase sólo una parte de la serie temporal (figura 4.27), lo que indica múltiples procesos en esta localización.

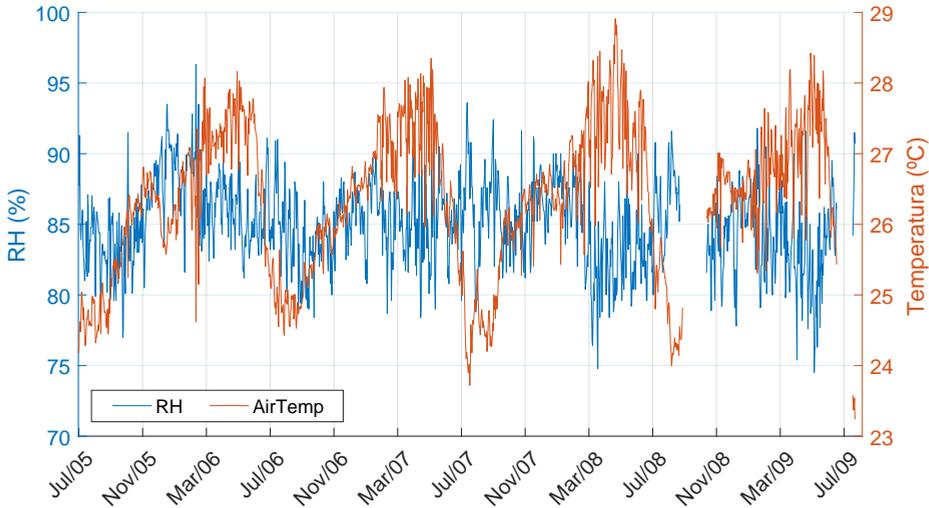


Figura 4.27

Series temporales diarias de la temperatura del aire y humedad relativa en $0^{\circ}\text{N} - 23^{\circ}\text{W}$.

Como se esperaba, el aire y la temperatura del mar en superficie presentan patrones similares y un comportamiento opuesto a la velocidad del viento. Por tanto, cuando la velocidad del viento es alta (baja), se produce un menor (mayor) valor de SST y temperatura del aire (figura 4.28).

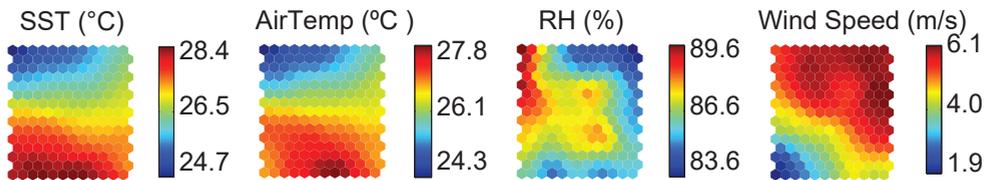


Figura 4.28

Mapas SOM de la SST, RH, temperatura del aire y velocidad del viento en la boya $0^{\circ}\text{N} - 23^{\circ}\text{W}$.

4.4.3. Análisis de las boyas $8^{\circ}\text{N} - 38^{\circ}\text{W}$ y $8^{\circ}\text{S} - 30^{\circ}\text{W}$

Los planos de componentes obtenidos para las boyas $8^{\circ}\text{N} - 38^{\circ}\text{W}$ y $8^{\circ}\text{S} - 30^{\circ}\text{W}$ se muestran en las figuras 4.29 y 4.30.

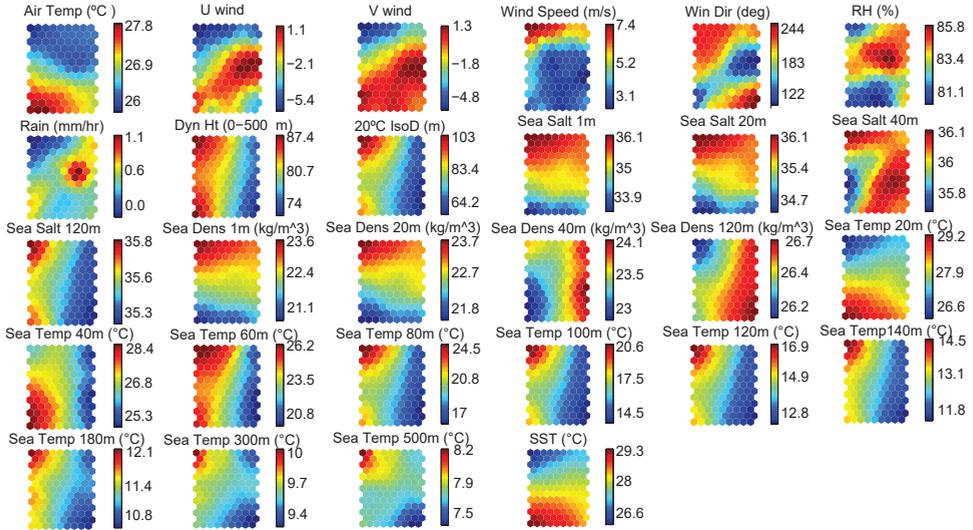


Figura 4.29

Mapas SOM de las variables meteo-oceanográficas para la boya PIRATA en 8°N - 38°W. El tamaño del mapa es 13 x 9 neuronas.

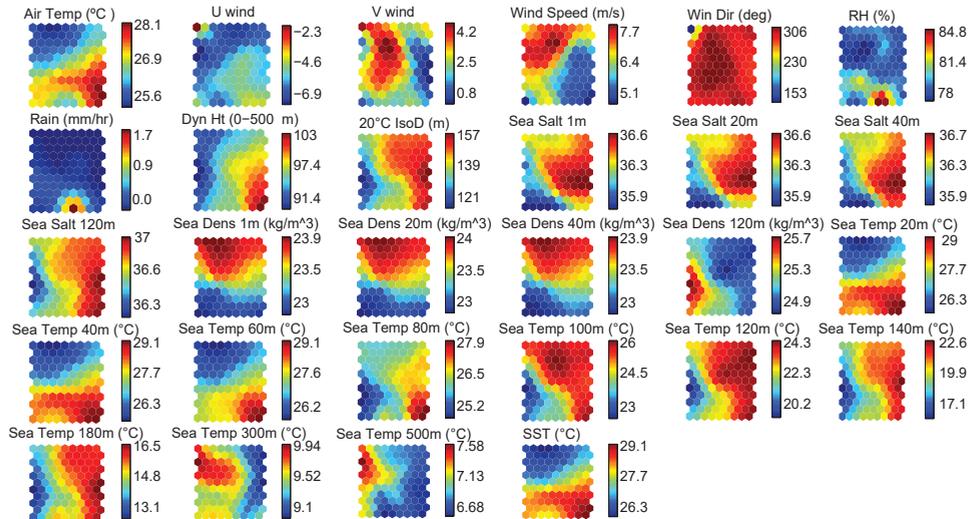


Figura 4.30

Mapas SOM de las variables meteo-oceanográficas para la boya PIRATA en 8°S - 30°W. El tamaño del mapa es 11 x 9 neuronas.

4.4.3.1. Análisis de la SST y temperatura en subsuperficie

Las boyas al norte y sur del ecuador, $8^{\circ}\text{N} - 38^{\circ}\text{W}$ y $8^{\circ}\text{S} - 30^{\circ}\text{W}$ respectivamente, presentan una variabilidad anual de la temperatura en superficie. Los mayores valores de temperatura se producen en septiembre-octubre (abril-mayo) y los menores en abril-mayo (septiembre-octubre) en el hemisferio norte (sur). Este ciclo anual se observa en las series temporales de SST y de temperatura a 20 m en ambas localizaciones (figuras 4.31 y 4.32). Además, en los mapas SOM de ambas boyas se observan patrones idénticos entre el mapa de SST y el de temperatura a 20 m (figura 4.33), con una correlación de 0,99 para la boya $8^{\circ}\text{N} - 38^{\circ}\text{W}$ y de 1 para la boya $8^{\circ}\text{S} - 30^{\circ}\text{W}$ (tabla 4.7). Como se muestra en la figura 4.32, en la boya $8^{\circ}\text{S} - 30^{\circ}\text{W}$, la similitud entre la temperatura en subsuperficie y la SST se extiende hasta los 60 m de profundidad. Este hecho también se aprecia en los mapas SOM (figura 4.33), mostrando patrones muy similares los mapas de SST, *Sea Temp 20m*, *Sea Temp 40m* y *Sea Temp 60m*. La correlación entre la SST y *Sea Temp 40m* presenta un valor de 1, mientras que la correlación entre la SST y *Sea Temp 60m* es de 0,97 (tabla 4.7).

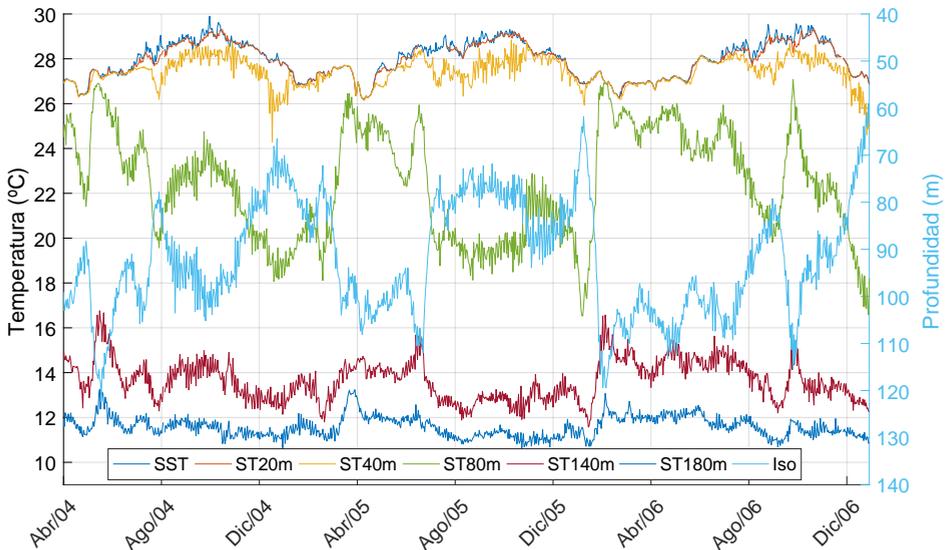


Figura 4.31

Temperatura diaria del océano Atlántico (eje izquierdo) a 1 m (SST), 40 m, 80 m, 140 m, 180 m y serie temporal de la profundidad de la isoterma a 20°C (eje derecho) en $8^{\circ}\text{N} - 38^{\circ}\text{W}$.

Por otra parte, en la boya $8^{\circ}\text{N} - 38^{\circ}\text{W}$ la temperatura a 40 m presenta un comportamiento notablemente diferente al de superficie (tabla 4.7, figura 4.33). Esta diferencia puede deberse a la influencia de la ITCZ, con una migración meridional entre 14°N y 4°S sobre esta localización (Ferreira y Mello., 2005). Los mapas de SST (figura 4.33) muestran un calentamiento más intenso en la localización norte, con un menor número de nodos azules (temperaturas bajas) y un mayor número de nodos rojos

(temperaturas altas) que en la boya sur. Este mayor calentamiento promueve una mayor estratificación, produciendo una capa de mezcla más delgada y un desacoplo entre la superficie y subsuperficie.

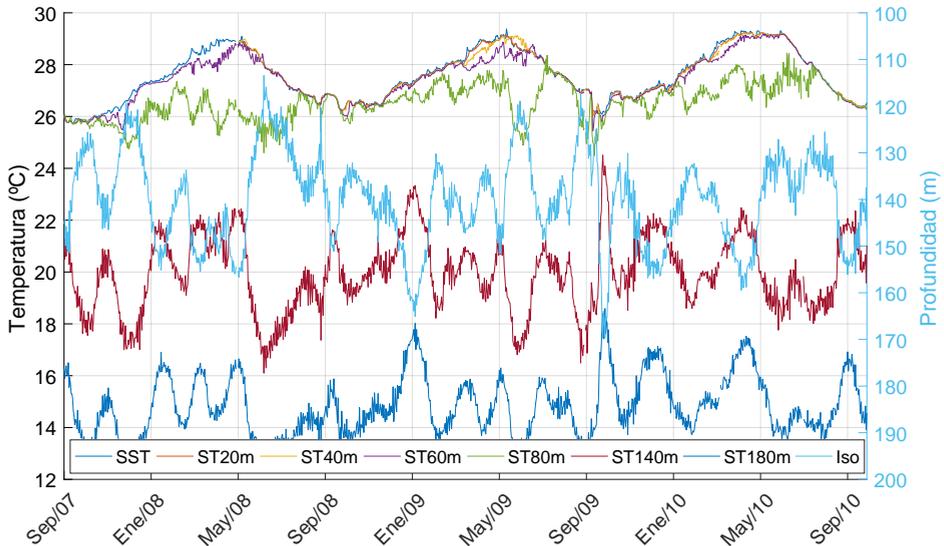
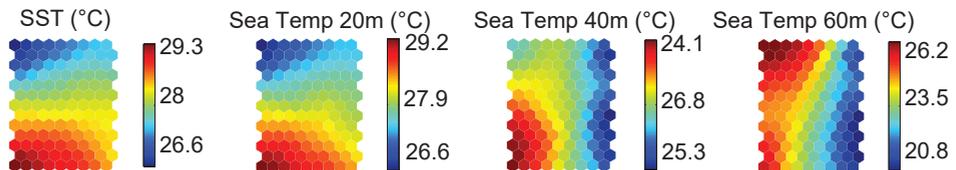


Figura 4.32

Temperatura diaria del océano Atlántico (eje izquierdo) a 1 m (SST), 80 m, 100 m, 140 m y serie temporal de la profundidad de la isoterma a 20 °C (eje derecho) en 8°S - 30°W.

Boya 8°N-38°W



Boya 8°S-30°W

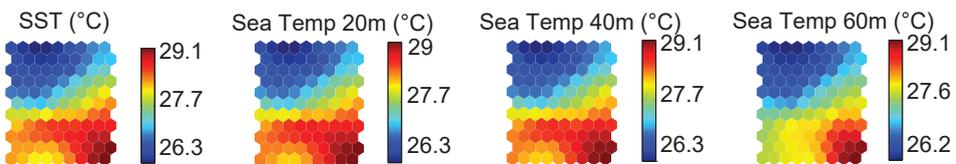


Figura 4.33

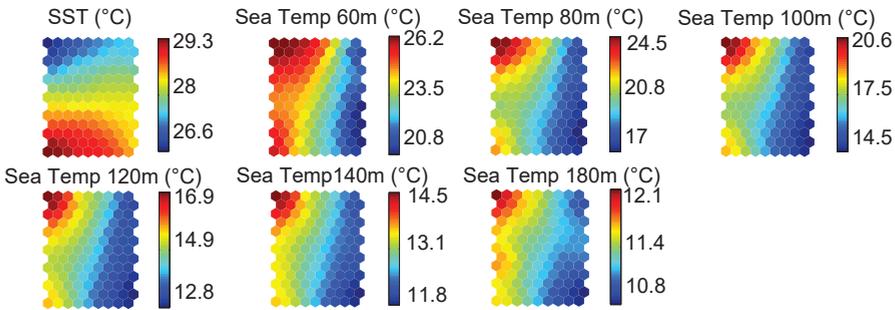
Mapas SOM de las boyas 8°N - 38°W y 8°S - 30°W para las variables (SST) y temperatura en la profundidad 20, 40, y 60 m.

Tabla 4.7

Estadísticos básicos ($^{\circ}\text{C}$) y correlación de *Perason* entre la SST y la temperatura en subsuperficie diaria a diferentes profundidades para las boyas en $8^{\circ}\text{N} - 38^{\circ}\text{W}$ y $8^{\circ}\text{S} - 30^{\circ}\text{W}$; DE: Desviación Estándar.

	SST	20 m	40m	60 m	80 m	100 m	120 m	140 m	180 m	300 m	500 m
$8^{\circ}\text{N} - 38^{\circ}\text{W}$											
Mín.	24,9	24,9	21,5	16,8	14,6	12,4	11,5	10,8	10,1	8,7	7,1
Máx.	30,0	29,8	29,2	28,1	27,8	27,3	27,3	22,8	14,9	10,8	8,7
Media	27,7	27,6	26,9	24,7	21,6	18,0	15,5	13,4	11,5	9,6	7,8
DE	0,9	0,9	1,0	2,2	2,9	2,7	2,0	1,2	0,6	0,3	0,3
<i>Pearson</i>	-	0,99	0,43	-0,15	-0,29	-0,30	-0,29	-0,28	-0,19	-0,18	-0,27
$8^{\circ}\text{S} - 30^{\circ}\text{W}$											
Mín.	25,8	26,0	25,7	25,4	24,2	21,9	19,0	16,1	12,5	8,6	5,7
Máx.	29,5	29,2	29,2	29,2	28,7	27,2	26,1	24,5	18,6	10,5	8,0
Media	27,6	27,6	27,5	27,4	26,7	25,1	23,0	20,3	15,1	9,7	7,1
DE	0,9	0,9	0,9	0,8	0,7	0,9	1,2	1,5	1,2	0,3	0,4
<i>Pearson</i>	-	1,00	1,00	0,97	0,48	-0,11	0,07	0,10	0,19	-0,09	-0,34

A mayor profundidad en la columna de agua, aumenta la diferencia en el comportamiento de la temperatura con respecto a la SST. Para la boya $8^{\circ}\text{N} - 38^{\circ}\text{W}$, entre los 60 y 180 m, las series de temperatura se encuentran en fase (figura 4.31), presentando sus mapas SOM una distribución de colores o patrones muy similares (figura 4.34), que es diferente a la distribución en la SST. La mayor variación en el rango de temperatura se da entre 80 - 100 m y disminuye a mayores profundidades (figura 4.31). Por otra parte, en la boya $8^{\circ}\text{S} - 30^{\circ}\text{N}$ la temperatura también está en fase entre los 100 y 180 m de profundidad, siendo distinta a la de la SST. En este caso no se observa una disminución clara en la amplitud de la variación de temperatura a medida que aumenta la profundidad (figura 4.32 y figura 4.35).

**Figura 4.34**

Mapas SOM para las variables (SST) y temperaturas en las profundidades desde 60 a 180 m en la boya $8^{\circ}\text{N} - 38^{\circ}\text{W}$.

La mayor disminución del rango de variación de temperatura en aguas profundas en la localización norte se asemeja al comportamiento de las boyas ecuatoriales, in-

dicando un mayor desacoplo entre las porciones de agua en superficie y subsuperficie en la posición norte en comparación con la sur. También indica que el espesor de la capa de mezcla difiere entre las diferentes localizaciones, lo que puede confirmarse con los perfiles climatológicos de temperatura y salinidad de la WOA de la región (figura 4.12.a, 4.12.b). Por tanto, las características de la serie de temperatura a lo largo de la columna de agua, combinado con los perfiles climatológicos, muestran que la capa de mezcla es generalmente más delgada en la posición $8^{\circ}\text{N} - 38^{\circ}\text{W}$, encontrándose esta localización bajo la influencia directa de la migración de la ITCZ y de la Contracorriente Ecuatorial Norte.

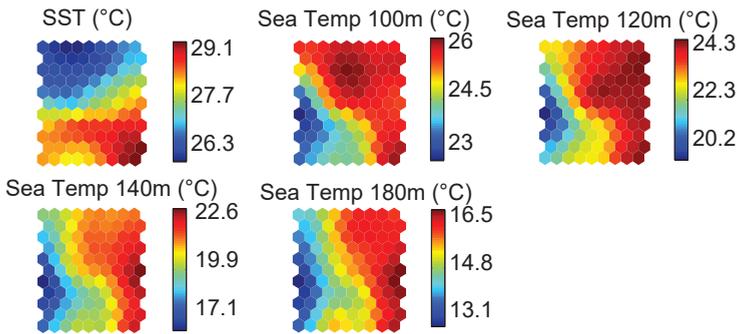


Figura 4.35

Mapas SOM para las variables (SST) y temperaturas en las profundidades desde 100 a 180 m en la boya $8^{\circ}\text{S} - 30^{\circ}\text{W}$.

4.4.3.2. Relación entre densidad, temperatura y salinidad

La salinidad en superficie y a 20 m presentan un mínimo valor aproximadamente en octubre (mayo) para la boya $8^{\circ}\text{N} - 38^{\circ}\text{W}$ ($8^{\circ}\text{S} - 30^{\circ}\text{W}$) Dessier y Donguy (1994). En ambas localizaciones, la variación de salinidad es debida principalmente a la advección producida por el transporte Ekman y el balance de Evaporación-Precipitación [E-P] (Foltz y McPhaden, 2008; Yu, 2011; Tzortzi et al., 2013). Por tanto, en periodos en los que el $E-P > 0$, hay una tendencia al aumento de la salinidad, se produce una mayor velocidad del viento y una mayor advección por el transporte Ekman. Además, la alta variabilidad y los relativamente bajos valores de salinidad registrados en la boya boreal, específicamente entre octubre y noviembre, sugieren que podría existir algún tipo de advección del agua dulce del río Amazonas debido a la Contracorriente Ecuatorial Norte, que es más fuerte en esta época del año.

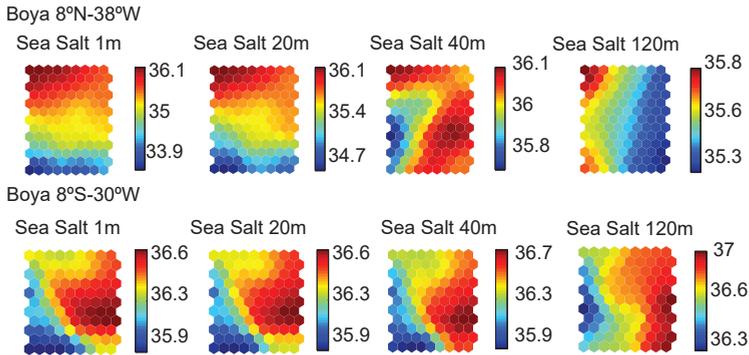
La salinidad en superficie en ambas localizaciones varía entre 31,2 y 37,1, dándose los mayores valores en la boya $8^{\circ}\text{S} - 30^{\circ}\text{W}$ (tabla 4.8). Para esta localización, la serie de salinidad a 40 m es similar a la salinidad a 1 y 20 m (figura 4.36: *Sea Salt 1m*, *Sea Salt 20m*, *Sea Salt 40m*). Sin embargo, esto no ocurre en la posición $8^{\circ}\text{N} - 38^{\circ}\text{W}$, donde existe una gran estratificación vertical de la salinidad en la porción

superior de la columna de agua (figura 4.36: *Sea Salt 1m*, *Sea Salt 20m*, *Sea Salt 40m*). A 120 m, la serie de salinidad en ambas boyas presentan valores entre 35,1 y 37,1, dándose de nuevo los valores más altos en la boya sur (tabla 4.8). A esta profundidad, la variabilidad en la salinidad se debe principalmente a las oscilaciones de alta frecuencia, sin una clara influencia del ciclo anual y, por tanto, difiere del comportamiento que se produce en la porción de superficie de la columna de agua (figura 4.36: *Sea Salt 1m*, *Sea Salt 20m*, *Sea Salt 40m* y *Sea Salt 120m*).

Tabla 4.8

Estadísticos básicos, Salinidad / Densidad (kg/m^3) diaria a diferentes profundidades para la boya PIRATA en $8^\circ N - 38^\circ W$ y $8^\circ S - 30^\circ W$; DE: Desviación Estándar.

	1m	20 m	40m	80 m	120 m
$8^\circ N - 38^\circ W$					
Mín.	31,2/18,9	32,3/20,0	34,8/22,2	35,7/23,6	35,1/23,3
Máx.	36,4/24,2	36,4/24,1	36,5/25,0	36,3/26,2	36,7/26,8
Media	35,4/22,8	35,6/22,9	36,0/23,5	36,0/25,3	35,6/26,4
DE	0,8/0,8	0,5/0,6	0,2/0,4	0,1/0,6	0,2/0,3
$8^\circ S - 30^\circ W$					
Mín.	35,2/22,2	35,4/22,4	35,5/22,6	-/-	36,0/24,4
Máx.	36,7/24,2	36,7/24,1	36,8/24,3	-/-	37,1/25,9
Media	36,3/23,5	36,3/23,5	36,3/23,6	-/-	36,7/25,2
DE	0,3 /0,4	0,2/0,3	0,2/0,3	-/-	0,2/0,2

**Figura 4.36**

Mapas SOM de salinidad para las boyas en $8^\circ N - 38^\circ W$ y $8^\circ S - 30^\circ W$.

En $8^\circ N - 38^\circ W$ la temperatura (salinidad) y la densidad se encuentran aproximadamente fuera (en) fase a 1, 20 y 40 m (figura 4.37: *Sea Dens 1m*, *Sea Dens 20m*, *Sea Dens 40m*, *Sea Salt 1m*, *Sea Salt 20m*, *Sea Salt 40m*, *Sea Temp 1m*, *Sea Temp 20m*, *Sea Temp 40m*). La variación relativamente grande que se observa en el rango de salinidad en superficie, posiblemente debido a la advección de agua fresca de la Contracorriente Ecuatorial Norte, indica que la salinidad juega un importante papel en la determinación de la densidad en esta porción de la columna de agua. A 120 m de profundidad el rango de variación de temperatura ($\sim 4^\circ C$) es mayor que el de la

salinidad ($\sim 0,5$). Por tanto, la densidad, en oposición de fase con la temperatura y la salinidad, viene determinada principalmente por la temperatura (figura 4.37: *Sea Dens 120m*, *Sea Salt 120m*, *Sea Temp 120m*).

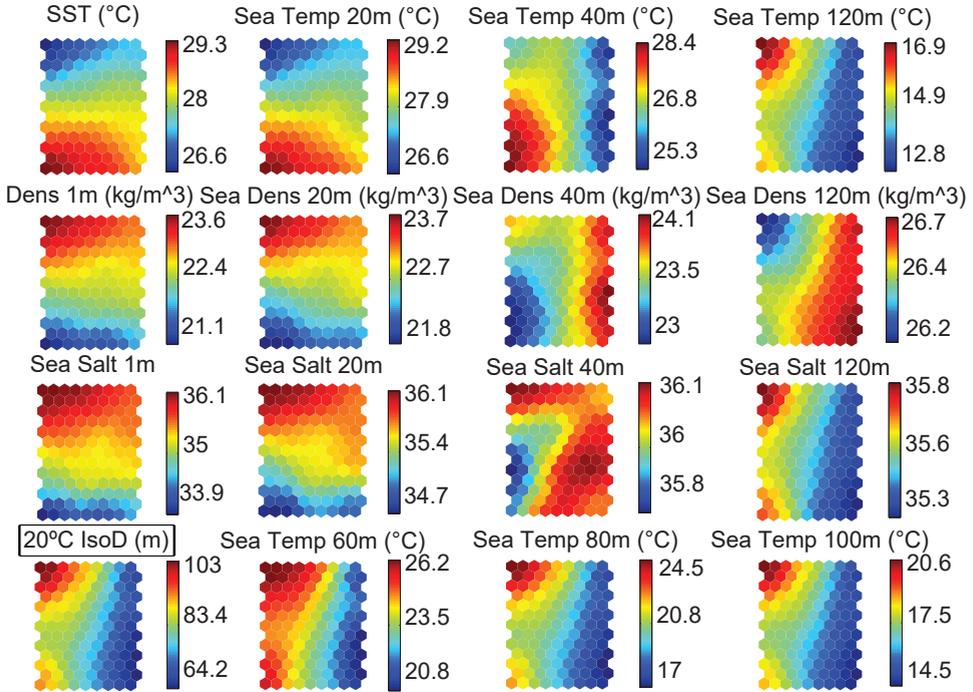


Figura 4.37
Mapas SOM de temperatura, salinidad, densidad y 20°C IsoD en la boya 8°N – 38°W.

En 8°S - 30°W la pequeña variación en el rango de salinidad en superficie, comparado con la localización norte, indica que la temperatura es también el principal determinante de la densidad. A pesar de ello, el hecho de que la densidad no se encuentre exactamente en oposición con la temperatura en los mapas SOM, sugiere que la salinidad también juega un papel importante en la determinación de la densidad (figura 4.38: *Sea Temp 1m*, *Sea Temp 20m*, *Sea Temp 40m*, *Sea Salt 1m*, *Sea Salt 20m*, *Sea Salt 40m*, *Sea Dens 1m*, *Sea Dens 20m*, *Sea Dens 40m*). En 120 m, los altos valores de densidad están asociados con valores bajos de temperatura y salinidad (figura 4.38: *Sea Temp 120m*, *Sea Salt 120m*, y *Sea Dens 120m*), indicando que cuando la capa de mezcla es más delgada, el núcleo de máxima salinidad se desplaza a niveles cercanos a la superficie. Esta observación queda respaldada por los perfiles climatológicos de la WOA en esta región (figura 4.12.a, 4.12.b).

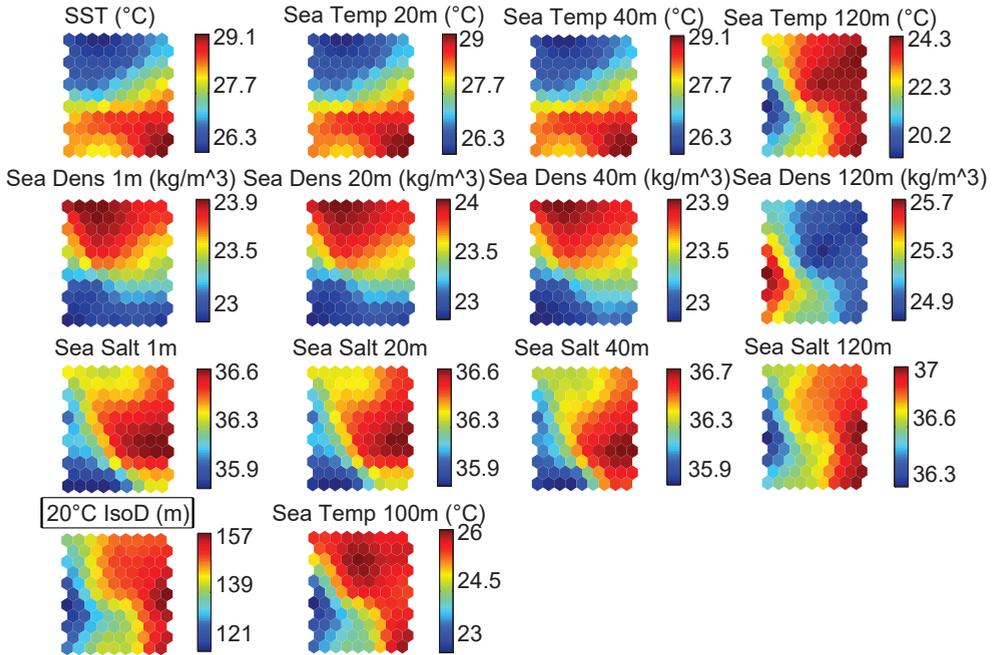


Figura 4.38

Mapas SOM de temperatura, salinidad, densidad y $20^{\circ}C$ IsoD en la boya $8^{\circ}S - 30^{\circ}W$.

4.4.3.3. Análisis de la profundidad de la isoterma a $20^{\circ}C$

La profundidad de la isoterma a $20^{\circ}C$ se encuentra, en general, más cerca de la superficie en la posición $8^{\circ}N - 38^{\circ}W$ (50 – 140 m) que en la posición $8^{\circ}S - 30^{\circ}W$ (115 – 170 m), demostrando que la capa de mezcla es más delgada en la zona norte (figura 4.31 y figura 4.32). En la localización $8^{\circ}N - 38^{\circ}W$, la serie de temporal de la profundidad de la isoterma a $20^{\circ}C$ presenta un ciclo anual, con una capa de mezcla más delgada aproximadamente en noviembre-diciembre y más gruesa aproximadamente en enero-febrero (figura 4.31), como resultado directo de la migración de la ITCZ. Además del ciclo anual, también es posible identificar una oscilación de alta frecuencia (aproximadamente tres meses) y baja amplitud en la serie temporal de la variable $20^{\circ}C$ IsoD, posiblemente relacionada con el flujo de la Contracorriente Norte del Ecuador. Cerca de la superficie (1 y 20 m), la densidad, salinidad y temperatura parecen presentar un desfase de tres meses con relación a la variable $20^{\circ}C$ IsoD (figura 4.37: $20^{\circ}C$ IsoD, Sea Dens 1m, Sea Dens 20m, Sea Salt 1m, Sea Salt 20m, Sea Temp 1m, Sea Temp 20m), de manera similar a lo observado en $0^{\circ}N - 23^{\circ}W$. Esto significa que, cuando disminuye la temperatura de la superficie, la salinidad y la densidad aumenta y, alrededor de tres meses después, la capa de mezcla alcanza su espesor máximo.

Entre los 60 y 120 m, la temperatura está en fase con la variable $20^{\circ}C$ IsoD (figura 4.31 y figura 4.37: *Sea Temp 60m*, *Sea Temp 80m*, *Sea Temp 100m*, *Sea Temp 120m*, $20^{\circ}C$ IsoD) y, a 120 m, la variable $20^{\circ}C$ IsoD está aproximadamente en fase con la salinidad y temperatura, así como en oposición de fase con la densidad (figura 4.37: *Sea Salt 120m*, *Sea Dens 120m*, $20^{\circ}C$ IsoD). Esto indica que cuando la capa de mezcla es más delgada, se observa un aumento de la densidad y temperatura, así como una disminución de la salinidad a esta profundidad, sugiriendo que el núcleo de máxima salinidad se desplaza hacia la parte superior de la columna de agua de manera similar a lo que ocurre en el ecuador. Además, el hecho de que la densidad aumente incluso cuando la salinidad disminuye, indica que la temperatura es el factor determinante en la densidad para esta profundidad. A 40 m de profundidad, la densidad, temperatura y salinidad presentan un comportamiento intermedio y, en los resultados obtenidos, no se ha observado una relación clara entre ellas y $20^{\circ}C$ IsoD (figura 4.37: *Sea Dens 40m*, *Sea Temp 40m*, *Sea Salt 40m*, $20^{\circ}C$ IsoD).

En $8^{\circ}S - 30^{\circ}W$ no se ha podido identificar un patrón anual de oscilación en la serie temporal $20^{\circ}C$ IsoD. Sin embargo, se ha identificado una oscilación de alta frecuencia (figura 4.32) que posiblemente esté influenciada por advección y/u olas de largo periodo. En esta localización ($8^{\circ}S - 30^{\circ}W$), entre la superficie y los 40 m de profundidad no se ha podido establecer una relación clara entre densidad y $20^{\circ}C$ IsoD (figura 4.38: $20^{\circ}C$ IsoD, *Sea Dens 1m*, *Sea Dens 20m*, *Sea Dens 40m*), pero a 120 m existe una oposición de fase (figura 4.38: $20^{\circ}C$ IsoD, *Sea Dens 120m*). Es decir, cuando la capa de mezcla se ensancha, la temperatura (densidad) a 120 m aumenta (reduce), de manera similar a lo observado en la boya $8^{\circ}N - 38^{\circ}W$ (figura 4.37: $20^{\circ}C$ IsoD, *Sea Dens 120m*, *Sea Temp 120m*).

Entre la superficie y los 40 m de profundidad, en $8^{\circ}S - 30^{\circ}W$ no se ha podido establecer una relación entre temperatura y $20^{\circ}C$ IsoD (figura 4.38: *Sea Temp 1m*, *Sea Temp 20m*, *Sea Temp 40m*, $20^{\circ}C$ IsoD), ya que la serie temporal de $20^{\circ}C$ IsoD presenta oscilaciones de alta frecuencia mientras que la de temperatura presenta principalmente oscilaciones anuales (figura 4.32). Sobre los 80 m, la temperatura comienza a presentar oscilaciones de mayor frecuencia, pareciéndose más a la serie de $20^{\circ}C$ IsoD y, entre 100 y 120 m, la salinidad, temperatura y $20^{\circ}C$ IsoD se encuentran en fase (figura 4.38: *Sea Temp 100m*, *Sea Temp 120m*, *Sea Salt 120m*, $20^{\circ}C$ IsoD) presentando oscilaciones de mayor frecuencia. De nuevo, esto sugiere la influencia de advección y/u olas de largo periodo en la región.

4.4.3.4. Análisis de la relación entre velocidad del viento, humedad relativa y temperatura del aire

En la localización $8^{\circ}N - 38^{\circ}W$ no se ha podido encontrar una relación clara entre la velocidad del viento y la profundidad de la isoterma a $20^{\circ}C$ (figura 4.39: *Wind Speed*, $20^{\circ}C$ IsoD), ya que la variable $20^{\circ}C$ IsoD presenta una variabilidad de frecuencia mucho más alta que la velocidad del viento. La SST, por otra parte, está aproximadamente en oposición de fase con respecto a la velocidad del viento (figura

4.39: *Wind Speed*, SST). Es decir, la SST es más alta (nodos rojos) cuando la velocidad del viento es más baja (nodos azules), indicando que el flujo de calor latente en la región es importante en el enfriamiento de las aguas superficiales (Carton y Zhou, 1997). La SST y la temperatura del aire presentan un comportamiento similar, con nodos azules en la mitad superior y nodos rojos en la mitad inferior de ambos mapas (figura 4.39: *Air Temp*, SST), existiendo una mayor variabilidad de la temperatura del aire en los momentos de mayor velocidad del viento. Esta variabilidad es de mayor frecuencia que en $8^{\circ}\text{S} - 30^{\circ}\text{W}$ (no se muestra la serie temporal). La temperatura del aire registrada en $8^{\circ}\text{N} - 38^{\circ}\text{W}$ también se encuentra aproximadamente en oposición de fase con la humedad relativa (figura 4.39: *Air Temp*, RH).

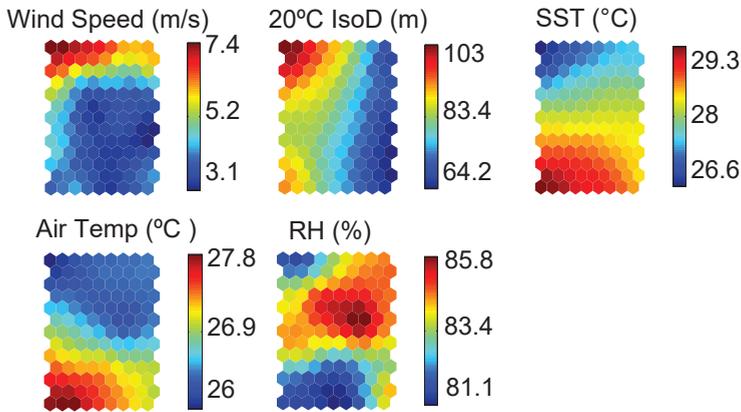


Figura 4.39

Mapas SOM para las variables *Wind Speed*, *20°C IsoD*, SST, *Air Temp* y RH en la boya $8^{\circ}\text{N} - 38^{\circ}\text{W}$.

En la boya $8^{\circ}\text{S} - 30^{\circ}\text{W}$ tampoco se ha podido identificar una relación clara entre la velocidad del viento y la profundidad de la isoterma a 20°C o la humedad relativa (figura 4.40: *20°C IsoD*, RH, *Wind Speed*). Este hecho corrobora la idea de que, en esta región, la variación del espesor de la capa de mezcla puede estar también relacionada con la advección, ondas o arrastre (Rugg et al., 2016). Las series temporales de SST y temperatura del aire están aproximadamente en fase y en oposición de fase con la velocidad del viento (figura 4.40: *Air Temp*, SST, *Wind Speed*), lo que significa que la mayor temperatura del aire y SST ocurre en los momentos de menor magnitud del viento. Esto indica que el flujo de calor latente tiene una gran relevancia en la variabilidad de la temperatura. Los mapas SOM para esta boya tampoco muestran una relación clara entre la temperatura del aire y la humedad relativa (figura 4.40: RH, *AirTemp*).

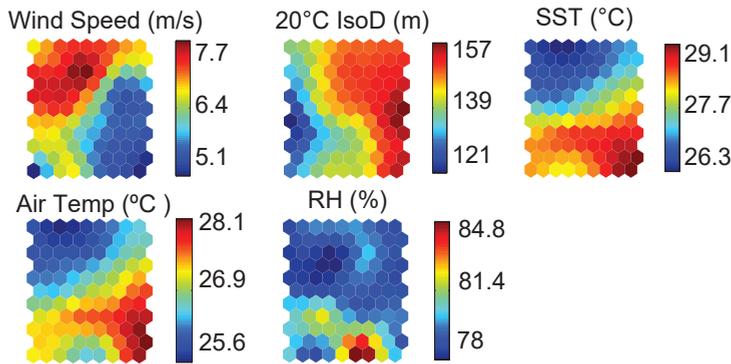


Figura 4.40

Mapas SOM para las variables *Wind Speed*, *20°C ISOD*, *SST*, *Air Temp* y *RH* en la boya 8°S - 30°W.

4.5. Conclusiones

En este capítulo de la Tesis Doctoral se han utilizado los SOMs para el reconocimiento de patrones en los datos meteo-oceanográficos diarios adquiridos por las boyas del proyecto PIRATA. Para el análisis se han escogido cuatro boyas en las coordenadas 0°N - 35°W, 0°N - 23°W, 8°N - 38°W y 8°S - 30°W del océano Atlántico tropical. Los mapas SOM han sido entrenados utilizando todas las variables disponibles en las boyas seleccionadas. Estas variables han sido analizadas simultáneamente, permitiendo visualizar conjuntos de datos complejos. Con ayuda de ciertas series temporales y estadísticas, ha sido posible extraer patrones y relaciones entre variables. Las conclusiones más importantes que se han obtenido mediante el análisis de los mapas SOM han sido:

- **Boya 0°N - 35°W:**

- Los mapas SOM muestran patrones similares entre la temperatura en superficie y a 20 m de profundidad, observándose una relación directa. Esta similitud disminuye al profundizar en la columna de agua.
- A 100 - 120 m se han observado los máximos valores en la amplitud de la temperatura, encontrándose estas series temporales en oposición de fase con respecto a la SST. Esta observación se ha manifestado como una relación inversa en los mapas SOM. Como han demostrado los patrones anuales de la isoterma a 20 °C, esta relación se debe a las fluctuaciones del espesor de la capa de mezcla y de la profundidad de la termoclina.
- Debido a que la termoclina separa las aguas superficiales de las aguas profundas, no se observaron relaciones directas o inversas entre la SST y la temperatura en altas profundidades.

- La salinidad y densidad presentaron un comportamiento similar entre la superficie y a 20 m de profundidad, disminuyendo también esta similitud a mayores profundidades. A 120 m los valores de salinidad coinciden con los de la temperatura, siendo en ambos casos valores opuestos a los de la SST. Por su parte, la densidad mostró patrones opuestos a los de temperatura y salinidad a 120 m.
- La temperatura del aire, la SST y la humedad relativa presentan patrones similares, siendo aproximadamente opuestos a los de la velocidad del viento.
- **Comparación entre la boya 0°N – 35°W y 0°N – 23°W:**
 - De manera similar a lo observado en 0°N – 35°W, en la boya 0°N – 23°W se identificó una relación directa entre los mapas SOM correspondientes a la SST y la temperatura a 20 m.
 - Mientras que en la boya 0°N – 35°W el máximo desfase observado (seis meses) entre la SST y la temperatura en subsuperficie se producía a 100 - 120 m de profundidad, en la boya 0°N – 23°W, el máximo desfase (tres meses) se produce a 60 - 80 m. En la posición 35°W este desfase se ha mostrado como una relación inversa entre los mapas SOM mientras que, en la posición 23°W, esta relación se ha reflejado como una rotación de 90° en los patrones de los mapas.
 - Las relaciones entre la isoterma a 20 °C y los valores de temperatura han mostrado que el espesor de la capa de mezcla disminuye hacia el este en la zona del ecuador.
 - La densidad presenta una relación inversa con la SST, que es similar en ambas localizaciones. Sin embargo, en la posición 23°W no se ha podido establecer una relación clara entre la salinidad y la densidad, ni entre la densidad y la SST, relaciones que sí se observaron en la posición 35°W.
 - A diferencia de lo observado en la boya 0°N – 35°W, en la posición 0°N – 23°W no se aprecia una relación entre la humedad relativa y la velocidad del viento, ni entre la humedad relativa y la temperatura del aire. Sin embargo, en ambas posiciones, los mapas SOM indicaron una relación directa entre la temperatura del aire y la SST, así como un comportamiento opuesto a la velocidad del viento.
 - Las diferencias encontradas entre ambas localizaciones sugieren que, hacia el este del ecuador, la variabilidad de la SST y la profundidad de la capa de mezcla están determinadas por diferentes procesos que en el oeste. Por ejemplo, además de la mezcla vertical promovida por el viento, en posiciones más hacia el este interviene la ascensión de vientos ecuatoriales.
- **Comparación entre la boya 8°N – 38°W y 8°S – 30°W:**

- En la boya $8^{\circ}\text{N} - 38^{\circ}\text{W}$ los patrones de temperatura son muy similares entre la superficie y los 20 m de profundidad, mientras que en la boya $8^{\circ}\text{S} - 30^{\circ}\text{W}$ esta similitud se extiende hasta los 60 m. Este hecho indica que la capa de mezcla es generalmente más delgada en la posición norte que en la posición sur debido, posiblemente, a la influencia de la migración de la ITCZ.
- Debido al mayor espesor de la capa de mezcla, las series de temperatura a 140 - 180 m de profundidad presentan un mayor rango de variación en la posición sur.
- La boya $8^{\circ}\text{S} - 30^{\circ}\text{W}$ presenta mayores valores de salinidad que la boya $8^{\circ}\text{N} - 38^{\circ}\text{W}$. Además, en la posición sur, los patrones de salinidad y densidad a 40 m son similares a los observados a 1 y a 20 m, mientras que en la boya $8^{\circ}\text{N} - 38^{\circ}\text{W}$ esta similitud se da únicamente entre 1 y 20 m. Sin embargo, entre 1 y 20 m, la relación entre temperatura, salinidad y densidad es más clara en la posición norte. A 120 m, en ambas localizaciones se observa una relación directa entre salinidad y temperatura, que es inversa con la densidad.
- No se ha podido establecer una relación clara entre velocidad del viento y la profundidad de la isoterma a 20°C en ninguna de las dos posiciones geográficas.
- La SST y la temperatura del aire presentan una relación directa, que es más clara en la boya $8^{\circ}\text{S} - 30^{\circ}\text{W}$. En ambas localizaciones, esta relación es inversa a la velocidad del viento.

Por tanto, en base al análisis cualitativo presentado en este capítulo, se puede concluir que el SOM tiene un alto potencial en el reconocimiento de patrones y extracción de características en datos meteo-oceanográficos, contribuyendo a una mejor comprensión de los procesos e interacciones oceánicas y atmosféricas.

Capítulo 5

Predicción de la SST mediante técnicas de aprendizaje automático

5.1. Introducción

La predicción de la SST es de gran importancia debido a que juega un papel fundamental en la comprensión y predicción del clima a nivel regional y global. La SST es uno de los indicadores climáticos con mayor influencia en los dos modos de variabilidad del sistema océano-atmósfera del océano Atlántico (modo ecuatorial y modo dipolo). Por una parte, las fases calientes del modo ecuatorial se caracterizan por unas temperaturas anómalamente altas de la SST, mientras que en los eventos fríos, la SST es anómalamente baja. Por otra parte, en el modo dipolo se producen variaciones en la SST entre los hemisferios norte y sur relacionadas con el desplazamiento de la ITCZ. Normalmente la ITCZ migra estacionalmente de su posición más al norte, aproximadamente 14°N entre agosto y septiembre, hacia posiciones más al sur $2^{\circ} - 4^{\circ}\text{S}$ aproximadamente entre febrero y abril. Este desplazamiento de la ITCZ está relacionado con los patrones de la SST sobre la cuenca del océano Atlántico tropical (Servain et al., 2000; Ferreira y Mello., 2005). La variabilidad de la SST que se produce en el modo ecuatorial tiene un alto impacto en el clima, afectando de manera económica y social a la región. Además, la oscilación de la SST en el modo meridional está fuertemente relacionada con el régimen de lluvias en el Nordeste de Brasil y en la región Oeste de África (Moura y Shukla, 1981; Servain, 1991).

El problema de las predicciones de la SST en los océanos Pacífico, Índico y Atlántico ha sido estudiado en numerosos trabajos. Por ejemplo, en el estudio de Tangang et al. (1997) utilizan redes neuronales de tipo SLFN (*Single Layer Feedforward Net-*

work) para obtener predicciones de anomalías mensuales de SST en la región del fenómeno ENSO del océano Pacífico ecuatorial. Los modelos son entrenados utilizando las primeras siete componentes de la descomposición EOF del estrés del viento. En este trabajo los autores concluyen que las redes neuronales tienen ventaja sobre la regresión lineal múltiple en las predicciones a largo plazo (de 9 a 12 meses). En trabajos posteriores, los mismos autores utilizan diferentes estrategias de predicción y comparan el rendimiento de las redes neuronales con modelos lineales (regresión lineal y análisis de correlación canónica) (Tang et al., 2000; Wu et al., 2006). En el primero de los trabajos utilizan las componentes EOF de las variables SLP y SST como predictores, no encontrando diferencias significativas en las predicciones de los modelos lineales y no lineales. En el segundo estudio emplean las componentes principales de SLP y de anomalías de SST para predecir la primera componente principal de la anomalía de SST. En este caso, obtienen mejores resultados con las redes neuronales en predicciones superiores a 6 meses. También en el océano Pacífico, los autores Aguilar-Martinez y Hsieh (2009) estudian las predicciones de anomalías de la SST mediante modelos no lineales (redes neuronales bayesianas y SVM) y lineales, utilizando como entradas las componentes principales de SLP y SST. Como en los trabajos anteriores, los modelos no lineales obtienen mejores predicciones en horizontes superiores a 6 meses. Además, no se encontraron diferencias de rendimiento comparando las redes neuronales bayesianas y modelos SVM. En todos los trabajos mencionados se utilizan datos proporcionados por satélites.

En el estudio de Breaker y Brewster (2009) se utilizan los modelos lineales autorregresivos ARIMA (*AutoRegressive Moving Average*), ARX (*Autoregressive Exogenous Inputs*) y ARMAX (*Autoregressive Moving Average Exogenous Inputs*) para predecir la SST en la bahía de Monterrey, situada en el Pacífico Norte. Los valores diarios de SST registrados en una estación marina se utilizan como entradas a los modelos, mientras que las salidas son los valores de SST registrados por dos boyas fijas. De los modelos comparados, ARMAX obtiene los mejores resultados.

En el océano Índico Norte, los autores Patil et al. (2013) realizan predicciones de SST mensuales mediante dos tipos de redes neuronales: SLFN y RNN (*Recurrent Neural Network*). En este caso, únicamente se utilizan muestras pasadas de SST para obtener las predicciones. Los experimentos realizados demuestran que la red neuronal autorregresiva obtiene mejores resultados que la red neuronal tradicional. En un estudio posterior (Patil et al., 2016), los mismos autores obtienen predicciones diarias, semanales y mensuales de SST utilizando redes neuronales. Como novedad, combinan los datos de SST con datos obtenidos por modelos numéricos de predicción para entrenar las redes neuronales, consiguiendo de esta manera mejorar los resultados.

En cuanto a estudios en el área de interés de esta Tesis Doctoral, el océano Atlántico, en los trabajos de Lins et al. (2010, 2013) se obtienen las predicciones de SST utilizando los datos de las boyas PIRATA y modelos basados en máquinas de vectores soporte. En el primer estudio (Lins et al., 2010) utilizan el modelo SVM en combinación con el algoritmo *Particle Swarm Optimization* (PSO) en los datos diarios de la boya 14°S - 32°W. Los modelos se entrenan utilizando la SST actual (y_t) para

obtener la SST futura con un horizonte de un día (y_{t+1}). En el trabajo posterior (Lins et al., 2013), se utiliza el modelo SVM y los datos diarios de SST en las boyas 8°N - 38°W y 10°S - 10°W. En este estudio el horizonte de predicción es de 365 días y, en lugar de utilizar la SST actual (y_t) para obtener la SST futura, se emplea el promedio de los valores de SST en los años anteriores ($y_t = (y_t + y_{t-365} + y_{t-2*365} + \dots)/n$ años). Además, también se comparan los resultados que se obtienen al diferenciar las entradas (primera y segunda derivada). El estudio concluye que aplicando la primera derivada se obtienen los peores resultados, mientras que utilizando las series originales y aplicando la segunda derivada, se obtienen resultados similares. Finalmente, el estudio de (Salles et al., 2016) evalúa las predicciones de SST mediante modelos ARIMA y aplicando agregación temporal. Esta metodología consiste en obtener predicciones de baja frecuencia (semanales o mensuales) a partir de las predicciones de alta frecuencia (datos diarios). Los resultados de estos experimentos indican que, en general, se obtienen mejores predicciones semanales y mensuales a partir de los datos de baja frecuencia que utilizando los datos diarios.

La revisión bibliográfica realizada pone de manifiesto que existen diferentes estrategias para resolver el problema de las predicciones diarias y mensuales de la SST. Mientras que unos autores optan por utilizar la variable estrés del viento o la variable SLP junto a la SST como predictores, otros autores utilizan únicamente la SST actual para predecir los valores futuros de la SST. En cuanto a los modelos de predicción, la literatura recoge el uso de modelos de regresión lineal, modelos autorregresivos, redes neuronales y SVM. En comparación a lo expuesto anteriormente, en este capítulo de la Tesis Doctoral se propone una comparativa más exhaustiva sobre distintas estrategias y métodos de aprendizaje automático para obtener las predicciones de la SST en el océano Atlántico.

5.2. Objetivos

El objetivo de este capítulo de la Tesis Doctoral es obtener predicciones diarias y mensuales de la SST utilizando los datos procedentes de la red de boyas del proyecto PIRATA. Con la finalidad de obtener las mejores predicciones posibles, se va a realizar una comparativa de diferentes técnicas de extracción y selección de características, así como de una gran variedad de modelos de aprendizaje automático. El desglose de los objetivos de este capítulo son:

- Analizar los efectos de la tendencia y la componente estacional de las series temporales de SST en la precisión de las predicciones.
- Reducir la dimensionalidad de los conjuntos de datos y descubrir los predictores con mayor impacto en las predicciones de la SST utilizando métodos de extracción y selección de características. Para ello se utilizará la PCA, el algoritmo *RReliefF* y el Análisis de Correlación (AC).

- Obtener predicciones diarias y mensuales de la SST en las diferentes boyas haciendo una revisión del estado del arte en cuanto a modelos de predicción basados en aprendizaje automático. Además, en el caso de las predicciones mensuales, estos modelos se compararán con el sistema de pronóstico climático NCEP-CFSv2. Los modelos evaluados son los siguientes:
 1. Modelo trivial (*Naïve model*).
 2. Media diaria/mensual o climatología.
 3. Regresión lineal múltiple robusta (*Robust Multiple Linear Regression*, RMLR).
 4. Perceptrón multicapa (*Multilayer Perceptron*, MLP).
 5. Máquinas de aprendizaje extremo (*Extreme Learning Machine*, ELM).
 6. Redes neuronales de regresión generalizada (*Generalized Regression Neural Networks*, GRNN).
 7. Árboles de regresión aleatorios (*Random Forest Bagging Trees*, RFBT).
 8. Máquinas de vectores soporte en su versión ajustada por mínimos cuadrados (*Least Squares - Support Vector Machines*, LS-SVM).
 9. RNN con unidades LSTM (*Long Short-Term Memory*, LSTM).
 10. Modelo de pronóstico climático NCEP-CFSv2.
- Analizar y comparar el rendimiento de los diferentes modelos en función de las variables de entrada utilizadas para obtener las predicciones.

Puesto que en este capítulo se va a presentar una gran cantidad de experimentos, en la figura 5.1 se muestra un esquema sobre la exposición de los resultados. En primer lugar se describirán los resultados de los experimentos preliminares basados en el análisis de la tendencia y la componente estacionaria. Posteriormente se evalúa el comportamiento de los modelos trivial y climatología en función del horizonte de predicción. Seguidamente se detallarán los resultados de los experimentos realizados utilizando los datos diarios. En el esquema se muestra el orden en el que se exponen los resultados obtenidos mediante las diferentes variables de entrada a los modelos. En cuanto a los modelos mensuales, se seguirá el mismo orden de exposición que en los modelos diarios, pero los resultados se explicarán de manera más resumida.

Como se detallará posteriormente, los patrones con registros ausentes se eliminarán o se sustituirán dependiendo del caso. En los experimentos con las redes recurrentes LSTM, los datos ausentes serán remplazados por valores promedio. Puesto que los conjuntos de entrada no son iguales a los utilizados en el resto de pruebas, los resultados obtenidos con LSTM se mostrarán en una sección individual. Por último se llevará a cabo un análisis estadístico de los resultados para comprobar si existen diferencias significativas en el rendimiento de los modelos evaluados.

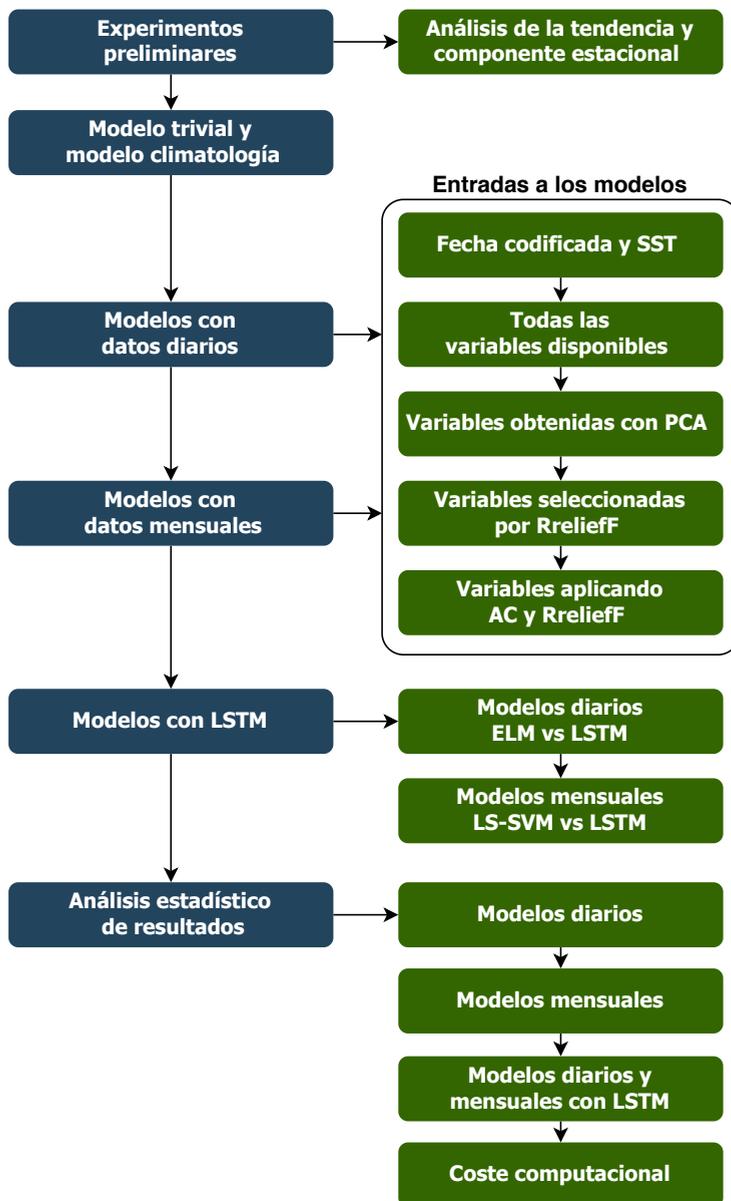


Figura 5.1

Diagrama de la presentación de los resultados obtenidos.

5.3. Materiales y métodos

5.3.1. Datos utilizados y procesado

Al igual que en el capítulo 2, las series de datos se obtuvieron de la página web de la NOAA (GTMBA, 2016). Sin embargo, para realizar los experimentos de este capítulo se han descargado los datos de las 21 boyas disponibles. Además, ahora las series de datos contienen registros desde el inicio del proyecto (11/07/1997) hasta diciembre de 2018. El método para procesar los datos es similar al utilizado para entrenar las redes SOM, pero con algunas etapas adicionales (figura 5.2). Tras descargas y generar las matrices, en la etapa de análisis de datos ausentes se excluyeron las boyas 2°N - 10°W, 2°S - 10°W, 5°S - 10°W y 6°S - 8°E por contener un número de ausentes mayor al 80% en todas sus variables. Además, de las cuatro boyas eliminadas, las tres primeras constan como inactivas en la web del proyecto PIRATA (GTMBA, 2016). La media (desviación estándar) de valores ausentes en el resto de boyas fue del 46,3% (12,3%). A continuación se generaron las salidas deseadas para las predicciones en los diferentes horizontes (datos diarios: $t + 7$, $t + 15$ y $t + 30$; datos mensuales: $t + 1$, $t + 2$ y $t + 3$) y, después, se procesaron los patrones con datos ausentes. Estos patrones fueron eliminados en los modelos diarios (excepto LSTM). En los modelos mensuales, salvo en los conjuntos con fecha y SST, se substituyeron por valores promedio debido al bajo número de patrones disponible. Además, se aplicó esta substitución en las pruebas con datos diarios y mensuales relacionadas con las redes LSTM, ya que son modelos recurrentes y utilizan muestras anteriores en el entrenamiento.

Posteriormente, la fecha (año, día y mes) se codificó en día del año (1-365) y se transformó en variables cíclicas aplicando el coseno y seno para evitar discontinuidades (ecuaciones 5.1 y 5.2). En el caso de los datos mensuales se codificó la fecha aplicando el coseno y seno al mes. Por último, todas las variables de entrada fueron estandarizadas (media cero y varianza unidad).

$$FechaCod1 = \cos\left(\frac{2 \cdot \pi \cdot FechaCod}{N}\right) \quad (5.1)$$

$$FechaCod2 = \sin\left(\frac{2 \cdot \pi \cdot FechaCod}{N}\right), \quad (5.2)$$

donde $N = 365$ para datos diarios y $N = 12$ para datos mensuales.

En la tabla 5.1 se muestra el número de patrones que contiene la variable SST y el número de patrones que dispone de todas las variables en cada boya, así como las fechas de inicio y fin de los datos tras aplicar el procesado. El número de patrones con todas las variables se ve reducido significativamente con respecto a los patrones de SST debido a los datos ausentes de las diferentes variables. Por otra parte, en la tabla 5.2 se pueden comprobar las variables disponibles en cada boya. La definición de las variables se puede consultar en la sección 2.2.1 del capítulo 2.

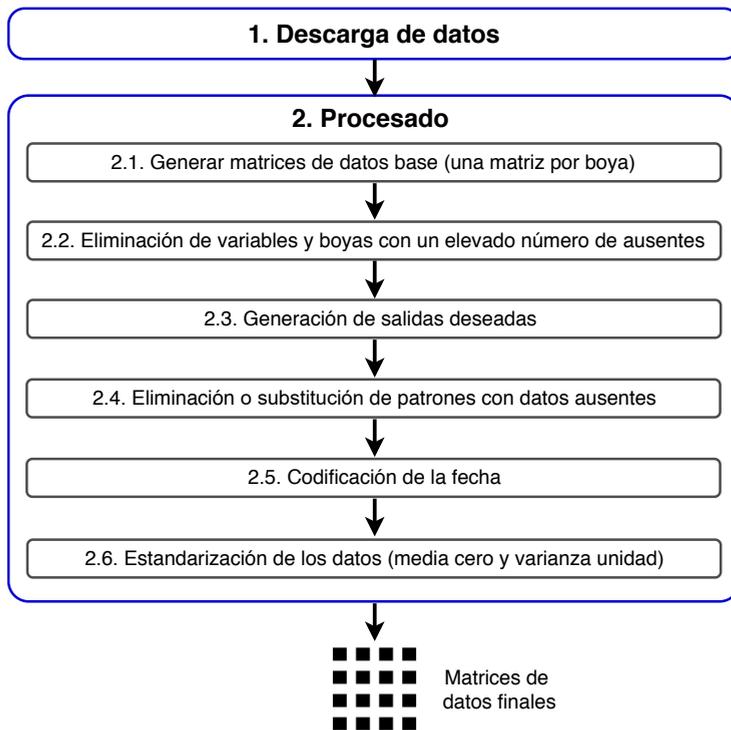


Figura 5.2
Esquema del procesado de datos.

Tabla 5.1
Número de patrones que disponen de la SST, patrones con todas las variables y fechas de los datos de cada boya.

Boya	N diario (SST)	N diario (Todas)	N Mensual (SST)	Fecha inicio	Fecha fin
0N0E	5125	2216	158	02/02/1998	31/10/2018
0N10W	5721	1990	180	29/01/1999	18/06/2018
0N23W	6430	1910	204	07/03/1999	31/10/2018
0N35W	5257	1233	162	22/01/1998	21/12/2017
4N23W	3895	1312	124	12/06/2006	31/10/2018
4N38W	4110	1985	121	21/02/1999	31/10/2018
6S10W	6056	2216	194	15/03/2000	12/04/2018
8N38W	5719	1468	177	31/01/1998	31/10/2018
8S30W	3512	2353	103	21/08/2005	23/10/2018
10S10W	6979	2922	225	11/09/1997	31/10/2018
12N23W	3662	1280	117	08/06/2006	26/05/2018
12N38W	5249	2022	161	04/02/1999	31/10/2018
14S32W	4069	2475	128	24/08/2005	07/02/2018
15N38W	6454	1757	206	28/01/1998	31/10/2018
19S34W	4610	1910	150	01/09/2005	29/10/2018
20N38W	3936	1795	127	23/05/2007	31/10/2018
21N23W	3632	1368	115	19/05/2007	03/09/2018

Tabla 5.2

Variables disponibles en cada boya. W, *Wind*; Sw, *Short wave*; RH, *Relative Humidity*; DYN, *Dynamic Height*; 20 °C IsoD, *20 °C isotherm depth*; S, *Salinity*; D, *Density*; SST, *Sea Surface Temperature*, ST, *Sea Temperature*. Las X indican datos ausentes.

Variables	0N0E	0N10W	0N23W	0N35W	4N23W	4N38W	6S10W	8N38W	8S30W	10S10W	12N23W	12N38W	14S32W	15N38W	19S34W	20N38W	21N23W
Year	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Month	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Day	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Air temp (°C)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
U wind	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
V wind	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
W speed (m/s)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
W dir (deg)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Sw rad (W/m ²)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
RH (%)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Rain (mm/hr)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	X
DYN (0-500m)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
20 °C IsoD (m)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	X	✓	X	✓	✓
S 1m	X	✓	✓	X	X	X	X	X	✓	✓	X	X	✓	✓	✓	X	X
S 10m	X	X	X	X	✓	X	X	X	X	X	✓	X	X	X	X	✓	X
S 20m	✓	✓	✓	✓	✓	✓	✓	✓	X	✓	✓	X	✓	✓	✓	✓	X
S 40m	✓	✓	✓	✓	✓	X	✓	X	✓	✓	✓	✓	✓	✓	✓	✓	X
S 60m	X	X	X	X	✓	X	X	X	X	X	✓	X	X	X	X	✓	X
S 80m	X	X	X	X	X	X	X	X	X	X	✓	X	X	X	X	X	X
S 120m	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	X
D 1m (kg/m ³)	X	✓	✓	X	X	X	X	✓	✓	✓	X	X	✓	✓	✓	X	X
D 10m (kg/m ³)	X	X	X	X	✓	X	X	X	X	X	✓	X	X	X	X	✓	✓
D 20m (kg/m ³)	✓	✓	✓	✓	✓	✓	✓	✓	X	✓	✓	✓	✓	✓	✓	✓	✓
D 40m (kg/m ³)	✓	✓	✓	✓	✓	X	✓	X	✓	✓	✓	✓	✓	✓	✓	✓	X
D 60m (kg/m ³)	X	X	X	X	✓	X	X	X	X	X	✓	X	X	X	X	✓	✓
D 80m (kg/m ³)	X	X	X	X	X	X	X	X	X	X	✓	X	X	X	X	X	X
D 120m (kg/m ³)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	X	✓	✓
SST 1m (°C)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
ST 5m (°C)	X	X	X	X	X	X	X	X	X	X	✓	X	X	X	X	X	X
ST 10m (°C)	X	X	X	X	✓	X	X	X	X	X	✓	X	X	X	X	✓	✓
ST 13m (°C)	X	X	X	X	✓	X	X	X	X	X	✓	X	X	X	X	✓	✓
ST 20m (°C)	✓	✓	✓	✓	✓	✓	✓	✓	X	✓	✓	X	✓	✓	✓	✓	✓
ST 40m (°C)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
ST 60m (°C)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
ST 80m (°C)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	X	✓	✓	X	✓	✓
ST 100m (°C)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
ST 120m (°C)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
ST 140m (°C)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
ST 180m (°C)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
ST 300m (°C)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
ST 500m (°C)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

5.3.2. Metodología de los experimentos preliminares

Antes de realizar los experimentos principales con todos los modelos, se llevaron a cabo una serie de análisis y pruebas preliminares utilizando los datos diarios de la SST. La finalidad de estas pruebas es estudiar el efecto de la tendencia, la estacionalidad y los valores pasados de SST en la precisión de las predicciones.

Según el análisis clásico de series temporales, la tendencia y la estacionalidad son las principales medidas descriptivas que caracterizan a una serie temporal (Hyndman, 2018):

- La tendencia se identifica como un cambio en la serie a largo plazo con respecto a su nivel medio.
- La estacionalidad se define como la presencia de una periodicidad o variación de la serie con un determinado periodo (mensual, anual, etc.)

La tendencia se ha calculado para la serie de SST de cada boya mediante regresión lineal. Además, se ha comprobado el efecto de eliminar la tendencia aplicando la primera y segunda derivada a la serie de SST. En cuanto a la estacionalidad, puesto que los datos a analizar son series temporales de tipo climático, es previsible la existencia de una componente estacional anual, además de una fuerte correlación entre muestras actuales y pasadas. Para analizar estas características se ha utilizado la autocorrelación, así como las funciones de autocorrelación (*AutoCorrelation Function*, ACF) y autocorrelación parcial (*Partial AutoCorrelation Function*, PACF) (Hyndman, 2018). La desestacionalización de la serie temporal se ha realizado aplicando la diferencia estacional, es decir, calculando la diferencia entre el valor de la serie de SST en un día concreto y el valor promedio de ese mismo día a lo largo de la serie histórica disponible (climatología).

Una vez analizada la tendencia y componente estacional, se ha utilizado la ELM para obtener las predicciones de SST en $t + 7$, $t + 15$ y $t + 30$ días. Para estos experimentos se realizó un barrido del número de neuronas de cinco a 50 en incrementos de cinco, y los resultados mostrados corresponden a la mejor arquitectura seleccionada por el criterio de información de Akaike (AIC). Tanto el modelo ELM como el criterio AIC se definirán posteriormente en la sección 5.3.4. Los experimentos preliminares realizados han sido:

1. Utilizar como variable de entrada únicamente la SST actual.
2. Junto a la SST actual, añadir valores anteriores de la SST como entradas ($t - 1$, $t - 2$, $t - 7$ y $t - 15$).
3. Aplicar la primera y segunda derivada a la SST para eliminar la tendencia.
4. Eliminar la componente estacional restando la climatología a la SST de entrada/salida.
5. Añadir la fecha sin codificar como variable de entrada.
6. Añadir la fecha codificada como variable de entrada.
7. Combinar el uso de la fecha con la SST actual, valores pasadas y eliminación de la componente estacional.

5.3.3. Reducción de la dimensionalidad. Selección y extracción de características

Las técnicas de selección y extracción de características (variables de entrada o atributos) tienen como principal objetivo reducir la dimensionalidad del conjunto de datos. Es decir, obtener el subconjunto de características más compacto e informativo posible que permita resolver de manera óptima el proceso de aprendizaje. Además, estas técnicas pueden ser utilizadas para (Guyon y Elisseeff, 2006):

- Reducir el número de datos para limitar los requerimientos de almacenamiento e incrementar la eficiencia del procesamiento aumentando, por lo tanto, la velocidad de los algoritmos.
- Reducir el número de variables para simplificar la siguiente recogida de datos.
- Incrementar la precisión de los modelos.
- Ganar conocimiento sobre el proceso de generación de los datos facilitando la visualización de los mismos.
- Reducir los problemas que se producen al analizar, visualizar u organizar conjuntos de datos de alta dimensionalidad.

Los métodos de extracción de características generan un conjunto de datos con nuevas variables a partir de combinaciones de las variables originales. Estas combinaciones, que pueden ser lineales o no lineales según el tipo de proyección utilizado, permiten obtener variables con un mayor poder de representación de los datos que los atributos originales. El método de extracción más conocido es la técnica lineal y no supervisada PCA (Alpaydin, 2010).

En cuanto a los métodos de selección de características, su finalidad es encontrar las variables más relevantes e informativas de un conjunto de datos. Existen tres principales aproximaciones (Guyon y Elisseeff, 2006, 2003):

- **Filtros.** Los métodos basados en filtros realizan una ordenación de los predictores basándose en criterios de relevancia, por ejemplo, coeficientes de correlación o tests estadísticos. El algoritmo *RReliefF* es un ejemplo clásico de este tipo de métodos.
- **Wrappers.** Estos métodos utilizan el algoritmo de aprendizaje automático de interés como una caja negra para evaluar el rendimiento de diferentes subconjuntos de variables de acuerdo a su poder de predicción.
- **Métodos embebidos (*embedded methods*).** Este tipo de técnicas realizan la selección de variables en el proceso de ajuste del modelo y suelen ser específicas del método de aprendizaje máquina. Por ejemplo, dentro de esta categoría se encuentra el algoritmo *Random Forest*.

5.3.3.1. Análisis de componentes principales

Es una de las técnicas estadísticas multivariante más conocida y es utilizada habitualmente en tareas de reducción de la dimensionalidad. De manera similar a otros métodos de extracción de características, esta técnica persigue obtener la información más importante de un conjunto de datos, reducir el tamaño y simplificar la descripción de los datos, así como analizar la estructura de las variables y observaciones (Abdi y Williams, 2010).

La técnica PCA obtiene nuevas variables, denominadas componentes principales (CPs), mediante combinaciones lineales de las variables originales. La primera CP es, de entre todas las combinaciones lineales de los predictores, la combinación lineal que explica la mayor varianza posible de los datos. Las siguientes CPs se calculan bajo la condición de presentar la mayor varianza restante y de no estar correlacionadas con las CPs anteriores (Kuhn y Johnson, 2013).

Metodología. Para reducir la dimensionalidad de los datos se aplicó la técnica PCA a los conjuntos de entrada, seleccionando el número de CPs que representasen, al menos, el 90 % de la varianza. Mediante diagramas de Pareto se examinó la varianza acumulada en función del número de CPs. Este tipo de diagrama presenta el porcentaje de la varianza que explica cada CP (barras) y el porcentaje total acumulado (curva). Además, se examinaron los valores de los coeficientes, o pesos, para determinar el nivel de contribución de cada predictor en la determinación de los CPs (Abdi y Williams, 2010).

5.3.3.2. Algoritmo *RReliefF*

Es una técnica de selección de características basada en el algoritmo de k-vecinos que realiza una ordenación o *ranking* de los atributos en función de su importancia para predecir las salidas. Presentado por Kira y Rendell (Kira y Rendell, 1992) para problemas de clasificación de dos clases, posteriormente fue mejorado por Kononenko (Kononenko, 1994) y Robnik-Sikonja (Robnik-Šikonja, 2003) para ser utilizado en problemas multiclase y de regresión. El algoritmo original se basa en estimar la importancia de los atributos en función de lo bien que distinguen datos muy cercanos y que pertenecen a clases distintas. Para realizar esta estimación, en primer lugar se inicializa aleatoriamente el caso seleccionado R_i y se buscan sus dos vecinos más cercanos, llamados acierto cercano H y fallo cercano M . A continuación se actualiza el vector de estimación de importancia $W[A]$ para todos los atributos A dependientes de los valores R_i , M y H . Si los casos R_i y H tienen valores diferentes al atributo A , quiere decir que A separa dos casos que pertenecen a la misma clase y, por tanto, se reduce la importancia del estimador $W[A]$. Por otra parte, si los casos R_i y M tienen valores diferentes del atributo A , quiere decir que A separa dos casos que pertenecen a distintas clases y, por tanto, se incrementa la importancia del estima-

donde $W[A]$. El pseudo-código del algoritmo *Relief* es el siguiente (Kira y Rendell, 1992):

Entrada. Para cada observación un vector con los valores de los atributos y un vector con los valores de las clases.

Salida. El vector W de las estimaciones de la importancia de los atributos.

1. Iniciar los pesos $W[A] := 0.0$
2. **for** $i := 1$ **to** m **do begin**
3. iniciar los valores aleatoriamente de R_i ;
4. encontrar acierto más cercano (H) y fallo más cercano (M);
5. **for** $A := 1$ **to** a **do**
6. $W[A] := W[A] - diff(A, R_i, H)/m + diff(A, R_i, M)/m$;
7. **end;**

Una de las limitaciones más importantes del algoritmo *Relief* es su sensibilidad ante atributos duplicados (o muy correlacionados). Si un problema contiene variables duplicadas, *Relief* les asignará la misma importancia (Robnik-Šikonja, 2003). En cuanto a los diferentes algoritmos de la familia *Relief*, el algoritmo *ReliefF* es una extensión del algoritmo base *Relief* que no está limitado a problemas de clasificación de dos clases, es más robusto y más inmune al ruido (Kononenko, 1994). Finalmente, el algoritmo *RReliefF* es una nueva modificación para poder ser usado en problemas de regresión (Robnik-Šikonja, 2003).

Metodología. Para obtener la importancia de los predictores, se aplicó el algoritmo *RReliefF* a las variables de entrada de cada boya fijando el número de k vecinos en 10. A pesar de que *RReliefF* ordena las variables según su importancia, no existe una regla fija a la hora de seleccionar el número de variables para entrenar los modelos. En primer lugar se analizó la posibilidad de eliminar aquellas variables cuyo peso fuese negativo. Sin embargo, en algunos casos este método eliminaba pocas variables. Finalmente se decidió eliminar aquellas variables con un peso 10 veces inferior al mayor peso, ya que este método mostró un buen equilibrio entre el número de variables eliminadas y la predicción obtenida en la mayoría de los casos.

5.3.3.3. Análisis de correlación

La función de correlación conocida como coeficiente de correlación de *Pearson*, permite determinar la posible relación lineal existente entre dos variables. Los funda-

mentos matemáticos de este tipo de correlación fueron expuestos en el apartado 3.3.3 del capítulo 3.

Cuando dos predictores presentan una correlación alta entre ellos se dice que existe colinealidad. Si la correlación se da entre varios predictores existe multicolinealidad. La multicolinealidad afecta a las tareas de aprendizaje en términos de estimación y explicación. A medida que aumenta la multicolinealidad, disminuye la varianza total explicada disminuyendo, por tanto, la capacidad de predicción de cada variable (estimación). Además, al compartirse la varianza entre un mayor número de variables, aumenta la dificultad para analizar el efecto que produce cada uno de los predictores en la estimación (explicación) (Hair et al., 2010).

La colinealidad añade complejidad a los modelos y empeora el rendimiento de las predicciones. Por tanto, es importante realizar el análisis de correlación para eliminar, o reducir, la información redundante. El procedimiento empleado para eliminar la colinealidad ha sido el siguiente (Kuhn y Johnson, 2013):

1. Calcular la matriz de correlación de las variables de entrada. Esta matriz proporciona el coeficiente de correlación de cada pareja de predictores, permitiendo evaluar todas las posibles combinaciones.
2. Buscar la pareja de predictores con el mayor valor absoluto de correlación (variables A y B).
3. Calcular la correlación media entre A y el resto de variables de entrada. Repetir el procedimiento con la variable B.
4. Eliminar la variable (A o B) con el mayor valor de correlación promedio.
5. Repetir los pasos 2 a 4 hasta que el mayor valor absoluto de correlación esté por debajo del umbral fijado.

Metodología. En la anterior sección se ha explicado que el algoritmo *RReliefF* es sensible a las variables correlacionadas. Por este motivo, se ha comprobado el efecto de aplicar el análisis de correlación como etapa previa a la aplicación de *RReliefF*. Para eliminar el efecto de multicolinealidad, se han eliminado recursivamente las variables más correlacionadas hasta que el mayor valor absoluto de correlación se encuentre por debajo de 0,8.

5.3.4. Metodología y definición de los modelos de predicción

Implementación. Los modelos LSTM se han implementado utilizando la versión R2019a de *Matlab*, mientras que para el resto modelos se ha empleado la versión R2016a. Con los modelos RMLR y RFBT se ha utilizado la librería *Statistics and Machine Learning*, con los modelos GRNN y MLP la librería *Neural Networks* y, para

los modelos LSTM, la librería *Deep Learning Toolbox*. Por otra parte, se ha empleado la librería LS-SVMlab1.8 para los modelos basados en LS-SVM (De Brabanter et al., 2010). Por último, para el modelo ELM se ha utilizado una implementación de código propia.

Los programas se han ejecutado en un servidor Linux con procesador Intel® Xeon® CPU E5-2620 v4 @2,10GHz x 16 con 251GB de RAM, un servidor Linux Intel® Xeon® CPU E5-2620 v4 @2,10GHz con 128GB de RAM, un servidor Linux Intel® Xeon® CPU E5530 @2,4GHz x16 con 70Gb de RAM y un ordenador personal Windows con procesador Intel® Core™ i5-2500 CPU @3,30GHz con 12GB de RAM.

Estrategia de predicción. Los modelos de predicción multipaso pueden ser implementados siguiendo diferentes estrategias. Las más conocidas son estrategia iterativa (o recursiva) (Chevillon, 2007), estrategia directa (Cox, 1961) y estrategia MIMO (*Multi-Input Multi-Output*, Múltiples-Entradas Múltiples-Salidas) (Ben Taieb et al., 2009). En los trabajos de Taieb et al. (2012) y Bao et al. (2014) se puede consultar una revisión de las diferentes estrategias de predicción multipaso. De manera resumida, la estrategia iterativa construye un modelo para realizar una predicción de un único paso hacia delante y, después, emplea este resultado como entrada para calcular la siguiente predicción. Este último paso se repite hasta alcanzar el horizonte de predicción deseado. Este método propaga el error de una predicción a la siguiente, por lo que el error puede aumentar rápidamente (Taieb et al., 2012). Por otra parte, la estrategia directa construye un modelo distinto para cada horizonte de predicción, utilizando únicamente observaciones pasadas. Este método consigue mejorar las predicciones de la estrategia iterativa ya que no propaga el error (Bao et al., 2014). Finalmente, la estrategia MIMO emplea un único modelo que calcula múltiples predicciones a partir de las observaciones pasadas de manera directa. Esta estrategia consigue preservar la dependencia temporal estocástica presente en las series temporales (Bao et al., 2014). En nuestro caso, las predicciones de la SST se han obtenido mediante la estrategia directa en los modelos RMLR y RFBT, mientras que en los modelos ELM, LS-SVM, GRNN y MLP se ha utilizado la estrategia MIMO. Las predicciones con redes LSTM se han obtenido mediante las estrategias recursiva y MIMO.

Evaluación de los modelos. Mediante los diferentes modelos se han obtenido predicciones de la SST en tres horizontes distintos: $t+7$, $t+15$ y $t+30$ días para los modelos diarios y $t+1$, $t+2$ y $t+3$ meses para los modelos mensuales. De manera previa a la generación de los conjuntos de entrenamiento y validación, se reservaron las últimas 365 muestras (36 en los modelos mensuales) de cada boya para formar los conjuntos de test. Utilizando el resto de datos, el ajuste de las arquitecturas de los modelos se ha realizado mediante el método de validación cruzada, dividiendo el conjunto de datos de manera aleatoria en 10 bloques de igual tamaño (*10-fold cross-validation*). En cada una de las 10 repeticiones se utilizan 9 bloques distintos de datos para entrenar el modelo (conjunto de entrenamiento) y el bloque restante

para validarlo (conjunto de validación). Para aumentar la robustez de los resultados se ha repetido 20 veces este procedimiento, por lo que el proceso de entrenamiento y validación se repite un total de 200 veces en cada arquitectura.

Puesto que en el caso de las redes LSTM es necesario preservar el orden cronológico de las muestras, el ajuste se realizó dividiendo los datos en 2/3 para entrenamiento y 1/3 para validación. Finalmente, en todo los casos se utilizaron los conjuntos de test para comprobar la generalización de los modelos en datos no utilizados y, además, posteriores en fecha a los empleados en el proceso de entrenamiento y validación.

Los errores cometidos por los modelos en las predicciones de SST se han estimado utilizando el índice de error MAE (*Mean Absolute Error*):

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (5.3)$$

donde y_i es la señal deseada, \hat{y}_i es la predicción y n el número de datos. Por tanto, este índice muestra la media de las diferencias entre la señal deseada y la señal predicha.

Ajuste de los parámetros. En los apartados correspondientes a la descripción de los modelos de predicción, se detalla también la búsqueda de la arquitectura o parámetros óptimos realizada en cada caso particular.

Además de realizar un barrido del número de neuronas, en los modelos MLP y ELM se ha utilizado el criterio AIC (Akaike, 1974). Los criterios de selección como AIC se basan en el error cometido por el modelo analizado (primer término del criterio) y en un término que penaliza el número de parámetros (p) (segundo término del criterio). A medida que aumenta el error cometido y el número de parámetros (por ejemplo neuronas) del modelo, aumenta el valor del índice AIC. Por lo tanto, el menor valor del índice representa el modelo con la mejor relación entre error cometido y número de parámetros. La expresión matemática del índice AIC es:

$$AIC = N \times \left(\log \left(\frac{2\pi Error}{N} \right) + 1 \right) + 2 \times p \quad (5.4)$$

En los siguientes apartados se describen los modelos matemáticos utilizados para realizar las predicciones de la SST.

5.3.4.1. Modelo trivial

El modelo trivial es el más sencillo de todos y consiste en considerar el valor actual de la variable como el resultado de la predicción:

- $y(t + 7) = x(t)$, $y(t + 15) = x(t)$ e $y(t + 30) = x(t)$ en los datos diarios.

- $y(t + 1) = x(t)$, $y(t + 2) = x(t)$ e $y(t + 3) = x(t)$ en los datos mensuales.

5.3.4.2. Modelo climatología

El siguiente modelo en cuanto a complejidad es la media de los datos históricos o climatología, que puede ser diaria o mensual dependiendo de los datos utilizados. Esta media se obtiene calculando el valor medio de la SST de cada día del año, o de cada mes, empleando todos los datos disponibles de la serie histórica. Estos dos modelos se utilizan como referencia para comprobar la bondad de los ajustes obtenidos con el resto de modelos de predicción. Para poder compararla con los demás modelos, la climatología se calculó sin utilizar los datos de test. Además de esta comparación, el modelo basado en la media también resulta útil para analizar la componente estacionaria de las series temporales tal y como se detalló en la sección de experimentos preliminares (5.3.2).

5.3.4.3. Regresión lineal múltiple robusta

Realiza un ajuste mediante un modelo que es lineal en sus coeficientes, el cual se calcula normalmente por mínimos cuadrados. El modelo general de MLR para un conjunto de datos de entrada con dimensión p y n observaciones viene dado por la siguiente ecuación Weisberg (2014):

$$y_i = b_0 + \sum_{k=1}^p b_k X_{(i,k)} + \epsilon_i, \quad i = 1, \dots, n, \quad (5.5)$$

donde b_k son los coeficientes de regresión, $X_{n \times p}$ es el vector de entrada, y_i es la i -th muestra de la variable de salida y ϵ es el error cometido en el ajuste, que puede ser determinado mediante los residuos (diferencias entre los valores observados y los predichos):

$$\hat{\epsilon}_i = \hat{y}_i - y_i \quad (5.6)$$

El ajuste por mínimos cuadrados es uno de los tipos de regresión lineal más utilizados. La versión robusta de la función de regresión utiliza un algoritmo iterativo de ajuste por mínimos cuadrados, donde los pesos se calculan en cada iteración aplicando la función bicuadrada a los residuos de la anterior iteración. Este algoritmo permite asignar menor peso a los puntos que no se ajustan bien al modelo. Por lo tanto el modelo resultante es menos sensible a los valores atípicos (*outliers*) de los datos que la regresión lineal por mínimos cuadrados estándar (Holland y Welsch, 1977).

5.3.4.4. Perceptrón multicapa

Una red neuronal artificial (ANN) es una técnica de aprendizaje automático no lineal inspirada en el funcionamiento del cerebro humano. De forma similar a como procesa el cerebro la información, una ANN resuelve problemas de aprendizaje modelando la relación entre las señales de entrada y de salida mediante una red interconectada de neuronas artificiales (Ripley, 1996; Bishop, 2006; Titterington, 2010). El modelo de ANN más ampliamente utilizado es el perceptrón multicapa ya que, debido a su gran versatilidad, puede ser utilizado para resolver todo tipo de problemas de aprendizaje (regresión, clasificación, reconocimiento de patrones, etc.) con excelentes resultados (Haykin, 2009).

Las neuronas son las unidades no lineales de procesamiento de las ANN y están formadas por un sumador y una función de activación como se muestra en la figura 5.3. En el sumador se computa la suma del producto entre las entradas \mathbf{x}_k a la neurona y los coeficientes de adaptación \mathbf{w}_k , llamados pesos sinápticos, que representan la conectividad entre las neuronas. La salida de la i -ésima neurona viene dada por (Bishop, 2006):

$$o_i = \phi\left(\sum_{k=0}^m \mathbf{w}_{ik} \mathbf{x}_k + b_i\right) \quad (5.7)$$

donde ϕ es la función de activación no lineal y b_i es el sesgo de la i -ésima neurona oculta.

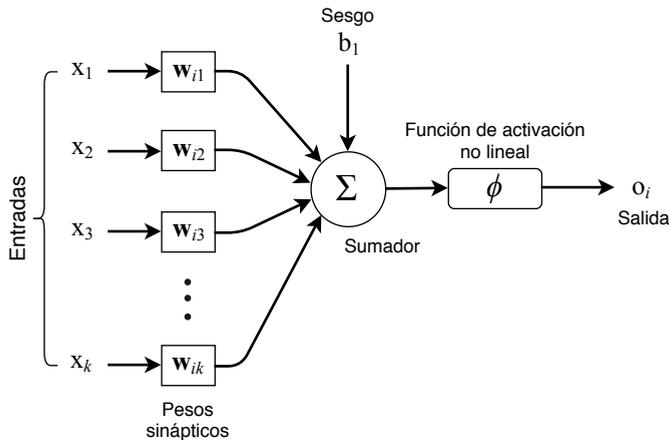


Figura 5.3
Representación esquemática de una neurona artificial.

La función de activación es una función que determina el modo en que la ANN procesa la información. Las funciones más utilizadas son: umbral, lineal, sigmoide,

tangente hiperbólica y gaussiana. La función umbral fue la primera función de activación utilizada en un modelo ANN (Arbib, 2003). Si el valor de entrada es negativo la salida de la neurona es 0 y 1, si el valor de entrada es positivo. La función umbral se define como:

$$f(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1 & \text{si } x \geq 0 \end{cases} \quad (5.8)$$

La función sigmoide es la función de activación más utilizada en las ANN ya que presenta un comportamiento equilibrado para datos lineales y no lineales. Además, a diferencia de la función umbral, es diferenciable (derivable en todo el rango de entradas). Matemáticamente, la función se expresa como:

$$f(x) = \frac{1}{1 + e^{-x}} \quad (5.9)$$

La versión bipolar de la función sigmoide, es la tangente hiperbólica definida como:

$$f(x) = a \frac{1 - e^{-bx}}{1 + e^{-bx}} \quad (5.10)$$

donde a es la amplitud y b , la pendiente de la función.

La arquitectura de un MLP está formada por una capa de entrada, una o varias capas ocultas y una capa de salida. Cada una de estas capas está compuesta por, al menos, una neurona artificial. La señal de entrada se propaga capa a capa hacia la salida de la red a través de las neuronas. Mientras que el número de neuronas de entrada/salida es igual al número de variables de entrada/salida del problema a resolver, el número de neuronas de las capas ocultas tiene que ser configurado por el diseñador de la ANN (Arbib, 2003). En la figura 5.4 se muestra un esquema de un MLP formado por dos capas ocultas.

El procedimiento utilizado para ajustar los parámetros de la red neuronal se denomina algoritmo de aprendizaje. Este algoritmo adapta el valor de los pesos sinápticos de la ANN para resolver un problema determinado, minimizando una medida del error cometido (Haykin, 2009). El algoritmo de retropropagación (*Backpropagation Algorithm*, BP) fue el primer algoritmo utilizado para adaptar los pesos de la ANN (Werbos, 1974); BP es un algoritmo de descenso por gradiente basado en la ecuación 5.11 que adapta el valor de los parámetros para hallar el mínimo de la función de coste J :

$$W_{n+1} = W_n - \alpha \nabla_{w_n} J \quad (5.11)$$

donde W son los pesos sinápticos, α la constante de adaptación y J la función de coste. El algoritmo BP funciona en dos fases:

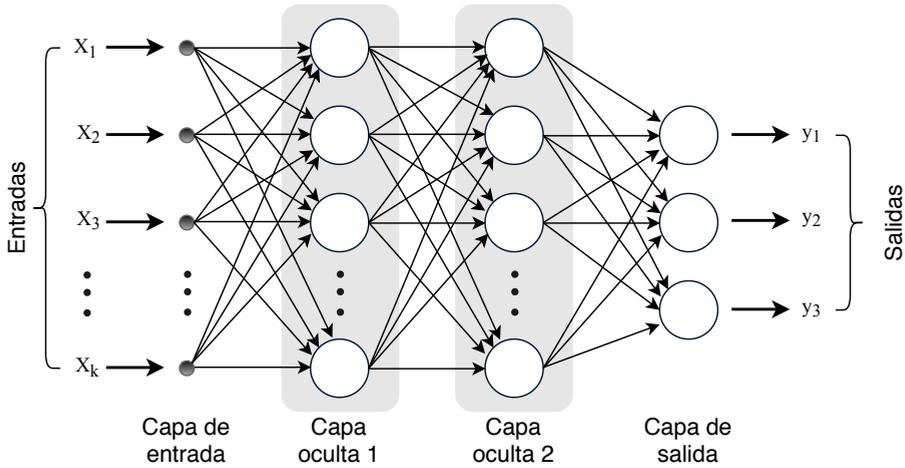


Figura 5.4

Arquitectura del perceptrón multicapa con dos capas ocultas.

1. **Propagación hacia adelante.** En esta fase las señales se propagan desde la capa de entrada hacia la de salida, aplicándose en cada caso los pesos y funciones de activación en cada neurona. El error se calcula comparando la señal de salida con la señal deseada.
2. **Propagación hacia atrás.** En función del error obtenido, el algoritmo optimiza recursivamente los valores de los pesos propagando la señal de error desde la capa de salida hacia la de entrada.

Una mejora del algoritmo BP, es el algoritmo *Levenberg-Marquardt* (LM). LM supone un compromiso entre el método de Newton (es rápido encontrando el mínimo local, o global, pero puede diverger) y el descenso por gradiente (se asegura la convergencia, pero es lento). LM es un método que mejora la velocidad de entrenamiento del descenso por gradiente evitando calcular la matriz hessiana (Haykin, 2009).

Ajuste de la arquitectura. La selección de la arquitectura del MLP se ha realizado mediante un barrido del número de neuronas desde dos hasta 10 neuronas en pasos de dos, con una y con dos capas ocultas. Cada arquitectura se ha inicializado 20 veces con pesos aleatorios para evitar el problema de mínimo local. El algoritmo de entrenamiento utilizado es el conocido como *Levenberg-Marquardt* (Haykin, 2009).

En la figura 5.5 se muestra el error (índice MAE) cometido en función del número de neuronas para las boyas $19^\circ\text{S} - 34^\circ\text{W}$ y $0^\circ\text{N} - 23^\circ\text{W}$. En la figura 5.5.a se puede observar que con cuatro neuronas los errores de entrenamiento, validación y test son menores que utilizando dos neuronas. Al aumentar las neuronas, los índices MAE

de entrenamiento y validación siguen disminuyendo progresivamente. Sin embargo, el índice MAE de test aumenta, indicando que se está produciendo sobreentrenamiento. Para este caso, el criterio AIC seleccionó la arquitectura con cuatro neuronas. En el segundo ejemplo (figura 5.5.b), el error de entrenamiento disminuye al aumentar el número de nodos, pero el error de validación aumenta a partir de seis neuronas. Por tanto, la arquitectura escogida según el mínimo MAE de validación o según el criterio AIC son similares para este caso (seis y cuatro respectivamente). La tabla 5.3 muestra la arquitectura del MLP que consigue un menor error de validación para cada boya. Se descartaron los modelos con dos capas ocultas ya que no consiguieron mejorar los resultados obtenidos por los modelos con una sola capa.

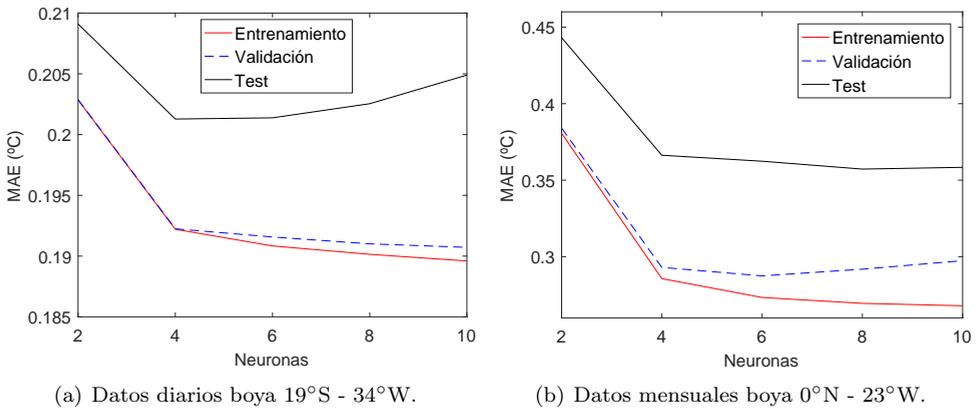


Figura 5.5 Índices MAE (°C) de entrenamiento, validación y test en función del número de neuronas del MLP utilizando como entradas SST y fecha codificada.

Tabla 5.3 Mejor arquitectura del MLP para cada boya utilizando como entradas la SST y fecha, así como utilizando todas las variables de entrada. Los resultados muestran el número de neuronas en una capa.

Boya		0N0E	0N10W	0N23W	0N35W	4N23W	4N38W	6S10W	8N38W	8S30W	10S10W	12N23W	12N38W	14S32W	15N38W	19S34W	20N38W	21N23W
Diarios	SST	6	6	8	6	8	4	10	10	10	10	8	6	8	6	6	6	4
	Todas	4	2	4	4	6	4	6	4	6	10	4	6	6	6	6	6	6
Mensuales	SST	4	4	4	4	6	4	4	4	6	4	4	8	4	4	10	6	4
	Todas	2	2	4	4	2	2	2	2	2	2	2	2	2	2	2	2	2

5.3.4.5. Máquinas de aprendizaje extremo

Son modelos neuronales compuestos por una única capa oculta (SLFN). En la figura 5.6 se muestra un esquema de este tipo de red neuronal. Si consideramos un conjunto de datos con N puntos $(\mathbf{x}_i, \mathbf{t}_i) \in (\mathbb{R}^n, \mathbb{R}^m)$, donde $i = 1, \dots, N$, la salida de una SLFN con L neuronas ocultas y función de activación $g(x)$ se puede expresar matemáticamente como:

$$\mathbf{d}_j = \sum_{i=1}^L \beta_i g(\mathbf{w}_i \mathbf{x}_i + b_i) \quad j = 1, \dots, N \quad (5.12)$$

donde \mathbf{w}_i es el vector de pesos que conectan las entradas con la i -ésima neurona oculta, β_i es el vector de pesos que conecta la i -ésima neurona oculta y la salida de las neuronas, b_i es el sesgo de la i -ésima neurona oculta, y \mathbf{d}_j es la salida dada por la red para una muestra j de los datos (figura 5.6). Una SLFN estándar con L neuronas ocultas y función de activación $g(x)$ puede aproximar N muestras con error cero en el caso ideal, inicializando aleatoriamente los pesos de entrada y el sesgo de las neuronas ocultas. Esta asignación aleatoria se puede llevar a cabo si $g(x)$ es infinitamente diferenciable (Huang et al., 2006).

El algoritmo ELM se fundamenta en este hecho, tras inicializar aleatoriamente los pesos de entrada considera la salida de la SLFN como un sistema lineal y los pesos de salida se pueden obtener, entonces, mediante la pseudo-inversa de la matriz de salidas de la capa oculta (Huang et al., 2006). De esta forma, dado un conjunto de N vectores de entrada, la salida de la SLFN aproxima estos N casos con error nulo $\sum_{j=1}^N \|\mathbf{d}_j - \mathbf{t}_j\| = 0$ y, por tanto, existe una β_i , \mathbf{w}_i y una b_i tal que,

$$\sum_{i=1}^L \beta_i g(\mathbf{w}_i \mathbf{x}_j + b_i) = \mathbf{t}_j \quad j = 1, \dots, N \quad (5.13)$$

Las anteriores ecuaciones se pueden expresar de manera compacta como:

$$\mathbf{H}\beta = \mathbf{T} \quad (5.14)$$

donde

$$\mathbf{H} = \begin{pmatrix} g(\mathbf{w}_1 \mathbf{x}_1 + b_1) & \dots & g(\mathbf{w}_L \mathbf{x}_1 + b_L) \\ \vdots & \dots & \vdots \\ g(\mathbf{w}_1 \mathbf{x}_N + b_1) & \dots & g(\mathbf{w}_L \mathbf{x}_N + b_L) \end{pmatrix}_{N \times L}$$

$$\beta = \begin{pmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{pmatrix}_{L \times m} \quad \text{y} \quad \mathbf{T} = \begin{pmatrix} \mathbf{t}_1^T \\ \vdots \\ \mathbf{t}_N^T \end{pmatrix}_{N \times m}$$

La matriz \mathbf{H} es la matriz de pesos de salida de la capa oculta de la SLFN, siendo sus dimensiones $N \times L$, donde N es el número de casos y L el número de neuronas en la capa oculta, β es el vector de pesos de salida de las neuronas y T el vector de las salidas deseadas. Por otra parte el sistema de ecuaciones tiene que ser resuelto mediante métodos numéricos, concretamente resolviendo:

$$\min_{\beta} \|\mathbf{H}\beta - \mathbf{T}\| \tag{5.15}$$

La solución a este problema con la mínima norma viene dada por:

$$\hat{\beta} = \mathbf{H}^\dagger \mathbf{T} \tag{5.16}$$

donde \mathbf{H}^\dagger es la pseudo-inversa de *Moore-Penrose* (Serre, 2002). Una matriz \mathbf{G} de orden $n \times m$ es la pseudo-inversa generalizada de *Moore-Penrose* de una matriz \mathbf{A} de orden $n \times m$ si,

$$\mathbf{AGA} = \mathbf{A}, \quad \mathbf{GAG} = \mathbf{G}, \quad (\mathbf{AG})^T = \mathbf{AG}, \quad (\mathbf{GA})^T = \mathbf{GA} \tag{5.17}$$

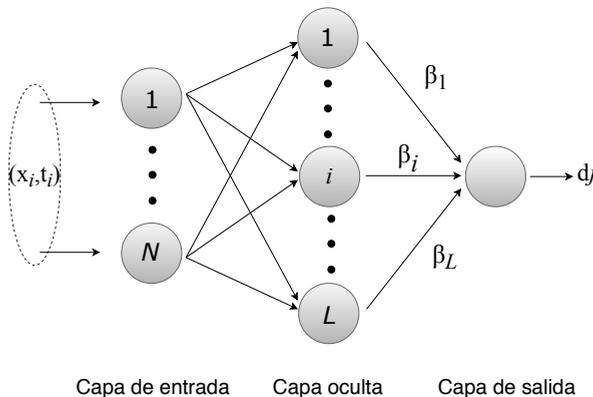


Figura 5.6 Representación esquemática de una ELM con N datos de entrada y L neuronas en la capa oculta.

Las características más importantes de esta solución son (Huang et al., 2006):

- Minimizar el error de entrenamiento.
- Obtener la norma más pequeña y el mejor rendimiento de generalización.
- La mínima norma de mínimos cuadrados de la solución $\mathbf{H}\beta = \mathbf{T}$ es única y es $\hat{\beta} = \mathbf{H}^\dagger \mathbf{T}$.

Por lo tanto el algoritmo ELM para SLFNs puede resumirse en los siguientes pasos:

Dado un conjunto de datos $(\mathbf{x}_i, \mathbf{t}_i) \in (\mathbb{R}^n, \mathbb{R}^m)$, $i = 1, \dots, N$ función de activación $g(x)$ y L neuronas ocultas:

1. Asignar aleatoriamente los pesos de entrada \mathbf{w}_i y sesgos b_i , $i = 1, \dots, L$.
2. Calcular la matriz de pesos de salida de la capa oculta \mathbf{H} .
3. Calcular los pesos de salida β según

$$\beta = \mathbf{H}^\dagger \mathbf{T}$$

donde \mathbf{H} , β y \mathbf{T} ya han sido definidos previamente.

El principal problema de la ELM es el ajuste del número de neuronas de la capa oculta. Este problema no tiene una única solución y depende del número de variables, del número de patrones y de la propia naturaleza de los datos (Huang et al., 2006). Para resolver este problema existen dos grandes estrategias posibles, por un lado entrenar los modelos con un gran número de neuronas e ir “podando” (*pruning*) las neuronas hasta conseguir el modelo óptimo. La otra estrategia consiste en empezar con un número pequeño de neuronas e ir incrementando ese número (*growing*) hasta obtener la mejor relación posible entre el número de neuronas y el error cometido.

Ajuste de la arquitectura. Al igual que en el MLP, la arquitectura de la ELM se ha fijado mediante un barrido del número de neuronas. Para los modelos con SST y fecha como entradas se ha utilizado un barrido de cinco en cinco hasta 50 neuronas, mientras que para el resto de modelos (con todas las variables, con PCA, etc.) se ha realizado un barrido de 20 en 20 hasta 200 neuronas. En la figura 5.7 se muestra el índice MAE de validación en función del número de nodos de la ELM con datos diarios. En la figura 5.7.a se observa que el error desciende rápidamente hasta las 20 neuronas, permaneciendo constante al seguir incrementando el número de neuronas. Sin embargo, utilizando todas las variables de entrada (figura 5.7.b) se observa que el error sigue descendiendo de manera progresiva al aumentar el número de neuronas. En la tabla 5.4 se muestra la arquitectura de la ELM con la que se ha obtenido un menor valor del índice MAE de validación en cada boya.

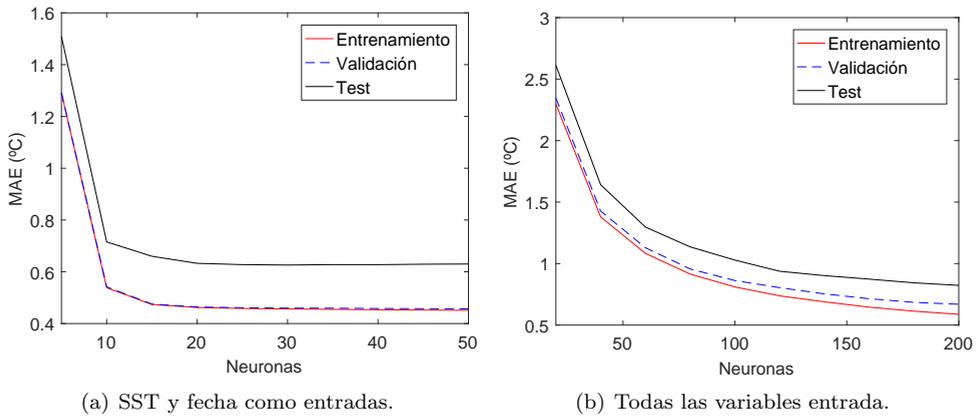


Figura 5.7
Índice MAE (°C) en función del número de neuronas de la ELM para la boya 0°N - 0°E.

Tabla 5.4
Mejor arquitectura (número de neuronas) de la ELM en cada boya para datos diarios y mensuales utilizando como entradas SST y fecha codificada.

Boya		0N0E	0N10W	0N23W	0N35W	4N23W	4N38W	6S10W	8N38W	8S30W	10S10W	12N23W	12N38W	14S32W	15N38W	19S34W	20N38W	21N23W
Dia.	SST	20	15	20	25	20	25	20	25	20	20	20	20	20	20	20	20	15
	Todas	160	140	180	120	120	180	200	160	200	200	140	180	200	180	160	200	140
Men.	SST	15	15	20	20	20	20	10	20	20	20	15	15	20	15	20	20	20
	Todas	200	200	80	100	200	200	200	200	200	100	200	200	200	100	200	200	200

5.3.4.6. Redes neuronales de regresión generalizada

Fue propuesta por Donald F. Specht en 1991 (Specht, 1991) y es una variante de la red neuronal de base radial (*Radial Basis Function Network*, RBFNN). Se utiliza habitualmente como aproximador de funciones y está formada por tres capas: una capa de entrada, una capa oculta y una capa de salida. La capa oculta está formada por neuronas de base radial, mientras que la capa de salida está formada por una, o varias neuronas con función de transferencia lineal. La salida de la red se obtiene aplicando la RBF, escalada por un factor de ensanchamiento (*spread factor*), a los datos de entrada. La RBF es la función gaussiana multivariante más utilizada y se define como:

$$G(\mathbf{x}, \mathbf{x}_i) = \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{x} - \mathbf{x}_i\|^2\right) \tag{5.18}$$

donde \mathbf{x}_i y σ son los centros y anchos, respectivamente, de las funciones gaussianas.

Dado m pares de entradas-salidas $\{\mathbf{x}_i, \mathbf{y}_j\} \in \mathfrak{R}^n \times \mathfrak{R}^i, i = 1, 2, \dots, m$, la salida de la GRNN para la muestra de test, $\mathbf{x} \in \mathfrak{R}^n$, se define como:

$$\hat{y}(\mathbf{x}) = \sum_{i=1}^n w_i y_i \quad (5.19)$$

donde,

$$w_i = \frac{\exp\left(-\frac{\|\mathbf{x}-\mathbf{x}_i\|^2}{2\sigma^2}\right)}{\sum_{k=1}^m \exp\left(-\frac{\|\mathbf{x}-\mathbf{x}_k\|^2}{2\sigma^2}\right)} \quad (5.20)$$

La GRNN, en su aproximación clásica, tiene un número de neuronas igual al número de patrones o muestras de entrada, por lo que suele requerir un mayor número de neuronas que una red neuronal estándar de propagación hacia atrás de errores o retropropagación. Sin embargo, el coste computacional para diseñar una GRNN es generalmente menor que el necesario para diseñar una red de retropropagación.

El único parámetro ajustable de la GRNN es el *spread factor*. Cuando este parámetro es pequeño, la función RBF es estrecha y sólo un número reducido de muestras de entrada contribuyen en la obtención de la salida de la red, produciéndose una salida abrupta. Por otra parte, a medida que aumenta el valor de este parámetro, aumenta el ancho de la RBF y un mayor número de muestras contribuyen en la obtención de la salida obteniéndose, por tanto, una salida más suave. Consecuentemente, el *spread factor* determina el rendimiento de la GRNN. En la figura 5.8 se muestra la función de base radial resultante para varios valores del *spread factor*.

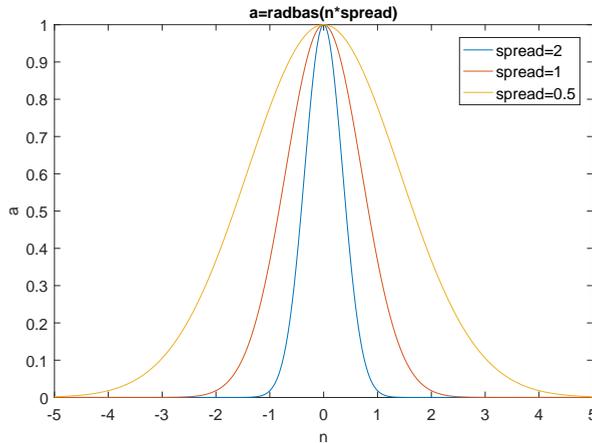


Figura 5.8

Función de base radial (RBF) para varios valores del factor de ensanchamiento.

Ajuste del *spread factor*. El ajuste se ha realizado mediante un barrido de este parámetro desde 0,01 hasta 0,1 en incrementos de 0,01 (10 arquitecturas) para los datos diarios y desde 0,05 hasta 0,5 en incrementos de 0,05 para los mensuales. En la figura 5.9 se muestran dos ejemplos del error en función del *spread factor*. La tabla 5.5 muestra el valor del *spread factor* óptimo para cada boya.

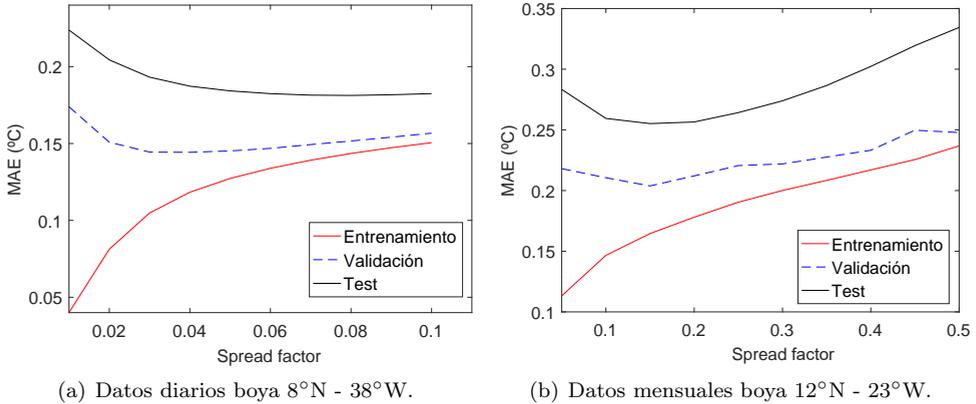


Figura 5.9
Índices MAE (°C) en función del *spread factor* de la GRNN.

Tabla 5.5
Factor de ensanchamiento de la GRNN con menor error de validación para SST y fecha como variables de entrada.

Boya	0N0E	0N10W	0N23W	0N35W	4N23W	4N38W	6S10W	8N38W	8S30W	10S10W	12N23W	12N38W	14S32W	15N38W	19S34W	20N38W	21N23W
Dia.	0,07	0,06	0,07	0,07	0,08	0,09	0,04	0,05	0,04	0,03	0,04	0,05	0,04	0,04	0,04	0,04	0,03
Men.	0,25	0,20	0,20	0,35	0,40	0,35	0,15	0,15	0,30	0,35	0,15	0,20	0,35	0,30	0,15	0,25	0,25

5.3.4.7. Árboles de regresión aleatorios

Los modelos basados en árboles de regresión se construyen dividiendo recursivamente el conjunto de datos mediante decisiones lógicas usando teoría de la información. Partiendo del nodo raíz, los patrones de entrenamiento se dividen en dos subconjuntos en función de los valores de un atributo de entrada. Esta división genera dos ramas con dos nuevos nodos. En cada nuevo nodo generado se aplicará una nueva decisión que volverá a segmentar el conjunto de datos. Este proceso se repite hasta alcanzar el límite de divisiones definido. El final de cada sección del árbol se denomina hoja, y contiene el valor de la predicción (Breiman et al., 1984; Witten et al., 2011). El resultado de este proceso se puede representar mediante un diagrama

como el mostrado en la figura 5.10, en el que se pueden observar las diferentes ramas, nodos y hojas generadas en un árbol de regresión aplicado a la boya 6°S - 8°E.

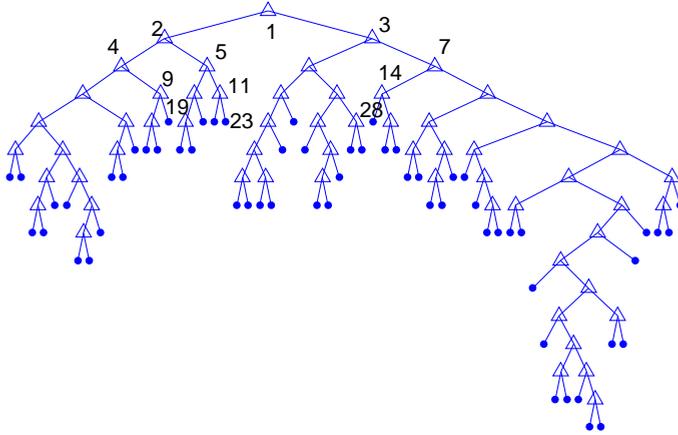


Figura 5.10

Ejemplo de árbol de decisión (boya 6°S - 8°E).

Las decisiones lógicas tomadas en los nodos numerados en la figura 5.10 son las siguientes:

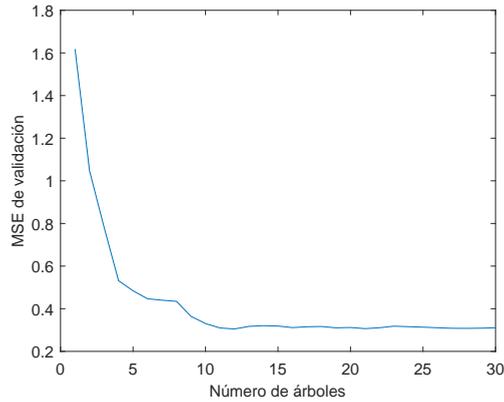
- *Nodo 1: if FechaCod1 < -0.78 then nodo 2 elseif FechaCod1 >= -0.78 then nodo 3 else 26.48*
- *Nodo 2: if FechaCod2 < 0.71 then nodo 4 elseif FechaCod2 >= 0.71 then nodo 5 else 23.59*
- *Nodo 3: if FechaCod2 < -1.07 then nodo 6 elseif FechaCod2 >= -1.07 then nodo 7 else 27.92*
- *Nodo 4: if FechaCod1 < -1.02 then nodo 8 elseif FechaCod1 >= -1.02 then nodo 9 else 23.14*
- *Nodo 5: if SST180 < 1.39 then nodo 10 elseif SST180 >= 1.39 then nodo 11 else 24.88*
- *Nodo 7: if FechaCod1 < -0.40 then nodo 14 elseif FechaCod1 >= -0.40 then nodo 15 else 28.89*
- *Nodo 9: if FechaCod1 < -0.84 then nodo 18 elseif FechaCod1 >= -0.841 then nodo 19 else 23.72*
- *Nodo 11: if SLP < -0.57 then nodo 22 elseif SLP >= -0.57 then nodo 23 else 25.54*

- *Nodo 14: if Salinity80m < 1.52 then nodo 28 elseif Salinity80m >= 1.52 then nodo 29 else 27.75*
- *Nodo 19: Predicción = 24.01*
- *Nodo 23: Predicción = 25.29*
- *Nodo 28: Predicción = 28.51*

Los árboles crecen (se ajustan) seleccionando con reemplazo el conjunto de entrenamiento mediante la técnica *Bootstrap AGGregatING* o *Bagging* (Breiman, 1996). Puesto que las predicciones con un único árbol pueden no ser óptimas, se agregan árboles al modelo para mejorar los resultados. Promediando los resultados del conjunto (o bosque) de árboles resultante, se obtiene la predicción final. Además, mediante la técnica *Random Subspace* se selecciona de manera aleatoria los predictores que se utilizan para dividir el conjunto de datos en cada decisión del árbol. Por tanto, los denominados árboles de regresión aleatorios o RFBT integran las técnicas árboles de regresión, *Bagging* y *Random Subspace* (Breiman, 1996, 2001). Los pasos que sigue el algoritmo RFBT son los siguientes:

1. Crear de manera aleatoria el conjunto de entrenamiento mediante la técnica *Bagging*. Al generarse con reemplazo, aproximadamente 2/3 de las muestras se utilizarán en el entrenamiento, mientras que el 1/3 restante se usará para la validación (muestras *out-of-bag*).
2. El conjunto creado mediante *Bagging* se asigna a un árbol de decisión. En cada nodo de decisión se realiza la división de los datos a partir de un subconjunto con m características escogidas de manera aleatoria. Para obtener la mejor división en cada nodo es necesario implementar una función objetivo (entropía o índice Gini).
3. Los pasos anteriores se repiten hasta alcanzar el número de árboles definido en el conjunto. Cada árbol se corresponderá con un conjunto aleatorio distinto.
4. La predicción final se obtiene promediando las predicciones de cada árbol.

Ajuste de los parámetros. El modelo de árbol de decisión escogido utiliza las técnicas de combinación de modelos *Bagging* y *Random Forest*. El algoritmo de *Bagging* se ha configurado para seleccionar aleatoriamente con reemplazo las muestras del conjunto de entrenamiento. Al utilizar reemplazo, aproximadamente el 33% de las muestras no se utilizan en el entrenamiento, reservándose para la validación (datos *Out of Bag*). Por otra parte el número de árboles del ensamblado se ha fijado en 30. Por último, para escoger los predictores en cada división de los datos se ha escogido el algoritmo *Random Forest*. En la figura 5.11 se muestra el error en función del número de árboles del conjunto. Se puede observar que a partir de 15 árboles el error se mantiene prácticamente constante.

**Figura 5.11**

Error medio cuadrático (MSE) en función del número de árboles para la boya $0^\circ\text{N} - 0^\circ\text{E}$.

5.3.4.8. Máquinas de vectores soporte

En problemas de clasificación, las máquinas de vectores soporte tratan de convertir un problema no separable en el espacio de entrada en uno separable en el denominado espacio de características, mediante la búsqueda del hiperplano de separación óptimo en este nuevo espacio de alta dimensionalidad (Vapnik, 1998). Cuando se utiliza en problemas de regresión, el objetivo es definir un intervalo en el que la diferencia entre la función de ajuste de los datos y el valor deseado sea menor que ϵ . Mediante este intervalo fijado por el valor de ϵ se define una región (“tubo”) alrededor de la salida deseada en la que deben encontrarse las predicciones (figura 5.12). Las salidas, por tanto, podrán encontrarse dentro o fuera del “tubo”, pudiéndose tratar el problema de regresión como uno de clasificación (Bishop, 2006).

Dados los patrones de entrada $\{x_i, t_i\}_{i=1}^N$, $\mathbf{x}_i \in \mathbb{R}^d$, $t_i \in \mathbb{R}$, una función lineal $y(x)$ en el espacio de características se define como (Bishop, 2006):

$$y(x) = \boldsymbol{\omega}^T \boldsymbol{\phi}_i(x) + b, \quad (5.21)$$

donde ϕ_i es la i -th característica, $\boldsymbol{\omega}$ es el vector de pesos y b es el sesgo. Introduciendo las variables de holgura ξ_i , $\hat{\xi}_i$ (*slack variables*) que indican cuando un punto está dentro o fuera del tubo, la función de error para el SVM de regresión puede definirse mediante la siguiente expresión:

$$C \sum_{i=1}^N (\xi_i + \hat{\xi}_i) + \frac{1}{2} \|\boldsymbol{\omega}\|^2 \quad (5.22)$$

donde C es el parámetro de regularización.

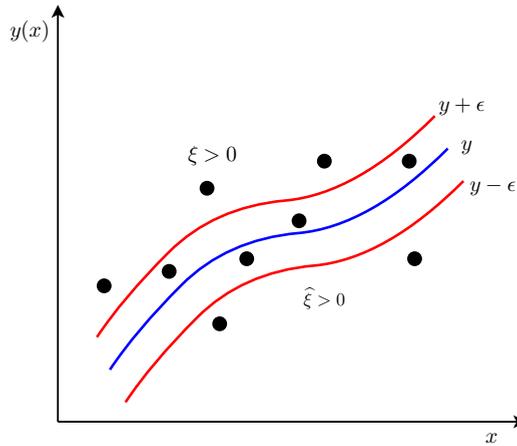


Figura 5.12

Esquema de la SVM para regresión. La curva azul representa la curva de regresión. Las curvas rojas marcan los límites del “tubo”, los puntos dentro del tubo tienen unos valores de ξ_i , $\hat{\xi}_i = 0$.

Esta expresión debe ser minimizada cumpliendo las siguientes condiciones:

$$t_i \leq y(\mathbf{x}_i) + \epsilon + \xi_i \tag{5.23}$$

$$t_i \geq y(\mathbf{x}_i) - \epsilon - \hat{\xi}_i \tag{5.24}$$

$$\xi_i, \hat{\xi}_i \geq 0 \tag{5.25}$$

Este problema de optimización con restricciones, puede resolverse introduciendo los multiplicadores de Lagrange $\alpha_i \geq 0$ y $\hat{\alpha}_i \geq 0$ en el denominado problema dual definido por la siguiente ecuación (Bishop, 2006):

$$\begin{aligned} \tilde{L}(\alpha, \hat{\alpha}) = & -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (\alpha_i - \hat{\alpha}_i)(\alpha_j - \hat{\alpha}_j) k(\mathbf{x}_i, \mathbf{x}_j) \\ & - \epsilon \sum_{i=1}^N (\alpha_i + \hat{\alpha}_i) + \sum_{i=1}^N (\alpha_i + \hat{\alpha}_i) t_i \end{aligned} \tag{5.26}$$

En esta expresión se ha introducido además el término correspondiente al *kernel* $k(\mathbf{x}_i, \mathbf{x}_j)$. Utilizando los *kernels* se consigue implementar una relación no lineal en las variables entre el espacio de entrada y el espacio de características. El *kernel* más utilizado es el gaussiano, definido en la ecuación 5.18.

Finalmente, las predicciones para un nuevo patrón de entrada se pueden obtener mediante la siguiente ecuación:

$$y(x) = \sum_{i=1}^N (\alpha_i - \hat{\alpha}_i) k(\mathbf{x}, \mathbf{x}_i) + b \quad (5.27)$$

La versión LS-SVM se propuso para obtener la solución al sistema lineal de ecuaciones presente en la SVM original mediante el ajuste por mínimos cuadrados, en lugar de utilizar *Convex Quadratic Programming* (QP) (Suykens et al., 2002).

Ajuste de los hiperparámetros. El ajuste de γ y σ^2 se realiza en dos pasos:

1. Mediante el algoritmo *Coupled Simulated Annealing* se determinan los parámetros de ajuste iniciales más adecuados.
2. La rutina de optimización *simplex* utiliza estos valores iniciales para realizar un ajuste fino de los parámetros (De Brabanter et al., 2010).

El parámetro de regularización γ determina el equilibrio entre minimización del error de entrenamiento y suavidad de la función estimada. Por otra parte, σ^2 es el parámetro de anchura de la función *kernel*. La función kernel utilizada es la RBF debido a su mejor rendimiento en comparación al *kernel* lineal. En la tabla 5.6 se muestran los estadísticos descriptivos de los hiperparámetros obtenidos en las predicciones mensuales de SST.

Tabla 5.6

Valor mediano de los hiperparámetros (γ y σ^2) de la LS-SVM para las predicciones mensuales.

	γ			σ^2		
	t+1	t+2	t+3	t+1	t+2	t+3
0N0E	32,5	1660,2	200,6	6,3	18,3	15,8
0N10W	10,2	47,1	2228,5	2,0	5,2	80,1
0N23W	85,2	5,0	279,4	4,6	3,5	22,5
0N35W	175,5	3,3	2,9	58,9	3,9	2,2
4N23W	2,6	1,0	20,8	1,7	1,2	6,4
4N38W	94,5	165,6	1748,6	36,3	10,3	34,8
6S10W	451,7	2804,2	6048,5	28,0	42,0	156,6
8N38W	10,4	4,0	28,0	1,0	1,7	2,4
8S30W	3431,6	131,5	3,9	354,8	56,4	9,4
10S10W	16,6	257,4	254,7	1,7	17,9	33,2
12N23W	19,8	15,0	81,5	2,3	4,1	4,8
12N38W	2263,4	785,1	1554,5	3697,5	8,5	9,0
14S32W	4701,3	1742,2	6230,9	102,5	42,4	54,8
15N38W	37,5	196,4	2015,3	2,9	4,3	9,9
19S34W	434,5	1069,9	3847,1	18,9	25,5	10,6
20N38W	223,6	278,7	2735,9	3,4	17,7	57,6
21N23W	55,2	3093,6	2893,4	12,6	60,2	62,7

5.3.4.9. Redes neuronales recurrentes LSTM

A diferencia de las redes neuronales tradicionales como el MLP, las redes neuronales recurrentes pueden almacenar representaciones de entradas recientes gracias a su esquema de conexiones retroalimentadas. Esta característica permite a este tipo de redes presentar un comportamiento temporal dinámico muy útil en tareas de reconocimiento de voz, composición musical o predicción de series temporales (Hochreiter y Schmidhuber, 1997). Sin embargo, la capacidad de aprendizaje de las RNN se ve limitada por los procedimientos de entrenamiento basados en descenso por gradiente como la retropropagación a través del tiempo (*Back-Propagation Through Time*) o el aprendizaje recurrente en tiempo real (*Real-Time Recurrent Learning*). Con estos algoritmos, la evolución de las señales de error retropropagadas en la red dependen de manera exponencial de la magnitud de los pesos y, por tanto, la retropropagación del error tiende rápidamente a desvanecerse o a aumentar sin control. Por este motivo, en la práctica, las RNN no logran funcionar adecuadamente en problemas que impliquen una relación entre entradas y salidas que supere los 5 - 10 retardos (*short-term memory*) (Bengio et al., 1994; Hochreiter y Schmidhuber, 1997; Gers et al., 2000).

Con la finalidad de remediar este problema, Hochreiter and Schmidhuber presentaron las redes LSTM en 1997. Las redes LSTM están compuestas por unidades de procesamiento denominadas celdas que permiten controlar el flujo del error de manera constante, extendiendo la capacidad de aprendizaje de la red hasta los 1000 retardos aproximadamente (*long-term memory*). El centro de estas celdas lo forma una conexión autorrecurrente (*constant error carousel*) cuyo acceso está controlado por una puerta de entrada (*input gate*) y una puerta de salida (*output gate*) (Hochreiter y Schmidhuber, 1997). Estas celdas fueron mejoradas posteriormente añadiendo la denominada puerta olvidar (*forget gate*) para permitir resetear el estado de la celda (Gers et al., 2000).

En la figura 5.13 se muestra como el bucle o recursividad de una LSTM puede ser entendido como una celda copiada múltiples veces, en la que la información de instantes temporales anteriores se transmite a las siguientes celdas. En esta figura también se detalla de manera gráfica los diferentes componentes de las unidades LSTM. Se puede observar que, a diferencia de una neurona clásica formada por un sumador y una función de activación, una celda LSTM está formada por múltiples elementos interconectados. El elemento principal es el denominado estado o memoria de la celda (C_t), representado por la línea horizontal superior del esquema. Esta memoria se retroalimenta de manera recursiva, y la modificación de la información que contiene está regulada por el resto de elementos de la celda. En el instante t , el vector de entrada X_t se suma con el vector de salida o estado oculto h_{t-1} . Esta suma se propaga a las puertas olvidar f_t , entrada i_t y salida o_t . La puerta olvidar, compuesta por una función de activación sigmoide y un multiplicador elemento a elemento (*Hadamard product*), controla la información de instantes anteriores que debe ser mantenida o eliminada de la memoria.

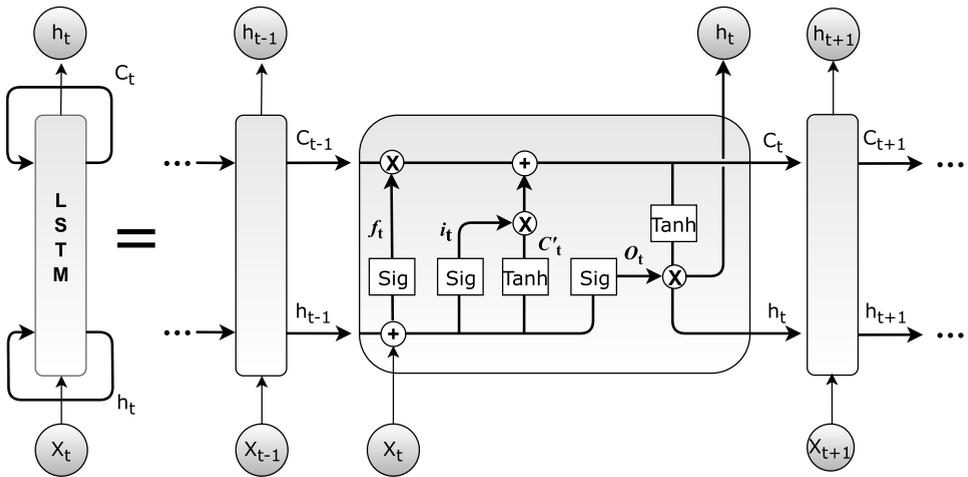


Figura 5.13

Esquema de una celda LSTM y propagación de la información a través de la misma. X_t es el vector de entrada, h_t el vector de salidas y C_t el estado o memoria de la celda. f_t , i_t , C'_t y O_t representan a la puerta olvidar, puerta de entrada, valor estado candidato y puerta de salida, respectivamente.

La salida en el instante t de la puerta olvidar se define como (Gers et al., 2000):

$$f_t = \sigma_g(W_f \mathbf{x}_t + R_f \mathbf{h}_{t-1} + b_f), \quad (5.28)$$

donde σ_g es una función de activación sigmoide, W es la matriz de pesos de entrada, R la matriz de pesos recurrentes y b el sesgo.

Por su parte, la puerta de entrada controla la información actual que modificará a la memoria de la celda y está formada por el vector i_t y el vector candidato C'_t :

$$i_t = \sigma_g(W_i \mathbf{x}_t + R_i \mathbf{h}_{t-1} + b_i) \quad (5.29)$$

$$C'_t = \sigma_c(W_C \mathbf{x}_t + R_C \mathbf{h}_{t-1} + b_C), \quad (5.30)$$

donde σ_c es una función de activación tangente hiperbólica.

Una vez obtenidas f_t , i_t y C'_t , estas señales se combinan para actualizar el estado de la celda C_t de la siguiente forma:

$$C_t = f_t \odot C_{t-1} + i_t \odot C'_t, \quad (5.31)$$

donde \odot es una multiplicación elemento a elemento.

La puerta de salida controla el valor h_t de salida que se obtiene en la celda, combinando el vector o_t con el estado C_t como sigue:

$$o_t = \sigma_g(W_o \mathbf{x}_t + R_o \mathbf{h}_{t-1} + b_o) \tag{5.32}$$

$$h_t = o_t \odot \sigma_c(C_t) \tag{5.33}$$

Por último, la salida h_t se retroalimenta en la celda y se propaga hacia la siguiente capa.

Ajuste de la arquitectura. La capa de entrada de la red (*sequenceInputLayer*) se ha configurado con un tamaño igual al número de entradas. El ajuste de la capa oculta (*lstmLayer*) se ha realizado mediante un barrido del número de unidades LSTM. Para los datos mensuales, el barrido se ha realizado desde dos hasta 10 en pasos de dos, mientras que para los datos diarios se ha realizado de cinco a 45 en pasos de cinco. Adicionalmente, se comprobaron los resultados diarios con dos y cuatro unidades. En la tabla 5.7 se muestran las arquitecturas con el menor error de validación en cada caso. Finalmente, la salida de la red está formada por una capa totalmente conectada (*fullyConnectedLayer*) de tamaño igual al número de salidas y una capa de regresión (*regressionLayer*).

La inicialización de los pesos de entrada se ha realizado mediante el inicializador *glorot* (Glorot y Bengio, 2010), mientras que para los pesos recurrentes se ha empleado una matriz de pesos aleatorios ortogonal. En cuanto a la configuración de los parámetros de entrenamiento, se ha escogido el optimizador adam (*adaptive moment estimation*) (Kingma y Ba, 2015), con un número máximo de iteraciones de 1000, un umbral de gradiente de 1 y una tasa de aprendizaje de 0,005. Por último, el entrenamiento se ha configurado de manera que, cada de 10 iteraciones del modelo, se compruebe el error cometido en los datos de validación y finalice el proceso de entrenamiento si, en tres comprobaciones consecutivas, no disminuye este error.

Tabla 5.7

Número óptimo de unidades LSTM para cada boya en los conjuntos mensuales y diarios de SST entrenando la red en modo directo y modo recurrente.

Boya		0N0E	0N10W	0N23W	0N35W	4N23W	4N38W	6S10W	8N38W	8S30W	10S10W	12N23W	12N38W	14S32W	15N38W	19S34W	20N38W	21N23W
Diarios	Directo	4	4	10	2	10	5	4	10	2	2	5	10	4	2	2	5	2
	Recurrente	4	4	4	4	4	4	2	4	4	2	4	4	4	4	4	4	4
Mensuales	Directo	10	10	8	4	2	2	2	2	4	8	6	4	6	2	2	2	2
	Recurrente	4	4	4	2	2	4	6	2	2	4	6	8	6	4	10	6	6

5.3.4.10. Sistema de pronóstico climático NCEP-CFSv2

El *North American Multi-Model Ensemble* (NMME) es un sistema de predicción climática estacional que combina los pronósticos de los principales modelos climáticos norteamericanos. Este proyecto multiagencia está formado por la NOAA, la *National Science Foundation* (NSF), la *National Aeronautics and Space Administration* (NASA) y por el *U.S. Department of Energy* (DOE). La NOAA comenzó en 2011 la primera fase del proyecto, uniéndose en 2012 el resto de agencias para iniciar la segunda fase. En las pruebas realizadas se ha demostrado que este sistema, gracias a la diversidad de modelos climáticos utilizados, presenta una habilidad superior a la de otros sistemas de predicción en el pronóstico estacional (Kirtman et al., 2014).

Uno de los modelos que integra el NMME es el sistema de pronóstico climático del *National Center for Environmental Prediction* (NCEP), conocido como NCEP-CFSv2. La primera versión (CFSv1) fue implementada en agosto de 2004, siendo el primer sistema de carácter casi global y formado por modelos de interacción océano-tierra-atmósfera plenamente acoplados que se utilizó en el NCEP para las predicciones estacionales. La versión mejorada (CFSv2) comenzó a utilizarse en marzo de 2011 y permite obtener predicciones mensuales de la SST, la precipitación total acumulada o el pronóstico del tiempo con un horizonte de 9 meses y una resolución zonal de 1° (Saha et al., 2006, 2014).

Al igual que otros métodos numéricos, CFSv2 utiliza información promedio de grandes extensiones espaciales, por lo que existe un cierto error entre los valores reales observados y las predicciones en puntos concretos, como por ejemplo, en las posiciones de la boyas (Patil et al., 2016). En el apartado 5.4 se compararán las predicciones mensuales de la SST proporcionadas por el modelo CFSv2 con las obtenidas mediante los diferentes métodos de aprendizaje automático.

Los datos de este modelo se han descargado de la página web de la *International Research Institute for Climate and Society* (IRI) (IRI, 2019). En esta misma página es posible descargar las predicciones del resto de modelos del NMME. Para obtener los datos de SST correspondientes a las posiciones de las boyas, se descargaron en primer lugar todos los datos comprendidos en el área $20^\circ\text{S} - 22^\circ\text{N}$ de latitud y $39^\circ\text{W} - 1^\circ\text{E}$ de longitud. Posteriormente, se extrajeron los valores de SST de las coordenadas correspondientes a las posiciones de las boyas y se promediaron las predicciones de los 28 miembros que forman el modelo CFSv2. Además, los datos se procesaron teniendo en cuenta que los registros constan de dos partes:

1. Desde enero de 1982 hasta marzo de 2011 las predicciones son retrospectivas (denominadas *hindcast*). Estas predicciones se utilizan para calibrar y validar los modelos.
2. Desde abril 2011 hasta la actualidad las predicciones son prospectivas (denominadas *forecast*).

5.4. Resultados

5.4.1. Experimentos preliminares

Las rectas de ajuste obtenidas mediante regresión lineal presentaron una tendencia promedio (en valor absoluto) en las 17 boyas de $6,71 \cdot 10^{-5}$ °C/día, por lo que para una predicción a 30 días, la tendencia promedio sería de 0,002 °C. Teniendo en cuenta estos resultados, la tendencia no debería tener efectos significativos en las predicciones.

En cuanto a la estacionalidad, en la figura 5.14 se puede observar la autocorrelación de las 2000 primeras muestras de SST de la boya 21°N - 23°W. Tal y como se preveía, la gráfica muestra que existe una periodicidad en los datos que se repite cada 365 muestras aproximadamente.

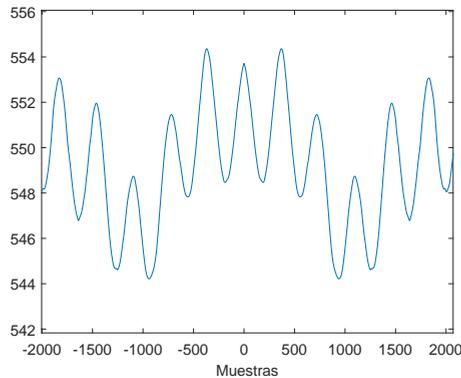
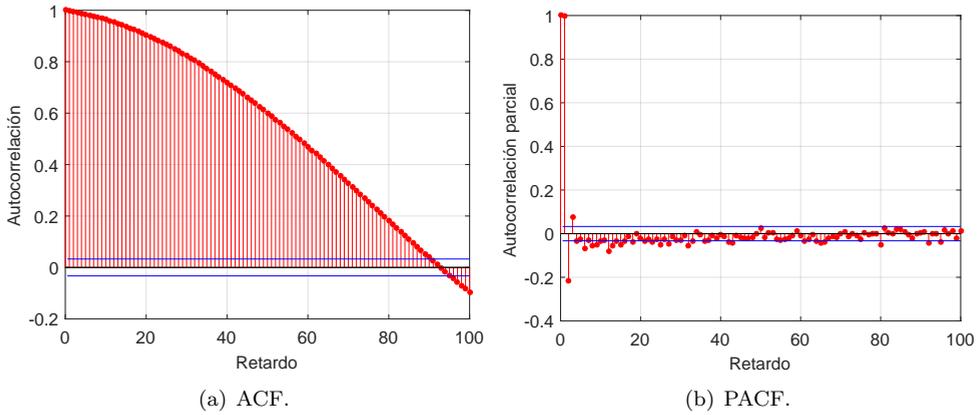


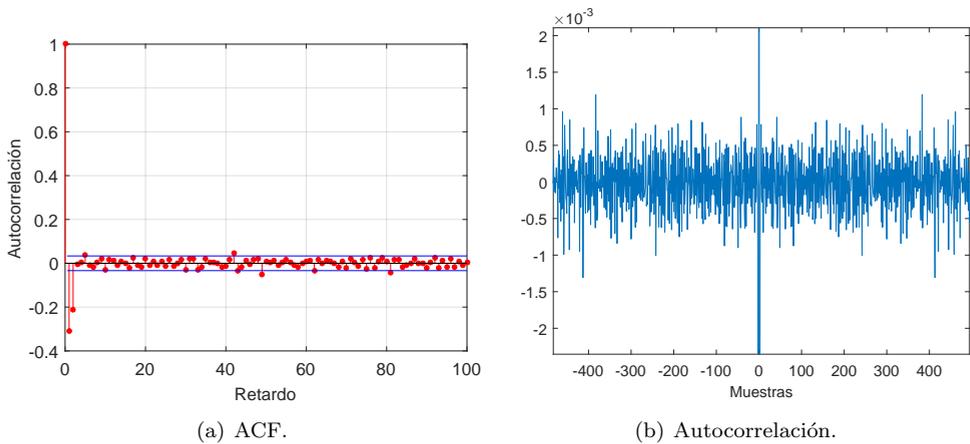
Figura 5.14
Autocorrelación de la variable SST en la boya 21°N - 23°W.

Por otro lado, como se muestra en la figura 5.15.a, el valor de la ACF para la SST en la boya 21°N - 23°W decae de manera gradual, indicando que la serie es no estacionaria y que existe una fuerte correlación entre la muestra actual y las muestras anteriores. Además, en la figura 5.15.b se observa que el valor de PACF es 1 para el primer retardo y, a pesar de que los retardos siguientes continúan presentando correlaciones significativas (en el exterior de los intervalos de confianza), la correlación decae bruscamente en el segundo retardo. El resto de boyas presentaron un comportamiento similar. Con estos resultados, y según el análisis clásico de series temporales, las predicciones se podrían realizar mediante modelos autorregresivos de tipo AR(1).

En la figuras 5.16.a y 5.16.b se muestra el efecto de aplicar la primera derivada a la SST en la boya 21°N - 23°W. Comparado con la figura 5.15.a, ahora la ACF decae bruscamente y presenta un comportamiento aleatorio tras unos pocos retardos, indicando que la serie es estacionaria. Además, en la figura 5.16.b se puede ver que ha desaparecido la periodicidad o ciclo anual.

**Figura 5.15**

ACF y PACF de la SST en la boya 21°N - 23°W.

**Figura 5.16**

ACF y autocorrelación de la SST tras diferenciar (boya 21°N - 23°W).

En las figuras 5.17.a, 5.18.a y 5.19.a se muestra el valor promedio y desviación estándar de la SST a lo largo del año para las boyas 20°N - 38°W, 0°N - 10°W y 19°S - 34°W. Se puede observar el comportamiento contrario que presenta la boya localizada al norte (temperaturas máximas en octubre y mínimas en marzo) comparada con la situada en el sur (temperaturas máximas en marzo/abril y mínimas en septiembre/octubre). También se aprecia que la boya situada en el ecuador (figura 5.18.a) alcanza unas temperaturas máximas y mínimas superiores a las otras dos boyas. Por su parte, en las figuras 5.17.b, 5.18.b y 5.19.b se muestran las series temporales diarias

de la SST superpuestas con la climatología diaria. Restando estas climatologías a las series de SST se consigue atenuar o eliminar la componente estacional. Además, las series de climatología diaria se pueden utilizar como modelo básico para realizar las predicciones de SST.

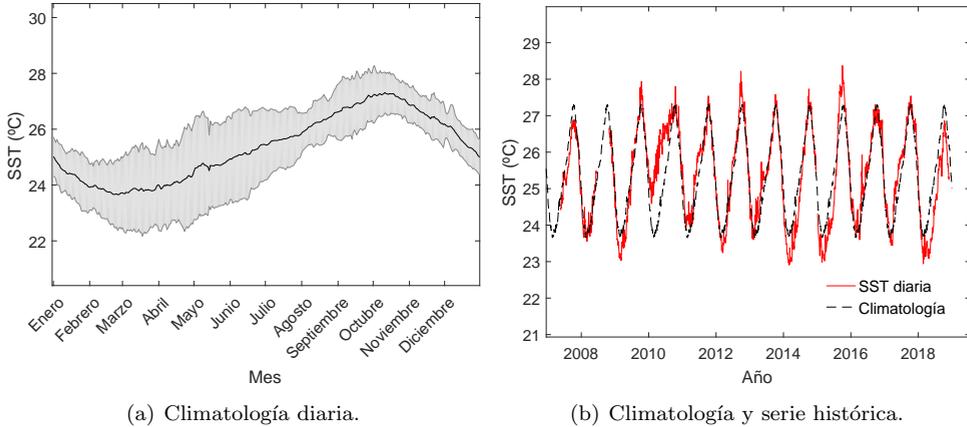


Figura 5.17
Climatología diaria y serie histórica de la SST para la boya 20°N - 38°W.

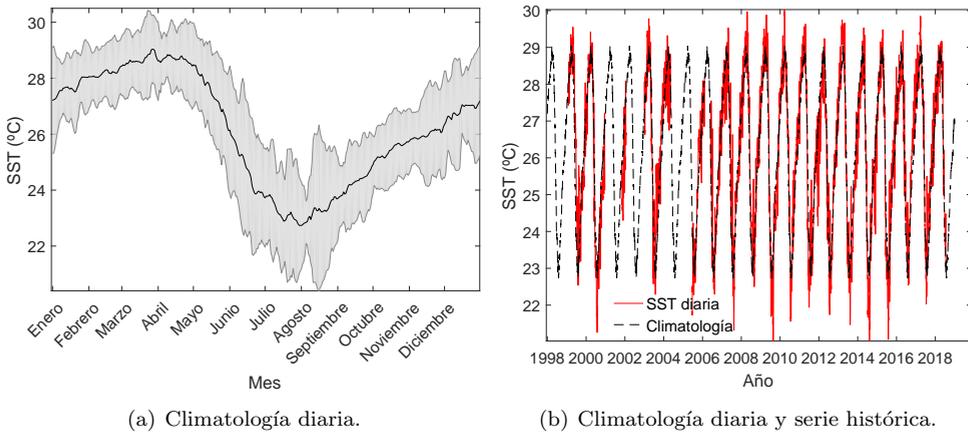
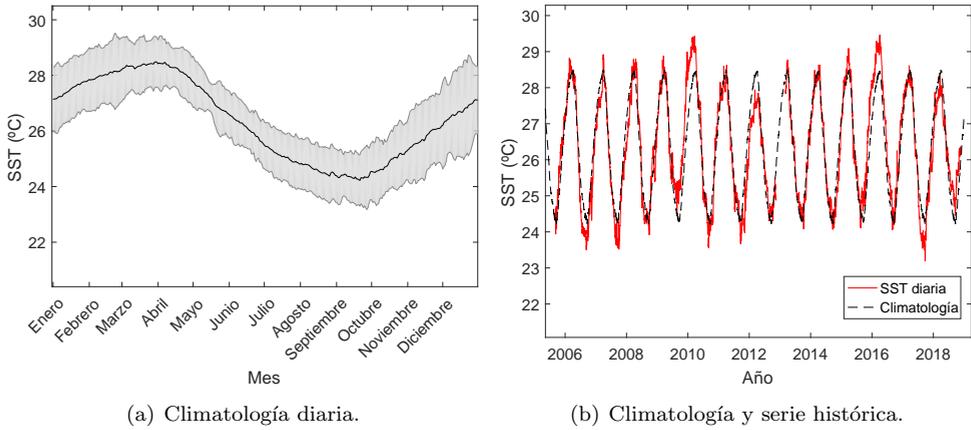


Figura 5.18
Climatología y serie histórica de la SST para la boya 0°N - 10°W.

Las medianas de los resultados obtenidos en los experimentos preliminares se detallan en la tabla 5.8. En esta tabla se observa que, aunque el análisis mediante ACF y PACF indicaba una alta correlación con muestras anteriores, añadiendo retardos de la SST como entradas no se consigue una reducción relevante en el índice MAE.

**Figura 5.19**

Climatología diaria y serie histórica de la SST para la boya 19°S - 34°W.

Tabla 5.8

Índices MAE (°C) obtenidos en datos diarios de test en las pruebas preliminares con la ELM (mediana de las 17 boyas).

	T+7	T+15	T+30
Solo SST	0,22	0,37	0,59
SST + 1 retardo	0,22	0,36	0,56
SST + 2 retardos	0,22	0,36	0,55
SST + 7 retardos	0,24	0,35	0,55
SST + 15 retardos	0,31	0,41	0,53
SST sin tendencia (primera derivada)	0,22	0,37	0,62
SST sin tendencia (segunda derivada)	0,23	0,38	0,63
SST sin comp. estacional (clima.)	0,21	0,27	0,34
Fecha (año, mes y día) + SST	0,27	0,39	0,48
Fecha (mes y día) + SST	0,20	0,28	0,37
Fecha cod	0,44	0,44	0,46
Fecha cod + SST	0,21	0,28	0,33
Fecha cod + SST + 2 retardos	0,21	0,29	0,33
Fecha cod + SST, sin comp. esta. (clima.)	0,21	0,28	0,35

En los siguientes experimentos, aplicando la primera y segunda derivada para eliminar la tendencia, tampoco se consiguió reducir el error cometido. Sin embargo, eliminando la componente estacional restando la climatología a la serie de SST, sí que se obtuvo una reducción del error, sobre todo en las predicciones en los horizontes $t + 15$ y $t + 30$. A continuación se comprobó la influencia de añadir la fecha en las predicciones.

El mejor resultado se obtuvo utilizando la SST y codificando la fecha como se detalla en la sección 5.3.1 (día del año (1-365) y aplicando coseno y seno). Finalmente, eliminar la componente estacional de la SST o añadir retardos no resultó en una mejora de las predicciones. Analizando los resultados se puede deducir que, utilizando

modelos de regresión, la época del año (componente estacionaria) tiene una mayor influencia sobre las predicciones que los valores anteriores de la serie. De hecho, la fecha es más importante que los propios valores actuales de la SST en predicciones a largo plazo (en $t + 30$, MAE = 0,59 °C sólo con SST y MAE = 0,46 °C sólo con fecha codificada). Además, los resultados muestran que añadir la fecha codificada como entrada aporta información sobre la estacionariedad de los datos al modelo, consiguiéndose el mismo resultado que eliminando la componente estacionaria en el caso de utilizar únicamente la SST como entrada.

5.4.2. Modelo trivial y modelo climatología

En esta sección se analizan los resultados de los modelos trivial y climatología en los datos diarios. En la tabla 5.9 se detallan los índices MAE que se obtienen aplicando estos modelos en los datos de test de las 17 boyas. En el caso del modelo climático se obtiene una mediana del índice MAE de 0,45 °C. Puesto que este método se basa en utilizar el valor promedio como la predicción, el error obtenido se mantiene constante a pesar de aumentar el horizonte de predicción. En cuanto al modelo trivial, debido a que utiliza el valor actual como predicción, el error cometido aumenta al incrementar el horizonte de predicción. Se puede observar en la tabla 5.9 que el error obtenido con el modelo trivial es inferior al del modelo climático hasta el horizonte $t + 15$ y, en $t + 30$, el modelo climatología obtiene menor error que el modelo trivial.

Tabla 5.9

Modelo climatología y modelo trivial diario en el conjunto de test.

	Clima.	T+1	T+2	T+3	T+4	T+5	T+6	T+7	T+15	T+30
0N0E	0,69	0,16	0,28	0,39	0,49	0,57	0,62	0,84	0,65	0,97
0N10W	0,54	0,15	0,27	0,37	0,46	0,54	0,61	0,68	0,81	1,12
0N23W	0,45	0,13	0,22	0,28	0,32	0,36	0,38	0,42	0,51	0,78
0N35W	0,24	0,09	0,15	0,19	0,22	0,25	0,27	0,26	0,32	0,42
4N23W	0,41	0,12	0,17	0,20	0,23	0,24	0,25	0,25	0,32	0,44
4N38W	0,28	0,06	0,09	0,11	0,12	0,13	0,14	0,19	0,24	0,26
6S10W	0,31	0,06	0,09	0,12	0,15	0,18	0,20	0,23	0,41	0,72
8N38W	0,29	0,08	0,12	0,15	0,17	0,19	0,21	0,22	0,33	0,50
8S30W	0,19	0,04	0,06	0,08	0,09	0,11	0,12	0,15	0,25	0,42
10S10W	0,38	0,05	0,07	0,10	0,12	0,14	0,16	0,17	0,32	0,59
12N23W	0,49	0,09	0,14	0,17	0,21	0,24	0,27	0,29	0,51	0,90
12N38W	0,57	0,07	0,11	0,13	0,16	0,18	0,20	0,21	0,40	0,64
14S32W	0,25	0,06	0,09	0,11	0,13	0,15	0,17	0,18	0,30	0,49
15N38W	0,52	0,04	0,07	0,10	0,12	0,14	0,16	0,17	0,31	0,59
19S34W	0,49	0,08	0,12	0,15	0,17	0,19	0,20	0,23	0,38	0,61
20N38W	0,71	0,06	0,09	0,12	0,15	0,17	0,19	0,21	0,38	0,66
21N23W	0,64	0,08	0,13	0,17	0,21	0,24	0,28	0,31	0,56	1,02
Mediana	0,45	0,08	0,12	0,15	0,17	0,19	0,20	0,23	0,38	0,61

5.4.3. Modelos diarios

5.4.3.1. Fecha codificada y SST como variables de entrada

Teniendo en cuenta los resultados de los experimentos preliminares, en primer lugar se obtuvieron las predicciones de SST utilizando la SST observada y la fecha codificada como variables de entrada. En la tabla 5.10 se muestran los resultados obtenidos para cada boya y cada modelo para las predicciones en el horizonte $t+7$. En todos los casos se muestran los índices MAE de los conjuntos de test correspondientes a la arquitectura (o ajuste de parámetros) con los mejores resultados en los datos de validación. Analizando el valor mediano de las 17 boyas (última fila de la tabla) se observa que el modelo RMLR obtiene el menor índice de error (0,19 °C), seguido por los modelos ELM, RFBT y MLP (0,20 °C), trivial (0,23 °C), GRNN y LS-SVM (0,26 °C) y, por último, el modelo climatología (0,45 °C). Aunque el modelo RMLR obtiene el mejor resultado global, analizando las boyas por separado se observa que los modelos ELM, RFBT y MLP consiguen igualar o mejorar el error en 10, 6 y 8 de las boyas respectivamente.

Tabla 5.10

Índices MAE (°C) de test obtenidos en las predicciones $t+7$ para cada boya utilizando la fecha codificada y la SST como variables de entrada. Los mejores resultados aparecen resaltados en negrita.

	Trivial	Clima.	RMLR	ELM	LS-SVM	GRNN	RFBT	MLP
0N0E	0,84	0,69	0,71	0,63	0,67	0,65	0,66	0,64
0N10W	0,68	0,54	0,60	0,52	0,59	0,59	0,56	0,52
0N23W	0,42	0,45	0,39	0,37	0,42	0,42	0,38	0,37
0N35W	0,26	0,24	0,23	0,23	0,26	0,25	0,23	0,22
4N23W	0,25	0,41	0,24	0,24	0,26	0,26	0,24	0,24
4N38W	0,19	0,28	0,18	0,18	0,19	0,19	0,18	0,18
6S10W	0,23	0,31	0,18	0,18	0,19	0,19	0,17	0,18
8N38W	0,22	0,29	0,20	0,18	0,21	0,20	0,19	0,19
8S30W	0,15	0,19	0,11	0,11	0,15	0,15	0,11	0,11
10S10W	0,17	0,38	0,14	0,15	0,18	0,18	0,15	0,15
12N23W	0,29	0,49	0,25	0,25	0,40	0,36	0,28	0,26
12N38W	0,21	0,57	0,19	0,20	0,33	0,27	0,20	0,21
14S32W	0,18	0,25	0,16	0,16	0,21	0,22	0,16	0,16
15N38W	0,17	0,52	0,13	0,12	0,18	0,16	0,12	0,12
19S34W	0,23	0,49	0,19	0,20	0,25	0,26	0,22	0,20
20N38W	0,21	0,71	0,18	0,20	0,36	0,28	0,19	0,19
21N23W	0,31	0,64	0,22	0,23	0,40	0,32	0,26	0,23
Mediana	0,23	0,45	0,19	0,20	0,26	0,26	0,20	0,20

Estos resultados indican que todos los modelos han conseguido mejorar al modelo climatología. Sin embargo, los modelos GRNN y LS-SVM no han conseguido mejorar la predicción del modelo trivial. Para este horizonte, la diferencia mediana entre el mejor modelo global (RMLR) y el modelo trivial es de tan solo 0,04 °C, mientras que la mayor diferencia con respecto al modelo trivial se da en la boya 0°N - 0°E comparando con el modelo ELM (0,21 °C). También cabe destacar la gran diferencia que existe en los valores MAE de la diferentes boyas. Por ejemplo, según el modelo trivial, la boya 8°S - 30°W presenta un MAE de 0,15 °C y la boya 0°N - 0°E un MAE

de 0,84 °C. Este hecho refleja que cada boya presenta una variabilidad diaria y/o un rango de valores de SST diferentes.

Las predicciones en $t + 15$ se muestran en la tabla 5.11. Al aumentar el horizonte de predicción aumenta el valor del índice MAE en todos los modelos, excepto en el modelo climatología. El mayor aumento del índice MAE se ha dado en el modelo trivial (0,15 °C). Los modelos ELM y MLP obtienen el menor error (0,28 °C), seguido por los modelos RMLR (0,29 °C), los RFBT (0,30 °C), GRNN (0,32 °C), LS-SVM (0,33 °C), el modelo trivial (0,38 °C) y el modelo climatología (0,45 °C). En este horizonte de predicción todos los modelos consiguen mejorar a los modelos trivial y climatología, siendo la diferencia mediana con respecto a los mejores modelos (ELM y MLP) de 0,10 °C y 0,17 °C respectivamente. Como en el caso anterior, los modelos con las menores medianas no son los modelos con el menor error en todas las boyas.

Tabla 5.11

Índices MAE (°C) de test obtenidos en las predicciones $t + 15$ para cada boya utilizando la fecha codificada y la SST como variables de entrada. Los mejores resultados aparecen resaltados en negrita.

	Trivial	Clima.	RMLR	ELM	LS-SVM	GRNN	RFBT	MLP
0N0E	0,65	0,69	0,56	0,52	0,62	0,59	0,54	0,52
0N10W	0,81	0,54	0,69	0,56	0,62	0,61	0,60	0,56
0N23W	0,51	0,45	0,45	0,39	0,43	0,44	0,41	0,40
0N35W	0,32	0,24	0,26	0,24	0,26	0,27	0,25	0,24
4N23W	0,32	0,41	0,31	0,29	0,33	0,35	0,30	0,29
4N38W	0,24	0,28	0,21	0,21	0,24	0,23	0,21	0,20
6S10W	0,41	0,31	0,29	0,27	0,30	0,29	0,26	0,27
8N38W	0,33	0,29	0,29	0,26	0,25	0,26	0,25	0,26
8S30W	0,25	0,19	0,15	0,15	0,20	0,19	0,16	0,15
10S10W	0,32	0,38	0,22	0,23	0,25	0,26	0,24	0,23
12N23W	0,51	0,49	0,35	0,37	0,57	0,51	0,42	0,37
12N38W	0,40	0,57	0,30	0,34	0,49	0,45	0,36	0,34
14S32W	0,30	0,25	0,24	0,21	0,30	0,29	0,22	0,21
15N38W	0,31	0,52	0,20	0,18	0,27	0,22	0,20	0,17
19S34W	0,38	0,49	0,27	0,28	0,33	0,32	0,32	0,28
20N38W	0,38	0,71	0,27	0,30	0,43	0,33	0,32	0,28
21N23W	0,56	0,64	0,34	0,33	0,47	0,37	0,39	0,32
Mediana	0,38	0,45	0,29	0,28	0,33	0,32	0,30	0,28

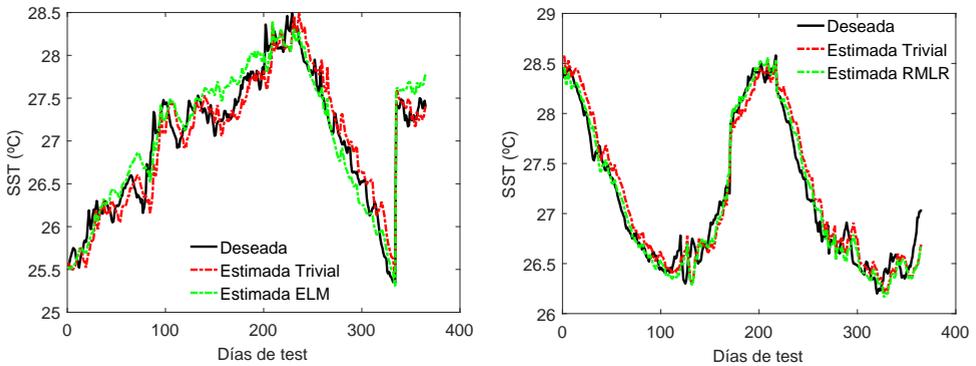
Por último, en la tabla 5.12 se detallan los resultados en el horizonte $t + 30$. Nuevamente, los índices de error han aumentado en todos los casos y, los modelos ELM y MLP obtienen los mejores resultados (0,34 °C). El orden del resto de modelos es: RFBT (0,35 °C), GRNN (0,36 °C), LS-SVM (0,38 °C), RMLR (0,39 °C), climatología (0,45 °C) y el modelo trivial (0,61 °C). En este horizonte, todos los modelos de predicción obtienen menor índice de error que el modelo trivial y con mayor margen (0,27 °C comparado con los mejores modelos). Por lo tanto, a mayor horizonte, los modelos de predicción obtienen un mejor resultado frente al modelo trivial. Sin embargo, puesto que el modelo climatología presenta siempre el mismo error, la ventaja del resto de modelos se ve reducida. Esta diferencia era de 0,26 °C en $t + 7$ y se ha reducido a 0,11 °C en $t + 30$. Puesto que existe poca diferencia en el error cometido por los modelos, en la sección posterior de análisis de resultados (sección 5.4.6) se estudiará si estas diferencias son estadísticamente significativas.

Tabla 5.12

Índices MAE ($^{\circ}\text{C}$) de test obtenidos en las predicciones $t + 30$ para cada boya utilizando la fecha codificada y la SST como variables de entrada. Los mejores resultados aparecen resaltados en negrita.

	Trivial	Clima.	RMLR	ELM	LS-SVM	GRNN	RFBT	MLP
0N0E	0,97	0,69	0,68	0,55	0,61	0,61	0,57	0,54
0N10W	1,12	0,54	0,75	0,57	0,66	0,65	0,61	0,57
0N23W	0,78	0,45	0,62	0,45	0,45	0,47	0,44	0,44
0N35W	0,42	0,24	0,30	0,25	0,24	0,26	0,25	0,25
4N23W	0,44	0,41	0,43	0,31	0,38	0,38	0,31	0,31
4N38W	0,26	0,28	0,20	0,20	0,22	0,22	0,20	0,20
6S10W	0,72	0,31	0,38	0,31	0,35	0,33	0,29	0,32
8N38W	0,50	0,29	0,41	0,34	0,32	0,32	0,32	0,34
8S30W	0,42	0,19	0,21	0,21	0,25	0,24	0,21	0,21
10S10W	0,59	0,38	0,34	0,34	0,38	0,35	0,35	0,35
12N23W	0,90	0,49	0,47	0,44	0,60	0,59	0,51	0,47
12N38W	0,64	0,57	0,39	0,43	0,53	0,52	0,43	0,44
14S32W	0,49	0,25	0,30	0,22	0,30	0,30	0,22	0,22
15N38W	0,59	0,52	0,28	0,22	0,30	0,24	0,21	0,22
19S34W	0,61	0,49	0,32	0,33	0,34	0,36	0,37	0,34
20N38W	0,66	0,71	0,41	0,42	0,55	0,43	0,42	0,42
21N23W	1,02	0,64	0,50	0,44	0,67	0,49	0,47	0,44
Mediana	0,61	0,45	0,39	0,34	0,38	0,36	0,35	0,34

En las figuras 5.20, 5.21 y 5.22 se muestran algunos ejemplos de series temporales observadas de SST y las predicciones obtenidas. Como indicaban los resultados obtenidos en el horizonte $t + 7$ (tabla 5.10), se puede observar en las figuras 5.20.a y 5.20.b que existe poca diferencia entre los errores cometidos por los modelos de predicción y el modelo trivial.

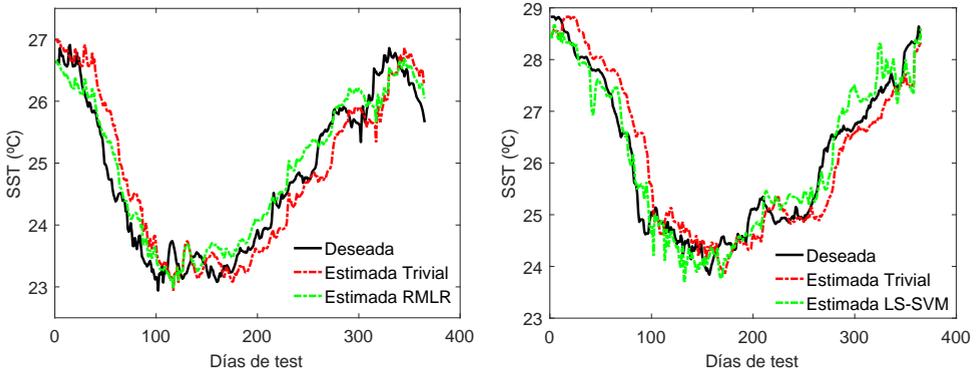


(a) Boya 14S32W. MAE ($^{\circ}\text{C}$): trivial=0,18; (b) Boya 8S30W. MAE ($^{\circ}\text{C}$): trivial=0,15; ELM=0,16. Fecha: 2016/09/02 a 2018/02/14. RMLR=0,11. Fecha: 2016/06/02 a 2018/10/30.

Figura 5.20

Ejemplos de series temporales diarias deseadas y estimadas de test para la SST en el horizonte $t + 7$ días. Resultados con la SST y fecha codificada como variables de entrada a los modelos.

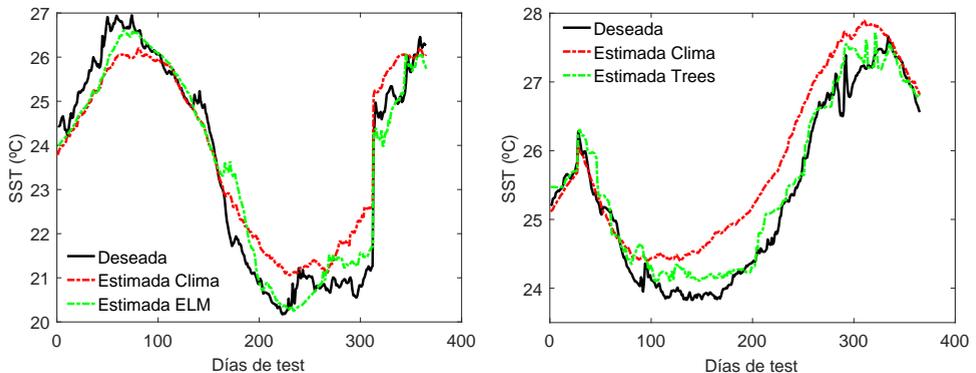
En el horizonte $t + 15$ aumenta esta diferencia con el modelo trivial (figuras 5.21.a y 5.21.b). Por último, las figuras 5.22.a y 5.22.b muestran dos ejemplos en los que los modelos de predicción mejoran claramente al modelo climatología en $t + 30$. En general, estos ejemplos muestran que los modelos consiguen predecir con bastante exactitud el ciclo anual, aunque no logran ajustarse con tanta precisión al ciclo diario debido a su alta variabilidad.



(a) Boya 20N38W. MAE (°C): trivial=0,38; (b) Boya 6S10W. MAE (°C): trivial=0,41; LS-RMLR=0,27. Fecha: 2017/11/16 a 2018/11/15. SVM=0,26. Fecha: 2017/04/16 a 2018/04/27.

Figura 5.21

Ejemplos de series temporales diarias deseadas y estimadas de test para la SST en el horizonte $t + 15$ días. Resultados con la SST y fecha codificada como variables de entrada a los modelos.



(a) Boya 21N23W. MAE (°C): clima.=0,64; (b) Boya 15N38W. MAE (°C): clima.=0,52; ELM=0,44. Fecha: 2017/07/09 a 2018/10/03. RFBT=0,21. Fecha: 2016/06/11 a 2018/11/30.

Figura 5.22

Ejemplos de series temporales diarias deseadas y estimadas de test para la SST en el horizonte $t + 30$ días. Resultados con la SST y fecha codificada como variables de entrada a los modelos.

5.4.3.2. Modelos con todas las variables de entrada

Tras analizar los resultados utilizando únicamente la fecha y la SST como entradas, se obtuvieron las predicciones de SST utilizando todas las variables de entrada disponibles en cada boya para intentar mejorar la predicciones. En la tabla 5.13 se indica el número de variables de cada conjunto de datos (ver también tabla 5.2) y, en la tabla 5.14, se exponen los resultados obtenidos. A diferencia de los resultados presentados en el anterior apartado, ahora se muestra únicamente el índice de error mediano de las 17 boyas.

Tabla 5.13

Número de variables en cada boya.

Boya	0N0E	0N10W	0N23W	0N35W	4N23W	4N38W	6S10W	8N38W	8S30W	10S10W	12N23W	12N38W	14S32W	15N38W	19S34W	20N38W	21N23W
Nvars	30	31	30	28	36	30	30	30	26	30	36	26	31	31	27	33	33

Empleando todas las variables disponibles, el modelo RMLR obtuvo el menor índice de error en los tres horizontes de predicción, seguido por los RFBT, MLP, LS-SVM, GRNN, trivial, climatología y ELM. En $t + 7$, únicamente el modelo RMLR mejora al modelo trivial. En $t + 15$, mejoran al modelo trivial los modelos RMLR y RFBT mientras que, en $t + 30$, todos los modelos excepto ELM mejoran al modelo trivial. Además, en $t + 30$, los modelo trivial, ELM y GRNN, no mejoran al modelo climatología. Comparando estos resultados con los obtenidos con fecha y SST como entradas, los índices de error del modelo RMLR se mantienen prácticamente sin cambios, mientras que en el resto de modelos empeoran significativamente. Como se ve, de todos los modelos, la ELM es el que mayor aumento del error presenta, indicando que es el menos robusto frente a los problemas que presentan los conjuntos de datos de alta dimensionalidad (variables correlacionadas, poco informativas, etc.).

Tabla 5.14

Índices MAE (°C) de test (mediana de las 17 boyas) obtenidos en los tres horizontes de predicción utilizando todas las variables de entrada. Los mejores resultados aparecen resaltados en negrita.

	Trivial	Clima.	RMLR	ELM	LS-SVM	GRNN	RFBT	MLP
t+7	0,24	0,51	0,23	0,59	0,39	0,44	0,30	0,34
t+15	0,39	0,52	0,33	0,66	0,47	0,48	0,36	0,41
t+30	0,64	0,51	0,40	0,68	0,48	0,56	0,46	0,50

5.4.3.3. Modelos aplicando PCA

En esta sección se muestran los resultados del análisis PCA y el rendimiento de los modelos aplicando PCA a los predictores. En la figura 5.23 se muestra un diagrama de Pareto de las componentes principales de la PCA aplicada al conjunto de datos

de la boya 0°N - 0°E. Se puede observar que, con las 9 primeras componentes de la PCA, se obtiene el 95 % de la varianza explicada. Por otra parte, en la tabla 5.15 se muestran los coeficientes de las primeras 9 componentes para la boya 0°N - 0°E. La primera componente (Comp1) representa la combinación de los predictores que explica la mayor variabilidad que, en este caso, es del 39,7%. Las variables *Air temp*, *Dynamic height* (DH), *20°C isotherm depth* (20IC), *Salinity* 120 m (S 120m), *Density* en 40 y 120 m así como la temperatura del mar desde 1 a 140 m son las que más contribuyen en la primera componente (pesos por encima de 0,20 en valor absoluto). Por otra parte las variables relativas al viento y la lluvia, son la que menos aportan.

Aplicando la PCA y seleccionando el número de componentes que representasen, al menos, el 90 % de la varianza, la reducción media (desviación estándar y [mínimo-maximo]) del número de variables ha sido de 19,47 (2,90) [15-24].

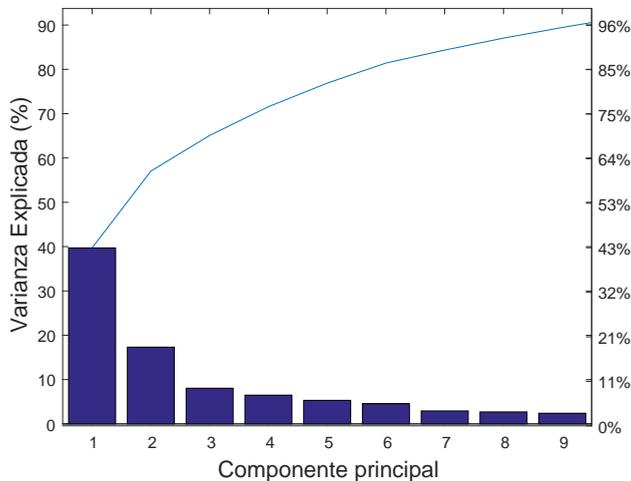


Figura 5.23

Diagrama de Pareto de la varianza explicada por cada componenete de la PCA para la boya 0°N - 0°E.

Una vez finalizado el análisis de PCA se entrenaron los modelos de predicción con los nuevos conjuntos de datos. En la tabla 5.16 se muestra el resumen de estos resultados. Al igual que utilizando todas las variables de entrada, el modelo RMLR obtuvo el mejor resultado en los tres horizontes de predicción, presentando unos errores semejantes a los mostrados en anteriores experimentos. Los modelos LS-SVM, GRNN y MLP también mantienen errores similares, mientras que los RFBT presentan el mayor aumento del error. Por último, el modelo ELM muestra la mayor reducción en el índice de error, pero continúa obteniendo peores resultados que el resto de modelos, no consiguiendo mejorar a los modelos trivial y climatología.

Tabla 5.15

Coefficientes de las 9 primeras componentes principales de la PCA para la boya 0°N - 0°E. A mayor valor del peso (en valor absoluto), mayor es la aportación del predictor en la determinación del componente.

Variable	Comp1	Comp2	Comp3	Comp4	Comp5	Comp6	Comp7	Comp8	Comp9
Air temp	0,21	-0,28	-0,09	0,10	-0,06	-0,03	-0,19	-0,05	0,11
U wind	0,02	0,02	0,52	0,14	-0,40	0,09	0,00	0,02	-0,10
V wind	-0,08	0,24	-0,27	0,25	-0,31	-0,43	-0,06	0,04	-0,05
W speed	-0,07	0,24	-0,28	0,24	-0,30	-0,44	-0,07	0,03	-0,04
W dir	-0,01	-0,02	-0,48	-0,16	0,45	-0,08	-0,01	-0,02	0,22
Sw rad	0,01	0,06	-0,22	0,45	0,06	0,44	-0,03	-0,06	-0,15
RH	-0,01	0,18	0,10	-0,41	0,01	-0,21	-0,51	0,08	0,13
Rain	0,05	-0,06	0,21	-0,42	0,06	-0,35	0,29	-0,20	-0,25
DH	0,29	0,05	0,07	0,10	0,10	-0,08	0,00	0,02	0,05
20°C IsoD	0,28	0,08	0,09	0,10	0,15	-0,06	-0,13	0,00	-0,04
S 20m	-0,10	0,35	0,03	-0,07	0,03	0,19	-0,25	-0,06	-0,01
S 40m	0,01	-0,34	-0,17	-0,14	-0,18	0,11	-0,15	-0,10	0,06
S 120m	0,26	0,13	-0,14	-0,14	-0,16	0,11	0,16	-0,05	-0,07
D 20m	-0,19	0,32	-0,01	-0,12	0,04	0,20	-0,13	-0,02	-0,10
D 40m	-0,25	-0,13	-0,17	-0,14	-0,20	0,09	0,14	-0,08	0,00
D 120m	-0,25	-0,10	0,16	0,17	0,21	-0,12	-0,04	0,18	0,10
ST 20m	0,22	-0,25	0,04	0,15	-0,05	-0,19	0,04	-0,02	0,15
ST 40m	0,27	0,06	0,15	0,12	0,17	-0,07	-0,19	0,07	0,02
ST 60m	0,28	0,09	0,07	0,07	0,14	-0,03	-0,16	0,00	-0,08
ST 80m	0,27	0,13	-0,02	0,00	0,10	0,02	-0,09	-0,12	-0,15
ST 100m	0,26	0,15	-0,07	-0,06	-0,02	0,05	-0,01	-0,21	-0,15
ST 120m	0,26	0,12	-0,16	-0,16	-0,19	0,11	0,09	-0,13	-0,09
ST 140m	0,23	0,07	-0,18	-0,20	-0,26	0,13	0,17	0,13	0,09
ST 180m	0,17	0,08	0,00	-0,14	-0,18	0,16	0,03	0,65	0,41
ST 300m	0,06	0,26	0,04	0,10	0,26	-0,12	0,55	0,25	0,00
ST 500m	-0,02	0,26	0,17	0,09	-0,05	0,02	0,16	-0,55	0,72
SST 1m	0,21	-0,30	-0,05	0,09	-0,06	-0,08	-0,05	-0,07	0,10

Tabla 5.16

Índices MAE (°C) de test (promedio de las 17 boyas) obtenidos en los tres horizontes de predicción aplicando PCA a las variables de entrada. Los mejores resultados aparecen resaltados en negrita.

	Trivial	Clima.	RMLR	ELM	LS-SVM	GRNN	RFBT	MLP
t+7	0,24	0,51	0,25	0,51	0,36	0,45	0,38	0,30
t+15	0,39	0,52	0,31	0,52	0,47	0,52	0,44	0,35
t+30	0,64	0,51	0,38	0,58	0,52	0,57	0,45	0,44

5.4.3.4. Modelos aplicando *RReliefF*

En este apartado se detallan los resultados del algoritmo *RReliefF* y el rendimiento de los modelos obtenido utilizando los predictores más importantes. En la figura 5.24 se muestran las variables de la boya 0°N - 0°E ordenadas según su importancia. Se observa que la variable Fecha Codificada 2 (FCod2) es la de mayor importancia, seguida por las temperaturas del mar (ST) en 1, 20 y 180 m. Por otra parte, las variables ST a 120 m y la profundidad de la isoterma a 20 °C (*20°C IsoD*) apenas contribuyen en la predicción. Además, las variables *Wind speed* y las posteriores tienen pesos negativos, indicando que no contribuyen en la resolución del problema. En la figura 5.25 se puede ver otro ejemplo del resultado del algoritmo, aplicado esta vez en la boya 0°N - 35°W. No se obtienen los mismos resultados que con la anterior

boya, sin embargo, se observan características comunes: las variables relacionadas con el viento y la lluvia aparecen en las últimas posiciones, mientras que las fechas y algunas variables de SST aparecen en los primeros puestos.

Con *RReliefF* la reducción media en el número de predictores ha sido de 8,71 (3,98) [2-17]. Por último, en la tabla 5.17 se muestran las variables ordenadas según su importancia para la boyas $0^{\circ}\text{N} - 0^{\circ}\text{E}$, $0^{\circ}\text{N} - 10^{\circ}\text{W}$, $0^{\circ}\text{N} - 23^{\circ}\text{W}$, $0^{\circ}\text{N} - 35^{\circ}\text{W}$, $15^{\circ}\text{N} - 38^{\circ}\text{W}$ y $19^{\circ}\text{S} - 34^{\circ}\text{W}$, donde se han resaltado en negrita las variables eliminadas en cada caso. Para estas boyas, las variables relativas a las fechas y la temperatura a ciertas profundidades aparecen siempre en los primeros puestos. Por el contrario, las variables relativas al viento y lluvia suelen aparecer en las últimas posiciones. También se puede observar que en la boyas $15^{\circ}\text{N} - 38^{\circ}\text{W}$ y $19^{\circ}\text{S} - 34^{\circ}\text{W}$ únicamente se consiguen eliminar dos variables.

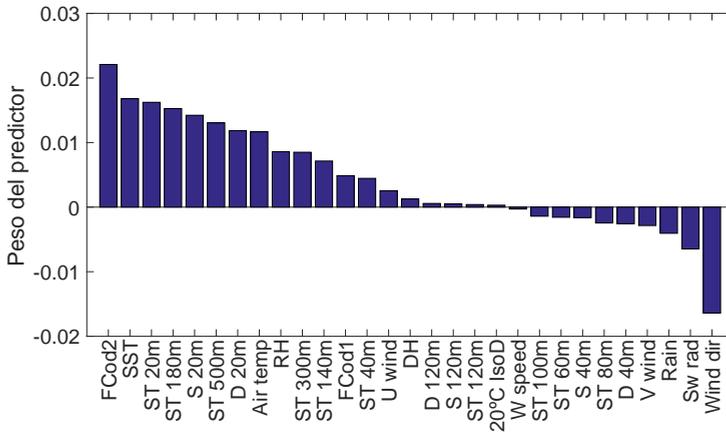


Figura 5.24

Gráfica de la importancia de las variables para la predicción según *RReliefF* (boya $0^{\circ}\text{N} - 0^{\circ}\text{E}$).

La tabla 5.18 muestra los resultados de las predicciones aplicando *RReliefF*. De nuevo, el modelo RMLR presenta los mejores resultados y con errores comparables a los de anteriores pruebas. En general, el resto de modelos muestran similares o peores resultados que los obtenidos con todas las variables y mediante PCA. Los resultados de los RFBT mejoran respecto a PCA, pero no respecto a utilizar todas las variables.

5.4.3.5. Modelos aplicando AC y *RReliefF*

En la sección 5.3.3.2 se explicó que el algoritmo *RReliefF* era sensible a las variables correlacionadas. Por este motivo y, debido al bajo número de variables que se ha conseguido eliminar con este método, en esta sección se presenta el análisis de correlación como etapa previa a la aplicación de *RReliefF*.

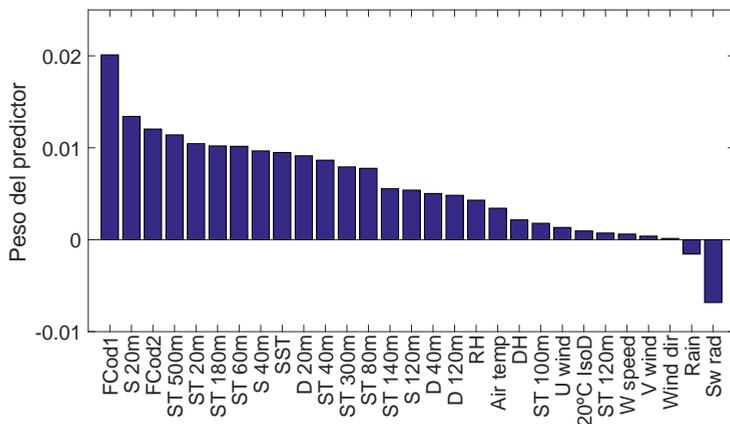


Figura 5.25

Gráfica de la importancia de las variables para la predicción según *RReliefF* (boya 0°N - 35°W).

Tabla 5.17

Variables ordenadas por nivel de importancia según el algoritmo *RReliefF*, las variables eliminadas aparecen resaltadas en grinta.

0N0E	0N10W	0N23W	0N35W	15N38W	19S34W
FCod2	FCod1	FCod2	FCod1	FCod1	FCod1
SST	FCod2	FCod1	S 20m	FCod2	ST 300m
ST 20m	S 1m	S 1m	FCod2	ST 60m	FCod2
ST 180m	S 40m	S 20m	ST 500m	ST 20m	ST 500m
S 20m	Air temp	S 40m	ST 20m	SST	S 40m
ST 500m	RH	RH	ST 180m	D 1m	ST 180m
D 20m	ST 180m	ST 500m	ST 60m	D 20m	S 20m
Air temp	SST	SST	S 40m	ST 40m	ST 140m
RH	S 20m	ST 20m	SST	ST 500m	D 1m
ST 300m	ST 300m	D 1m	D 20m	D 40m	DH
ST 140m	U wind	ST 180m	ST 40m	Air temp	D 20m
FCod1	D 1m	Air temp	ST 300m	RH	RH
ST 40m	ST 20m	ST 60m	ST 80m	S 1m	S 1m
U wind	ST 40m	ST 40m	ST 140m	ST 300m	ST 120m
DH	D 20m	D 20m	S 120m	S 20m	D 40m
D 120m	D 40m	D 40m	D 40m	ST 80m	ST 100m
S 120m	ST 140m	ST 300m	D 120m	DH	SST
ST 120m	S 120m	ST 140m	RH	S 40m	ST 40m
20 °C IsoD	ST 500m	Rain	Air temp	S 120m	ST 20m
Wind speed	DH	ST 80m	DH	D 120m	ST 60m
ST 100m	D 120m	20 °C IsoD	ST 100m	U wind	Air temp
ST 60m	Rain	Sw rad	U wind	ST 140m	Sw rad
S 40m	ST 120m	S 120m	20 °C IsoD	ST 180m	Wind dir
ST 80m	W speed	ST 100m	ST 120m	ST 120m	Rain
D 40m	ST 100m	U wind	Wind speed	20 °C IsoD	U wind
V wind	ST 60m	DH	V wind	ST 100m	V wind
Rain	ST 80m	D 120m	Wind dir	Sw rad	Wind speed
Sw rad	20 °C IsoD	W dir	Rain	V wind	
Wind dir	V wind	V wind	Sw rad	Rain	
	Wind dir	Wind speed		Wind dir	
	Sw rad	ST 120m		Wind speed	

Tabla 5.18

Índices MAE ($^{\circ}\text{C}$) de test (mediana de las 17 boyas) obtenidos en los tres horizontes de predicción aplicando *RReliefF* a las variables de entrada. Los mejores resultados aparecen resaltados en negrita.

	Trivial	Clima.	RMLR	ELM	LS-SVM	GRNN	RFBT	MLP
t+7	0,24	0,51	0,23	0,56	0,41	0,41	0,30	0,35
t+15	0,39	0,52	0,34	0,66	0,51	0,47	0,37	0,40
t+30	0,64	0,51	0,41	0,67	0,55	0,52	0,46	0,51

La figura 5.26 muestra la matriz de correlación para las variables de la boya 0°N - 0°E . Esta matriz presenta las correlaciones entre todas las parejas de variables (ver 5.3.3.3). Las correlaciones aparecen representadas de tres maneras simultáneamente: el área coloreada dentro de cada cuadro es proporcional al nivel de correlación; el tono de azul indica correlaciones positivas y los tonos de rojo negativas; además, el coeficiente de correlación aparece dentro de cada cuadro. Se puede observar que el nivel de variables altamente correlacionadas es muy elevado. Existe una alta correlación entre las variables *V wind* y *W Speed* (0,84), entre DH y 20°C IsoD (0,98), entre S 20m y D 20m (0,89), entre diferentes variables de ST y densidades, así como entre las variables de ST a diferentes profundidades. Para eliminar el efecto de multicolinealidad se aplicó el procedimiento explicado en la sección 5.3.3.3. Se eliminaron recursivamente las variables más correlacionadas hasta que el mayor valor absoluto de correlación quedó por debajo de 0,8. En la figura 5.27 se muestra la matriz de correlación tras aplicar el análisis de correlación a las variables de entrada. Se puede comprobar que se ha eliminado un gran número de variables y que el nivel de correlación se sitúa por debajo de 0,8.

En la tabla 5.19 se muestra el número de variables seleccionado en cada boya mediante AC + *RReliefF* comparado con el resto de métodos. Mediante este método la reducción media en el número de patrones ha sido de 18,47 (296) [12-22], siendo una reducción menor al del método PCA. Finalmente, la tabla 5.20 muestra las variables seleccionadas con el AC y ordenadas por importancia con *RReliefF* para la boyas 0°N - 0°E , 0°N - 10°W , 0°N - 23°W , 0°N - 35°W , 15°N - 38°W y 19°S - 34°W . Comparado con los resultados aplicando sólo el algoritmo *RReliefF* (ver tabla 5.17), se observa que el número de variables se ha reducido considerablemente. Nuevamente, en las primeras posiciones aparecen las variables de las fechas, así como predictores de SST, densidad y salinidad a diferentes profundidades. Las variables relacionados con el viento (*U wind*, *V wind*, *W dir*, *W speed*), la variable *Rain* y *Sw Rad* aparecen en las últimas posiciones.

La tabla 5.21 muestra el rendimiento de los modelos en las predicciones aplicando AC + *RReliefF*. Como en los casos anteriores, los mejores resultados (excepto en $t + 30$) han sido obtenidos por el modelo RMLR. Sin embargo, con este método, los índices de error han resultado ser mayores que en anteriores experimentos. En el resto de modelos también aumenta el error cometido respecto a utilizar todas las variables. Este hecho indica que el método AC + *RReliefF* ha eliminado variables que resultaban útiles en las predicciones.

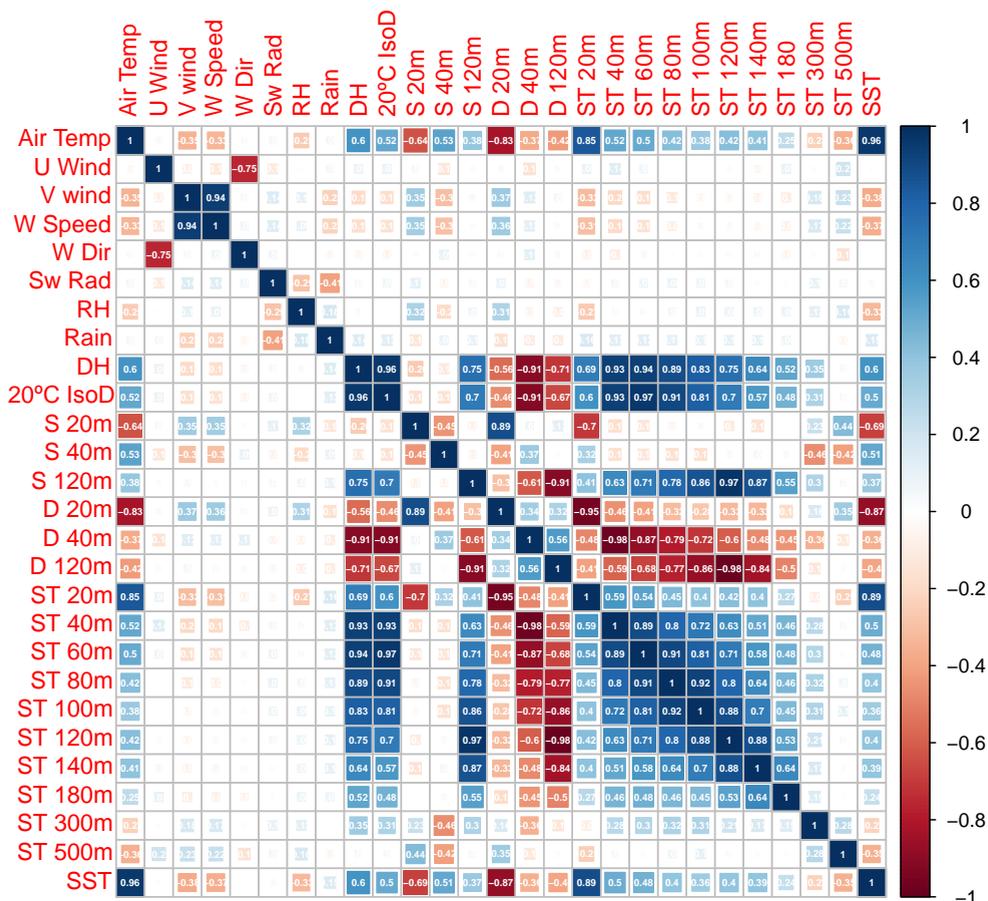


Figura 5.26

Visualización de la matriz de correlación de los predictores de la boya 0°N - 0°E; se puede observar que existe un gran número de variables con altos valores de correlación.

Tabla 5.19

Número de variables de cada boya aplicando análisis de correlación y *RReliefF* a las entradas.

Boya	0N0E	0N10W	0N23W	0N35W	4N23W	4N38W	6S10W	8N38W	8S30W	10S10W	12N23W	12N38W	14S32W	15N38W	19S34W	20N38W	21N23W
Todas	30	31	30	28	36	30	30	30	26	30	36	26	31	31	27	33	33
PCA	12	11	13	12	12	11	11	10	11	9	12	10	11	11	11	10	10
<i>RReliefF</i>	13	19	16	18	26	22	18	24	15	19	30	18	23	28	22	31	28
AC + <i>RReliefF</i>	10	9	11	16	14	12	9	14	10	9	16	12	15	11	11	13	12

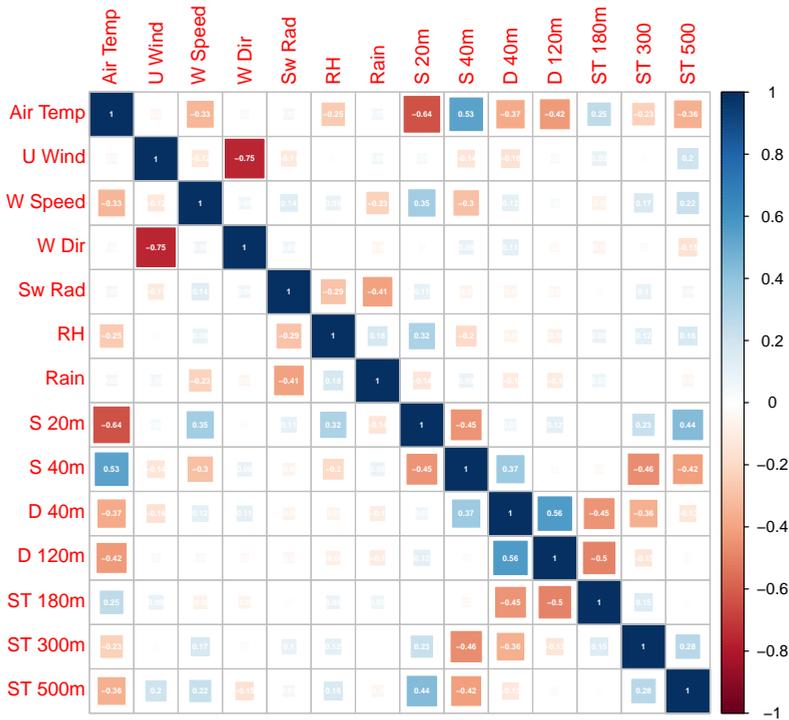


Figura 5.27

Visualización de la matriz de correlación de los predictores de la boya 0°N - 0°E tras aplicar el análisis de correlación.

Tabla 5.20

Variables seleccionadas por el AC y ordenadas según importancia mediante *RReliefF*, en negrita se muestran las variables eliminadas.

0N0E	0N10W	0N23W	0N35W	15N38W	19S34W
SST	FCod1	SST	SST	FCod2	FCod1
S 20m	SST	S 1m	S 40m	D 40m	S 20m
ST 180m	S 20m	S 40m	ST 60m	S 120m	ST 300m
ST 40m	S 40m	ST 40m	FCod2	S 1m	ST 140m
ST 500m	D 40m	FCod1	ST 40m	ST 60m	FCod2
RH	ST 180m	ST 180m	ST 80m	FCod1	ST 60m
ST 300m	RH	ST 60m	ST 500m	ST 500m	ST 500m
D 120m	ST 140m	RH	ST 180m	ST 80m	ST 100m
FCod1	ST 300m	ST 500m	D 40m	ST 300m	RH
U wind	ST 100m	ST 300m	ST 140m	W dir	Rain
W speed	ST 500m	ST 140m	ST 300m	Rain	Sw rad
S 40m	U wind	S 120m	Air temp	RH	W dir
W dir	W dir	Rain	D 120m	W speed	U wind
Rain	W speed	D 120m	W dir	V wind	V wind
Sw rad	V wind	W dir	RH	Sw rad	W speed
	Sw rad	Sw rad	U wind		
		V wind	Rain		
		U wind	Sw rad		
		W speed			

Tabla 5.21

Índices MAE ($^{\circ}\text{C}$) de test (mediana de las 17 boyas) obtenidos en los tres horizontes de predicción aplicando análisis de correlación y *RReliefF* a las variables de entrada. Los mejores resultados aparecen resaltados en negrita.

	Trivial	Clima.	RMLR	ELM	LS-SVM	GRNN	RFBT	MLP
t+7	0,24	0,51	0,23	0,55	0,48	0,44	0,31	0,32
t+15	0,39	0,52	0,32	0,61	0,56	0,51	0,40	0,40
t+30	0,64	0,51	0,39	0,68	0,55	0,58	0,47	0,51

5.4.4. Modelos mensuales

Siguiendo la metodología utilizada en los conjuntos de datos diarios, en primer lugar se obtuvieron las predicciones mensuales de SST utilizando como entradas la SST observada y la fecha codificada. Los resultados para las predicciones en $t + 1$ se detallan en la tabla 5.22. El valor mediano de las 17 boyas indica que el modelo LS-SVM consigue el menor MAE (0,24 $^{\circ}\text{C}$), seguido por el MLP (0,26 $^{\circ}\text{C}$), GRNN (0,27 $^{\circ}\text{C}$), ELM y RFBT (0,28 $^{\circ}\text{C}$), RMLR (0,29 $^{\circ}\text{C}$), climatología (0,35 $^{\circ}\text{C}$), CFSv2 (0,45 $^{\circ}\text{C}$) y trivial (0,62 $^{\circ}\text{C}$). Mientras que en los datos diarios obtenía los peores resultados, ahora el modelo LS-SVM obtiene el menor MAE. Por el contrario, el modelo RMLR ha empeorado con respecto a los experimentos con datos diarios. Por otra parte, ahora el modelo trivial presenta un error muy superior al resto de los modelos. Este hecho refleja que, en los datos mensuales, la diferencia que existe entre la SST actual y la futura es mayor que en los datos diarios. Sin embargo, el modelo climatología presenta unos índices MAE menores que en los datos diarios, indicando que la SST mensual presenta una menor variabilidad interanual que la SST diaria. En este horizonte, todos los modelos de predicción presentan una mediana de error menor a la de los modelos trivial, climatología y CFSv2. El modelo con el menor error (LS-SVM) mejora de manera global en 0,38 $^{\circ}\text{C}$ al modelo trivial, en 0,11 $^{\circ}\text{C}$ al modelo climatología y en 0,21 $^{\circ}\text{C}$ al modelo CFSv2. Aunque según la mediana únicamente mejora al modelo trivial, CFSv2 obtiene el menor error en la boya 21 $^{\circ}\text{N}$ - 23 $^{\circ}\text{W}$, mejorando en 0,07 $^{\circ}\text{C}$ al modelo LS-SVM.

En cuanto a las predicciones en $t + 2$ meses, la tabla 5.23 muestra que los resultados son similares a los anteriores, obteniendo el menor índice de error el modelo LS-SVM (0,30 $^{\circ}\text{C}$). Además, todos los modelos de predicción mejoran a los modelos trivial, climatología y CFSv2. Al igual que con los datos diarios, los modelos obtienen mayor ventaja respecto al modelo trivial al aumentar el horizonte de predicción. Sin embargo, esta ventaja se reduce frente al modelo climatología. En $t + 2$, el modelo climatología obtiene los mejores resultados en las boyas 12 $^{\circ}\text{N}$ - 38 $^{\circ}\text{W}$ y 14 $^{\circ}\text{S}$ - 32 $^{\circ}\text{W}$. Nuevamente, el modelo CFSv2 consigue el menor error en la boya 21 $^{\circ}\text{N}$ - 23 $^{\circ}\text{W}$.

Por último, la tabla 5.24 recoge los resultados en el horizonte $t + 3$ meses. Estos resultados indican que el modelo LS-SVM consigue nuevamente la menor mediana de error (0,33 $^{\circ}\text{C}$). Excepto RMLR, todos los modelos mejoran a los modelos trivial, climatología y CFSv2. Manteniéndose la tendencia de las pruebas anteriores, aumenta

la ventaja de los modelos respecto a los modelos trivial y CFSv2, pero se reduce frente al modelo climatología. En este horizonte, el modelo climatología obtiene el menor MAE en las boyas 12°N - 23°W (empatado con los RFBT), 12°N - 38°W y 14°S - 32°W. Además, ahora el modelo RMLR mejora al modelo CFSv2 en la boya 21°N - 23°W.

Tabla 5.22

Índices MAE (°C) de test obtenidos en las predicciones mensuales en $t + 1$ para cada boya utilizando la fecha codificada y la SST como variables de entrada. Los mejores resultados aparecen resaltados en negrita. *No se muestra el resultado por disponer solo de nueve datos.

	Trivial	Clima.	CFSv2	RMLR	ELM	LS-SVM	GRNN	RFBT	MLP
0N0E	0,79	0,38	0,70	0,42	0,30	0,29	0,33	0,32	0,32
0N10W	0,96	0,37	0,70	0,56	0,33	0,32	0,37	0,36	0,34
0N23W	0,73	0,41	0,85	0,49	0,36	0,35	0,34	0,36	0,37
0N35W	0,33	0,26	1,05	0,26	0,20	0,19	0,21	0,20	0,19
4N23W	0,39	0,33	0,45	0,35	0,28	0,25	0,27	0,31	0,26
4N38W	0,30	0,32	*	0,22	0,28	0,20	0,22	0,24	0,21
6S10W	0,73	0,25	0,33	0,26	0,18	0,19	0,19	0,20	0,20
8N38W	0,55	0,33	0,45	0,37	0,28	0,24	0,27	0,28	0,28
8S30W	0,42	0,24	1,31	0,14	0,16	0,14	0,15	0,21	0,16
10S10W	0,68	0,31	0,29	0,23	0,24	0,24	0,26	0,27	0,23
12N23W	0,90	0,37	0,46	0,36	0,33	0,30	0,33	0,42	0,33
12N38W	0,60	0,35	0,30	0,29	0,30	0,29	0,31	0,31	0,30
14S32W	0,47	0,18	0,76	0,20	0,17	0,16	0,16	0,21	0,17
15N38W	0,62	0,46	0,42	0,29	0,23	0,22	0,23	0,24	0,24
19S34W	0,69	0,46	0,39	0,24	0,25	0,24	0,27	0,29	0,26
20N38W	0,56	0,36	0,26	0,28	0,23	0,22	0,22	0,25	0,25
21N23W	0,94	0,52	0,29	0,38	0,39	0,36	0,36	0,42	0,42
Mediana	0,62	0,35	0,45	0,29	0,28	0,24	0,27	0,28	0,26

Tabla 5.23

Índices MAE (°C) de test obtenidos en las predicciones mensuales en $t + 2$ para cada boya utilizando la fecha codificada y la SST como variables de entrada. Los mejores resultados aparecen resaltados en negrita. *No se muestra el resultado por disponer solo de nueve datos.

	Trivial	Clima.	CFSv2	RMLR	ELM	LS-SVM	GRNN	RFBT	MLP
0N0E	1,48	0,40	0,98	0,54	0,37	0,36	0,39	0,39	0,37
0N10W	1,72	0,37	1,06	0,66	0,35	0,33	0,37	0,38	0,36
0N23W	1,25	0,42	1,20	0,56	0,40	0,38	0,40	0,38	0,39
0N35W	0,53	0,26	1,23	0,34	0,22	0,22	0,24	0,24	0,23
4N23W	0,58	0,36	0,62	0,46	0,31	0,30	0,32	0,33	0,32
4N38W	0,47	0,30	*	0,27	0,31	0,24	0,24	0,25	0,25
6S10W	1,38	0,25	0,60	0,35	0,22	0,22	0,23	0,22	0,23
8N38W	0,90	0,32	0,55	0,40	0,30	0,29	0,32	0,29	0,32
8S30W	0,75	0,23	0,56	0,18	0,21	0,17	0,20	0,20	0,19
10S10W	1,33	0,31	0,52	0,30	0,30	0,29	0,30	0,30	0,28
12N23W	1,64	0,38	0,70	0,51	0,40	0,38	0,35	0,48	0,40
12N38W	1,00	0,35	0,47	0,35	0,38	0,36	0,38	0,37	0,37
14S32W	0,87	0,18	0,72	0,26	0,19	0,18	0,18	0,19	0,19
15N38W	1,17	0,46	0,70	0,41	0,35	0,34	0,31	0,34	0,34
19S34W	1,29	0,46	0,52	0,35	0,33	0,33	0,37	0,38	0,32
20N38W	1,10	0,37	0,45	0,40	0,34	0,29	0,30	0,36	0,32
21N23W	1,72	0,49	0,41	0,43	0,52	0,48	0,51	0,47	0,52
Mediana	1,17	0,36	0,61	0,40	0,33	0,30	0,32	0,34	0,32

Tabla 5.24

Índices MAE ($^{\circ}\text{C}$) de test obtenidos en las predicciones mensuales en $t+3$ para cada boya utilizando la fecha codificada y la SST como variables de entrada. Los mejores resultados aparecen resaltados en negrita. *No se muestra el resultado por disponer solo de nueve datos.

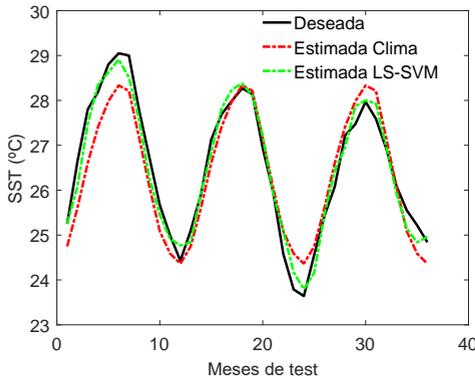
	Trivial	Clima.	CFSv2	RMLR	ELM	LS-SVM	GRNN	RFBT	MLP
0N0E	1,90	0,41	1,15	0,55	0,38	0,37	0,43	0,38	0,39
0N10W	2,30	0,35	1,05	0,52	0,35	0,33	0,34	0,37	0,35
0N23W	1,59	0,40	1,12	0,57	0,37	0,37	0,45	0,37	0,38
0N35W	0,65	0,27	1,08	0,35	0,24	0,23	0,26	0,25	0,26
4N23W	0,64	0,38	0,68	0,51	0,36	0,33	0,39	0,37	0,36
4N38W	0,56	0,29	*	0,28	0,34	0,26	0,26	0,27	0,27
6S10W	1,98	0,27	0,79	0,35	0,24	0,24	0,25	0,25	0,25
8N38W	1,09	0,33	0,55	0,37	0,31	0,30	0,31	0,30	0,31
8S30W	1,03	0,22	0,43	0,22	0,27	0,21	0,24	0,22	0,23
10S10W	1,91	0,31	0,73	0,33	0,31	0,29	0,32	0,32	0,30
12N23W	2,32	0,38	0,85	0,54	0,41	0,39	0,40	0,38	0,43
12N38W	1,38	0,38	0,61	0,39	0,40	0,39	0,42	0,42	0,40
14S32W	1,19	0,18	0,63	0,26	0,20	0,19	0,19	0,21	0,20
15N38W	1,62	0,46	0,84	0,44	0,40	0,38	0,36	0,39	0,39
19S34W	1,79	0,47	0,60	0,39	0,35	0,34	0,41	0,43	0,36
20N38W	1,52	0,38	0,53	0,44	0,42	0,37	0,36	0,36	0,39
21N23W	2,37	0,47	0,54	0,44	0,53	0,49	0,55	0,45	0,53
Mediana	1,59	0,38	0,70	0,39	0,35	0,33	0,36	0,37	0,36

En la figura 5.28.a se muestra un ejemplo de SST mensual observada y las predicciones en $t+1$ según el modelo climatología y el modelo LS-SVM. Representar tres años de test (36 meses) permite apreciar que, por ejemplo, las temperaturas máximas y mínimas presentan cierta variabilidad interanual. En el primer año de test la SST alcanza una temperatura máxima de $29,1^{\circ}\text{C}$, mientras que en el segundo se reduce a $28,4^{\circ}\text{C}$. Por otra parte, la temperatura mínima presenta un valor de $24,4^{\circ}\text{C}$ el primer año y de $23,6^{\circ}\text{C}$ el segundo. Además, comparado con la climatología, también se aprecia que la SST observada del primer año se adelanta en la primera mitad del año y se retrasa en la segunda parte. Es decir, las temperaturas máximas se alcanzan más pronto, y las mínimas más tarde, que en un año promedio. En el segundo año, a excepción de la temperatura mínima, se puede ver que las señales de SST observada y climatología aparecen más solapadas que en el primer año.

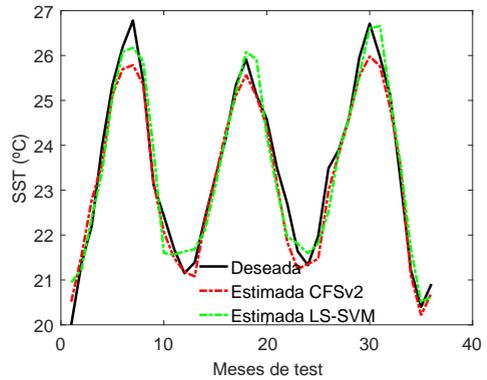
En cuanto al modelo LS-SVM, se puede apreciar que las predicciones se ajustan mucho a la SST deseada ($\text{MAE} = 0,24^{\circ}\text{C}$), siendo capaz de predecir tanto el ciclo mensual como la variabilidad interanual. Los resultados comentados anteriormente mostraban que el modelo CFSv2 mejoraba al modelo LS-SVM en la boya $21^{\circ}\text{N} - 23^{\circ}\text{W}$. Comparando ambas predicciones con la SST deseada (figura 5.28.b), se observa que el modelo CFSv2 consigue ajustarse mejor en las temperaturas mínimas de cada ciclo mientras que LS-SVM se ajusta mejor en las temperaturas máximas. No obstante, ambos modelos consiguen predecir con poco error la SST deseada.

La figura 5.29.a muestra un ejemplo en el que se muestran dos predicciones en $t+2$ con idéntico MAE (modelo trivial y LS-SVM). Otro ejemplo interesante es el de la boya $4^{\circ}\text{N} - 23^{\circ}\text{W}$ (figura 5.29.b), que está situada más cerca del ecuador que las

boyas vistas anteriormente. Las boyas situadas más al norte o más al sur del ecuador muestran unas transiciones suaves entre las temperaturas mínimas del invierno y las máximas del verano. En la boya 4°N - 23°W se aprecia una temperatura media más alta con transiciones más abruptas. Los modelos consiguen captar el ciclo anual, pero presentan dificultades con las variaciones mensuales. Por último, en las figuras 5.30.a y 5.30.b aparecen dos ejemplos para las predicciones en $t + 3$ meses.



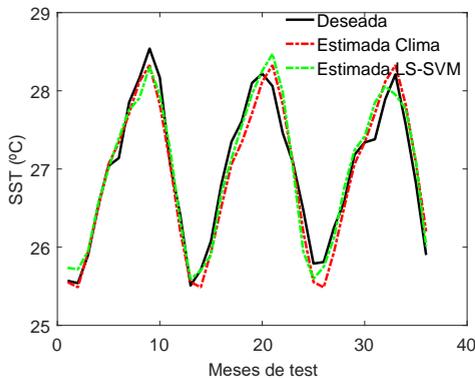
(a) Boya 19S34W. MAE (°C): clima.=0,46; LS-SVM=0,24. Fecha: 2015/10 a 2018/09.



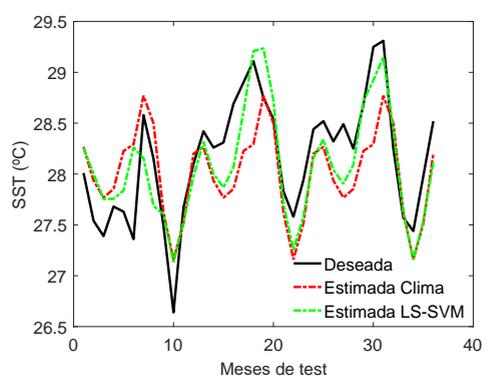
(b) Boya 21N23W. MAE (°C): CFSv2=0,29; LS-SVM=0,36. Fecha: 2015/04 a 2018/03.

Figura 5.28

Ejemplos de series temporales mensuales deseadas y estimadas de test para la SST en el horizonte $t + 1$ mes. Resultados con la SST y fecha codificada como variables de entrada a los modelos.



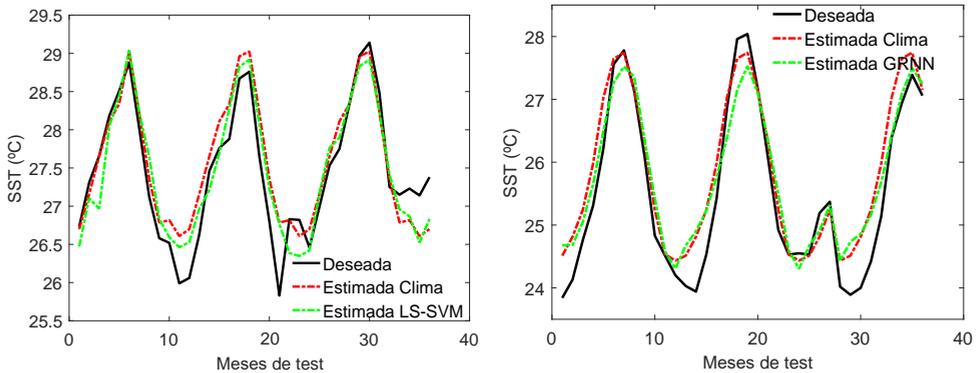
(a) Boya 14S32W. MAE (°C): clima.=0,18; LS-SVM=0,18. Fecha: 2014/08 a 2017/07.



(b) Boya 4N23W. MAE (°C): clima.=0,36; LS-SVM=0,30. Fecha: 2014/11 a 2018/10.

Figura 5.29

Ejemplos de series temporales mensuales deseadas y estimadas de test para la SST en el horizonte $t + 2$ meses. Resultados con la SST y fecha codificada como variables de entrada a los modelos.



(a) Boya 8N38W. MAE ($^{\circ}\text{C}$): clima.=0,33; LS-SVM=0,30. Fecha: 2012/04 a 2016/04. (b) Boya 15N38W. MAE ($^{\circ}\text{C}$): clima.=0,46; GRNN=0,36. Fecha: 2014/04 a 2018/11.

Figura 5.30

Ejemplos de series temporales deseada y estimada de test para la SST en el horizonte $t + 3$ meses. Resultados con la SST y fecha codificada como variables de entrada a los modelos.

Tras analizar los resultados de las predicciones utilizando únicamente la SST y fecha como entradas, se estudió el desempeño de los modelos mensuales utilizando las técnicas de selección y extracción de características vistas en los modelos diarios. Inicialmente, el número promedio de patrones que disponían de la variable SST era de 150 (ver tabla 5.1). Debido al gran número de datos ausentes que presentan las distintas variables, este promedio quedaba reducido a 50. Para evitar esta pérdida de datos, se substituyeron los valores ausentes por sus valores históricos promedio (climatología). Además, se seleccionaron periodos de test con el mayor número posible de datos consecutivos sin valores ausentes. Para poder comparar resultados y, puesto que los conjuntos de test no son iguales a los de las anteriores pruebas mensuales, se muestran también los nuevos índices MAE obtenidos con la SST y fecha como entradas.

El resumen de los índices MAE obtenidos en estos experimentos se muestran en la tabla 5.25. Nuevamente, los resultados indican que el modelo LS-SVM obtiene los menores índices MAE. Puesto que algunos conjuntos de test contienen valores de la climatología o promedio de la serie, a excepción del modelo trivial, todos los modelos han reducido su error con respecto a los experimentos mensuales anteriores (SST y fecha de entradas). Evidentemente, la mayor reducción del índice MAE se produce en el modelo climatología. Comparando los resultados de las diferentes técnicas de selección y extracción de características, el modelo RMLR obtiene los menores índices MAE en los experimentos con $AC + RRelieff$. El resto de modelos consiguen el menor error en los experimentos con la SST y fecha codificada como entradas.

Tabla 5.25

Índices MAE (°C) de test (mediana de las 17 boyas) obtenidos en los datos mensuales utilizando los diferentes tipos de variables de entrada.

		Trivial	Clima.	RMLR	ELM	LS-SVM	GRNN	RFBT	MLP
SST	t+1	0,62	0,29	0,28	0,22	0,21	0,24	0,24	0,22
	t+2	1,10	0,29	0,37	0,30	0,28	0,30	0,30	0,29
	t+3	1,53	0,30	0,39	0,31	0,30	0,32	0,32	0,31
Todas	t+1	0,62	0,29	0,32	0,96	0,24	0,67	0,27	0,35
	t+2	1,10	0,29	0,44	1,02	0,32	0,86	0,36	0,37
	t+3	1,53	0,30	0,46	1,09	0,36	0,89	0,38	0,43
PCA	t+1	0,62	0,29	0,28	0,66	0,26	0,55	0,29	0,26
	t+2	1,10	0,29	0,32	0,70	0,34	0,60	0,33	0,33
	t+3	1,53	0,30	0,35	0,72	0,36	0,60	0,37	0,36
<i>RReliefF</i>	t+1	0,62	0,29	0,27	0,50	0,23	0,38	0,25	0,27
	t+2	1,10	0,29	0,34	0,56	0,30	0,46	0,33	0,35
	t+3	1,53	0,30	0,36	0,60	0,34	0,62	0,32	0,40
AC + <i>RReliefF</i>	t+1	0,62	0,29	0,26	0,42	0,22	0,39	0,24	0,25
	t+2	1,10	0,29	0,33	0,51	0,30	0,43	0,32	0,32
	t+3	1,53	0,30	0,37	0,54	0,34	0,47	0,34	0,36

5.4.5. Modelos diarios y mensuales con LSTM

En este apartado se muestran los resultados obtenidos en los experimentos con LSTM comparados con ELM y LS-SVM. Utilizando el procedimiento explicado anteriormente, en estas pruebas se han substituido los valores ausentes por sus valores promedio para poder comparar resultados. En la tabla 5.26 se encuentran los índices de error de la LSTM y de la ELM en los conjuntos de datos diarios, utilizando la SST y fecha como entradas a los modelos. Según lo explicado en la sección 5.3.4, la estrategia MIMO consiste en entrenar un único modelo con varias entradas para obtener varias salidas o predicciones. Mediante esta estrategia, ambos modelos obtienen similares resultados, 0,22 °C en $t + 7$ y 0,33 °C en $t + 30$. En $t + 15$, la mediana del modelo LSTM es de 0,29 °C frente a 0,28 °C del modelo ELM. Analizando los resultados por boya, la LSTM obtiene menor MAE en 25 de las 51 pruebas (17 boyas por 3 predicciones), e igual MAE en 13 de ellas. La ELM consigue, por tanto, el mejor resultado en las 13 restantes. Además de la estrategia MIMO, con la LSTM se experimentó la estrategia de predicción recursiva. En la primera prueba se obtuvieron las predicciones de SST con un horizonte de 365 días. Es decir, utilizando el último dato de validación disponible, se predijeron los siguientes 365 días de SST sin actualizar el modelo con nuevos datos, lográndose un MAE mediano de 0,42 °C. En el segundo experimento, el horizonte de la predicción fue de 30 días, obteniéndose un MAE de 0,31 °C. Promediando los resultados del modelo LSTM en $t + 7$, $t + 15$ y $t + 30$ se obtiene un MAE de 0,28 °C, valor inferior a los conseguidos con las predicciones recursivas. El último experimento con los datos diarios consistió en entrenar los modelos LSTM utilizando las variables seleccionadas por el método AC + *RReliefF*. Las medianas de error fueron: 0,26 °C en $t + 7$, 0,34 °C en $t + 15$ y 0,41 °C en $t + 30$, no consiguiendo mejorar los resultados utilizando la SST y fecha como entradas.

Tabla 5.26

Índices MAE ($^{\circ}\text{C}$) de test obtenidos en las predicciones diarias con la LSTM utilizando la SST y fecha como entradas.

	ELM			LSTM				
	Predicción MIMO			Predicción MIMO			Recursiva	
	t+7	t+15	t+30	t+7	t+15	t+30	365	30
0N0E	0,43	0,35	0,41	0,38	0,34	0,40	0,50	0,39
0N10W	0,52	0,54	0,57	0,51	0,52	0,57	0,53	0,55
0N23W	0,29	0,32	0,39	0,28	0,29	0,34	0,41	0,38
0N35W	0,23	0,24	0,26	0,21	0,22	0,25	0,24	0,22
4N23W	0,31	0,37	0,38	0,28	0,32	0,33	0,52	0,34
4N38W	0,12	0,13	0,16	0,12	0,14	0,17	0,16	0,16
6S10W	0,19	0,27	0,32	0,20	0,27	0,33	0,36	0,31
8N38W	0,20	0,27	0,31	0,20	0,26	0,30	0,32	0,25
8S30W	0,12	0,15	0,20	0,12	0,15	0,19	0,25	0,18
10S10W	0,16	0,24	0,33	0,15	0,21	0,29	0,44	0,23
12N23W	0,28	0,39	0,49	0,31	0,39	0,50	0,57	0,33
12N38W	0,23	0,33	0,39	0,23	0,29	0,36	0,42	0,30
14S32W	0,16	0,21	0,22	0,18	0,20	0,22	0,37	0,20
15N38W	0,17	0,22	0,27	0,18	0,24	0,28	0,37	0,24
19S34W	0,22	0,30	0,33	0,22	0,30	0,33	0,53	0,32
20N38W	0,21	0,30	0,43	0,24	0,34	0,45	0,56	0,36
21N23W	0,22	0,33	0,45	0,22	0,32	0,44	0,71	0,35
Mediana	0,22	0,30	0,33	0,22	0,29	0,33	0,42	0,31

Los resultados de las predicciones mensuales de la LSTM comparados con la LS-SVM se detallan en la tabla 5.27. Ahora, el modelo LS-SVM obtiene un menor MAE mediano que la LSTM en los tres horizontes de predicción. Además, por separado, la LSTM solo mejora al modelo LS-SVM en tres de los casos. En cuanto a los resultados utilizando las predicciones recursivas, la LSTM obtuvo un MAE de $0,31\text{ }^{\circ}\text{C}$ en las predicciones a 36 días y $0,29\text{ }^{\circ}\text{C}$ en las predicciones a 3 días. El MAE promedio en $t + 1$, $t + 2$ y $t + 3$ son $0,26\text{ }^{\circ}\text{C}$ para LS-SVM y $0,32\text{ }^{\circ}\text{C}$ para la LSTM. Por tanto, las predicciones recursivas mejoran a las predicciones MIMO en la LSTM, pero no mejoran los resultados MIMO del modelo LS-SVM. Por último, se entrenaron los modelos mensuales con las variables escogidas por $AC + RR_{liefF}$, obteniendo un índice de error de $0,37\text{ }^{\circ}\text{C}$ en $t + 1$, $0,38\text{ }^{\circ}\text{C}$ en $t + 2$ y $0,41\text{ }^{\circ}\text{C}$ en $t + 3$. Como en el caso de los datos diarios, tampoco se consiguió reducir del error de los experimentos con la SST y fecha.

5.4.6. Análisis estadístico de los resultados

En las anteriores secciones se ha evaluado el rendimiento de los modelos de predicción en los diferentes conjuntos de datos y con los distintos métodos de selección y extracción de características. En esta sección se presenta la comparación estadística de estos resultados para determinar si existen diferencias significativas entre las distintas pruebas y modelos. Además, se evalúa el coste computacional de los diferentes algoritmos.

Tabla 5.27

Índices MAE ($^{\circ}\text{C}$) de test obtenidos en las predicciones mensuales con LSTM utilizando la SST y fecha como entradas.

	LS-SVM			LSTM			Recursiva	
	Predicción MIMO			Predicción MIMO			36	3
	t+1	t+2	t+3	t+1	t+2	t+3		
0N0E	0,25	0,30	0,35	0,28	0,32	0,34	0,36	0,29
0N10W	0,32	0,34	0,34	0,35	0,35	0,35	0,40	0,37
0N23W	0,25	0,29	0,30	0,29	0,31	0,36	0,32	0,29
0N35W	0,19	0,22	0,24	0,22	0,26	0,28	0,27	0,26
4N23W	0,16	0,24	0,27	0,29	0,36	0,40	0,29	0,24
4N38W	0,20	0,26	0,28	0,31	0,34	0,35	0,42	0,30
6S10W	0,19	0,22	0,24	0,21	0,29	0,31	0,22	0,21
8N38W	0,19	0,25	0,29	0,21	0,27	0,28	0,29	0,23
8S30W	0,13	0,17	0,20	0,22	0,25	0,27	0,22	0,20
10S10W	0,19	0,24	0,26	0,21	0,25	0,25	0,27	0,23
12N23W	0,26	0,30	0,31	0,37	0,58	0,57	0,38	0,33
12N38W	0,22	0,28	0,30	0,28	0,30	0,32	0,30	0,29
14S32W	0,16	0,18	0,18	0,20	0,21	0,24	0,21	0,20
15N38W	0,21	0,31	0,33	0,28	0,35	0,35	0,31	0,30
19S34W	0,24	0,33	0,34	0,34	0,35	0,40	0,40	0,34
20N38W	0,22	0,29	0,37	0,34	0,39	0,49	0,61	0,51
21N23W	0,40	0,51	0,47	0,59	0,66	0,72	0,55	0,48
Mediana	0,21	0,28	0,30	0,28	0,32	0,35	0,31	0,29

5.4.6.1. Análisis estadístico modelos diarios

Puesto que los modelos han sido entrenados con 20 repeticiones aleatorias de los conjuntos de datos, se han considerado métodos estadísticos para muestras relacionadas en la comparación entre modelos. Para el análisis de las diferencias entre métodos de selección de variables se han considerado métodos para muestras no relacionadas, ya que los conjuntos de datos son diferentes (Hothorn et al., 2005; Eugster et al., 2008).

El análisis estadístico se ha realizado con el programa IBM SPSS Statistics (versión 22.0, Armonk, NY: IBM Corp). Para analizar las diferencias entre los modelos y los distintos métodos de reducción de variables empleados, se utilizó un análisis de varianza (ANOVA) de dos factores con medidas repetidas en el factor modelo. Se aplicó la corrección de Bonferroni para evitar el error tipo I al realizar las comparaciones múltiples cuando el análisis ANOVA indicaba diferencias significativas. Se consideró significación estadística con $p < 0,05$.

En la tabla 5.28 se muestran los valores medios y desviación estándar del índice MAE (promediando los tres horizontes de predicción y las 17 boyas) para cada modelo. Según estos valores, los menores índices MAE se obtienen en las pruebas con la SST y fecha como entradas. Además, los modelos ELM y MLP obtienen los menores errores [MAE=0,30 (0,13) $^{\circ}\text{C}$]. Los resultados de los efectos principales del modelo ANOVA mostraron diferencias significativas en el factor modelo [F(3,29; 2,40)=139,28; $p < 0,001$; $\eta_p^2=0,36$], factor prueba [F(4,70; 4,00); $p=0,004$; $\eta_p^2=0,06$] y

en la interacción modelo*prueba [$F(13,15; 0,23)=13,46; p<0,001; \eta_p^2=0,18$]. En la tabla 5.29 se muestran los resultados de las comparaciones múltiples entre pruebas que resultaron significativas. En el modelo ELM existen diferencias significativas entre el conjunto 1 (SST y fecha como entradas) y el resto de pruebas, siendo siempre inferior el error utilizando únicamente la SST y fecha. Esto indica que la ELM es muy sensible a las variables seleccionadas en la etapa de entrenamiento. En cuanto a LS-SVM, aunque el menor error se obtiene utilizando la SST y fecha como entradas, únicamente es significativa la diferencia con el análisis AC + *RReliefF*. El rendimiento de LS-SVM por tanto, se ve menos afectado por la etapa de selección de variables. Al igual que con la ELM, en el modelo GRNN se han encontrado diferencias entre la SST y fecha con todas las demás pruebas. En el caso de los RFBT, comparado con la prueba utilizando SST y fecha, únicamente es significativa la diferencia con el análisis PCA. Por último, en el MLP, las diferencias han sido significativas con todas la pruebas excepto con el análisis PCA. Los modelos trivial, climatología y RMLR no aparecen en la tabla puesto que no se encontraron diferencias significativas entre las diferentes pruebas, destacando que el modelo RMLR obtiene un índice MAE bajo en todos los casos.

Tabla 5.28

Índices MAE promedio y desviación estándar según el método de selección de variables para cada modelo en los datos diarios (medias de los tres horizontes y las 17 boyas).

	SST	Todas	PCA	Relief	ACrelief
Trivial	0,45 (0,24)	0,46 (0,22)	0,46 (0,22)	0,46 (0,22)	0,46 (0,22)
Clima.	0,44 (0,16)	0,48 (0,16)	0,48 (0,16)	0,48 (0,16)	0,48 (0,16)
RMLR	0,33 (0,16)	0,33 (0,15)	0,33 (0,14)	0,33 (0,14)	0,33 (0,15)
ELM	0,30 (0,13)	0,63 (0,18)	0,54 (0,19)	0,62 (0,20)	0,59 (0,19)
LS-SVM	0,37 (0,15)	0,44 (0,17)	0,45 (0,17)	0,45 (0,16)	0,50 (0,19)
GRNN	0,35 (0,14)	0,46 (0,15)	0,48 (0,16)	0,45 (0,16)	0,51 (0,21)
RFBT	0,31 (0,14)	0,37 (0,13)	0,44 (0,16)	0,37 (0,13)	0,37 (0,14)
MLP	0,30 (0,13)	0,39 (0,15)	0,38 (0,15)	0,39 (0,16)	0,40 (0,18)

En cuanto a la comparación entre modelos, se analizaron únicamente las diferencias en la prueba que obtuvo los mejores resultados (SST y fecha). Los pares de diferencias significativas se muestran en la tabla 5.30. Estos resultados indican que todos los modelos de predicción obtuvieron índices MAE significativamente inferiores a los obtenidos por los modelos Trivial y climatología. Además, el modelo LS-SVM obtuvo peores resultados que los modelos ELM, RFBT y MLP. Por último, no se encontraron diferencias entre los modelos ELM, RMLR, GRNN, RFBT y MLP.

5.4.6.2. Análisis estadístico modelos mensuales

Tras analizar los modelos diarios y, puesto que los mejores resultados se obtuvieron con la SST y fecha como entradas, el análisis estadístico en los datos mensuales se muestra únicamente para los experimentos utilizando la SST y fecha como entradas. En este caso, se analizaron las diferencias entre los distintos modelos de predicción utilizando un análisis ANOVA de medidas repetidas de un único factor.

Tabla 5.29

Análisis estadístico de las diferencias entre pruebas con datos diarios. Únicamente se muestran los modelos y pruebas con diferencias significativas. C1: Conjunto 1; C2: Conjunto 2; IC: Intervalo de Confianza; LI: Límite Inferior; LS: Límite Superior.

Modelo	C1	C2	Diferencia de medias (1-2)	Error estándar	Valor p	95 % IC	
						LI	LS
ELM	SST	Todas	-0,33	0,04	<0,001	-0,43	-0,23
		PCA	-0,24	0,04	<0,001	-0,34	-0,14
		<i>RReliefF</i>	-0,32	0,04	<0,001	-0,42	-0,22
		AC + <i>RReliefF</i>	-0,29	0,04	<0,001	-0,39	-0,19
LSSVM	SST	AC + <i>RReliefF</i>	-0,13	0,03	<0,001	-0,23	-0,04
GRNN	SST	Todas	-0,11	0,03	0,010	-0,20	-0,02
		PCA	-0,14	0,03	<0,001	-0,23	-0,04
		<i>RReliefF</i>	-0,10	0,03	0,028	-0,19	-0,01
		AC + <i>RReliefF</i>	-0,16	0,03	<0,001	-0,25	-0,07
RFBT	SST	PCA	-0,12	0,03	<0,001	-0,20	-0,05
MLP	SST	Todas	-0,09	0,03	0,039	-0,18	-0,01
		<i>RReliefF</i>	-0,09	0,03	0,029	-0,18	-0,01
		AC + <i>RReliefF</i>	-0,10	0,03	0,012	-0,19	-0,01

Tabla 5.30

Análisis estadístico de las diferencias entre modelos para el método SST más fecha codificada en los conjuntos diarios. Únicamente se muestran los resultados con diferencias significativas. IC: Intervalo de Confianza; LI: Límite Inferior; LS: Límite Superior.

Modelo 1	Modelo 2	Diferencia de medias (1-2)	Error estándar	Valor p	95 % IC	
					LI	LS
Trivial	RMLR	0,13	0,02	<0,001	0,07	0,18
	ELM	0,15	0,02	<0,001	0,08	0,23
	LS-SVM	0,09	0,02	0,013	0,01	0,16
	GRNN	0,11	0,02	<0,001	0,03	0,19
	RFBT	0,14	0,02	<0,001	0,07	0,21
	MLP	0,15	0,02	<0,001	0,09	0,22
Clima.	RMLR	0,11	0,02	<0,001	0,04	0,18
	ELM	0,14	0,02	<0,001	0,07	0,21
	LS-SVM	0,07	0,02	0,006	0,01	0,13
	GRNN	0,09	0,02	<0,001	0,04	0,15
	RFBT	0,13	0,02	<0,001	0,07	0,18
	MLP	0,14	0,02	<0,001	0,07	0,20
LS-SVM	ELM	0,07	0,01	0,000	0,03	0,10
	RFBT	0,06	0,01	<0,001	0,02	0,09
	MLP	0,07	0,01	<0,001	0,03	0,10

Los valores medios y desviaciones estándar de cada modelo se muestran en la tabla 5.31. El efecto principal del ANOVA muestra diferencias significativas en los índices MAE obtenidos por los modelos de predicción [$F(1,61; 18,29) = 95,96; p < 0,001; \eta_p^2 = 0,66$]. El análisis en detalle de las comparaciones múltiples se pueden observar en la tabla 5.32. El modelo trivial presenta un error significativamente mayor que el resto de modelos. Por su parte, el modelo climatología mejora significativamente al modelo trivial y a CFSv2, pero no mejora a los modelos ELM, LS-SVM, GRNN, RFBT y MLP. Además, su desempeño es similar al del modelo RMLR, ya que las

diferencias entre ambos modelos no son significativas. Consecuentemente, el modelo RMLR obtiene peores resultados que los demás modelos de predicción (excepto con modelo trivial y CFSv2). En cuanto al modelo LS-SVM, el análisis muestra que su índice MAE promedio es significativamente menor que el del resto de modelos. Por último, puesto que no presentan diferencias entre ellos, los modelos ELM, GRNN, RFBT y MLP son comparables en los errores cometidos en las predicciones.

Tabla 5.31

Índices MAE promedio y desviación estándar (DE) de cada modelo en los datos mensuales utilizando SST y fecha como entradas (medias de los tres horizontes y las 17 boyas).

	Trivial	Clima.	CFSv2	RMLR	ELM	LS-SVM	GRNN	RFBT	MLP
Media	1,09	0,35	0,67	0,37	0,31	0,29	0,31	0,32	0,31
DE	(0,56)	(0,08)	(0,27)	(0,12)	(0,08)	(0,08)	(0,09)	(0,08)	(0,08)

Tabla 5.32

Análisis estadístico de las diferencias entre modelos mensuales para el método SST y fecha codificada. Únicamente se muestran los resultados con diferencias significativas. IC: Intervalo de Confianza; LI: Límite Inferior; LS: Límite Superior.

Modelo 1	Modelo 2	Diferencia de medias (1-2)	Error estándar	Valor p	95 % IC	
					LI	LS
Trivial	Clima.	0,74	0,07	<0,001	0,49	1,00
	CFSv2	0,42	0,08	<,001	0,15	0,69
	RMLR	0,72	0,07	<0,001	0,48	0,96
	ELM	0,78	0,07	<0,001	0,54	1,02
	LS-SVM	0,80	0,07	<0,001	0,56	1,04
	GRNN	0,78	0,07	<0,001	0,54	1,02
	RFBT	0,77	0,07	<0,001	0,53	1,02
	MLP	0,78	0,07	<0,001	0,54	1,02
Clima.	CFSv2	-0,32	0,04	<0,001	-0,46	-0,18
	ELM	0,04	0,01	0,002	0,01	0,06
	LS-SVM	0,06	0,01	<0,001	0,03	0,08
	GRNN	0,04	0,01	0,001	0,01	0,07
	RFBT	0,03	0,01	0,003	0,01	0,06
	MLP	0,04	0,01	<0,001	0,01	0,07
CFSv2	RMLR	0,30	0,03	<0,001	0,18	0,42
	ELM	0,36	0,04	<0,001	0,23	0,49
	LS-SVM	0,38	0,04	<0,001	0,25	0,51
	GRNN	0,36	0,04	<0,001	0,23	0,49
	RFBT	0,35	0,04	<0,001	0,22	0,49
	MLP	0,36	0,04	<0,001	0,23	0,50
RMLR	ELM	0,06	0,01	<0,001	0,02	0,10
	LS-SVM	0,08	0,01	<0,001	0,05	0,12
	GRNN	0,06	0,01	<0,001	0,03	0,10
	RFBT	0,06	0,01	<0,001	0,02	0,09
	MLP	0,07	0,01	<0,001	0,03	0,10
LS-SVM	ELM	-0,02	0,00	<0,001	-0,03	-0,01
	GRNN	-0,02	0,00	<0,001	-0,03	-0,01
	RFBT	-0,03	0,00	<0,001	-0,04	-0,01
	MLP	-0,02	0,00	<0,001	-0,02	-0,01

5.4.6.3. Análisis estadístico modelos LSTM

En los datos diarios y, con SST y fecha como entradas, se compararon los resultados obtenidos por los modelos ELM y LSTM mediante un ANOVA de medidas repetidas con un factor. Los índices MAE promedio se muestran en la tabla 5.33. Las comparaciones por parejas señalan que las predicciones recursivas a 365 pasos con LSTM son significativamente peores ($p = 0,001$) que la del resto de modelos. Además, no se encontraron diferencias entre la ELM, LSTM MIMO y LSTM recursiva a 30 pasos.

Tabla 5.33

Índices MAE promedio y desviación estándar para los modelos diarios ELM y LSTM (medias de los tres horizontes y las 17 boyas).

ELM	LSTM MIMO	LSTM 365	LSTM 30
0,29 (0,10)	0,29 (0,09)	0,43 (0,14)	0,30 (0,10)

Para los datos mensuales se utilizó un análisis similar, comparando los modelos LS-SVM y los tres modelos LSTM mediante una ANOVA. En la tabla 5.34 se detallan los índices MAE de estos modelos. Según el análisis de comparaciones múltiples, los resultados obtenidos con LS-SVM son estadísticamente mejores que los obtenidos por la LSTM con las estrategias MIMO y recursiva a 36 meses. No se encontraron diferencias entre LS-SVM y LSTM recursivo a tres días, ni entre LSTM MIMO y LSTM recursivo a 36 días.

Tabla 5.34

Índices MAE promedio y desviación estándar para los modelos mensuales LS-SVM y LSTM (medias de los tres horizontes y las 17 boyas).

LS-SVM	LSTM MIMO	LSTM 36	LSTM 3
0,26 (0,07)	0,34 (0,11)	0,34 (0,11)	0,30 (0,09)

5.4.6.4. Coste computacional

Además de la precisión, otro factor importante a tener en cuenta a la hora de evaluar un modelo es su coste computacional, ya que un tiempo de ejecución demasiado elevado puede hacer que el modelo no sea utilizable en la práctica. También puede ocurrir que la mejora en precisión no compense debido al aumento en el tiempo de cálculo. Por estos motivos, se presenta un análisis del coste computacional de los modelos utilizados en las predicciones de SST.

Debido a las características de los diferentes modelos (arquitecturas y ajustes de parámetros) es difícil establecer una comparación de tiempos exacta. Para realizar este análisis se ha tenido en cuenta una única repetición de la etapa de entrenamiento, utilizando el número de neuronas, hiperparámetros o el factor de ensanchamiento óptimo en cada caso. No se ha contabilizado, por lo tanto, la etapa de ajuste y

validación de los modelos. Como excepción, en los RFBT sí se incluye el tiempo de ajuste de la arquitectura. Puesto que los algoritmos RMLR y RFBT necesitan un modelo por horizonte de predicción, se han sumado los tiempos de los tres modelos. Además, todas las pruebas se han realizado en la misma máquina para evitar las diferencias debidas a la potencia de cálculo.

En la tabla 5.35 se muestra el coste computacional medio en las 17 boyas para cada modelo en las pruebas diarias y mensuales utilizando como entradas la SST y la fecha. Con los datos diarios, el modelo más rápido es la ELM (24 ms), mientras que el más lento es el modelo LSTM (43,7 s). En los datos mensuales, puesto que el número de patrones es muy inferior, los tiempos de cálculo se reducen en todos los casos. Los modelos ELM y LS-SVM son ahora los más rápidos (1 ms), siendo el modelo LSTM el más lento (2,28 s). No obstante, aunque en todos los casos el tiempo de ejecución pueda parecer bajo, en la práctica, la búsqueda de la arquitectura óptima mediante validación cruzada y el barrido del número de neuronas incrementa estos tiempos notablemente. Además, en el caso la LS-SVM, el tiempo para la obtención de los hiperparámetros depende del número de patrones con una relación de N^2 , por lo que este proceso es muy rápido con los datos mensuales, pero muy costoso en los datos diarios.

Tabla 5.35

Coste computacional en segundos (media y desviación estándar de las 17 boyas) de cada modelo utilizando como entradas la SST y la fecha codificada.

	RMLR	ELM	LS-SVM	GRNN	RFBT	MLP	LSTM
Diarios	0,034 (0,004)	0,024 (0,007)	5,572 (3,716)	0,044 (0,033)	0,766 (0,042)	0,498 (0,210)	43,733 (25,585)
Mensuales	0,005 (0,001)	0,001 (0,001)	0,001 (0,001)	0,041 (0,010)	0,353 (0,005)	0,061 (0,014)	2,280 (0,998)

5.5. Conclusiones

En este capítulo de la Tesis Doctoral se ha abordado el problema de las predicciones diarias y mensuales de la SST en 17 de las boyas que forman la red del proyecto PIRATA. Para la extracción y selección de características se ha propuesto el análisis PCA, el algoritmo *RReliefF*, así como la combinación entre el análisis de correlación y el algoritmo *RReliefF*. Además, se han comparado los resultados de estos métodos con utilizar únicamente la SST y fecha como entradas a los modelos. En cuanto a las predicciones de SST, los modelos propuestos han sido. modelo trivial, modelo climatología, RMLR, MLP, ELM, GRNN, RFBT, SVM y LSTM. En el caso de los datos mensuales, además, el desempeño de los modelos ha sido comparado con las predicciones del modelo numérico CFSv2. Del análisis de los resultados se pueden extraer las siguientes conclusiones:

- La inclusión de la fecha codificada ayuda a los modelos a predecir la componente estacional de las series de SST mejorando, por tanto, las predicciones.
- Tanto en los modelos diarios como en los mensuales, los menores índices MAE se obtienen empleando únicamente la fecha codificada y la SST actual como predictores.
- Los métodos de extracción y selección de características no consiguieron obtener un subconjunto de datos que permitiese lograr mejores resultados que utilizando exclusivamente la fecha codificada y la SST.
- Comparando la precisión de los diferentes modelos diarios con fecha y SST, los menores índices MAE promedio se obtuvieron con los modelos ELM, MLP, RFBT, RMLR y GRNN, no encontrándose diferencias estadísticamente significativas.
- En los datos mensuales, el modelo LS-SVM obtuvo resultados significativamente mejores a los logrados por los demás modelos de predicción.
- Los modelos LSTM diarios entrenados con estrategias MIMO y recursiva a 30 pasos, obtuvieron resultados similares a los modelos ELM. En los datos mensuales, LS-SVM obtuvo resultados estadísticamente mejores que los modelos LSTM MIMO y LSTM recursivo a 36 pasos. Sin embargo, LS-SVM y LSTM recursivo a tres días mostraron índices de error similares.
- Como conclusión final, la ELM obtuvo el menor error promedio y con el menor tiempo de entrenamiento en los datos diarios. Por otra parte, el modelo LS-SVM obtuvo el mejor rendimiento en los datos mensuales.

Capítulo 6

Visualización y predicción de datos PIRATA con Processing

6.1. Introducción

En capítulos anteriores se ha puesto de manifiesto que las fluctuaciones océano-atmosféricas que afectan al océano Atlántico tienen una gran repercusión socioeconómica en los países de las regiones tropicales (Servain et al., 1998). Debido a este hecho, existen diversos proyectos científicos enfocados al estudio de estos fenómenos climáticos. Por ejemplo, el proyecto PIRATA utiliza una red de boyas fijas para adquirir variables atmosféricas y oceánicas con en el objetivo de observar los dos modos principales de variabilidad del océano Atlántico (Bourlès et al., 2008). Por una parte, durante las fases cálidas del modo ecuatorial, se producen vientos alisios débiles y valores de la SST anormalmente altos, mientras que en las fases frías, se observan vientos fuertes y valores de la SST anormalmente bajos. Por otra parte, el modo dipolo está relacionado con la migración estacional de la ITCZ y se caracteriza por una oscilación en el gradiente interhemisférico norte-sur de la SST (Servain, 1991). Por tanto, la SST es uno de los parámetros climáticos con mayor influencia en estos dos modos de variabilidad, por lo que su predicción es de gran importancia para la comprensión del clima a nivel regional y global (Servain et al., 2000).

Con la finalidad de facilitar su observación y análisis, los datos relativos a la SST y a otras variables climáticas se encuentran disponibles tanto para la comunidad científica como para el público general. Por ejemplo, los datos de los proyectos TAO/TRITON, PIRATA y RAMA se pueden descargar y visualizar en la aplicación web de la GTMBA perteneciente a la NOAA (GTMBA, 2016). En la figura 6.1 se puede observar el aspecto que presenta esta aplicación y, en la figura 6.2, se muestra una ejemplo del tipo de gráfica que permite generar.

Data Display and Delivery

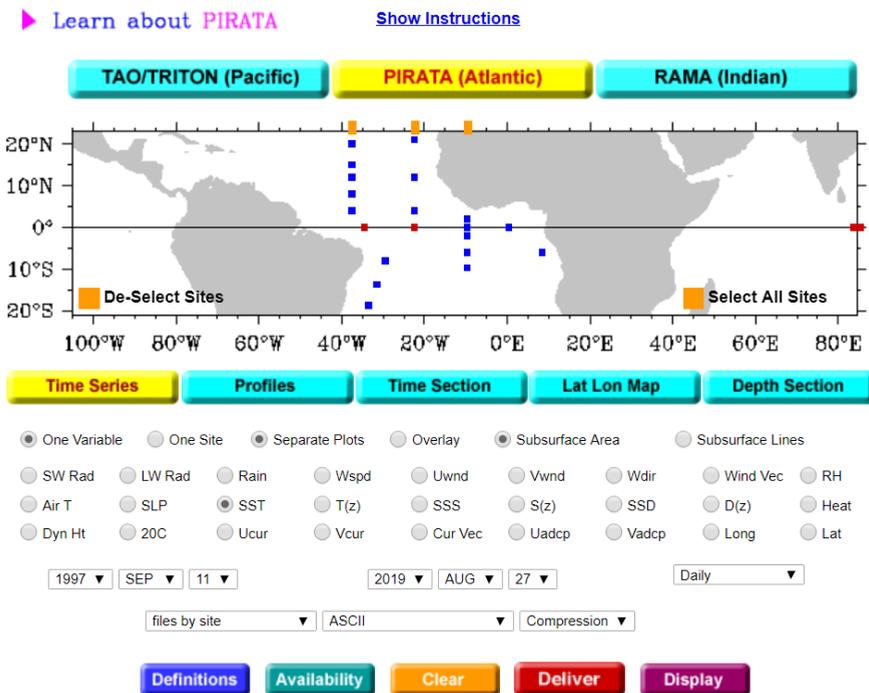


Figura 6.1

Aplicación web de la *Global Tropical Moored Buoy Array Project Office* para la descarga y visualización de datos de los proyectos TAO/TRITON, PIRATA y RAMA (GTMBA, 2016).

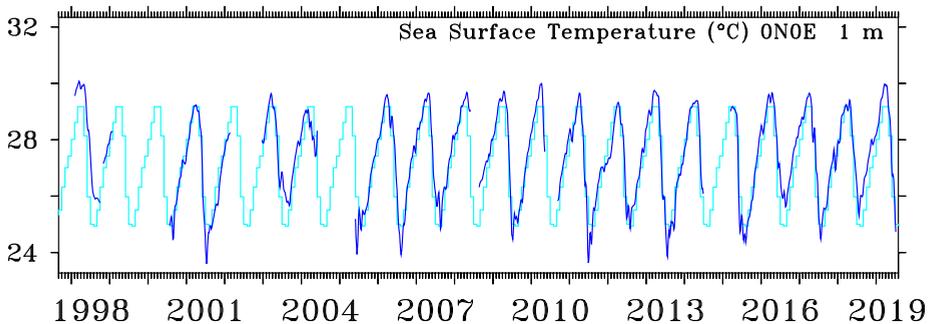


Figura 6.2

Ejemplo de gráfica de las series mensuales de SST y climatología generada con la aplicación de la GTMBA para la boya 0°N - 0°E.

Además de la aplicación de la GTMBA, existen otras muchas herramientas para el análisis de variables climáticas. Algunos ejemplos de las más importantes son:

- **NOAA View Data Exploration Tool.** Esta aplicación web de la NOAA permite visualizar sobre un mapamundi la evolución en el tiempo de más de 100 variables ambientales utilizando datos de satélites, modelos climáticos y otros dispositivos de observación (NOAA, 2019).
- **Ferret.** Es un entorno de programación para la visualización y el análisis de grandes conjuntos de datos climáticos diseñado para oceanógrafos y meteorólogos. Aunque su funcionamiento se basa en el uso de funciones o inserción de instrucciones mediante línea de comandos, también dispone de una interfaz gráfica de usuario (Ferret, 2019).
- **Live Access to the National Virtual Ocean Data System (NVODS).** Mediante esta aplicación web, cuyo motor se basa en *Ferret*, es posible visualizar de manera global una gran cantidad de variables océano-atmosféricas procedentes de las principales bases de datos climáticas.

En general, las aplicaciones mencionadas anteriormente permiten visualizar variables climáticas a nivel global haciendo uso de múltiples fuentes de datos. En cuanto a la aplicación web de la GTMBA para la descarga y visualización de datos PIRATA, las gráficas generadas son imágenes estáticas y, por tanto, no son interactivas. En este capítulo de la Tesis Doctoral se plantea la creación de una herramienta específicamente diseñada para explorar los datos de las boyas PIRATA. Utilizando el entorno de programación *Processing*, se pretende conseguir una aplicación totalmente personalizada y que presente una experiencia de usuario sencilla e intuitiva. Además de descargar y visualizar los datos PIRATA, se pretende que la herramienta sea capaz de obtener predicciones de la SST y permita estudiar el fenómeno del modo dipolo.

6.2. Objetivos

En este capítulo de la Tesis Doctoral se propone una aplicación de usuario diseñada con el objetivo de interactuar con las series de datos del proyecto PIRATA. Esta herramienta ha sido desarrollada en el entorno de programación *Processing* y la implementación de varias de sus funcionalidades principales está basada en las conclusiones obtenidas en anteriores capítulos. Las características deseables que debe presentar la aplicación son las siguientes:

- Interfaz gráfica de usuario atractiva, fluida y fácil de utilizar.
- Descarga rápida de datos PIRATA mediante la conexión al servidor de la NOAA (*TAO PMEL FTP*).

- Visualización interactiva de las series temporales de datos descargadas.
- Predicciones *on line* de la SST mensual mediante un modelo de aprendizaje automático.
- Obtención del dipolo a partir de la selección de las boyas de interés.
- Posibilidad de exportar gráficas y datos numéricos.

Además de describir en profundidad la aplicación y su funcionamiento interno, en la sección de resultados se mostrarán una serie de casos prácticos de estudio.

6.3. Materiales y métodos

6.3.1. Entorno de programación

La implementación de una interfaz gráfica de usuario se puede llevar a cabo mediante diferentes plataformas de desarrollo. Por ejemplo, son ampliamente utilizadas las tecnologías WPF y UWP de *Microsoft*, las librerías AWT y *Swing* de *Java* o la herramienta GUIDE de *Matlab*. Además, también existen librerías de programación específicas para el desarrollo de aplicaciones orientadas a la visualización interactiva de datos como *D3.js*, *Bokeh* o *Processing*, basadas en *JavaScript*, *Python* y *Java*, respectivamente. Debido a sus características y gran versatilidad, la aplicación que se presenta en este capítulo ha sido desarrollada en la plataforma *Processing*.

Processing es un entorno de desarrollo y un lenguaje de programación de libre distribución, creado por Ben Fry y Casey Reas en el *MIT Media Lab* en 2001 (Reas y Fry, 2010). Inicialmente, *Processing* se diseñó con la finalidad de ser utilizado en la enseñanza de los fundamentos de programación en el contexto del diseño gráfico. Gracias a la implicación de los usuarios de la comunidad *online*, este entorno de programación ha evolucionado rápidamente. Actualmente dispone de una gran variedad de librerías creadas por los propios usuarios, que permiten simplificar la programación en campos como la visión artificial, animaciones en 2D y 3D, visualización de datos, interfaces gráficas de usuario, aplicaciones para dispositivos *Android*, etc. (Shiffman, 2008). *Processing* emplea las mismas estructuras y conceptos de programación que otros lenguajes (bucles, sentencias, tipos de variables, etc.), hecho que facilita el aprendizaje a usuarios con experiencia previa en programación. En particular, *Processing* está basado en *Java*, por lo que es posible utilizar las librerías desarrolladas en este lenguaje de programación.

Puesto que *Processing* es un lenguaje pensado para la creación de animaciones e interacciones con el usuario, se genera una realimentación visual inmediata entre el código y los gráficos mostrados por pantalla. Esto facilita la comprensión del código y motiva al programador a mejorar sus diseños. La condición de multiplataforma de *Processing* es otra de sus grandes ventajas, ya que es capaz de funcionar en los sistemas

operativos *MAC*, *Windows* y *Linux*. Además, los programas generados pueden ser exportados como aplicaciones web o como aplicaciones independientes (*standalone applications*) a cualquiera de las plataformas mencionadas (Reas y Fry, 2010).

6.3.2. Descripción de la aplicación

La interfaz de usuario se ha diseñado con la finalidad de resultar sencilla e intuitiva, por lo que se ha minimizado al máximo el número de controles, comandos y opciones de configuración. A excepción de la función de exportar a PDF, únicamente es necesario el uso del ratón para acceder a las distintas funcionalidades. La figura 6.3 muestra el aspecto que presenta la ventana inicial de la aplicación. En esta pantalla se muestra un mapamundi centrado en el Atlántico tropical y las diferentes boyas que componen la red PIRATA. En la parte superior de la pantalla está situado el menú desplegable de opciones que permite seleccionar la variable a descargar y visualizar, así como el periodo de tiempo de la serie (mensual / diario). Por último, el botón “Exportar” permite exportar los datos a una hoja de cálculo.

Por defecto, aparece seleccionada la variable SST y el periodo mensual, por lo que únicamente es necesario seleccionar una boya para poder descargar y visualizar los datos correspondientes. Esta selección se realiza pulsando el botón izquierdo del ratón sobre la boya de interés. Tras unos pocos segundos se muestran los datos seleccionados. En la figura 6.4 se puede observar un ejemplo para la boya 4°N - 38°W. La gráfica generada ocupa la mayor parte de la pantalla, quedando el mapa de selección de las boyas en la parte superior de la misma. En el caso de las SSTs mensuales, además de la serie temporal (línea azul), se muestra la media histórica o climatología (en rojo) y las predicciones a 12 meses (en verde). Encima de la esquina superior izquierda de la gráfica se muestra la variable y la boya seleccionada. Al desplazar el ratón por encima de las series, aparece un cursor vertical con información acerca del punto seleccionado. En esta información se muestra el valor observado de la SST, de la climatología, de la anomalía (diferencia entre la SST y la climatología) y la fecha.

Una vez mostradas las gráficas en pantalla, es posible exportar sus datos numéricos en una hoja de cálculo mediante el botón “Exportar”. Al pulsar este botón se muestra una pantalla para seleccionar el destino y el nombre del archivo. Además, es posible guardar las gráficas en un documento PDF pulsando la tecla “s”. En este caso, el archivo se guarda en la carpeta de instalación con la fecha y hora de guardado como nombre. En la parte inferior central de la pantalla se encuentra el control “Rango datos” que permite desplazarse por la serie y seleccionar el dato inicial / final a visualizar. Puesto que las series de datos se ajustan siempre al tamaño del área gráfica, reduciendo el número de datos en pantalla se consigue realizar una ampliación de la serie. Manteniendo pulsado el botón izquierdo del ratón sobre el triángulo negro de la esquina inferior derecha, es posible reducir o aumentar el tamaño de la gráfica. Por otra parte, manteniendo pulsado el botón izquierdo del ratón sobre el cuadro negro de la esquina superior izquierda, es posible cambiar la posición de la gráfica.

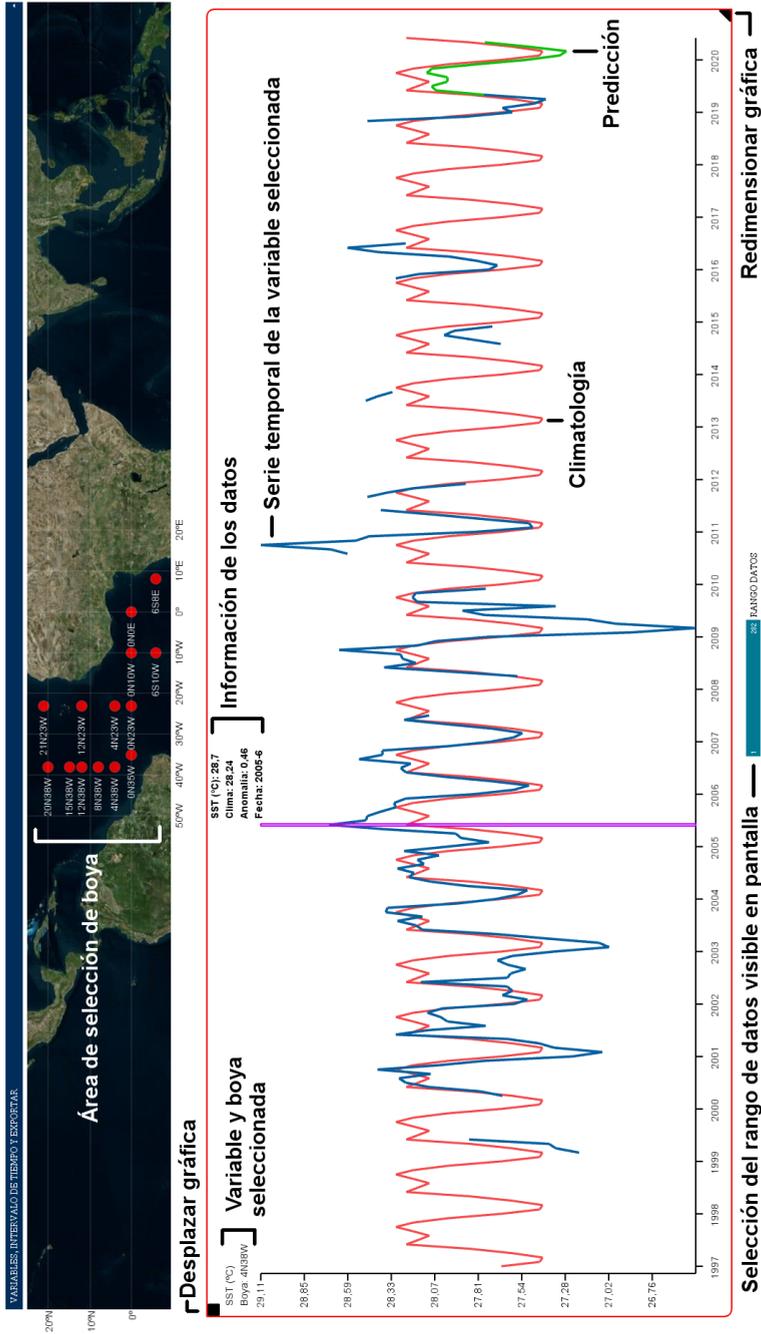


Figura 6.4
Visualización de datos mensuales.

En la figura 6.5 se muestra el “modo dipolo” (únicamente para SSTs mensuales). Una vez seleccionada una boya con el botón izquierdo del ratón, este modo se activa seleccionando una segunda boya con el botón derecho. En la pantalla principal aparecerán entonces tres gráficas: en la parte superior izquierda se muestra la gráfica correspondiente a la boya seleccionada con el botón izquierdo, en la parte inferior izquierda la boya seleccionada con el botón derecho y, en la parte superior derecha, el dipolo resultante. Pulsando el botón central del ratón se reinicia el área gráfica y el mapa de selección de boyas vuelve a ocupar toda la pantalla de la aplicación. De manera similar al anterior ejemplo (figura 6.4), en las dos gráficas correspondientes a las boyas se muestran las series históricas de SST, la climatología y las predicciones. En la gráfica del dipolo se muestra una línea roja de referencia en 0 °C y el dipolo (diferencia entre las anomalías de temperatura entre ambas boyas), calculado a partir de los datos observados y de los datos predichos (en verde). El tamaño y posición de estas gráficas se puede modificar siguiendo el procedimiento explicado anteriormente (cuadrado para desplazar y triángulo para ampliar/reducir). El control “Rango datos” modifica los datos visibles en pantalla de manera sincronizada en las tres gráficas. En este modo también es posible exportar los datos y gráficas. En la figura 6.6 se muestra un ejemplo de datos exportados a un documento .xlsx.

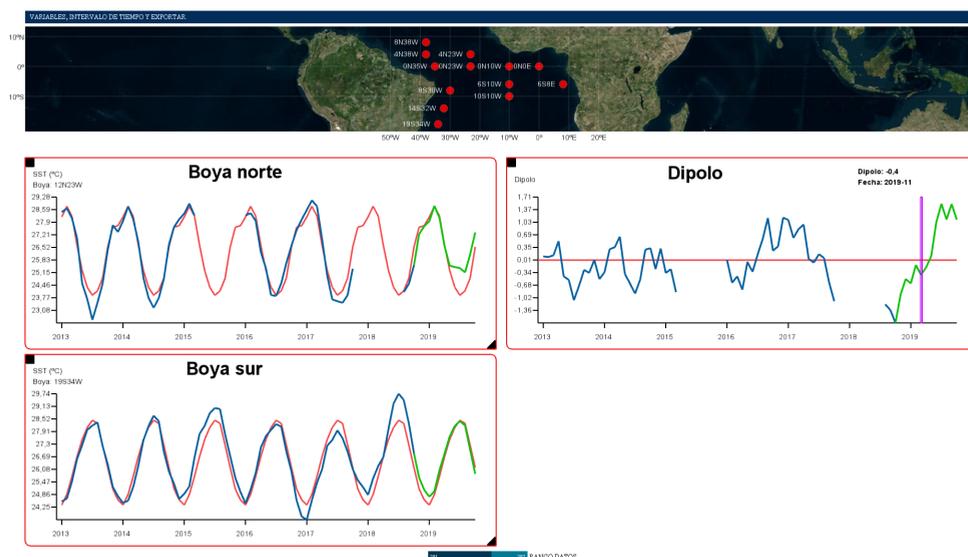


Figura 6.5
Visualización del modo dipolo.

Las variables viento, salinidad, densidad y temperatura del mar disponen de múltiples variables, tanto en el periodo diario como en el mensual, por lo que es necesario un menú desplegable adicional para seleccionar la variable o profundidad de interés. En la figura 6.7 se muestra un ejemplo para la variable viento en la boya 0°N - 23°W.

	A	B	C	D	E	F	G	H	I	J
1	Año	Mes	Día	SST1	Clima	Anomalía	SST2	Clima	Anomalía	Dipolo
122	2007	1	1	26.13	26.17	-0.04	27.80	27.57	0.23	-0.26
123	2007	2	1	25.77	25.55	0.22	28.27	28.05	0.22	0.00
124	2007	3	1	25.12	25.28	-0.16	28.50	28.55	-0.05	-0.11
125	2007	4	1	25.16	25.49	-0.33	28.66	28.85	-0.19	-0.13
126	2007	5	1	25.59	25.88	-0.29	28.44	28.69	-0.25	-0.04
127	2007	6	1	26.58	26.42	0.16	27.91	28.19	-0.28	0.44
128	2007	7	1	26.99	27.23	-0.24	27.09	27.48	-0.39	0.15
129	2007	8	1	27.56	28.05	-0.49	26.38	26.82	-0.44	-0.05
130	2007	9	1	27.68	28.41	-0.73	25.85	26.39	-0.54	-0.19
131	2007	10	1	27.35	28.22	-0.87	25.98	26.44	-0.46	-0.41
132	2007	11	1	27.38	27.70	-0.32	26.36	26.66	-0.30	-0.02
133	2007	12	1	26.57	27.16	-0.59	27.11	27.18	-0.07	-0.52
134	2008	1	1	25.51	26.17	-0.66	27.47	27.57	-0.10	-0.55
135	2008	2	1	24.65	25.55	-0.90	28.09	28.05	0.04	-0.94
136	2008	3	1	25.37	25.28	0.09	28.57	28.55	0.02	0.07
137	2008	4	1	25.50	25.49	0.01	28.92	28.85	0.07	-0.05
138	2008	5	1	25.84	25.88	-0.04	28.69	28.69	0.00	-0.04
139	2008	6	1	26.63	26.42	0.21	27.83	28.19	-0.36	0.57
140	2008	7	1	27.48	27.23	0.25	27.14	27.48	-0.34	0.59
141	2008	8	1	28.30	28.05	0.25	26.71	26.82	-0.11	0.36
142	2008	9	1	28.68	28.41	0.27	26.62	26.39	0.23	0.04
143	2008	10	1	28.30	28.22	0.08	26.52	26.44	0.08	0.00
144	2008	11	1	27.74	27.70	0.04	26.58	26.66	-0.08	0.12
145	2008	12	1	27.65	27.16	0.49	27.22	27.18	0.04	0.45

Figura 6.6
Ejemplo de datos (“modo dipolo”) exportados a una hoja de cálculo.

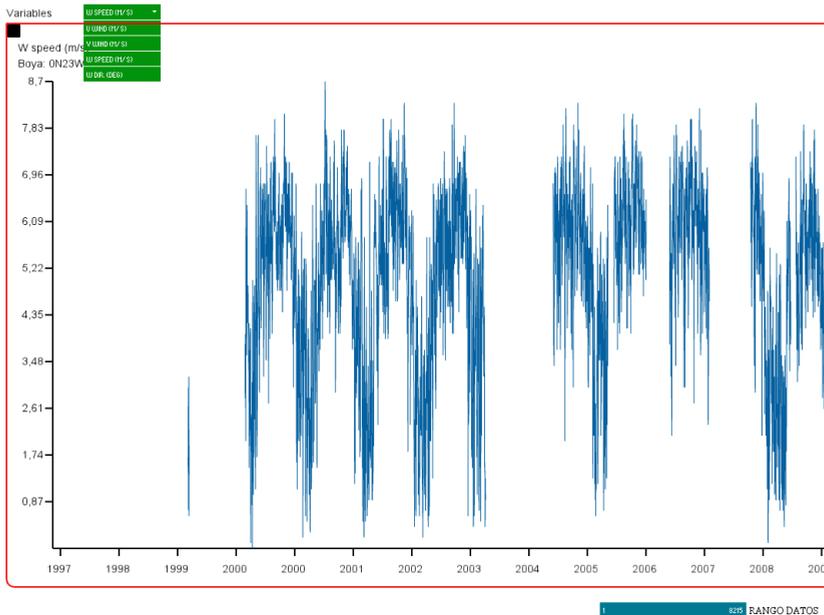


Figura 6.7
Menú de selección de variables del viento.

En el menú desplegable se pueden seleccionar las distintas variables disponibles: componentes este-oeste y norte-sur, velocidad y dirección. En cuanto a las variables con diferentes profundidades, se muestra un ejemplo para la boya 8°S - 30°W en la figura 6.8. En este caso, el menú permite seleccionar la profundidad de registro de la temperatura del mar. Para las variables salinidad y densidad se muestra un menú similar, aunque con un menor número de profundidades disponible.

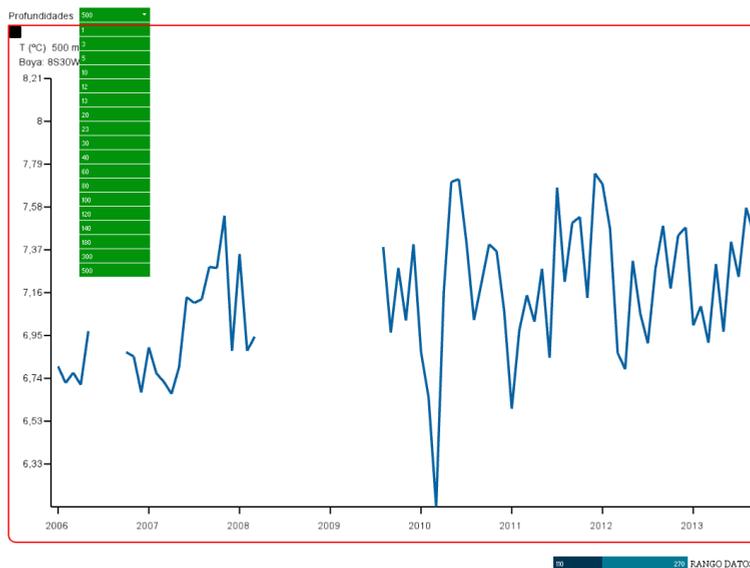


Figura 6.8
Menú de selección de profundidades.

6.3.3. Funcionamiento interno

El código de la aplicación se ha diseñado siguiendo una estructura modular para facilitar la legibilidad del código y la corrección de errores. Además, esta estructura permite reutilizar los módulos de código o funciones de manera más eficiente. El diagrama de bloques de la aplicación se muestra en la figura 6.9. Los módulos principales que forman la aplicación son: carga de datos, procesado, predicciones y visualización.

6.3.3.1. Carga de fichero de datos

La aplicación es compatible con los archivos de datos *.ascii* que proporciona el proyecto PIRATA en su página web y en su servidor FTP (*File Transfer Protocol*). No obstante, la aplicación desarrollada descarga los datos de manera automática mediante FTP, por lo que no es necesario descargarlos previamente.

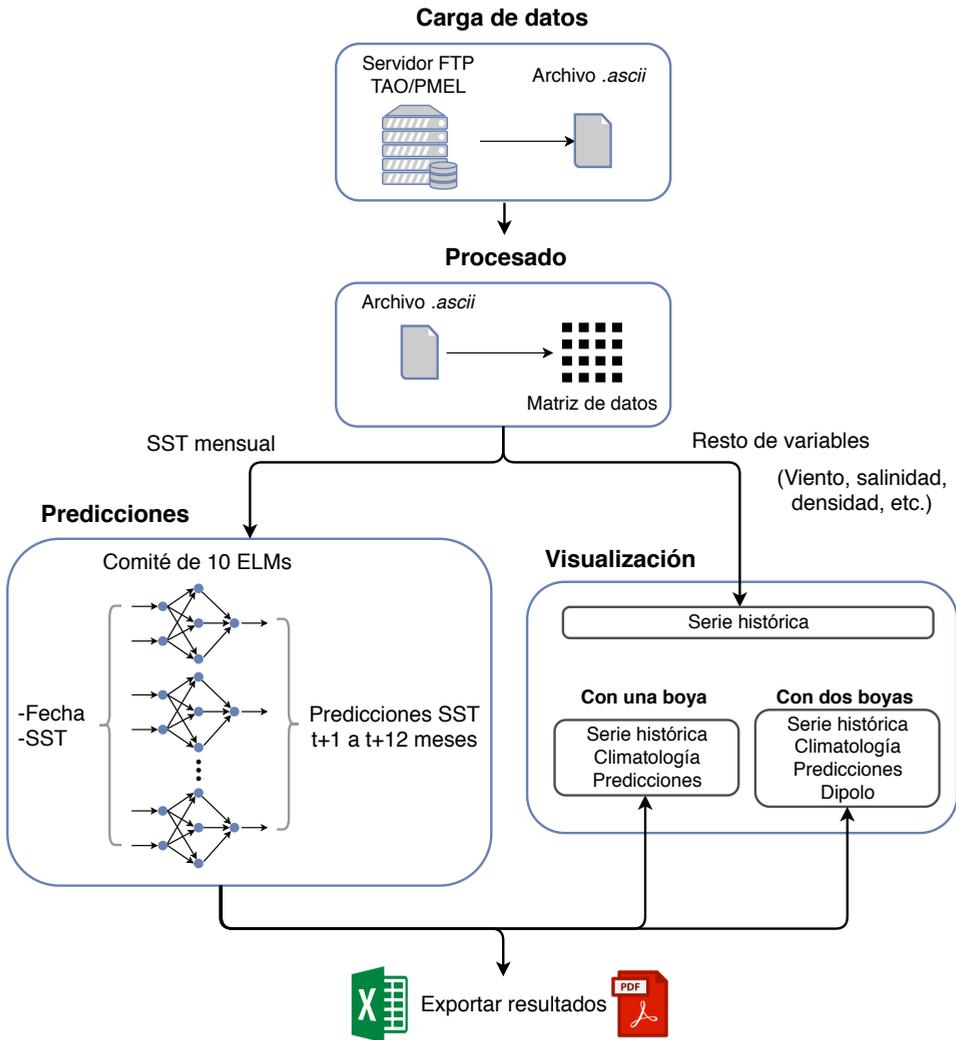


Figura 6.9
Esquema de funcionamiento y módulos principales de la aplicación.

Al seleccionar una variable, periodo y boya de interés, el programa busca el archivo en el servidor y en la carpeta local “Datos”, situada en el directorio de instalación de la aplicación. En el caso de no encontrarse el fichero en la carpeta local, el archivo se descarga del servidor FTP. Si el servidor dispone de un archivo más reciente que el disponible en la carpeta local, el fichero se descarga y reemplaza al anterior. Para realizar la conexión con el servidor se utiliza la librería *Java edtFTPj* (Blackshaw, 2016). Tras obtener el fichero, se aplica la etapa de procesado de datos.

6.3.3.2. Procesado de datos

Cada tipo de variable se proporciona en un fichero individual que presenta una estructura de datos por bloques de fechas. Estos archivos contienen cabeceras de información que preceden a las series de datos en cada uno de los bloques. En las cabeceras se muestra información acerca de la variable medida, número de registros y número de datos del bloque, así como de las fechas e índices de inicio y fin de los datos del bloque. El número de profundidades registradas es diferente para cada variable y boya seleccionada, pudiendo variar además de un bloque de datos al siguiente. Para poder procesar estas variables, por tanto, es necesario utilizar los datos que ofrecen las cabeceras sobre el número de profundidades registradas en cada caso. Por otra parte, los registros de los ficheros aparecen ordenados de manera cronológica de más antiguo a más reciente. Sin embargo, debido a fallos de los sensores o tareas de mantenimiento, pueden existir saltos temporales de días o meses entre el último dato de uno de los bloques y el primero dato del siguiente. En cuanto a las variables registradas en cada archivo, su número varía en función del tipo de variable. Las series de datos estándar disponen de cuatro columnas con la fecha, hora, medición, calidad y fuente del dato registrado. Las variables temperatura del aire, lluvia o SST siguen esta estructura. En el caso de la variable viento, la medición se registra en cuatro subvariables. Por último, las variables salinidad, densidad y temperatura del mar contienen registros a diferentes profundidades.

Con la finalidad de conseguir una matriz de datos utilizable por los siguientes módulos del programa, en primer lugar se genera una matriz de datos base con registros inicializados a *NaN* (*Not a Number*). Esta matriz contiene un número de filas igual al número de fechas posibles desde el inicio del proyecto hasta la fecha actual, así como un número de columnas igual al mayor número posible de columnas que pueda contener el tipo de variable seleccionado. Posteriormente, se asignan los registros de los ficheros a la fila y columna correspondiente en la matriz base tomando como referencia la fecha (año y mes para datos mensuales; año, mes y día para datos diarios) y el tipo de variable. Puesto que los registros no utilizados de la matriz base permanecen en *NaN*, es posible representar las series mostrando la posición temporal real de cada dato.

En el caso de la variable SST mensual, el siguiente paso consiste en calcular la serie de climatología. Esta serie se obtiene calculando el valor promedio de SST en todos los meses de enero, febrero, marzo, etc. disponibles, obteniéndose una serie con 12 datos. Esta serie se replica un número de veces igual al número de años de la serie de SST para poder comparar cada dato real con el valor de la climatología. Finalizado este cálculo, el programa continúa con el módulo de obtención de predicciones.

6.3.3.3. Cálculo de las predicciones

Puesto que para estudiar el fenómeno del dipolo del Atlántico resulta de mayor interés la obtención de las predicciones mensuales, se decidió implementar únicamente

modelos mensuales. En cuanto al horizonte de predicción, se decidió extenderlo hasta $t + 12$, por lo que el modelo debe obtener las predicciones de la SST desde $t + 1$ a $t + 12$ meses. Además, para que el modelo pueda adaptarse a posibles cambios en el comportamiento histórico de las series, en lugar de implementar un modelo entrenado previamente (aprendizaje *off line*), se decidió implementar un modelo capaz de utilizar los datos más recientes para su entrenamiento y posterior obtención de las predicciones (aprendizaje *on line*).

Con el fin de escoger el modelo más adecuado para realizar estas predicciones, además de los resultados obtenidos en cuanto a error cometido y tiempo de cálculo, se ha tenido en cuenta la complejidad de su implementación. Los resultados mostrados en el capítulo 5 indicaron que el modelo LS-SVM presenta el menor índice MAE promedio [MAE = 0,29 (0,08) °C] y el menor tiempo de entrenamiento (1 ms). Sin embargo, para cumplir los requisitos expuestos anteriormente, su implementación requiere del cálculo de hiperparámetros diferentes para cada boya y para cada horizonte de predicción mediante validación cruzada y, además, estos valores deben actualizarse al disponer de nuevos datos.

El siguiente modelo con la mejor relación entre MAE y velocidad de ejecución es la ELM, con un MAE de 0,31 (0,08) °C e idéntico tiempo de entrenamiento que el modelo LS-SVM. El único parámetro de ajuste de este modelo es el número de neuronas de la capa oculta. A diferencia de los hiperparámetros del modelo LS-SVM, el número óptimo de neuronas de la ELM permanece prácticamente constante en las diferentes boyas, horizontes de predicción y al incluir nuevos datos en el entrenamiento. Por tanto, se decidió escoger el modelo ELM ya que presenta un desempeño similar a LS-SVM y su implementación resulta menos compleja.

Antes de incluir la ELM en la aplicación, se obtuvo el número de neuronas óptimo en *Matlab* siguiendo el mismo procedimiento que se detalló en el capítulo 5. Es decir, mediante validación cruzada, realizando un barrido del número de neuronas y escogiendo la arquitectura con menor error de validación. La única diferencia consistió en entrenar los modelos con 12 horizontes de predicción en lugar de con tres. El menor índice de validación promedio en las 17 boyas se obtuvo con 15 neuronas en la capa oculta, siendo el MAE de 0,31 (0,09) °C promediando los resultados de los horizontes $t + 1$ a $t + 3$ y, de 0,38 (0,09) °C, promediando los resultados de $t + 1$ a $t + 12$. El resultado de $t + 1$ a $t + 3$ es similar al obtenido con los modelos ELM entrenados con tres horizontes de predicción (tabla 5.31), por lo que ampliar el número de predicciones no ha empeorado su rendimiento.

Con el objetivo de obtener un resultado más robusto, se implementó un comité de 10 ELMs con igual arquitectura y función de activación, pero con diferentes inicializaciones aleatorias de los pesos y sesgos de la capa oculta. Las predicciones finales se obtienen promediando los resultados de las 10 ELMs. Aunque existen varias librerías en *Java* que incluyen el modelo ELM, se decidió realizar una implementación propia. No obstante, para el cálculo de la pseudo-inversa se utilizó la librería JAMA para cálculos matriciales versión 1.0.3 (Hicklin et al., 2012).

La primera etapa para entrenar la ELM consiste en generar la matriz de entradas y deseadas. Esta matriz está compuesta por la fecha y la SST en el instante t como entradas, así como la SST en los instantes $t + 1$ a $t + 12$ como deseadas. Tras generar la matriz, se eliminan los patrones con datos ausentes y se codifica la fecha con el seno y el coseno (ver sección 5.3.1 del capítulo 5). Finalmente, se estandarizan las entradas con media cero y varianza unidad. A continuación, para entrenar la red, se aplican los pasos detallados en la sección 5.3.4.5 del capítulo 5:

1. Inicializar los pesos y sesgos de entrada con valores aleatorios.
2. Calcular la matriz de pesos de salida de la capa oculta \mathbf{H} .
3. Obtener los pesos de salida $\hat{\beta}$ mediante la pseudo-inversa de la matriz \mathbf{H} y la matriz de salidas deseadas \mathbf{T}_d ($\hat{\beta} = \mathbf{H}^\dagger \mathbf{T}_d$).

Una vez finalizado el proceso de entrenamiento se obtienen las predicciones:

1. Estandarizar el patrón de datos de entrada más reciente disponible utilizando la media y desviación estándar de los datos de entrenamiento.
2. Obtener el vector de pesos de salida de la capa oculta.
3. Obtener las predicciones utilizando los pesos de salida calculados en la fase de entrenamiento ($\mathbf{T}_p = \mathbf{H}\hat{\beta}$).

Posteriormente, las predicciones se copian a la matriz de datos base. A continuación, se calculan las diferencias entre la serie de la SST observada y la climatología para obtener las anomalías. En el caso de haber seleccionado una segunda boya, se repiten las etapas de procesado y predicción en la nueva boya. Finalmente, el último de los cálculos consiste en obtener el dipolo a partir de las anomalías de ambas boyas.

6.3.3.4. Visualización

En *Processing*, el código relativo a la representación gráfica debe incluirse dentro de la función denominada “draw()”, que se ejecuta indefinidamente en cada refresco de pantalla. Esta tasa de refresco por defecto es de 60 Hz, por lo que es importante optimizar el código para que el procesador consiga realizar todas las operaciones necesarias dentro del tiempo de cada ciclo. Por tanto, con el fin de no sobrecargar el procesador y evitar cálculos redundantes, los tres módulos anteriores se calculan fuera de la función “draw()”.

Después de cargar los datos y realizar todos los cálculos necesarios se activa el último módulo, que se encarga de mostrar de manera gráfica los resultados. Las gráficas mostradas son de implementación propia, por lo que no se han utilizado librerías adicionales. De manera resumida, el proceso general para representar los datos está formado por los siguientes pasos:

1. Calcular el valor máximo y mínimo de la serie de datos.
2. Definir el tamaño y posición del área gráfica y dibujar ejes x e y.
3. Realizar un mapeo entre el valor numérico de cada dato y la posición que debe ocupar dentro del área gráfica.
4. Trazar una línea entre el dato actual y el dato anterior.

Para posibilitar la característica de interactividad de la gráfica es necesario repetir el proceso de dibujado de los datos en cada refresco de pantalla (función “draw()”). De esta manera, al posicionar el puntero del ratón sobre un dato de la gráfica, es posible adquirir el índice relativo a la posición de dicho dato y mostrar la información correspondiente.

En cuanto a la interfaz gráfica, ésta se ha diseñado utilizando los controles (botones, menús desplegables, etc.) de la librería *ControlP5* (Schlegel, 2015). Para poder mostrar el mapamundi se ha empleado la librería *unfoldingmaps* (Nagel, 2014). En la sección anterior (sección 6.3.2) se muestran ejemplos del aspecto que presenta la interfaz de usuario y las gráficas de las series de datos.

6.4. Resultados

En esta sección se va a mostrar una serie de casos prácticos de estudio utilizando la herramienta desarrollada. En concreto, se va a mostrar una comparación de la profundidad de la isoterma a 20 °C y de la temperatura del mar en subsuperficie, se van a analizar los valores de la SST y del dipolo durante los eventos meteorológicos que tuvieron lugar en el periodo 2009-2010 y se van a mostrar las predicciones de la SST, así como de los dipolos para los años 2019-2020.

6.4.1. Exploración de la profundidad de la isoterma a 20 °C y de la temperatura del mar

Como se describió en el capítulo 4, la profundidad de la termoclina y el espesor de la capa de mezcla van disminuyendo en dirección este a lo largo del ecuador. Con el fin analizar este fenómeno se ha explorado la profundidad de la isoterma a 20 °C, variable utilizada como referencia de la posición de la termoclina, en la boya situada más al oeste del ecuador (0°N - 35°W) y en la boya situada más al este (0°N - 0°E). En la figura 6.10 se puede observar que, para la boya 0°N - 35°W, la profundidad anual de la isoterma a 20 °C oscila entre los 65 y los 150 m aproximadamente. Por otra parte, la figura 6.11 confirma que en la boya situada al este del ecuador, la profundidad de la termoclina es menor, oscilando entre los 16 y los 100 m.

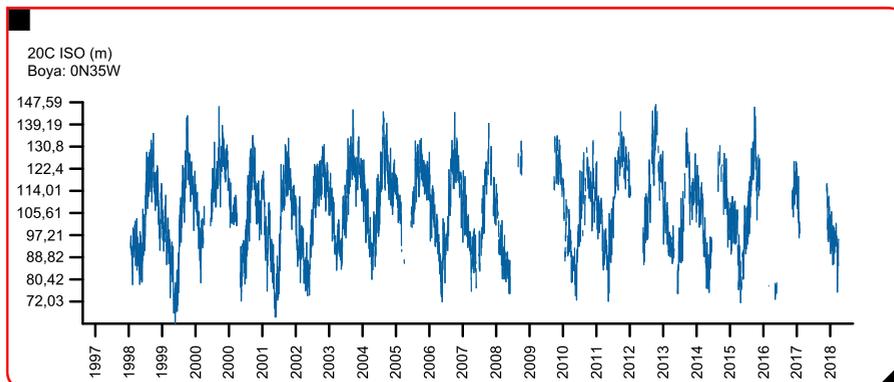


Figura 6.10

Serie temporal diaria de la profundidad de la isoterma a 20 °C en la boya 0°N - 35°W.

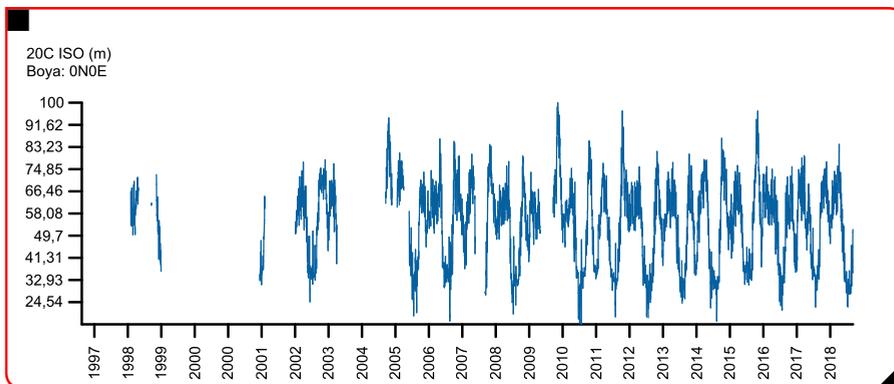
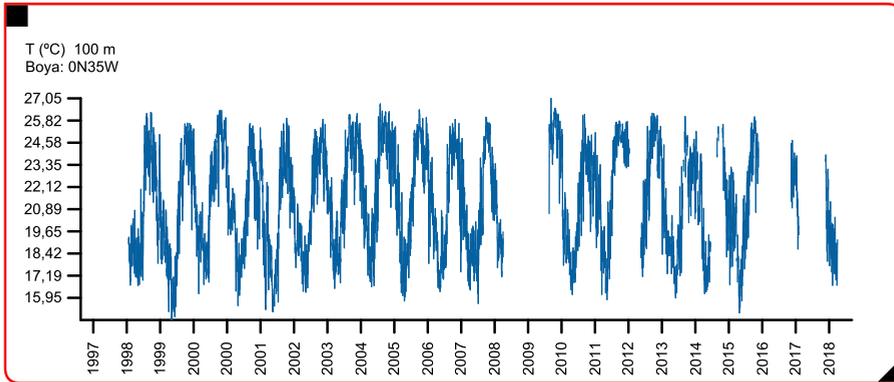


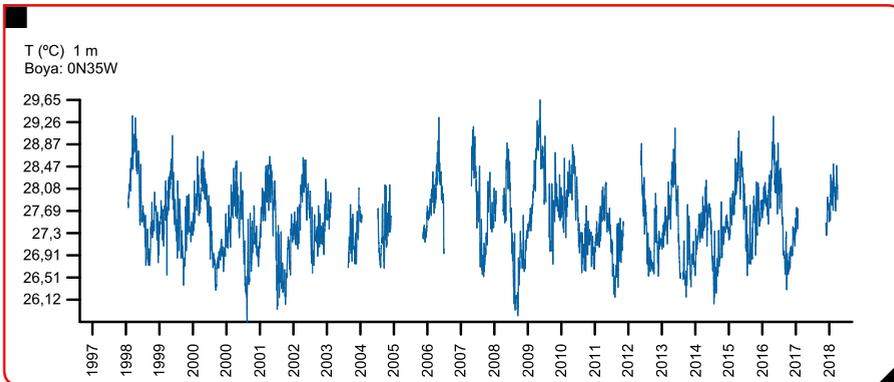
Figura 6.11

Serie temporal diaria de la profundidad de la isoterma a 20 °C en la boya 0°N - 0°E.

Además de la isoterma a 20 °C, también es posible determinar la posición media de la termoclina estudiando el comportamiento de la temperatura a diferentes profundidades. Por ejemplo, en la boya 0°N - 35°W la mayor variación de temperatura, unos 11 °C entre los valores mínimos y máximos, se da a 100 m de profundidad (figura 6.12), indicando que la termoclina se encuentra cerca de estas profundidades. Cuando la capa de mezcla presenta un menor espesor, la termoclina se sitúa por encima de los 100 m y la temperatura del agua a estas profundidades disminuye al mezclarse con aguas profundas más frías. Al aumentar el espesor de la capa de mezcla, la termoclina se sitúa por debajo de los 100 m, aumentando la temperatura del agua al mezclarse con las aguas calientes de la capa de mezcla. Este proceso explica la gran variación en la amplitud de la temperatura que se observa en la figura 6.12. Comparativamente, el rango de la SST es mucho menor (unos 3,5 °C) como muestra la figura 6.13.

**Figura 6.12**

Serie temporal diaria de la temperatura a 100 m de profundidad en la boya 0°N - 35°W.

**Figura 6.13**

Serie temporal diaria de la SST en la boya 0°N - 35°W.

Por otra parte, en la posición 0°N - 0°E, debido a que la capa de mezcla presenta un menor espesor y, por tanto, una menor profundidad de la termoclina, la temperatura en superficie y subsuperficie presenta un comportamiento distinto al observado en la boya 0°N - 35°W. En 0°N - 0°E la mayor variación en la temperatura, 12 °C aproximadamente, se produce a 40 m de profundidad (figura 6.14). Esta profundidad es mucho menor que la observada en la boya 0°N - 35°W. Además, como muestra la figura 6.15, la SST muestra una variabilidad de hasta 8 °C, siendo un rango de variación mucho mayor que en la posición más al oeste. Debido al bajo espesor de la capa de mezcla, la termoclina se sitúa cerca de la superficie, haciendo disminuir la temperatura de las aguas superficiales.

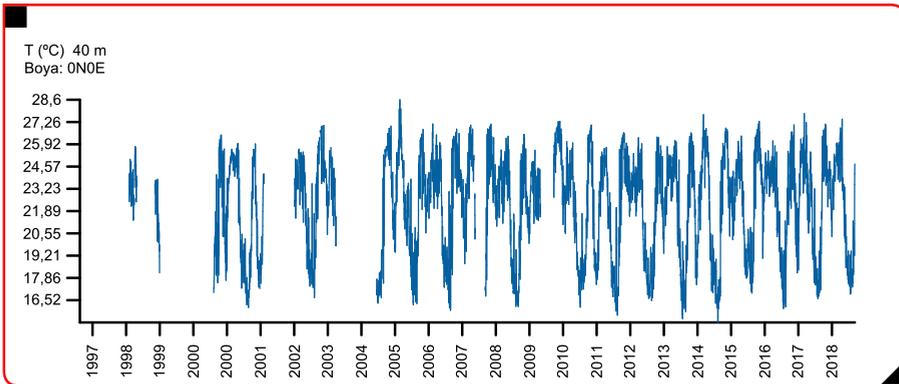


Figura 6.14

Serie temporal diaria de de la temperatura a 40 m de profundidad en la boya 0°N - 0°E.

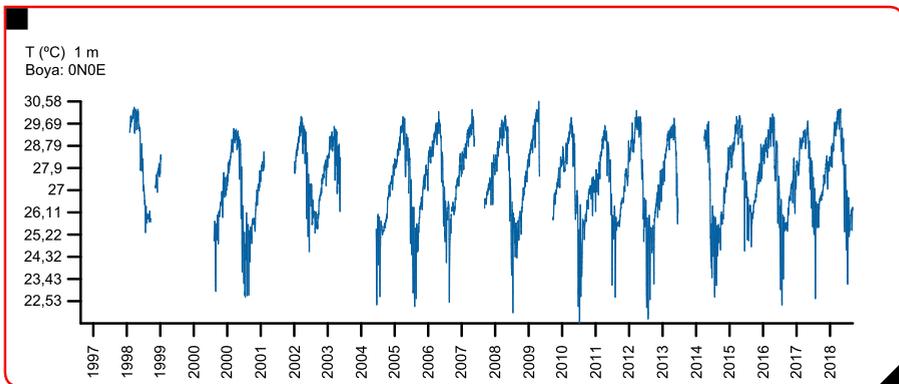


Figura 6.15

Serie temporal diaria de la SST en la boya 0°N - 0°E.

6.4.2. Dipolo negativo del año 2009

En el capítulo 2 se mostraron las anomalías de la SST que tuvieron lugar entre marzo y mayo del año 2009 en el océano Atlántico (figuras 2.6.a, 2.6.b y 2.6.c.). Durante estos meses la SST al sur del ecuador fue anormalmente alta y la SST al norte del ecuador fue anormalmente baja. El dipolo negativo generado tuvo un fuerte impacto en las lluvias del estado de Ceará, registrándose unos valores de precipitación acumulada mucho mayores de lo habitual en un año promedio (figura 2.5). Utilizando la aplicación diseñada, se van a comparar los valores de la SST registrados por diferentes boyas de interés durante este evento, así como sus dipolos correspondientes.

De acuerdo a las conclusiones obtenidas en el capítulo 3, se van a estudiar los dipolos PIRATA 12°N / 14°S y 8°N / 14°S por sus buenos resultados en las correlaciones

con las lluvias. Además, ya que la fundación FUNCEME los monitoriza de manera habitual, se ha incluido el dipolo “clásico” ($15^{\circ}\text{N} / 10^{\circ}\text{S}$) y el dipolo de Servain en esta comparativa. Finalmente, observando la distribución de la SST en las figuras 2.6.a, 2.6.b y 2.6.c., se decidió incluir también el dipolo formado por las boyas $12^{\circ}\text{N} - 23^{\circ}\text{W}$ y $6^{\circ}\text{S} - 10^{\circ}\text{W}$. A excepción del dipolo de Servain y los datos de lluvia, que fueron descargados de la página web de FUNCEME (2016), el resto de datos y gráficas se obtuvieron mediante la herramienta diseñada.

En la tabla 6.1 se muestran las anomalías de la SST registradas por las boyas de interés de enero a mayo de 2009. En la boyas $15^{\circ}\text{N} - 38^{\circ}\text{W}$, $12^{\circ}\text{N} - 23^{\circ}\text{W}$ y $8^{\circ}\text{N} - 38^{\circ}\text{W}$ comenzaron a registrarse valores de la SST menores de lo habitual a partir del mes de febrero. Sin embargo, en la boya $12^{\circ}\text{N} - 38^{\circ}\text{W}$, estas anomalías negativas de la SST no comenzaron hasta mayo. Puesto que las anomalías de la SST comenzaron junto a las costas de África, la boya $12^{\circ}\text{N} - 23^{\circ}\text{W}$ registró las mayores anomalías. En cuanto a las boyas situadas al sur del ecuador, las anomalías se mantuvieron positivas durante todo el periodo analizado en las tres boyas seleccionadas y, en general, la boya $6^{\circ}\text{S} - 10^{\circ}\text{W}$ presentó los mayores valores de anomalía.

Tabla 6.1

Valores de enero a mayo de 2009 de anomalías de temperatura para las boyas seleccionadas.

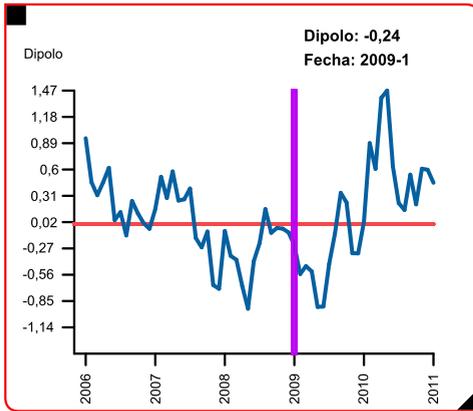
	Anomalía ($^{\circ}\text{C}$)				
	Enero	Febrero	Marzo	Abril	Mayo
15N38W	0,01	-0,42	-0,17	-0,34	-0,52
12N23W	0,00	-1,19	-1,51	-1,43	-0,91
12N38W	0,28	0,02	0,51	0,03	-0,61
8N38W	0,79	-0,47	-0,28	-0,46	-0,95
6S10W	0,53	0,31	0,45	0,28	0,35
10S10W	0,25	0,13	0,29	0,19	0,39
14S32W	0,15	0,14	0,14	0,31	0,09

Los valores de los dipolos resultantes se muestran en la tabla 6.2. Se puede observar que el dipolo de Servain se mantuvo con valores altos de anomalía negativa durante todo el periodo evaluado. El dipolo PIRATA $15^{\circ}\text{N} / 10^{\circ}\text{S}$, así como el formado por las boyas $12^{\circ}\text{N} - 23^{\circ}\text{W}$ y $6^{\circ}\text{S} - 10^{\circ}\text{W}$ también mostraron valores negativos durante los meses analizados, siendo mucho más extremos los valores de las anomalías en este segundo dipolo. Por su parte, el dipolo $8^{\circ}\text{N} / 14^{\circ}\text{S}$ también presenta valores negativos de anomalías, pero éstos empiezan en el mes de febrero y son menos extremos que en los dipolos mencionados. Por último, el dipolo $12^{\circ}\text{N} / 14^{\circ}\text{S}$ se muestra menos consistente, con valores menos negativos que el resto de dipolos y, además, con un valor positivo en marzo. En la figura 6.16 se muestran las series temporales de los dipolos PIRATA estudiados. Como indicaban los valores numéricos, se observa que el dipolo formado por las boyas $12^{\circ}\text{N} - 23^{\circ}\text{W}$ y $6^{\circ}\text{S} - 10^{\circ}\text{W}$ (figura 6.16.d) presenta los mayores valores de anomalía durante el periodo analizado del año 2009.

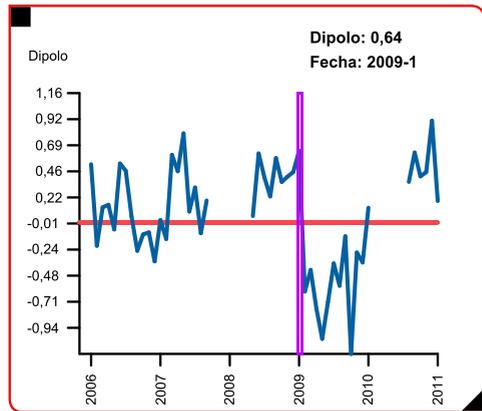
Tabla 6.2

Valores de enero a mayo de 2009 de los dipolos estudiados (°C).

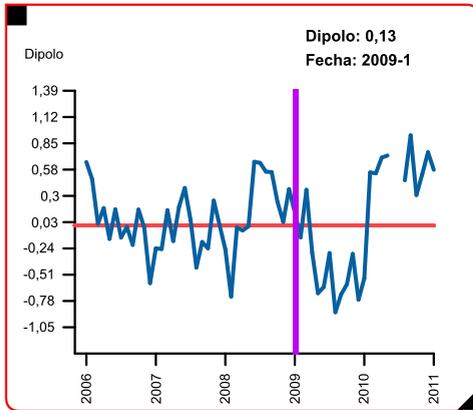
Dipolo	Enero	Febrero	Marzo	Abril	Mayo
Servain	-0,96	-1,30	-1,51	-1,41	-1,83
15N38W y 10S10W	-0,24	-0,55	-0,46	-0,52	-0,91
12N23W y 6S10W	-0,54	-1,50	-1,96	-1,71	-1,26
12N38W y 14S32W	0,13	-0,12	0,37	-0,28	-0,70
8N38W y 14S32W	0,64	-0,62	-0,43	-0,77	-1,04



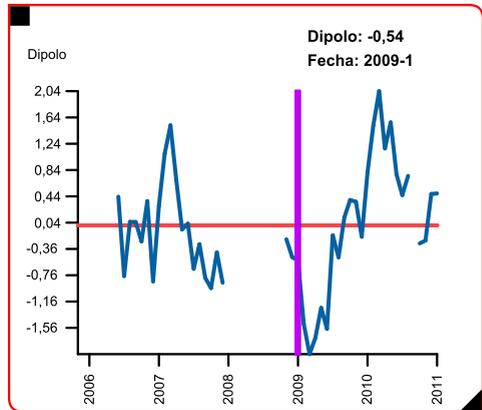
(a) Dipolo 15N38W y 10S10W



(b) Dipolo 8N38W y 14S32W



(c) Dipolo 12N38W y 14S32W



(d) Dipolo 12N23W y 6S10W

Figura 6.16

Series temporales de los dipolos PIRATA obtenidos mediante la aplicación desarrollada para analizar los eventos meteorológicos de los años 2009 y 2010.

6.4.3. Dipolo positivo del año 2010

En la figura 6.17 se puede observar que la precipitación acumulada durante el año 2010 fue muy inferior a la registrada en 2009 (figura 2.5). Aunque en los meses de octubre y diciembre se registraron precipitaciones por encima del promedio, en general, 2010 fue un año más seco de lo habitual. A excepción de la boya $8^{\circ}\text{N} - 14^{\circ}\text{S}$, que no dispone de datos en los meses de interés, para analizar el dipolo de 2010 se han utilizado las mismas boyas que en el anterior caso práctico. En la tabla 6.3 se muestran los valores de anomalías registradas entre enero y mayo de 2010 en las diferentes boyas. A diferencia de lo observado en el periodo de 2009, ahora las boyas situadas al norte del ecuador presentan anomalías de temperatura positivas, siendo la boya $12^{\circ}\text{N} - 23^{\circ}\text{W}$ la que registró las anomalías más altas.

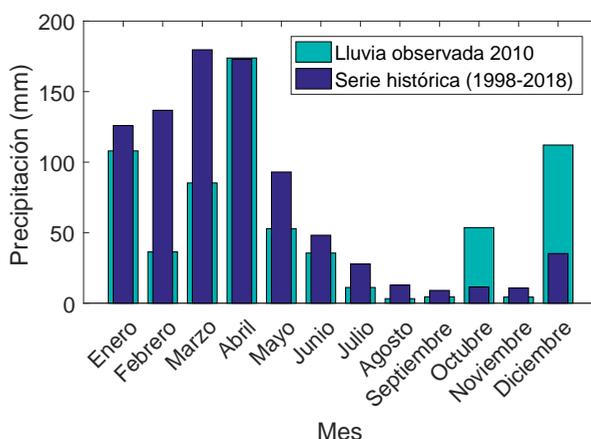


Figura 6.17

Precipitación acumulada durante 2010 con respecto a la media histórica en el estado de Ceará.

Tabla 6.3

Valores de enero a mayo de 2010 de anomalías de temperatura para las boyas seleccionadas.

	Anomalia ($^{\circ}\text{C}$)				
	Enero	Febrero	Marzo	Abril	Mayo
15N38W	0,63	1,03	1,30	1,75	1,75
12N23W	0,96	1,81	2,37	1,62	1,96
12N38W	-0,11	0,75	1,05	0,99	1,35
6S10W	0,15	0,29	0,33	0,45	0,40
10S10W	0,61	0,14	0,69	0,37	0,28
14S32W	0,44	0,20	0,51	0,29	0,63

En cuanto a los dipolos calculados (tabla 6.4), los valores más altos de anomalías se dan en el dipolo PIRATA formado por las boyas $12^{\circ}\text{N} - 23^{\circ}\text{W}$ y $6^{\circ}\text{S} - 10^{\circ}\text{W}$ y, a diferencia de lo observado en el periodo de 2009, estas anomalías son ahora positivas. Además, los valores de este dipolo y los del dipolo $15^{\circ}\text{N} / 10^{\circ}\text{S}$ son superiores a

los mostrados por el dipolo de Servain. Finalmente, utilizando las mismas figuras del anterior caso práctico (figura 6.16), se puede observar el comportamiento de las series temporales de los dipolos durante el año 2010. Los cuatro dipolos analizados son negativos durante la mayor parte de 2009 mientras que, al inicio de 2010, estos dipolos comienzan a presentar valores positivos.

Tabla 6.4

Valores de enero a mayo de 2010 de los dipolos estudiados ($^{\circ}\text{C}$).

	Enero	Febrero	Marzo	Abril	Mayo
Servain	-0,40	0,88	0,64	1,08	0,98
15N38W y 10S10W	0,02	0,89	0,61	1,39	1,47
12N23W y 6S10W	0,80	1,52	2,04	1,17	1,56
12N38W y 14S32W	-0,55	0,55	0,54	0,70	0,72

6.4.4. Predicciones para el periodo 2019 - 2020

Por último, en este apartado se muestran ejemplos de predicciones de la SST obtenidas mediante la aplicación. En primer lugar se seleccionaron las series de la SST para las boyas $15^{\circ}\text{N} - 38^{\circ}\text{W}$ y $10^{\circ}\text{S} - 10^{\circ}\text{W}$. En las figuras 6.18 y 6.19 se puede observar la SST mensual (azul), la climatología (rojo) y las predicciones (verde) de estas boyas. En el momento de la exploración, el dato más actual disponible en la boya $15^{\circ}\text{N} - 38^{\circ}\text{W}$ se corresponde con el de agosto de 2019, por lo que se obtienen las predicciones desde septiembre de 2019 hasta agosto de 2020 (12 meses). En agosto y los meses previos, la SST observada presenta valores muy cercanos a los de la climatología. Por este motivo, las predicciones continúan esta tendencia y muestran temperaturas futuras muy próximas a las que presenta la climatología. En el caso de la boya $10^{\circ}\text{S} - 10^{\circ}\text{W}$ (figura 6.19), el último dato disponible es de julio de 2019 y las predicciones también se muestran sin anomalías, con temperaturas muy cercanas a las de la climatología. Consecuentemente, el dipolo resultante (figura 6.20) presenta valores futuros próximos a 0°C .

Debido a que el último dato disponible en la boya $6^{\circ}\text{S} - 10^{\circ}\text{W}$ es de abril del año 2018, no se han podido obtener las predicciones del dipolo estudiado anteriormente, que estaba formado por esta boya y la situada en $12^{\circ}\text{N} - 23^{\circ}\text{W}$. Por otro lado, en las figuras 6.21, 6.22 y 6.23 se muestran las predicciones para las boyas $12^{\circ}\text{N} - 38^{\circ}\text{W}$ y $14^{\circ}\text{S} - 32^{\circ}\text{W}$, así como su dipolo correspondiente. Para estas boyas, ahora las predicciones muestran anomalías de temperatura con respecto al promedio de la serie histórica. En la boya situada al norte, la predicción en los siguientes tres meses muestra anomalías negativas, mientras que en la boya situada al sur muestra anomalías positivas. Debido a este hecho, en la figura 6.23 se puede observar como las predicciones indican que el dipolo permanecerá en valores negativos hasta mediados del 2020. Los valores numéricos de estas predicciones se muestran en la tabla 6.5.

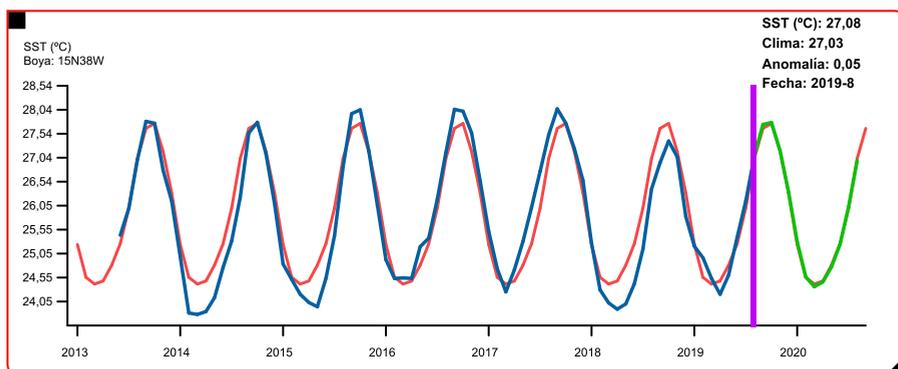


Figura 6.18
Serie temporal de la SST, climatología y predicciones para la boya 15°N - 38°W.

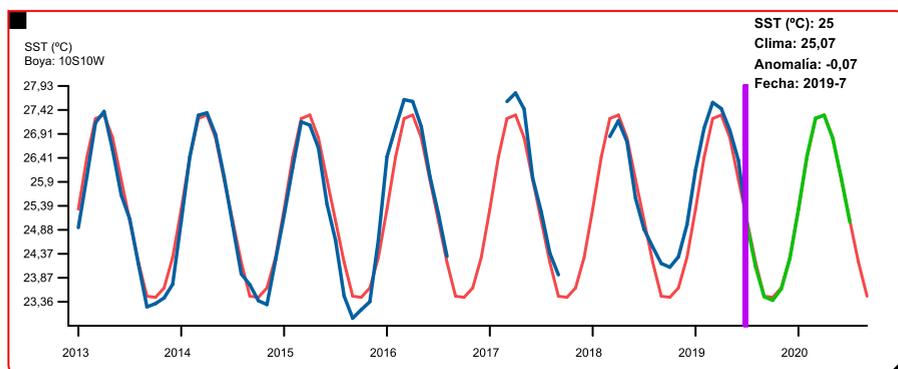


Figura 6.19
Serie temporal de la SST, climatología y predicciones para la boya 10°S - 10°W.

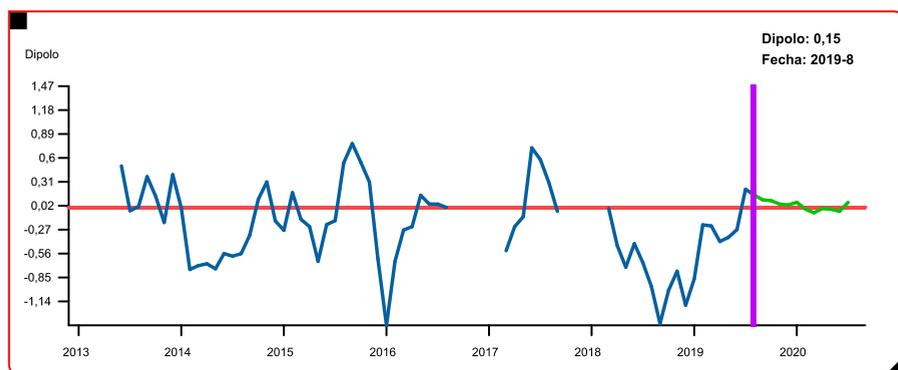


Figura 6.20
Dipolo formado por las boyas 15°N - 38°W y 10°S - 10°W.

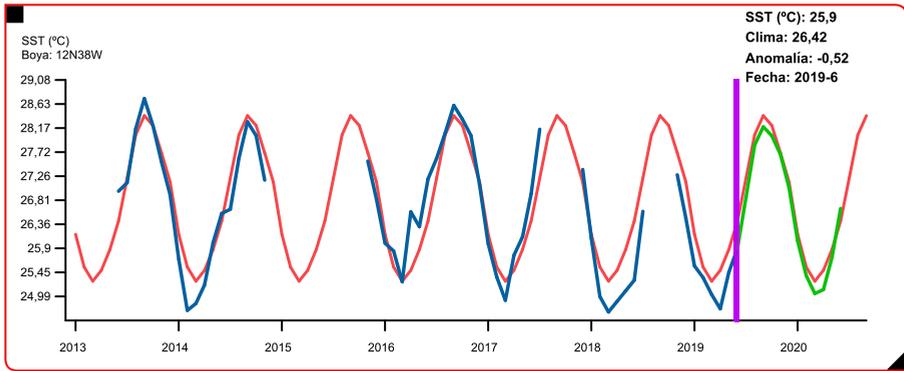


Figura 6.21
Serie temporal de la SST, climatología y predicciones para la boya 12°N - 38°W.

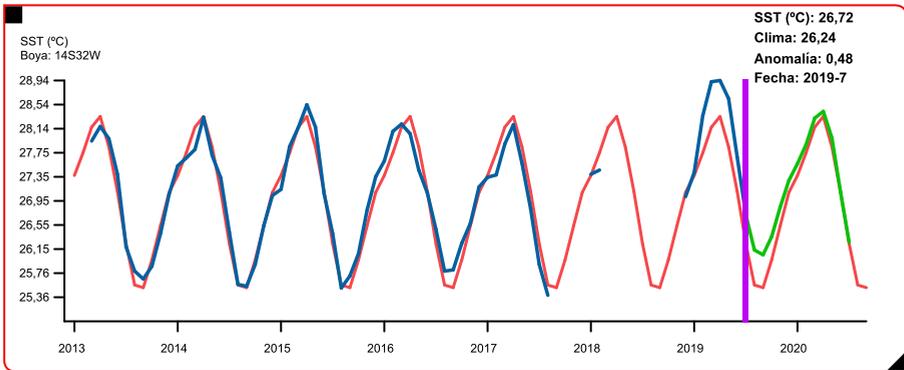


Figura 6.22
Serie temporal de la SST, climatología y predicciones para la boya 14°S - 32°W.

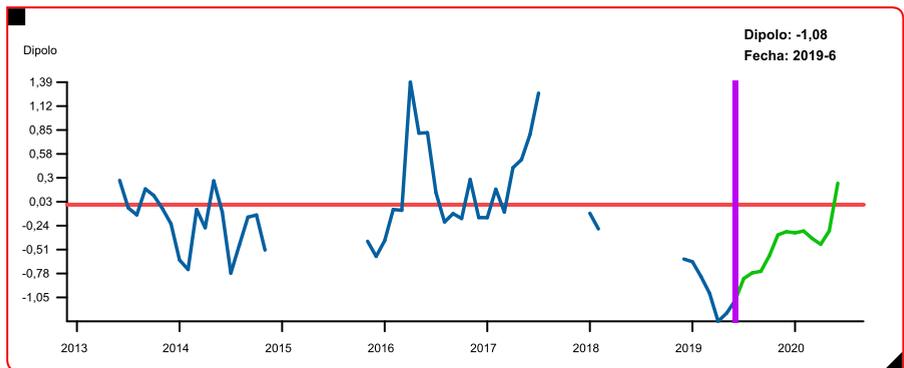


Figura 6.23
Dipolo formado por las boyas 12°N - 38°W y 14°S - 32°W.

Tabla 6.5

Valores de las predicciones (en negrita) de la SST ($^{\circ}\text{C}$), climatología, anomalías y dipolo de las boyas 12 $^{\circ}\text{N}$ - 38 $^{\circ}\text{W}$ y 14 $^{\circ}\text{S}$ - 32 $^{\circ}\text{W}$.

Año	Mes	12N38W	Clima.	Anomalía	14S32W	Clima.	Anomalía	Dipolo
2019	Mayo	25,45	25,88	-0,43	28,64	27,84	0,80	-1,23
	Junio	25,90	26,42	-0,52	27,67	27,10	0,57	-1,08
	Julio	26,87	27,23	-0,36	26,72	26,24	0,48	-0,84
	Agosto	27,86	28,05	-0,19	26,02	25,56	0,46	-0,65
	Septiembre	28,19	28,41	-0,22	25,98	25,52	0,46	-0,68
	Octubre	28,04	28,22	-0,18	26,33	25,98	0,35	-0,53
	Noviembre	27,69	27,70	-0,01	26,82	26,54	0,28	-0,30
	Diciembre	27,07	27,16	-0,08	27,31	27,09	0,22	-0,31
2020	Enero	26,07	26,17	-0,09	27,53	27,37	0,16	-0,26
	Febrero	25,41	25,55	-0,14	27,80	27,75	0,06	-0,19
	Marzo	25,04	25,28	-0,24	28,21	28,17	0,04	-0,29
	Abril	25,11	25,49	-0,37	28,32	28,35	-0,02	-0,35
	Mayo	25,72	25,88	-0,16	27,91	27,84	0,07	-0,23
	Junio	26,63	26,42	0,21	27,02	27,10	-0,09	0,30

6.5. Conclusiones

En este capítulo de la Tesis Doctoral se ha descrito una herramienta diseñada con la finalidad de descargar, visualizar y analizar los datos del proyecto PIRATA. Utilizando el entorno de programación *Processing* se ha implementado una interfaz gráfica de usuario atractiva visualmente y sencilla de utilizar. En base a las funcionalidades implementadas en la aplicación y a los casos de estudio analizados, las principales conclusiones de este capítulo son:

- Seleccionando la variable y boya PIRATA de interés, la aplicación descarga y muestra de forma gráfica la serie temporal solicitada.
- Las gráficas generadas son interactivas, permitiendo ampliar o reducir el rango de datos y facilitando la obtención de información de cada punto de la serie.
- En el caso de la SST mensual, además de la serie de datos, la aplicación muestra las predicciones a 12 meses y la climatología.
- La herramienta permite analizar el modo dipolo seleccionando dos boyas de interés.
- Todo el proceso de descarga, cálculos internos y visualización se realiza en unos pocos segundos, lo que permite una rápida exploración de los datos.
- Los datos y gráficas generadas son fácilmente exportables para su posterior análisis.

Capítulo 7

Conclusiones y proyección futura

7.1. Resumen y conclusiones generales

El trabajo desarrollado en esta Tesis Doctoral se justifica por la importancia de avanzar en la comprensión de los diversos procesos de interacción océano-atmósfera que tienen lugar en el océano Atlántico tropical. Los dos modos principales de variabilidad del ciclo estacional que presenta este océano, son los denominados modo ecuatorial y modo meridional o dipolo. Ambos modos de variabilidad presentan una fuerte influencia en las condiciones climáticas de los países tropicales de América y África, especialmente en la distribución e intensidad de las precipitaciones. El alto impacto social y económico que generan estas fluctuaciones del clima, justifica el gran interés de la comunidad científica en estudiar los diferentes procesos océano-atmosféricos del Atlántico. Con la finalidad de progresar en el conocimiento de estos fenómenos y mejorar los modelos de predicción del sistema climático del océano Atlántico, se implementó la red de boyas fijas del proyecto PIRATA. Estas boyas autónomas adquieren y transmiten vía satélite diferentes variables océano-atmosféricas, generando un gran volumen de datos que ha de ser procesado y analizado.

Esta Tesis Doctoral aporta conocimiento sobre dichos procesos mediante el análisis de los datos PIRATA, empleando para ello diferentes técnicas estadísticas y algoritmos de aprendizaje automático. Para llevar a cabo este cometido se han realizado diferentes análisis y experimentos. Se ha expuesto la relación que existe entre la posición de la ITCZ, la SST y el dipolo del Atlántico con el régimen de precipitaciones de la región Nordeste de Brasil. La posición de la ITCZ influye en gran medida en las temperaturas observadas sobre el océano Atlántico tropical, por lo que el análisis del dipolo proporciona información importante para la previsión del clima de la región.

No obstante, identificar las regiones más representativas del dipolo sigue siendo tema de estudio. Debido a este hecho, se ha presentado un análisis de correlación con el objetivo de obtener el dipolo o dipolos formado por las boyas con mayor potencial en el pronóstico de las precipitaciones en el NEB. Los resultados de estos experimentos mostraron que los dipolos $12^{\circ}\text{N} / 14^{\circ}\text{S}$ y $8^{\circ}\text{N} / 14^{\circ}\text{S}$, presentaban correlaciones más altas con el régimen de lluvias del NEB que los dipolos monitorizados habitualmente (dipolo “clásico” y dipolo de Servain).

Con el objetivo de descubrir patrones y relaciones entre las diferentes variables meteo-oceanográficas registradas por las boyas PIRATA, en esta Tesis Doctoral se ha planteado el uso de mapas auto-organizados. Los resultados de los experimentos mostraron que el SOM es una herramienta útil para estudiar los procesos que intervienen en la variabilidad de la SST, analizar los límites entre la capa de mezcla y la termoclina, establecer relaciones entre densidad, salinidad y temperaturas en superficie y subsuperficie, así como identificar patrones entre el viento, humedad relativa y temperatura del aire.

Debido a la gran importancia que presenta el comportamiento de la SST en la comprensión y predicción del clima a nivel regional y global, se ha propuesto una comparativa exhaustiva entre distintas estrategias y métodos de aprendizaje automático con la finalidad de obtener predicciones diarias y mensuales de la SST. Los resultados mostraron que los menores errores de predicción se obtienen utilizando únicamente la fecha codificada y la SST actual como entradas a los modelos. Además, el modelo ELM logró los mejores resultados en los datos diarios, mientras que el modelo LS-SVM obtuvo el mejor rendimiento en los datos mensuales.

Finalmente, en esta Tesis Doctoral se ha propuesto una herramienta diseñada para la exploración de los datos adquiridos por las boyas PIRATA. Haciendo uso de las librerías de *Processing*, se ha diseñado una aplicación que presenta una interfaz gráfica de usuario sencilla y de funcionamiento intuitivo. Las diferentes funcionalidades implementadas y los casos de estudio expuestos ponen de manifiesto que la aplicación permite la descarga y visualización interactiva de los datos, así como la exportación de tablas y gráficas generadas. Además de las funcionalidades básicas, la aplicación desarrollada presenta dos características novedosas: permite observar el modo dipolo y obtener predicciones mensuales de la SST utilizando un modelo de aprendizaje automático.

7.2. Conclusiones específicas

Al término de cada uno de los capítulos principales se han expuesto las conclusiones obtenidas. No obstante, en este apartado se recogen las conclusiones específicas de cada capítulo a modo de resumen.

7.2.1. Estudio de la relación entre el dipolo PIRATA del Atlántico y las anomalías de lluvias en el Nordeste de Brasil

Las conclusiones obtenidas en el capítulo 3, hacen referencia a los experimentos realizados con la finalidad de obtener el dipolo o dipolos más representativos para el pronóstico de las precipitaciones en el NEB:

- En este capítulo se han calculado los dipolos mensuales utilizando todas las combinaciones posibles entre las boyas PIRATA fondeadas en el Atlántico Norte (línea meridional 38°W) y en el Atlántico Sur (boya 10°S - 10°W y boyas de la PIRATA-SWE).
- Se han obtenido las correlaciones entre los diferentes dipolos y las anomalías de lluvias registradas en las diferentes regiones del NEB mediante los coeficientes *Pearson*, *Spearman* y *Kendall*.
- Los análisis han mostrado que los dipolos 12°N / 14°S y 8°N / 14 °S son los mejores candidatos para representar el dipolo del Atlántico, ya que obtuvieron el mayor número de correlaciones significativas con las anomalías registradas en el NEB y en sus diferentes estados.
- El dipolo “clásico” (15°N / 10°S) y el dipolo SST de Servain obtuvieron los peores resultados de entre los dipolos comparados, siendo necesarios futuros análisis para confirmar estos hallazgos.
- Además de estudiar el dipolo “clásico” y el dipolo de Servain, sería necesario añadir la monitorización de los nuevos dipolos 12°N / 14°S y 8°N / 14 °S, debido a que podrían ayudar a mejorar las previsiones de tiempo y clima en el NEB.

7.2.2. Identificación de patrones en los datos PIRATA mediante mapas auto-organizados

En el capítulo 4 se han empleado los SOMs para realizar un análisis visual multivariante de los datos proporcionados por las boyas 0°N – 35°W, 0°N – 23°W, 8°N – 38°W y 8°S – 30°W. La posición de estas boyas permite estudiar los sistemas de interacción océano-atmósfera en la zona ecuatorial, así como en las regiones norte y sur del Atlántico tropical. Los resultados más interesantes obtenidos fueron:

- **Boyas 0°N – 35°W y 0°N – 23°W:**
 - En ambas boyas, los mapas SOM de temperatura en superficie y a 20 m de profundidad mostraron patrones similares, observándose una relación directa.

- En la boya $0^{\circ}\text{N} - 35^{\circ}\text{W}$, la máxima variación en la amplitud de la temperatura se identificó a 100 - 120 m de profundidad, observándose un desfase de seis meses entre dichas series que se mostró como una relación inversa en los mapas SOM. En la posición $0^{\circ}\text{N} - 23^{\circ}\text{W}$, la mayor variación de amplitud se observó a 60 - 80 m, existiendo un desfase de tres meses con la SST que se identificó como una rotación de 90° en los patrones de los mapas.
 - Los análisis mostraron que el espesor de la capa de mezcla y la profundidad de la termoclina es menor hacia el este en la zona del ecuador.
 - Debido a que la termoclina separa las aguas superficiales de las aguas profundas, no se observaron relaciones directas o inversas entre la SST y la temperatura en altas profundidades.
 - En ambas boyas la densidad presenta una relación inversa con la SST. Sin embargo, en la boya $0^{\circ}\text{N} - 23^{\circ}\text{W}$ no se ha podido identificar una relación clara entre la salinidad y la densidad o entre la densidad y la SST, relaciones que sí se observaron en la boya $0^{\circ}\text{N} - 35^{\circ}\text{W}$.
 - En la boya $0^{\circ}\text{N} - 35^{\circ}\text{W}$ se identificaron relaciones claras entre la humedad relativa y la velocidad del viento, así como entre la humedad relativa y la temperatura del aire. Sin embargo, estas relaciones no se pudieron observar en la posición $0^{\circ}\text{N} - 23^{\circ}$. En ambas posiciones, los mapas SOM indicaron una relación directa entre la temperatura del aire y la SST, así como una relación inversa con la velocidad del viento.
 - Las diferencias en las relaciones entre variables observadas sugieren que, en la posición este del ecuador, el comportamiento de la SST y el espesor de la capa de mezcla están afectadas por diferentes interacciones que en la posición oeste. Por ejemplo, además de la mezcla vertical promovida por el viento, la ascensión de vientos ecuatoriales intervienen en los procesos que tienen lugar en la posición este.
- **Boya $8^{\circ}\text{N} - 38^{\circ}\text{W}$ y $8^{\circ}\text{S} - 30^{\circ}\text{W}$:**
- Mientras que en la posición $8^{\circ}\text{N} - 38^{\circ}\text{W}$, la similitud entre los patrones de temperatura en superficie y subsuperficie se observa hasta los 20 m, en la boya $8^{\circ}\text{S} - 30^{\circ}\text{W}$, esta semejanza se extiende hasta los 60 m. Por tanto, comparado con la posición sur, la capa de mezcla presenta generalmente un menor espesor en la posición norte debido, posiblemente, a la influencia de la posición de la ITCZ.
 - Debido al mayor espesor de la capa de mezcla que presenta la posición sur, las series de temperatura a 140 - 180 m de profundidad presentan un mayor rango de variación que en la posición norte.
 - En la posición sur se observaron mayores valores de salinidad que en la boya norte. Además, los patrones de salinidad y densidad presentan semejanza

hasta los 40 m en la posición sur, mientras que en la boya norte esta similitud se da únicamente hasta los 20 m. Sin embargo, entre 1 y 20 m, la relación entre temperatura, salinidad y densidad es más clara en la posición norte. A 120 m, en ambas boyas se identificó una relación directa entre salinidad y temperatura, que es inversa con la densidad.

- En cuanto a las variables velocidad del viento y profundidad de la isoterma a 20 °C, no se pudo identificar una relación clara entre sus mapas SOM en ninguna de las dos boyas.
- Entre los mapas de la SST y temperatura del aire se observó una relación directa, que es más clara en la boya 8°S – 30°W. En ambas boyas, esta relación es inversa a la velocidad del viento.
- Finalmente, los resultados obtenidos indicaron que los mapas SOM resultan útiles en el reconocimiento de patrones y extracción de características en datos meteo-oceanográficos, contribuyendo a mejorar el conocimiento de las interacciones entre el océano y la atmósfera.

7.2.3. Predicción de la SST mediante técnicas de aprendizaje automático

El capítulo 5 se ha centrado en resolver el problema de las predicciones diarias y mensuales de la SST. Comparado con estudios previos, se ha ampliado el análisis en varios aspectos:

- Se han obtenido predicciones diarias y mensuales con modelos independientes para cada una de las 17 boyas activas en el proyecto PIRATA.
- Las predicciones se han realizado en tres horizontes de predicción: a) $t+7$, $t+15$ y $t+30$ días para datos diarios y b) $t+1$, $t+2$ y $t+3$ meses para los datos mensuales.
- Además de emplear los valores de la SST para entrenar los modelos, se han analizado los conjuntos de datos completos que proporcionan las boyas.
- Para reducir la dimensionalidad de los conjuntos de datos y descubrir las variables con mayor impacto en las predicciones de la SST, se han utilizado métodos de extracción y selección de características. En concreto, se ha utilizado PCA, AC y el algoritmo *RReliefF*.
- Se ha comparado el rendimiento de una gran variedad de modelos de aprendizaje automático: RMLR, MLP, ELM, GRNN, RFBT, SVM y LSTM. Además, estos modelos han sido comparados con el modelo trivial, el modelo climatología y el modelo de pronóstico climático NCEP-CFSv2.

El análisis de los resultados obtenidos en los diferentes experimentos llevados a cabo en este capítulo, se resumen en las siguientes conclusiones específicas:

- La fecha codificada permite a los modelos captar la componente estacional de la SST y consigue disminuir el error de las predicciones.
- Los menores índices de error se obtienen utilizando como entradas únicamente la fecha codificada y la SST actual.
- Aunque el algoritmo *RReliefF* indicó que ciertas variables, como la SST a diferentes profundidades, contribuyen en la obtención de las predicciones, no se consiguió mejorar a los modelos entrenados únicamente con la fecha y la SST.
- Utilizando datos diarios con fecha y SST, los modelos ELM, MLP, RFBT, RMLR y GRNN obtuvieron los menores índices MAE promedio, sin diferencias estadísticamente significativas entre ellos. Todos los modelos mencionados y LS-SVM, mejoraron los resultados de los modelos trivial y climatología.
- En cuanto a los modelos con datos mensuales, el algoritmo con menor índice MAE promedio fue LS-SVM, siendo este resultado significativamente mejor que el obtenido por el resto de modelos de predicción.
- Los modelos LSTM diarios con estrategias de predicción MIMO y recursiva a 30 pasos mostraron índices MAE similares a los obtenidos por los modelos ELM. En los modelos mensuales, los resultados del modelo LS-SVM resultaron ser estadísticamente mejores que los obtenidos por los modelos LSTM MIMO y LSTM recursivo a 36 pasos. Sin embargo, no se encontraron diferencias entre LS-SVM y LSTM con estrategia recursiva a tres días.
- Resumiendo, de los modelos evaluados con datos diarios, la ELM logró el menor MAE promedio y con el menor coste computacional. En cuanto a los datos mensuales, la LS-SVM resultó ser el modelo con el menor índice MAE y menor tiempo de entrenamiento.

7.2.4. Visualización y predicción de datos PIRATA con Processing

Las principales conclusiones que se pueden extraer de la aplicación diseñada para la exploración de los datos PIRATA son:

- Mediante el entorno de *Processing* se ha podido implementar una interfaz gráfica de usuario atractiva y sencilla de utilizar.
- Seleccionando una variable, periodo y boya de interés, la herramienta descarga los datos más recientes disponibles en el servidor FTP de la NOAA.

- Las gráficas interactivas permiten la selección del rango de datos a mostrar y la obtención de información sobre cada punto de la serie.
- Para la SST mensual, además de la serie observada, la gráfica muestra la climatología y las predicciones a 12 meses.
- El bloque de predicción incluye un comité de 10 ELMs diseñadas con 15 neuronas en la capa oculta. La configuración *on line* elegida permite entrenar los modelos utilizando todos los datos disponibles, incluyendo los más recientes.
- Seleccionando una segunda boya, se muestra también el dipolo formado por la diferencia entre las anomalías de ambas boyas. Puesto que para la obtención del dipolo se utilizan los datos reales y las predicciones realizadas en ambas boyas, la gráfica muestra además la predicción del dipolo para los siguientes 12 meses.
- Los datos y gráficas generadas pueden ser exportados para posteriores análisis.
- Por último, cabe destacar que todo el proceso de descarga, cálculos y generación de gráficas se realiza en pocos segundos, permitiendo una rápida exploración de las series de datos.

7.3. Proyección futura

Una vez presentadas las conclusiones del trabajo realizado en esta Tesis Doctoral, a continuación se describen algunas de las posibles líneas futuras que podrían seguirse para mejorar las soluciones a los problemas planteados:

- Utilizar los registros diarios y mensuales de la SST obtenidos por satélites. Estos conjuntos de datos cuentan con unos 30 años de registros y presentan una resolución espacial de 4 x 4 km sobre el océano Atlántico.
- Estos datos de satélite, al cubrir áreas más grandes del océano y disponer de un mayor número de registros, podrían ayudar a mejorar o ampliar el análisis de la relación entre el dipolo y las lluvias en la región del NEB. También sería interesante ampliar el estudio a otras regiones costeras del Atlántico tropical como la costa Oeste de África. Además de utilizar métodos lineales de correlación, las relaciones entre el dipolo y las lluvias se podrían obtener mediante métodos no lineales, como estimadores de información mutua o el coeficiente de información máxima (MIC).
- El análisis mediante SOMs también podría ampliarse con los datos de satélite. Además, sería interesante estudiar la relación entre el ciclo anual, estacional e interdecadal de la SST y el resto de variables océano-atmosféricas para analizar las anomalías que presenta la SST. También podrían mejorarse los experimentos realizados utilizando variantes del SOM como GHSOM.

- En cuanto a las predicciones de SST, puesto que se ha comparado una gran variedad de modelos, estudios posteriores podrían enfocarse en la etapa de extracción y selección de características evaluando métodos como *Correlation-based Feature Subset Selection* (CFS), *Sequential Forward Selection* (SFS), *Sparse Logistic Regression* (SLR) o el algoritmo *Stepwise*. Las predicciones también podrían mejorarse utilizando comités de modelos. Además, se podría estudiar si el algoritmo LSTM es capaz de descubrir relaciones más complejas entre la SST y el resto de variables utilizando un mayor número de nodos y capas ocultas.
- En relación a la aplicación desarrollada, ésta podría mejorarse en diferentes aspectos y funcionalidades. Por ejemplo, permitiendo el acceso a los datos del dipolo de Servain, a los registros de satélite o las bases de datos de precipitaciones en el NEB. La herramienta también podría incluir un módulo para analizar la relación entre los datos de lluvia y el dipolo, así como un módulo para la visualización de los datos mediante SOMs. El módulo de predicciones podrían ampliarse permitiendo la predicción de más variables de interés, además de ofrecer la posibilidad de escoger y configurar otros modelos de predicción. Finalmente, sería necesario validar con la comunidad científica la utilidad de la aplicación.

Bibliografía

- Hervé Abdi and Lynne J. Williams. Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(4):433–459, 2010.
- Silvestre Aguilar-Martinez and William W. Hsieh. Forecasts of Tropical Pacific Sea Surface Temperatures by Neural Networks and Support Vector Regression. *International Journal of Oceanography*, 2009:1–13, 2009.
- H. Akaike. A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, December 1974.
- Ethem Alpaydin. *Introduction to Machine Learning*. The MIT Press, 2nd edition, 2010.
- Michael Arbib. *The Handbook of Brain Theory and Neural Networks*. Mit Press, 2003.
- Yukun Bao, Tao Xiong, and Zhongyi Hu. Multi-step-ahead time series prediction using multiple-output support vector regression. *Neurocomputing*, 129(0):482 – 493, 2014.
- S. Ben Taieb, G. Bontempi, A. Sorjamaa, and A. Lendasse. Long-term prediction of time series by combining direct and mimo strategies. In *Neural Networks, 2009. IJCNN 2009. International Joint Conference on*, pages 3054–3061, June 2009.
- Y. Bengio, P. Simard, and P. Frasconi. Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, 5(2):157–166, March 1994.
- Christopher Bishop. *Pattern Recognition and Machine Learning*. Springer-Verlag, New York, 1 edition, 2006.
- Bruce Blackshaw. edtfTPj/Free - Open-source FTP library for Java, 2016. URL <https://enterprisedt.com/products/edtfTPj/>.
- Bruno Blanke and Pascale Delecluse. Variability of the Tropical Atlantic Ocean Simulated by a General Circulation Model with Two Different Mixed-Layer Physics. *Journal of Physical Oceanography*, 23(7):1363–1388, July 1993.
- Bernard Bourlès, Rick Lumpkin, Michael J. McPhaden, Fabrice Hernandez, Paulo Nobre, Edmo Campos, Lisan Yu, Serge Planton, Antonio Busalacchi, Antonio D. Moura, Jacques Servain, and Janice Trotte. The pirata program: History, accomplishments, and future directions. *Bulletin of the American Meteorological Society*, 89(8):1111–1125, 2008.

- Laurence C. Breaker and Jodi K. Brewster. Predicting offshore temperatures in Monterey Bay based on coastal observations using linear forecast models. *Ocean Modelling*, 27(1-2):82–97, January 2009.
- Leo Breiman. Bagging predictors. *Machine Learning*, 24(2):123–140, 1996.
- Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.
- Leo Breiman, Jerome Friedman, Charles J Stone, and R. A. Olshen, editors. *Classification and regression trees*. Wadsworth International Group, 1 edition, 1984.
- Lantz Brett. *Machine Learning with R*. Packt Publishing, 2nd edition, 2015.
- James A. Carton and Zengxi Zhou. Annual cycle of sea surface temperature in the tropical Atlantic Ocean. *Journal of Geophysical Research: Oceans*, 102(C13):27813–27824, December 1997.
- Jorge P. Castello and Luiz C. Krug. *Introdução às Ciências do Mar*. Pelotas: Editora Textos, Brasil, 2015.
- P. Chang, T. Yamagata, P. Schopf, S. K. Behera, J. Carton, W. S. Kessler, G. Meyers, T. Qu, F. Schott, S. Shetye, and S.-P. Xie. Climate fluctuations of tropical coupled systems—the role of ocean dynamics. *Journal of Climate*, 19(20):5122–5174, 2006.
- Guillaume Chevillon. Direct multi-step estimation and forecasting. *Journal of Economic Surveys*, 21(4):746–785, 2007.
- D. R. Cox. Prediction by exponentially weighted moving averages and related methods. *Journal of the Royal Statistical Society. Series B (Methodological)*, 23(2):pp. 414–422, 1961.
- CPTEC/INPE. Centro de previsão de tempo e estudos climáticos, instituto nacional de pesquisas espaciais, 2018. URL <http://ftp.cptec.inpe.br/modelos/io/produtos/MERGE/>.
- G. T. Csanady. *Air-sea interaction: laws and mechanisms*. Cambridge University Press, Cambridge ; New York, 2001.
- Wayne W. Daniel. *Applied nonparametric statistics*. PWS-KENT, 2. ed edition, 1990.
- Sumit Das, Aritra Dey, Akash Pal, and Nabamita Roy. Article: Applications of artificial intelligence in machine learning: Review and prospect. *International Journal of Computer Applications*, 115(9):31–41, April 2015.
- K. De Brabanter, P. Karsmakers, F. Ojeda, C. Alzate, J. De Brabanter, K. Pelckmans, B. De Moor, J. Vandewalle, and J.A.K Suykens. *LS-SVMlab Toolbox User's Guide version 1.8*. Internal Report 10-146, ESAT-SISTA, K.U.Leuven, Leuven, Belgium, 2010.
- Alain Dessier and Jean René Donguy. The sea surface salinity in the tropical Atlantic between 10°S and 30°N—seasonal and interannual variations (1977–1989). *Deep Sea Research Part I: Oceanographic Research Papers*, 41(1):81–100, January 1994.

- Manuel J. A. Eugster, Torsten Hothorn, and Friedrich Leisch. Exploratory and inferential analysis of benchmark experiments, 2008.
- A. G. Ferreira and N. G. S. Mello. Principais sistemas atmosféricos atuantes sobre a região nordeste do brasil e a influência dos oceanos pacífico e atlântico no clima da região. *Revista Brasileira de Climatologia*, 1(1), 2005.
- Ferret. Program for analysis and graphics, noaa’s pacific marine environmental laboratory, 2019. URL <http://ferret.pmel.noaa.gov/Ferret/>.
- Gregory R. Foltz and Michael J. McPhaden. Seasonal mixed layer salinity balance of the tropical North Atlantic Ocean. *Journal of Geophysical Research*, 113(C2), February 2008.
- Gregory R. Foltz, Claudia Schmid, and Rick Lumpkin. Seasonal Cycle of the Mixed Layer Heat Budget in the Northeastern Tropical Atlantic Ocean. *Journal of Climate*, 26(20):8169–8188, October 2013.
- FUNCEME. Fundação cearense de meteorologia e recursos hídricos, 2016. URL <http://www.funceme.br>.
- Felix A. Gers, Jürgen Schmidhuber, and Fred Cummins. Learning to forget: Continual prediction with LSTM. *Neural Computation*, 12(10):2451–2471, 2000.
- Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In Yee Whye Teh and Mike Titterton, editors, *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, volume 9 of *Proceedings of Machine Learning Research*, pages 249–256, Chia Laguna Resort, Sardinia, Italy, 13–15 May 2010. PMLR.
- R.M. Gray. Vector quantization. *ASSP Magazine, IEEE*, 1(2):4–29, April 1984.
- GTMBA. Global tropical moored buoy array project office, noaa’s pacific marine environmental laboratory, 2016. URL <https://www.pmel.noaa.gov/gtmba/>.
- I. Guyon and A. Elisseeff. An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3:1157–1182, 2003.
- I. Guyon and A. Elisseeff. *Feature Extraction Foundations and Applications*. Springer, 2006.
- Joseph Hair, William Black, Barry Babin, and Rolph Anderson. *Multivariate Data Analysis*. Pearson Prentice Hall, 7 edition, 2010.
- S. P. Hayes, L. J. Mangum, J. Picaut, A. Sumi, and K. Takeuchi. Toga-tao: A moored array for real-time measurements in the tropical pacific ocean. *Bulletin of the American Meteorological Society*, 72(3):339–347, 1991.
- Simon Haykin. *Neural Networks and Learning Machines*. Prentice Hall, New York, 3 edition, 2009.

- Joe Hicklin, Cleve Moler, Peter Webb, Ronald F. Boisvert, Bruce Miller, Roldan Pozo, and Karin Remington. JAMA: A Java Matrix Package, 2012. URL <http://https://math.nist.gov/javanumerics/jama/#Related/>.
- Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, 1997.
- P. W. Holland and R. E. Welsch. Robust Regression Using Iteratively Reweighted Least-Squares. *Communications in Statistics: Theory and Methods*, A6:813–827, 1977.
- Torsten Hothorn, Friedrich Leisch, Achim Zeileis, and Kurt Hornik. The design and analysis of benchmark experiments. *Journal of Computational and Graphical Statistics*, 14(3):675–699, 2005.
- Robert W. Houghton. The relationship of sea surface temperature to thermocline depth at annual and interannual time scales in the tropical atlantic ocean. *Journal of Geophysical Research: Oceans*, 96(C8):15173–15185, 1991.
- G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew. Extreme Learning Machine: A New Learning Scheme of Feedforward Networks. *Neurocomputing*, 70:489–501, 2006.
- Rob J Hyndman. *Forecasting: principles and practice*. Otexts, Melbourne, Australia, 2018.
- IRI. International research institute for climate and society. data library: Models nme ncep-cfsv2, 2019. URL <https://iridl.ldeo.columbia.edu/SOURCES/.Models/.NMME/.NCEP-CFSv2/>.
- Nathaniel C. Johnson, Steven B. Feldstein, and Bruno Tremblay. The Continuum of Northern Hemisphere Teleconnection Patterns and a Description of the NAO Shift with the Use of Self-Organizing Maps. *Journal of Climate*, 21(23):6354–6371, December 2008.
- Maurice G. Kendall and Jean Dickinson Gibbons. *Rank correlation methods*. E. Arnold; Oxford University Press, 5th ed edition, 1990.
- Diederik Kingma and Jimmy Ba. Adam: a method for stochastic optimization (2014). *arXiv preprint arXiv:1412.6980*, 15, 2015.
- Kenji Kira and Larry A. Rendell. A Practical Approach to Feature Selection. In *Proceedings of the Ninth International Workshop on Machine Learning, ML '92*, pages 249–256, San Francisco, CA, USA, 1992. Morgan Kaufmann Publishers Inc.
- Ben P. Kirtman, Dughong Min, Johnna M. Infanti, James L. Kinter, Daniel A. Paolino, Qin Zhang, Huug van den Dool, Suranjana Saha, Malaquias Pena Mendez, Emily Becker, Peitao Peng, Patrick Tripp, Jin Huang, David G. DeWitt, Michael K. Tippett, Anthony G. Barnston, Shuhua Li, Anthony Rosati, Siegfried D. Schubert, Michele Rienecker, Max Suarez, Zhao E. Li, Jelena Marshak, Young-Kwon Lim, Joseph Tribbia, Kathleen Pegion, William J. Merryfield, Bertrand Denis, and Eric F. Wood. The north american multimodel ensemble: Phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction. *Bulletin of the American Meteorological Society*, 95(4):585–601, 2014.

- K. Kiviluoto. Topology preservation in self-organizing maps. In *Neural Networks, 1996., IEEE International Conference on*, volume 1, pages 294–299 vol.1, jun 1996.
- Teuvo Kohonen. *Self-Organizing Maps*. Springer Berlin Heidelberg, 3, extended edition edition, 2001.
- Igor Kononenko. Estimating Attributes: Analysis and Extensions of Relief. In *Proceedings of the European Conference on Machine Learning, ECML-94*, pages 171–182, Secaucus, NJ, USA, 1994. Springer-Verlag New York, Inc.
- Max Kuhn and Kjell Johnson. *Applied Predictive Modeling*. Springer New York, 2013.
- Xin Li, Chongyin Li, Jian Ling, and Yanke Tan. The relationship between contiguous el niño and la niña revealed by self-organizing maps. *Journal of Climate*, 28(20): 8118–8134, 2015.
- Isis Didier Lins, Márcio das Chagas Moura, Marcus André Silva, Enrique López Droguett, Dóris Veleda, Moacyr Araújo, and Carlos Magno Jacinto. Sea surface temperature prediction via support vector machines combined with particle swarm optimization. In *Proceedings of the 10th International Probabilistic Safety Assessment & Management Conference, at Seattle*, 2010.
- Isis Didier Lins, Moacyr Araujo, Márcio das Chagas Moura, Marcus André Silva, and Enrique López Droguett. Prediction of sea surface temperature in the tropical Atlantic by support vector machines. *Computational Statistics & Data Analysis*, 61:187–198, May 2013.
- Yonggang Liu, Robert H. Weisberg, and Ruoying He. Sea Surface Temperature Patterns on the West Florida Shelf Using Growing Hierarchical Self-Organizing Maps. *Journal of Atmospheric and Oceanic Technology*, 23(2):325–338, February 2006.
- R. A. Locarnini, A. V. Mishonov, J. I. Antonov, T. P. Boyer, H. E. Garcia, O. K. Baranova, M. M Zweng, C. R. Paver, J. R. Reagan, D. R. Johnson, M. Hamilton, and D. Seidov. Volume 1: Temperature. *World Ocean Atlas 2013*, 1:40, 2013.
- Michael J. McPhaden, Antonio J. Busalacchi, Robert Cheney, Jean René Donguy, Kenneth S. Gage, David Halpern, Ming Ji, Paul Julian, Gary Meyers, Gary T. Mitchum, Pearn P. Niiler, Joel Picaut, Richard W. Reynolds, Neville Smith, and Kensuke Takeuchi. The tropical ocean-global atmosphere observing system: A decade of progress. *Journal of Geophysical Research: Oceans*, 103(C7):14169–14240, 1998.
- J. Merle. Seasonal heat budget in the equatorial atlantic ocean. *Journal of Physical Oceanography*, 10(3):464–469, 1980.
- H. B. Milburn, P. D. McLain, and C. Meinig. Atlas buoy-reengineered for the next decade. In *OCEANS 96 MTS/IEEE Conference Proceedings. The Coastal Ocean - Prospects for the 21st Century*, volume 2, pages 698–702 vol.2, Sep 1996.
- G. Monterey and S. Levitus. Seasonal variability of mixed layer depth for the world ocean. *NOAA Atlas NESDIS 14*, page 96, 1997.

- Antonio D. Moura and Jagadish Shukla. On the dynamics of droughts in northeast Brazil: Observations, theory and numerical experiments with a general circulation model. *Journal of the Atmospheric Sciences*, 38(12):2653–2675, 1981.
- Antonio Divino Moura and Stefan Hastenrath. Climate prediction for Brazil's Nordeste: Performance of empirical and numerical modeling methods. *Journal of Climate*, 17(13):2667–2672, 2004.
- Till Nagel. Unfolding maps, 2014. URL <http://unfoldingmaps.org/>.
- P.P. Niiler and E.B. Kraus. *One Dimensional Models of the Upper Ocean, Modeling and Prediction of the Upper Layers of the Ocean*. E.B. Kraus, Ed., Pergamon Press, 2009.
- NOAA. What is thermocline? national ocean service website, 2018. URL <https://oceanservice.noaa.gov/facts/thermocline.html>.
- NOAA. View data exploration tool, 2019. URL <https://www.nnvl.noaa.gov/view/globaldata.html#SURF>.
- P. Nobre. The Pirata south western extension. In *Proceedings of 8th International Conference on Southern Hemisphere Meteorology and Oceanography, Foz do Iguaçu, Brasil*, pages 24–28, 2006.
- Kalpesh Patil, M. C. Deo, Subimal Ghosh, and M. Ravichandran. Predicting Sea Surface Temperatures in the North Indian Ocean with Nonlinear Autoregressive Neural Networks. *International Journal of Oceanography*, 2013:1–11, 2013.
- Kalpesh Patil, M. C. Deo, and M. Ravichandran. Prediction of sea surface temperature by combining numerical and neural techniques. *Journal of Atmospheric and Oceanic Technology*, 33(8):1715–1726, 2016.
- Joseph Pedlosky. *Geophysical Fluid Dynamics*. Springer-Verlag New York, 1987.
- Casey Reas and Ben Fry. *Getting Started with Processing*. O'Reilly Media, Sebastopol, USA, 2010.
- David B. Reusch, Richard B. Alley, and Bruce C. Hewitson. North Atlantic climate variability from a self-organizing map perspective. *Journal of Geophysical Research*, 112(D2), January 2007.
- G. Reverdin, D. Cayan, and Y. Kushnir. Decadal variability of hydrography in the upper northern North Atlantic in 1948–1990. *Journal of Geophysical Research: Oceans*, 102(C4):8505–8531, 1997.
- Lumpkin Rick, Bob Molinari, and Mike McPhaden. *A Northeast Extension of the PI-RATA Array, White Doc*. NOAA/AOML, Miami, FL, 2006.
- Brian D. Ripley. *Pattern Recognition and Neural Networks*. Cambridge University Press, 1 edition, 1996.
- Igor Robnik-Šikonja, Marko e Kononenko. Theoretical and Empirical Analysis of ReliefF and RReliefF. *Machine Learning*, 53(1-2):23–69, 2003.

- A. P. B. Rocha, E. M. Dantas, I. R. D. Morais, and M. S. Oliveira. *Geografia do Nordeste*. Universidade Federal do Rio Grande do Norte (UFRN), 2. ed edition, 2011.
- C. F. Ropelewski and M. S. Halpert. Global and regional scale precipitation patterns associated with the el niño/southern oscillation. *Monthly Weather Review*, 115(8): 1606–1626, 1987.
- D. Ross. *Introduction to Oceanography*. HarperCollins College Publishers, New York, 1995.
- M. Rouault. Pirata south-eastern extension white paper. In *UFCT Special Report, Cape Town, South Africa*, page 32, 2004.
- Patrick Royston. Approximating the shapiro-wilk w-test for non-normality. *Statistics and Computing*, 2(3):117–119, 1992.
- Allyson Rugg, Gregory R. Foltz, and Renellys C. Perez. Role of Mixed Layer Dynamics in Tropical North Atlantic Interannual Sea Surface Temperature Variability. *Journal of Climate*, 29(22):8083–8101, November 2016.
- S. Saha, S. Nadiga, C. Thiaw, J. Wang, W. Wang, Q. Zhang, H. M. Van den Dool, H.-L. Pan, S. Moorthi, D. Behringer, D. Stokes, M. Peña, S. Lord, G. White, W. Ebisuzaki, P. Peng, and P. Xie. The NCEP climate forecast system. *Journal of Climate*, 19(15): 3483–3517, 2006.
- Suranjana Saha, Shrinivas Moorthi, Xingren Wu, Jiande Wang, Sudhir Nadiga, Patrick Tripp, David Behringer, Yu-Tai Hou, Hui-ya Chuang, Mark Iredell, Michael Ek, Jesse Meng, Rongqian Yang, Malaquías Peña Mendez, Huug van den Dool, Qin Zhang, Wanqiu Wang, Mingyue Chen, and Emily Becker. The NCEP climate forecast system version 2. *Journal of Climate*, 27(6):2185–2208, 2014.
- Rebecca Salles, Patricia Mattos, Ana-Maria Dubois Iorgulescu, Eduardo Bezerra, Leonardo Lima, and Eduardo Ogasawara. Evaluating temporal aggregation for predicting the sea surface temperature of the atlantic ocean. *Ecological Informatics*, 36:94 – 105, 2016.
- Andreas Schlegel. ControlP5, 2015. URL <http://www.sojamo.de/libraries/controlP5/>.
- D. Serre. *Matrices: Theory and Applications*. Springer, 2002.
- Jacques Servain. Simple climatic indices for the tropical atlantic ocean and some applications. *Journal of Geophysical Research: Oceans*, 96(C8):15137–15146, 1991.
- Jacques Servain and David M. Legler. Empirical orthogonal function analyses of tropical atlantic sea surface temperature and wind stress: 1964–1979. *Journal of Geophysical Research: Oceans*, 91(C12):14181–14191, 1986.
- Jacques Servain, M. Seva, S. Lukas, and G. Rougier. Climatic atlas of the tropical Atlantic wind stress and sea surface temperature : 1980-1984. *Ocean-Air Interactions*, 1:109–182, 1987.

- Jacques Servain, Antonio J. Busalacchi, Michael J. McPhaden, Antonio D. Moura, Gilles Reverdin, Marcio Vianna, and Stephen E. Zebiak. A pilot research moored array in the tropical atlantic (pirata). *Bulletin of the American Meteorological Society*, 79(10): 2019–2031, 1998.
- Jacques Servain, Ilana Wainer, Julian P. McCreary, and Alain Dessier. Relationship between the equatorial and meridional modes of climatic variability in the tropical atlantic. *Geophysical Research Letters*, 26(4):485–488, 2 1999.
- Jacques Servain, Ilana Wainer, Hervé Ludos Ayina, and Hervé Roquet. The relationship between the simulated climatic variability modes of the tropical atlantic. *International Journal of Climatology*, 20(9):939–953, 2000.
- Jacques Servain, Gabriel Clauzet, and Ilana C. Wainer. Modes of tropical atlantic climate variability observed by pirata. *Geophysical Research Letters*, 30(5), 2003.
- Daniel Shiffman. *Learning Processing: A Beginner's Guide to Programming Images, Animation, and Interaction*. Morgan Kaufmann Publishers, Elsevier, Burlington, USA, 2008.
- A. C. Silva, B. Bourles, and M. Araujo. Circulation of the thermocline salinity maximum waters off the Northern Brazil as inferred from in situ measurements and numerical results. *Annales Geophysicae*, 27(5):1861–1873, May 2009.
- Everaldo Barreiros de Souza. Um estudo observacional sobre o padrão de dipolo de anomalias de temperatura da superfície do mar no Oceano Atlântico Tropical. *Revista Brasileira de Geofísica*, 17:210 – 211, 11 1999.
- D.F. Specht. A general regression neural network. *IEEE Transactions on Neural Networks*, 2(6):568–576, November 1991.
- Lothar Stramma, Monika Rhein, Peter Brandt, Marcus Dengler, Claus Böning, and Maren Walter. Upper ocean circulation in the western tropical Atlantic in boreal fall 2000. *Deep Sea Research Part I: Oceanographic Research Papers*, 52(2):221–240, February 2005.
- J. A. K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, and J. Vandewalle. *Least Squares Support Vector Machines*. World Scientific, Singapore, 2002.
- Souhaib Ben Taieb, Gianluca Bontempi, Amir F. Atiya, and Antti Sorjamaa. A review and comparison of strategies for multi-step ahead time series forecasting based on the {NN5} forecasting competition. *Expert Systems with Applications*, 39(8):7067 – 7083, 2012.
- Lynne D Talley. *Descriptive physical oceanography an introduction*. Academic Press, Amsterdam; Boston, 2011.
- Benyang Tang, William W. Hsieh, Adam H. Monahan, and Fredolin T. Tangang. Skill Comparisons between Neural Networks and Canonical Correlation Analysis in Predicting the Equatorial Pacific Sea Surface Temperatures. *Journal of Climate*, 13(1): 287–293, January 2000.

- F. T. Tangang, W. W. Hsieh, and B. Tang. Forecasting the equatorial Pacific sea surface temperatures by neural network models. *Climate Dynamics*, 13(2):135–147, February 1997.
- Michael Titterton. Neural networks. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(1):1–8, 2010.
- E. Tzortzi, S. A. Josey, M. Srokosz, and C. Gommenginger. Tropical Atlantic salinity variability: New insights from SMOS: Tropical atlantic SSS variability. *Geophysical Research Letters*, 40(10):2143–2147, May 2013.
- V. Vapnik. *Statistical Learning Theory*. Wiley, 1998.
- Juha Vesanto, Esa Alhoniemi, Johan Himberg, Kimmo Kiviluoto, and Jukka Parviainen. Self-organizing map for data mining in matlab: the som toolbox. *Simulation News Europe*, 1(25):54, March 1999.
- M. Cuiqin W. Hua and Z. Lijuan. A Brief Review of Machine Learning and Its Application. In *Information Engineering and Computer Science, 2009. ICIECS 2009*, pages 1–4, 2009.
- D.E. Waliser and X. Jiang. Tropical meteorology and climate — intertropical convergence zone. In Gerald R. North, John Pyle, and Fuqing Zhang, editors, *Encyclopedia of Atmospheric Sciences (Second Edition)*, pages 121 – 131. Academic Press, Oxford, second edition edition, 2015.
- Thomas J. Weingartner and Robert H. Weisberg. A Description of the Annual Cycle in Sea Surface Temperature and Upper Ocean Heat in the Equatorial Atlantic. *Journal of Physical Oceanography*, 21(1):83–96, January 1991.
- S. Weisberg. *Applied Linear Regression*. John Wiley, New York, 4th edition, 2014.
- Paul Werbos. *Beyond regression: new tools for prediction and analysis in the behavioral sciences*. PhD thesis, Harvard University, 1974.
- I. H. Witten, Eibe Frank, and Mark A. Hall. *Data mining: practical machine learning tools and techniques*. Morgan Kaufmann, Burlington, MA, 3rd ed edition, 2011.
- Aiming Wu, William W. Hsieh, and Benyang Tang. Neural network forecasts of the tropical Pacific sea surface temperatures. *Neural Networks*, 19(2):145–154, March 2006.
- Xiangbai Wu, Xiao-Hai Yan, Young-Heon Jo, and W. Timothy Liu. Estimation of Subsurface Temperature Anomaly in the North Atlantic Using a Self-Organizing Map Neural Network. *Journal of Atmospheric and Oceanic Technology*, 29(11):1675–1688, November 2012.
- Lisan Yu. A global relationship between the ocean water cycle and near-surface salinity. *Journal of Geophysical Research*, 116(C10), October 2011.
- Lisan Yu, Xiangze Jin, and Robert A. Weller. Role of Net Surface Heat Flux in Seasonal Variations of Sea Surface Temperature in the Tropical Atlantic Ocean. *Journal of Climate*, 19(23):6153–6169, December 2006.

M. M. Zweng, J. R. Reagan, J. I. Antonov, R. A. Locarnini, A. V. Mishonov, T. P. Boyer, H. E. Garcia, O. K. Baranova, D. R. Johnson, D. Seidov, and M. M. Biddle. Volume 2: Salinity. *World Ocean Atlas 2013*, page 39, 2013.

Visualización, predicción y análisis de variables
climáticas del océano Atlántico mediante técnicas de
aprendizaje automático

Juan José Carrasco Fernández, octubre de 2019

