GLORiA²
GLObal change Resilience in Aquaculture

Departamento de Ciencias del Mar y Biología Aplicada. Universidad de Alicante
Carretera San Vicente del Raspeig s/n 03690
San Vicente del Raspeig, Alicante
965909840

Análisis de resultados

Entregable 2.2 Análisis estadístico de cada una de las series temporales y datos estáticos asociados a los escapes y factores ambientales y gráficos de las granjas

Ana Juan Licián
Rosa Martínez Álvarez – Castellanos
Juan Carlos Sanz González

Gobierno de España · Vicepresidencia Tercera del Gobierno · Ministerio para la Transición Ecológica y el Reto Demográfico
Fundación Biodiversidad
Programa pleamar
Unión Europea · Fondo Europeo Marítimo y de Pesca (FEMP)
Universitat d'Alacant · Universidad de Alicante

Departamento de Ciencias del Mar y Biología Aplicada
Universidad de Alicante
Carretera San Vicente del Raspeig s/n 03690
San Vicente del Raspeig, Alicante

GLORiA²
GLObal change Resilience in Aquaculture

Índice

Departamento de Ciencias del Mar y Biología Aplicada
Universidad de Alicante
Carretera San Vicente del Raspeig s/n 03690
San Vicente del Raspeig, Alicante

1. Antecedentes

Uno de los principales problemas de la acuicultura *offshore* desarrollada en jaulas o viveros flotantes son los escapes de peces. Esta pérdida de biomasa de los viveros puede derivar de la merma de integridad de sus infraestructuras, fallos en los sistemas de fondeo o roturas de la red de cultivo, con frecuencia bajo el efecto de temporales o cazadores furtivos[1].

La ocurrencia de estos fenómenos medioambientales se da cada vez con mayor frecuencia debido al **cambio climático**, surgiendo la necesidad de disponer de modelos matemáticos predictivos que adelanten información sobre la ciclicidad de los temporales y que los pronostiquen con mayor anticipación de manera que apoyen la toma de decisiones por parte de las empresas y de las administraciones públicas.

En el Mediterráneo, se producen escapes de diferentes especies como la dorada (*Sparus aurata*), lubina (*Dicentrarchus labrax*) o corvina (*Argyrosomus regius*). Sus repercusiones desde el punto de vista económico implican pérdidas por merma directa de biomasa cuya cuantía depende de la magnitud de los escapes[2].

Desde el punto de vista medioambiental, los escapes se convierten en un elemento más del ecosistema marino pudiendo generar competencia con otras especies locales e incluso interacción genética de la fauna salvaje[3].

Sin embargo, estos escapes no solo afectan a la actividad acuícola y al medio ambiente. En el transcurso del proyecto GLORIA[4], se demostró que existe una correlación entre **escapes de peces** y las **capturas de pescadores** en las zonas cercanas a las instalaciones de acuicultura después de temporales[1].

Siguiendo estos resultados, en presente trabajo se va a profundizar en el desarrollo de herramientas predictivas que permitan establecer umbrales de las condiciones ambientales que causan los escapes y así ayudar a prevenir futuros desastres en las instalaciones offshore, las cuales son cada vez más recurrentes debido al cambio climático global.

[1] Informe Final y Modelos GLORIA, 2021. Resultados 3.1 y 3.3 de la Acción 3: Modelización de la frecuencia, magnitud y causas de los escapes.

[2] P. Arechavala-Lopez, K. Toledo-Guedes, D. Izquierdo-Gomez, T. Šegvić-Bubić & P. Sanchez-Jerez (2018) Implications of Sea Bream and Sea Bass Escapes for Sustainable Aquaculture Management: A Review of Interactions, Risks and Consequences, Reviews in Fisheries Science & Aquaculture, 26:2, 214-234, DOI: 10.1080/23308249.2017.1384789

[3] Memoria de sostenibilidad 2021. Acuicultura de España V1.02.07.21.

[4] GLORIA: GLObal change Resilience in Aquaculture (2021). Fundación Biodiversdiad https://fundacion-biodiversidad.es/es/content/gloria-global-change-resilience-aquaculture

2. Objetivos

En base a los antecedentes descritos anteriormente, el principal objetivo del presente estudio es realizar una mejora del modelo predictivo del proyecto GLORIA mediante el análisis de series temporales.

Para ello, se plantean las siguientes tareas:

1. Evaluación inicial de los datos accesibles en bases de datos estructuradas para adaptarla a los análisis que se van a realizar.
2. Análisis descriptivo del conjunto de datos recolectados.
3. Análisis predictivo de series temporales basadas en el histórico de datos para la elaboración de modelos analíticos, capaces de predecir las variables ambientales.

Continuando con el trabajo realizado en el Entregable 2.1, el presento documento aborda:

- Estudio de las posibles causas de los escapes de peces mediante el análisis predictivo mediante métodos tradicionales estadísticos ARIMA y SARIMA.
- Estudio de las posibles causas de los escapes de peces mediante el análisis predictivo mediante algoritmos de aprendizaje automático como LSTM.

El presente informe recoge los resultados obtenidos de la segunda tarea (2), centrada en el análisis estadístico de las series temporales y datos estáticos asociados a los escapes y factores ambientales, así como la obtención e identificación del algoritmo con mayor potencial de predicción para las series temporales mediante el testeo de, al menos, 2 algoritmos de predicción de escapes de peces y variables relacionadas.

En concreto, los modelos de predicción que se describen en este documento, así como los resultados, se centran en las provincias de Burriana y Santa Pola (Comunidad Valenciana) como muestras más significativas del conjunto de datos completo. De este modo, el documento sirve como base a un informe final que aúne el estudio completo.

3. Metodología

La metodología en la que se basa todo el trabajo desarrollado durante este proyecto es CRISP-DM (Cross Industry Standard Process for Data Mining). Una metodología enfocada a la minería de datos y que permite la extracción de conocimiento a partir de los datos disponibles.

Para ello, se han descrito unos objetivos que se mostrarán a continuación y en este primer informe se presentará el primero de ellos.

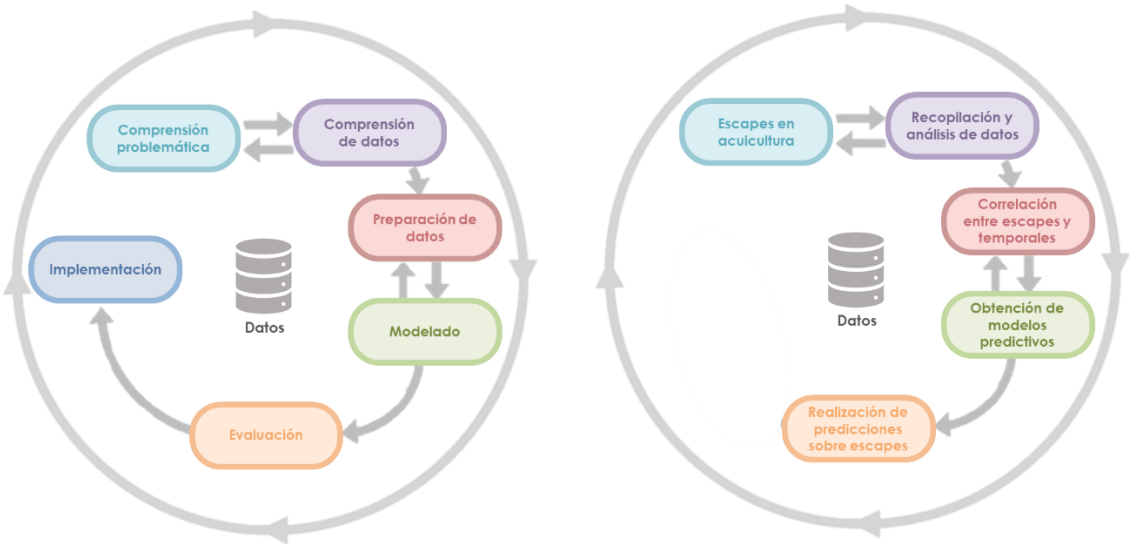

Figura 1. Metodología CRISP-DM aplicada al proyecto GLORIA 2

En este segundo informe nos centramos en el modelado de los datos y su evaluación. Así, tal y como se muestra en la figura de la derecha, estos dos puntos se traducen en:

- Obtención de modelos predictivos utilizando diferentes técnicas de machine learning aplicado a series temporales.
- Evaluación y análisis de los modelos diseñados.
- Realización de predicción sobre las variables ambientales (oleaje, velocidad del viento y mar de fondo).

Los siguientes apartados abordan en detalle cada uno de estos puntos.

4. Técnicas de series temporales

4.1. Modelos univariable

4.1.1. ARIMA

Los modelos ARIMA (Modelo Autorregresivo Integrado de la Media Móvil) son los más generales de las técnicas de predicción de series de tiempo. Se basan en la idea de transformar la serie temporal en estacionaria mediante un proceso de diferenciación. Donde se asume que la serie es estacionaria si sus propiedades estadísticas son constantes en el tiempo. Por lo tanto, la ecuación ARIMA para una serie de tiempo es una ecuación lineal en la que la entrada consiste en pequeños intervalos de la variable dependiente junto con intervalos del error de predicción. Estos modelos se denotan como $ARIMA\ (p,d,q)$, donde "p" representa el orden de la parte autorregresiva, "d" denota el grado de primera diferenciación involucrado y "q" denota el orden de la parte de la media móvil.

4.1.2. SARIMA y SARIMAX

Los modelos SARIMA[5] y SARIMAX[6] parten de los modelos ARIMA, se diferencian de estos ya que incluyen estacionalidad y variables exógenas (SARIMAX), convirtiéndolos así en modelos muy poderosos. Presentan un conjunto adicional de componentes autorregresivas y de media móvil y permiten diferenciar los datos por la frecuencia estacional. El modelo SARIMAX, tiene en cuenta las variables exógenas, es decir, usa datos externos en las previsiones haciendo que el modelo considere datos que otros modelos no son capaces de tratar y responda mucho más rápido. Estos modelos se denotan como $SARIMAX(p,d,q)(P,D,Q,s)$. Comparten una parte con ARIMA, pero incluyen otra parte para especificar el efecto de la estacionalidad, que se conoce como orden estacional. El parámetro "P" hace referencia al orden autorregresivo estacional, "D" representa el orden de integración estacional y "Q" denota el orden de la media móvil estacional. Por último, el cuarto termino "s" representa la duración del ciclo, es decir, el número de periodos necesarios para que la tendencia se repita.

4.2. Modelos multivariable

4.2.1. Redes de memoria a corto plazo (*LSTM)*

Las redes de memoria a corto plazo[7] son un tipo de red neuronal recurrente (RNN), usualmente empleadas para aprender dependencias de largo alcance en datos secuenciales sin conllevar por ello sobreajuste sustancial, ni un gasto computacional adicional. La arquitectura habitual de estas redes es una capa de entrada, una capa oculta conteniendo los bloques de memoria, y la capa de salida. La capa oculta incluye varias células de memorias encadenadas, y cada bloque de memoria incluye una puerta de entrada, otra de salida, y una puerta de olvida en su estructura interna. Estas puertas lógicas permiten que los bloques de memoria lean, escriban y borren (de su memoria) información y les da la capacidad en última instancia de capturar dependencias de largo alcance en la serie de datos.

[5] Modelo Autorregresivo Integrado Estacional de la Media Móvil
[6] Modelo Exógeno Autorregresivo Integrado Estacional de la Media Móvil
[7] XXXX ref 11

5. Implementación modelos predictivos

5.1. Procedimiento

Este documento se centra en las provincias de Santa Pola y Burriana por ser dos provincias con mayor muestras representativas de datos de capturas durante la serie temporal de 2004 a 2021. Además de las variables de captura, el conjunto de datos que hemos empleado se conforma de las variables de oleaje, velocidad de viento y corrientes, procedentes de los puntos SIMAR de Puertos del Estado que se encuentran más cercanos a las instalaciones acuícolas de cada provincia.

Sobre estos dos conjuntos de datos se han implementado un total de 4 modelos predictivos, de los cuales, 3 pertenecen al conjunto de métodos estadísticos clásicos (ARIMA, SARIMA y SARIMAX) y el restante a un modelo de predicción de *machine learning* (LSTM). Asimismo, estos modelos se han diferenciado para conjuntos de datos univariantes y multivariantes.

De este modo, el estudio se ha dividido en la implementación de modelos predictivos univariable y multivariable. Esto quiere decir que para los modelos univariantes únicamente se han realizado predicciones para cada una de las variables oceanográficas independientemente; mientras que para el modelo multivariante se han considerado el conjunto de variables disponibles (oceanográficas y de capturas) para la predicción de cada una de las variables.

Con todo, el procedimiento que se ha seguido (equivalente para cada una de las pruebas realizadas) ha consistido en la división del conjunto de datos en un 85% de los datos de entrenamiento y un 15% de datos para test. Sobre el conjunto de test se han realizado diferentes implementaciones para encontrar los hiperparámetros óptimos para cada algoritmo. Una vez identificados los parámetros que mejor ajustan el modelo, se ha realizado la predicción sobre el conjunto de test, del que se ha obtenido el error cuadrático medio.

Los resultados obtenidos para cada una de las implementaciones se muestran en los siguientes apartados y subapartados.

5.2. Implementación y evaluación

A continuación se presentan los resultados obtenidos mediante la implementación de cada uno de los modelos con los hiperparámetros óptimos proporcionados por las diferentes configuraciones testeadas. Asimismo, se representan las gráficas más significativas para cada modelo.

5.2.1. Modelo univariable

Estación	Modelo	Oleaje		Vel. Viento		Corriente	
		Config.	MAE	Config.	MSE	Config.	MAE
Sta. Pola	ARIMA	(1,1,2)	0.305	(0,1,2)	**1.157**	(1,1,1)	0.291
	SARIMA	(2,0,2)	**0.071**	(1,1,2)	2.676	(2,0,2)	**0.037**
	LSTM[8]	(1,2,1)	0.192	(1,2,1)	1.480	(1,2,1)	0.159
Burriana	ARIMA	(1,1,2)	0.183	(0,1,2)	**0.767**	(1,1,2)	0.201
	SARIMA	(1,1,2)	**0.036**	(2,0,2)	2.203	(2,1,2)	**0.024**
	LSTM	(1,2,1)	0.264	(1,2,1)	1.641	(1,2,1)	0.192

Tabla 1. Resultados obtenidos mediante la implementación de modelos predictivos univariantes mediante ARIMA, SARIMA y LSTM para Sta. Pola y Burriana.

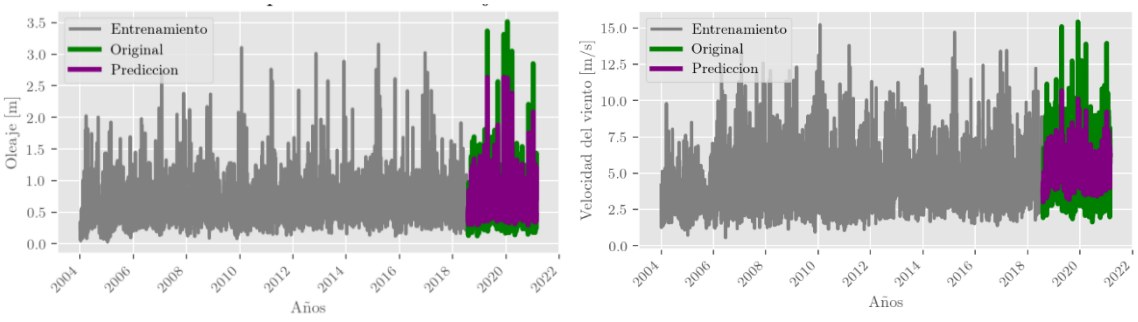

Figura 1. Predicciones obtenidas para las variables de oleaje con SARIMA (izq.) y velocidad del viento con ARIMA (dcha.) correspondientes a Sta. Pola

[8] En LSTM la configuración que se refleja en la tabla se corresponde con (nº capas entrada, nº bloques de memoria, nº capas de salida).

5.2.2. Modelo multivariable

Est.	Modelo	Variables (T/E)[9]				Error
		Oleaje	Corriente	Vel. Viento	Capturas[10]	MAE
Sta. Pola	SARIMAX (1,1,2),(2,1,2,12)	T	E	E	-	0.144
		E	T	E	-	0.323
		E	E	T	-	4.677
		E	E	E	T	282.651
	LSTM (n = 2)	T	E	E	-	**0.123**
		E	T	E	-	0.134
		E	E	T	-	1.056
		E	E	E	T	402.981
Burriana	SARIMAX (1,1,2),(2,1,2,12)	T	E	E	-	**0.068**
		E	T	E	-	0.163
		E	E	T	-	3.192
		E	E	E	T	388.871
	LSTM (n = 2)	T	E	E	-	0.098
		E	T	E	-	0.081
		E	E	T	-	1.007
		E	E	E	T	387.381

Tabla 2. Resultados obtenidos mediante la implementación de modelos predictivos multivariantes mediante SARIMAX y LSTM para Sta. Pola y Burriana

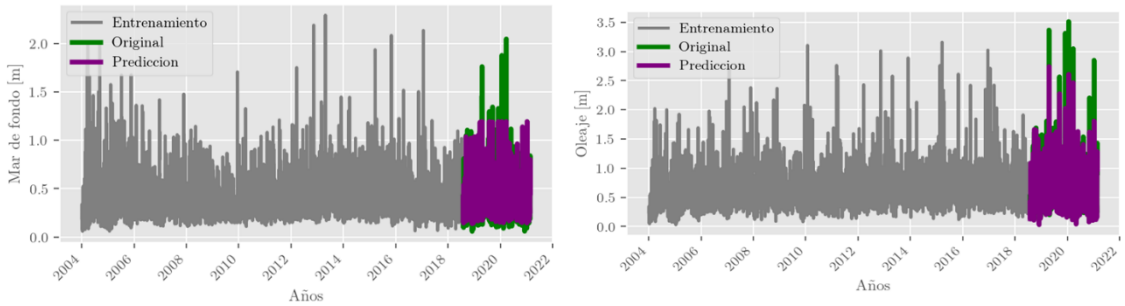

Figura 2. Predicciones obtenidas para las variables de mar de fondo con LSTM (izq.) y oleaje con SARIMAX (dcha.) correspondientes a Sta. Pola.

[9] Definimos variables "*target*" o "*exógenas*" en función de si aplicamos el modelo sobre la variable "*target*" considerando el resto de variables (*exógenas)*.

[10]Cabe destacar que en esta tabla se representa el resultado de predicciones sobre la variable de Capturas para Dorada únicamente como muestra representativa del conjunto de especies, debido a que los resultados que se obtienen para cada una de estas pruebas no resulta concluyente.

6. Conclusiones

En este informe queda reflejado el análisis predictivo de series temporales basadas en el histórico de datos llevado a cabo para la elaboración de modelos analíticos capaces de predecir las variables oceanográficas y su efecto sobre la actividad acuícola. En concreto, los datos que han sido utilizados para la elaboración de este documento se corresponden con los datos procedentes de las provincias de Santa Pola y Burriana.

Las conclusiones extraídas se exponen a continuación:

- Sobre los métodos estadísticos:
  - Se han llevado a cabo análisis predictivos sobre series temporales basados en métodos estadísticos clásicos y métodos de *machine-learning,* cuya implementación se ha diferenciado en su carácter uni/multi-variable.
  - Los métodos univariables que se han implementado se corresponden a los modelos de ARIMA, SARIMA (estadísticos clásicos) y LSTM (machine-learning). Los métodos multivariables se corresponden con los modelos de SARIMAX (estadísticos clásicos) y LSTM (machine-learning).
- Sobre el análisis predictivo realizado:
  - Se han realizado diferentes implementaciones de los modelos identificados para obtener la mejor configuración de cada uno de ellos mediante la obtención de sus hiperparámetros óptimos.
  - Para los modelos univariables, se han considerado únicamente las variables oceanográficas debido a la correlación que existe entre estas y las capturas (como se concluyó en el primer informe).
  - Para los modelos multivariables se han considerado diferentes grupos de variables exógenas y variables objetivo para cada una de las implementaciones.
  - Mediante la implementación de los modelos univariantes, se obtiene que:
    - Para las variables de oleaje y corrientes, tanto para Santa Pola como Burriana, el método que mejores resultados arroja es SARIMA, con un error de 0.071y 0.036, para oleaje y, 0.037 y 0.024, para corrientes.
    - Para la velocidad del viento, el mejor modelo es ARIMA con un error de 1.157 y 0.767, para Santa Pola y Burriana.
  - Mediante la implementación de los modelos multivariables, se obtiene que para Santa Pola y Burriana, la mejor combinaciones de variables exógenas es considerar las corrientes y la velocidad del viento, siendo la variable de oleaje el objetivo a predecir. Para el caso de Sta. Pola, el modelo que menor error arroja es LSTM (0.123), mientras que en Burriana es SARIMAX (0.068).

Con todo, en este informe se ha podido corroborar la capacidad predictiva de estos algoritmos que, si bien no predicen con una precisión excelente los eventos futuros, sí que son capaces de detectar los picos anómalos que se producen cuando ocurre un temporal.