

MODELAMIENTO ESTOCÁSTICO DEL MOVIMIENTO DE LARVAS DE LOCO Y ERIZO ROJO PARA EVALUAR EL EFECTO DE CAMBIOS EN LA SALINIDAD DEL AMBIENTE

**Memoria para optar al título profesional de
Ingeniero Civil Matemático**

FRANCISCO MATÍAS BERNAL FLORES

Profesores guía

Kerlyns Martínez Rodríguez

Departamento de Ingeniería Matemática, Universidad de Concepción

Héctor Olivero Quinteros

Instituto de Ingeniería Matemática, Universidad de Valparaíso

Miembros de la comisión

Rolando Rebolledo Berroeta

Instituto de Ingeniería Matemática, Universidad de Valparaíso

INSTITUTO DE INGENIERÍA MATEMÁTICA
VALPARAÍSO, 29 DE AGOSTO DE 2025

Resumen

La creciente instalación de plantas desalinizadoras en Chile plantea interrogantes sobre el impacto de las descargas hipersalinas en organismos marinos locales. En este trabajo se analiza el efecto de la variación de la salinidad en el movimiento de larvas de *Concholepas concholepas* (loco) y *Loxechinus albus* (erizo rojo), especies de alta relevancia ecológica y económica. A partir de datos obtenidos de registros en video de bioensayos experimentales, se construyó una base de datos de trayectorias bidimensionales que fueron procesadas y caracterizadas utilizando métricas cinemáticas.

En una primera etapa, se aplicaron análisis estadísticos clásicos y funcionales para identificar las variables de movimiento más sensibles a cambios en las condiciones salinas. Los resultados mostraron diferencias significativas entre niveles de salinidad, evidenciando que dichas variaciones alteran los patrones locomotores de manera dependiente de la especie y del contexto experimental.

Posteriormente, se propuso un modelo estocástico sencillo para describir la movilidad larval, basado en procesos de reversión a la media (Ornstein–Uhlenbeck y Cox–Ingersoll–Ross). La calibración de parámetros y las simulaciones obtenidas permitieron reproducir cualitativamente trayectorias con características levemente semejantes a las observadas experimentalmente, estableciendo un primer marco matemático para cuantificar el efecto de la salinidad en la dinámica larval. Este trabajo ofrece una base para futuros estudios orientados a complejizar los modelos, incorporando herramientas específicas a partir de los patrones de comportamiento revelados por los datos, refinando así la comprensión de la respuesta biológica frente a estas perturbaciones ambientales.

Agradecimientos

Quiero expresar mi sincero agradecimiento a todas las personas que estuvieron presentes durante el desarrollo de este trabajo. Sin su apoyo este logro no habría sido posible.

Quiero agradecer, en primer lugar, a los excelentes profesores que guiaron el desarrollo de esta memoria. Gracias a ambos por su paciencia, dedicación y sus estándares de exigencia. Gracias profesor Héctor por darme ánimo, por creer en mí y en que esto se podía sacar adelante. Gracias por enseñarme a no subestimar la dificultad de los problemas ni a sobrestimar nuestra capacidad para resolverlos, sin duda es uno de los aprendizajes más valiosos que me ha dejado, y que sigo integrando día a día. Gracias por su comprensión, buen trato y por hacer de esta etapa tan intensa algo un poco más ameno, así como por toda su ayuda en las diversas aristas del quehacer matemático-ingenieril.

Gracias profesora Kerlyns por su intachable calidad de trabajo, sus siempre acertados comentarios e indudable compromiso con el proyecto. Ver y entender sus ideas en acción ha sido inspirador y sin duda una fuente de motivación durante los altibajos de este proceso. Sin sus valiosas ideas esta memoria hubiese resultado muy diferente.

A ambos, infinitas gracias por su tiempo y por la paciencia que tuvieron conmigo, sobre todo en mis días más procrastinadores...

Gracias al profesor Rolando por aceptar ser parte de la Comisión, y por el tiempo dedicado a la lectura y revisión de este documento. Sin duda es un honor para mí recibir sus comentarios evaluativos sobre este trabajo, así como haber sido su estudiante.

Gracias a los representantes del proyecto IDRO en la Universidad de Valparaíso, en particular al profesor Patricio Winckler de la Escuela de Ingeniería Oceánica. Agradezco la confianza depositada en mí y la oportunidad de brindar un pequeño aporte dentro de un proyecto tan relevante para nuestros ecosistemas.

Agradezco también al profesor Joaquín Fontbona del Departamento de Ingeniería Matemática de la Universidad de Chile, por el financiamiento otorgado para la realización de esta memoria a través del proyecto FONDECYT 1242001.

Al equipo SWAM, Inria y a todas las personas que conocí en Francia, por expandir mis horizontes e inspirarme a ser mejor, como también por recordarme la importancia de valorar mis procesos y honrar mis tiempos. Mi total gratitud a quienes hicieron este viaje posible, especialmente a mi familia y a mis profesores guía por todo el apoyo.

Extiendo mis agradecimientos a todo el personal del Instituto de Ingeniería Matemática por su invaluable calidad y buenas relaciones. A cada profesor y profesora de la Universidad de quien fui alumno, muchas gracias. De cada curso, y de cada uno, aprendí lecciones valiosas y diferentes que han contribuido al profesional en el que me convierto hoy.

Gracias infinitas a mi mamá, Maritza, por su amor incondicional, por apoyarme en todas mis decisiones y acompañarme en cada paso que di en estos cinco años, aun cuando estaba lleno de miedo e incertidumbre. Gracias

por ser tú, por ser tan resiliente y por enseñarme que no hay que ser duro para ser fuerte, ni insensible para ser valiente. Te admiro y te amo con todo mi corazón. Esto también es por y para ti.

Gracias a mi papá, Juan, por ser parte de este proceso a pesar de que muchas veces no lo entendieras. Gracias por estimular mi curiosidad con las conversaciones que solo tú y yo podemos tener, por enseñarme el valor de los momentos genuinos y de la seguridad en uno mismo, y a siempre cuestionar lo importante de la vida.

Gracias a mi hermana menor pero mayor, Dana, por su infinita (infinita) paciencia y por simplemente ser tú. Siempre nos tendremos el uno al otro. Tenerte cerca todos los días me hace más feliz de lo que probablemente te lo digo. Te amo para siempre < 3.

A mi prima Amanda, quien se ha acostumbrado a escuchar mis ocurrencias y por su genuino interés cuando intento explicar cada pensamiento que pasa por mi cabeza. Sin duda es una tarea difícil, pero nuestras conversaciones cargadas de honestidad siempre me traen paz y claridad. Ha sido un gusto ser roomies estos últimos años. Te quiero mucho!

A Joche, por ser la sorpresa más bonita de este año. Gracias por estar ahí, por levantarme cuando he estado a punto de rendirme, por inspirarme a ser mejor y por el calor de tus abrazos que siempre calman a mi corazón.

A mis hermanos Jean, Martín, Lucas y Javi, por ser tan especiales como cada uno es y hacerme sentir parte de este gran clan que, aunque a veces bien esparcido, siempre nos une el cariño, la música y el Just Dance :).

A toda mi familia salamanquina por siempre, siempre estar ahí. Gracias por criarme, por enseñarme sus valores y transmitirme la importancia de la familia. Con ustedes lo tengo todo, y me hace feliz siempre tener un hogar al que volver. A mi familia viñamarina, por ser una red de apoyo desde nuestros primeros años de vuelta aquí. Gracias a todos mis tíos y tías, primos y primas, que han sido parte de este bonito, estresante pero enriquecedor período de mi vida.

A mis amigas de la vida, Javi y Marce, por quererme tal y como soy, y por regalarme tantos momentos. Los años de amistad y el cariño que les tengo es enorme y a prueba de meses sin hablar; lo hemos comprobado y me alegra saber que siempre estarán ahí para mí, así como yo para ustedes.

A todos los amigos y amigas que conocí desde el comienzo y a lo largo de este camino —algunos con quienes tomamos rutas distintas—, y con los que compartí en distintos momentos del recorrido: Javi L., Cate, Cristian, Seba, Nico, Kathryn, Cris, Aline, Sofi. Les deseo éxito en todos sus propósitos. De todos ustedes he aprendido algo.

Y así termina esto. Y así comienza.

¡Muchas gracias!

Y como dice mi profe guía:

¡No está muerto el que pelea!

F.

*A mi familia...
y al Fran de hace cinco años.*

Índice general

Resumen	2
Agradecimientos	6
Índice de figuras	8
Índice de tablas	8
1. Introducción	13
1.1. Objetivo general	15
1.2. Objetivos específicos	15
2. Preliminares	17
2.1. Antecedentes biológicos y ecológicos	17
2.1.1. <i>Concholepas concholepas</i>	17
2.1.2. <i>Loxechinus albus</i>	19
2.2. Proyecto IDRO y LabOcéano: Origen y caracterización de los datos experimentales . . .	21
2.3. Naturaleza y estructura de los datos	23
3. Procesamiento de datos y exploración inicial	25
3.1. Consolidación y preprocesamiento	25
3.2. Análisis de los tiempos de grabación	28
3.2.1. Criterio de optimización para una ventana temporal uniforme	30
3.3. Caracterización inicial del comportamiento locomotor	31
3.4. Análisis descriptivo	34
3.4.1. La importancia de la ubicación	34
3.4.2. Análisis de la condición inicial	36
3.4.3. Métricas entre salinidades	37
3.4.4. Correlaciones entre salinidades	41
4. Evaluación del efecto de la salinidad	44
4.1. Modelos de comparación y variabilidad grupal	44

4.1.1.	Análisis de la varianza	44
4.1.2.	Análisis de la varianza de Welch	45
4.1.3.	Prueba de Kruskal–Wallis	46
4.2.	Verificación de supuestos y selección del modelo	48
4.2.1.	Prueba de Shapiro–Wilk para normalidad	48
4.2.2.	Prueba de Bartlett para homocedasticidad	48
4.2.3.	Prueba de Levene para homocedasticidad	49
4.3.	Análisis post hoc	49
4.3.1.	Prueba HSD de Tukey	50
4.3.2.	Prueba de Games-Howell	50
4.3.3.	Prueba de Dunn y correcciones por múltiples comparaciones	50
4.4.	Medidas de tamaño del efecto	51
4.4.1.	Pruebas paramétricas (ANOVA y ANOVA de Welch)	51
4.4.2.	Pruebas no paramétricas (Kruskal–Wallis)	52
4.5.	Análisis sobre indicadores agregados	52
4.5.1.	Homogeneidad entre estadios larvales de <i>Loxechinus albus</i>	53
4.5.2.	Influencia de otros factores	54
4.6.	Resultados por grupo experimental	57
4.7.	Evaluación intra-nivel de salinidad	59
4.7.1.	<i>C. concholepas</i> : Recuperación	59
4.7.2.	<i>L. albus</i> : Exposición	60
4.8.	Análisis sobre estructuras funcionales	61
4.8.1.	Representación en bases de funciones	62
4.8.2.	Regularización L_2 en P-splines	65
4.9.	ANOVA funcional	66
4.9.1.	Estadístico de prueba	67
4.9.2.	Aplicación al conjunto de datos	67
4.10.	Implementación de FDA	68
4.10.1.	Estandarización temporal de trayectorias	68
4.10.2.	Representación mediante P-splines	69
4.10.3.	Optimización de hiperparámetros	69
4.10.4.	Análisis de sensibilidad	70
4.11.	Resultados funcionales por grupo experimental	70
4.11.1.	CC-R	70
4.11.2.	LA-E	76
4.12.	Variables angulares	87
4.12.1.	CC-R	87
4.12.2.	LA-E	89

5. Un primer modelo estocástico para la movilidad larval	94
5.1. Ecuaciones diferenciales estocásticas	94
5.1.1. Movimiento browniano	94
5.2. Modelos de reversión a la media	96
5.2.1. Modelo de Ornstein–Uhlenbeck	96
5.2.2. Modelo de Cox–Ingersoll–Ross	97
5.3. Un modelo para larvas de <i>C. concholepas</i> en recuperación	97
5.3.1. Posición radial: proceso CIR	98
5.3.2. Velocidad angular: proceso OU	98
5.4. Calibración	98
5.5. Esquema numérico de Euler-Maruyama	99
5.5.1. Aplicación al proceso CIR	100
5.5.2. Aplicación al proceso OU	100
5.6. Métodos de estimación	101
5.6.1. Variación cuadrática para la estimación de σ	101
5.6.2. Estimación de parámetros de reversión a la media	101
5.6.3. Estimación por máxima verosimilitud	103
5.7. Resultados	105
5.7.1. Prueba con datos simulados	105
5.7.2. Optimización de la log-verosimilitud	107
5.7.3. Visualización de parámetros estimados	109
5.8. Simulaciones	112
5.8.1. Simulaciones unidimensionales	113
5.8.2. Simulaciones bidimensionales	116
6. Conclusiones	119
6.1. Análisis de resultados	119
6.2. Trabajo futuro	120
A. Nota sobre las unidades prácticas de salinidad (PSU, <i>Practical Salinity Units</i>)	122

Índice de figuras

2.1. Surgencia	18
2.2. Ciclo de vida de <i>Concholepas concholepas</i>	19
2.3. Ciclo de vida de <i>Loxechinus albus</i>	20
2.4. Aparato experimental donde se realizaron los bioensayos.	22
3.1. Diagramas de caja de la duración de videos analizados.	28
3.2. Histograma de duraciones de videos.	29
3.3. Distribución de los pasos temporales Δt	30
3.4. Función de pérdida total de información.	31
3.5. Comparación media y mediana por diagramas de caja	33
3.6. Número de observaciones por grupo	34
3.7. Mapa de recolección de muestras	35
3.8. Diagramas de caja de r_0 por salinidad y grupo experimental	36
3.9. KDE del grupo CC-E	37
3.10. KDE del grupo CC-R	38
3.11. KDE del grupo LA-E	39
3.12. KDE del grupo LA-R	40
3.13. Correlaciones para CC-E	41
3.14. Correlaciones para CC-R	41
3.15. Correlaciones para LA-E	42
3.16. Correlaciones para LA-R	42
4.1. Flujo de trabajo para procesamiento de datos funcionales.	68
4.2. Gráfico de dispersión de los datos originales y suavizados por P-splines: $r(t)$, CC-R	71
4.3. Curvas medias: $r(t)$, CC-R	72
4.4. Trayectorias de $r(t)$ representadas en distintas bases funcionales, CC-R.	73
4.5. Gráfico de dispersión de datos originales y suavizados por P-splines: $\dot{r}(t)$, CC-R	75
4.6. Curvas medias: $\dot{r}(t)$, CC-R	75
4.7. Trayectorias de $\dot{r}(t)$ representadas en distintas bases funcionales, CC-R.	76
4.8. Gráfico de dispersión de datos originales y suavizados por P-splines: $a(t)$, LA-E	78
4.9. Curvas medias $a(t)$, LA-E	78

4.10. Trayectorias de $a(t)$ representadas en distintas bases funcionales, LA-E	79
4.11. Gráfico de dispersión de datos originales y suavizados por P-spline: $\dot{r}(t)$, LA-E	80
4.12. Curvas medias: $\dot{r}(t)$, LA-E	81
4.13. Trayectorias de $\dot{r}(t)$ representadas en distintas bases funcionales, LA-E.	82
4.14. Gráfico de dispersión de datos originales y suavizados por P-splines: $\alpha(t)$, LA-E	83
4.15. Curvas medias: $\alpha(t)$, LA-E	84
4.16. Trayectorias de $\alpha(t)$ representadas en distintas bases funcionales, LA-E.	85
4.17. Gráfico de dispersión de datos originales y suavizados por P-splines: $\theta_v(t)$, LA-E	86
4.18. Curvas medias: $\theta_v(t)$, LA-E	87
4.19. Gráfico de dispersión de datos originales y suavizados por P-splines: $\omega(t)$, CC-R	88
4.20. Curvas medias: $\omega(t)$, CC-R	89
4.21. Gráfico de dispersión de datos originales y suavizados por P-splines: $r(t)$, LA-E	90
4.22. Curvas medias: $r(t)$, LA-E	91
4.23. Gráfico de dispersión de datos originales y suavizados por P-splines: $\omega(t)$, LA-E	92
4.24. Curvas medias: $\omega(t)$, LA-E	92
5.1. Simulación de un proceso CIR	106
5.2. Simulación de un proceso OU	107
5.3. Optimización de la función de log-verosimilitud para el modelo CIR.	108
5.4. Optimización de la función de log-verosimilitud para el modelo OU.	109
5.5. Dispersión de los parámetros encontrados por salinidad del modelo CIR.	110
5.6. Gráfico de dispersión: α vs r_∞ , CIR	110
5.7. Dispersión de los parámetros encontrados por salinidad.	111
5.8. Gráfico de dispersión: α vs ω_∞ , OU	112
5.9. Ejemplos de trayectorias observadas comparadas con simulaciones bajo el modelo CIR.	113
5.10. Histograma del porcentaje de contención de bandas simuladas, CIR	114
5.11. Diagrama de caja del porcentaje de contención por nivel de salinidad bajo el modelo CIR.	114
5.12. Ejemplos de trayectorias observadas comparadas con simulaciones bajo el modelo OU.	115
5.13. Histograma de porcentaje de contención de trayectorias simuladas, OU	115
5.14. Diagrama de caja del porcentaje de contención, OU	116
5.15. Trayectorias bidimensionales simuladas en coordenadas polares	117

Índice de tablas

2.1. Estructura de categorización experimental.	23
2.2. Información de los experimentos realizados por ubicación.	23
2.3. Resumen de videos analizados por salinidad: <i>Concholepas concholepas</i>	24
2.4. Resumen de videos analizados por salinidad: <i>Loxechinus albus</i>	24
3.1. Grupos experimentales considerados según especie y tratamiento.	32
4.1. Variables diferenciadas entre estadios larvales en <i>Loxechinus albus</i>	53
4.2. Efecto de la presencia de vibraciones externas, CC-E	54
4.3. Efecto de la presencia de vibraciones externas, LA-E	55
4.4. Efecto de la presencia de vibraciones externas, LA-R	55
4.5. Efecto de la presencia de vibraciones externas, CC-E	55
4.6. Efecto del número de individuos, CC-R	56
4.7. Efecto del número de individuos, LA-E	56
4.8. Diferencias entre niveles de salinidad, CC-R	57
4.9. Diferencias entre niveles de salinidad, LA-E	57
4.10. Intra-nivel (normal): CC-R	59
4.11. Intra-nivel (alta): CC-R	60
4.12. Intra-nivel (baja): LA-E	60
4.13. Intra-nivel (alta): LA-E	60
4.14. Resultados fANOVA: $r(t)$, CC-R.	71
4.15. Resultados fANOVA: $\dot{r}(t)$, CC-R.	74
4.16. Resultados fANOVA: $a(t)$, LA-E.	77
4.17. Resultados fANOVA: $\dot{r}(t)$, LA-E.	80
4.18. Resultados fANOVA: $\alpha(t)$, LA-E.	83
4.19. Resultados fANOVA: $\theta_v(t)$, LA-E.	86
4.20. Resultados fANOVA: $\omega(t)$, CC-R.	88
4.21. Resultados fANOVA: $r(t)$, LA-E.	90
4.22. Resultados fANOVA: $\omega(t)$, LA-E.	91
5.1. Comparación de parámetros CIR	106

5.2. Comparación de parámetros OU	107
---	-----

Capítulo 1

Introducción

Las plantas desalinizadoras permiten obtener agua potable a partir de fuentes salinas como el agua de mar o agua salobre, lo cual contribuye a satisfacer las demandas hídricas tanto del consumo humano como de usos industriales, sin incrementar la presión sobre las fuentes naturales de agua dulce en regiones con escasez hídrica [1]. En este sentido, la desalinización representa una alternativa viable frente al déficit de agua en numerosas zonas [2]. Sin embargo, una consecuencia inevitable del funcionamiento de estas plantas es la descarga de salmueras hipersalinas en ambientes costeros o estuarinos cercanos, que puede generar impactos ambientales que varían desde insignificantes hasta altamente perjudiciales [3].

Durante la última década se han impulsado en Chile iniciativas orientadas a enfrentar la sequía, buscando establecer lineamientos y prioridades para el uso del agua de mar y la implementación de plantas de desalinización por osmosis inversa (OI), con el objetivo de mejorar la calidad de vida de las comunidades costeras del centro–norte del país, abasteciendo agua para consumo humano, riego agrícola y diversos procesos industriales, especialmente la minería. En esta zona, la costa se caracteriza por una actividad de surgencia¹ persistente durante todo el año y un campo de salinidad media constante en torno a 34 unidades prácticas de salinidad (PSU, por su sigla en inglés, ver [Apéndice A](#)) [4]. La variabilidad estacional de la salinidad es baja, sin superar una unidad [5, 6], con valores mínimos durante el verano y otoño debido a una menor afloración de aguas subsuperficiales de alta salinidad de origen ecuatorial [7].

Actualmente, Chile cuenta con 38 plantas desalinizadoras de osmosis inversa en operación, con capacidades variables que, en conjunto, alcanzan una producción total de 739 411 m³ por día [4]. Entre ellas, destacan las pequeñas plantas destinadas al abastecimiento rural, capaces de entregar entre 1 y 100 m³ diarios para uso humano en zonas costeras rurales. En este contexto, se hace necesaria una mayor disponibilidad de información sobre los posibles efectos de las descargas de salmuera en los organismos costeros locales. Hasta ahora, los estudios realizados en el país son limitados y han documentado impactos negativos de salinidades elevadas sobre macroalgas y algunas especies como la angiosperma marina *Zostera chilensis* (Blanco Murillo et al., 2023) [8] y el molusco *Mesoderma donacium*, comúnmente conocido como macha (Manríquez et al., 2024) [9]. Aunque se han identificado algunos efectos, aún persiste la necesidad de evaluar cómo las salmueras provenientes de plantas desalinizadoras por OI inciden

¹Ascenso a la superficie de masas de agua profundas, que son frías y ricas en nutrientes, debido al movimiento de aguas superficiales mar adentro.

en atributos ecológicos relevantes en especies modelo de invertebrados marinos locales [4, 10].

En el marco del proyecto “IDRO: Evaluación y modelo de impactos ambientales de plantas desalinizadoras por Osmosis Inversa (OI)”, el equipo de Biología Marina de la Universidad Andrés Bello (UNAB) llevó a cabo bioensayos en los que se expusieron larvas en diferentes etapas de desarrollo de *Concholepas concholepas* (loco) y *Loxechinus albus* (erizo rojo) a distintas concentraciones de salinidad, con el objetivo de emular su respuesta ante una descarga de salmuera generada por este tipo de instalaciones. Las muestras fueron recolectadas en agosto de 2023 en la planta desalinizadora de Minera Candelaria² y en octubre del mismo año en el Comité de Agua Potable Rural (APR) de Quintay³. Los ensayos, categorizados según la procedencia de las muestras como Candelaria, Quintay A y Quintay B (estos últimos diferenciados por la fecha de muestreo), fueron registrados en video y analizados mediante técnicas de procesamiento por el equipo del Laboratorio de Procesos Costeros (LabOcéano) de la Escuela de Ingeniería Oceánica de la Universidad de Valparaíso (UV) [11, 12].

Los resultados obtenidos en la segunda etapa —de un total de tres— de IDRO, “Efectos de la salinidad de la salmuera” [13], constituyen el punto de partida del presente trabajo. Este se desarrolla, además, en colaboración con el equipo SWAM (*Sea, Waves, And ecosystems: Stochastic models for perturbed marine environments*) [14], grupo interdisciplinario franco-chileno liderado por el equipo Calisto del instituto de investigación Inria, con participación de la UV, la Universidad de Concepción y otras instituciones, dedicado al desarrollo de modelos estocásticos y simulaciones de fluidodinámica computacional para el estudio de ambientes marinos perturbados.

Esta memoria de título tiene como objetivo analizar y cuantificar el efecto de la salinidad del ambiente en el movimiento larval de las especies bentónicas *Concholepas concholepas* y *Loxechinus albus*. Ambas son de alta relevancia ecológica, por su influencia en las comunidades marinas, y económica, al ser recursos de valor comercial; además, su dispersión y asentamiento dependen directamente de sus patrones de movimiento larval, por lo que cualquier alteración inducida por cambios en la salinidad podría tener consecuencias significativas en su supervivencia y dinámica poblacional. En este sentido, el presente estudio busca responder a las siguientes preguntas de investigación:

1. ¿Cómo afecta la variación de la concentración de salinidad del entorno en el movimiento de larvas de loco y erizo rojo?
2. ¿Qué variables observables pueden considerarse indicadores de estrés inducido por el ambiente en estos organismos?

Manríquez et al. (2024) [4] indican que para *Concholepas concholepas*:

- Salmueras indisueltas (56 PSU) no matan a las larvas recién eclosionadas.
- Luego de una exposición aguda, se observa una reducción de la movilidad larval sobre los 48 PSU.
- Luego de un período de recuperación, se observa una reducción de la movilidad larval sobre los 48 PSU.

²Ubicación geográfica: 27°03'23"S, 70°50'24"O (coordenadas decimales: -27.05629, -70.84011), región de Atacama.

³Ubicación geográfica: 33°11'37"S, 71°42'00"O (coordenadas decimales: -33.19380, -71.70009), región de Valparaíso.

- Luego de un período de recuperación, se observa un incremento en el metabolismo sobre los 38 PSU.

1.1. Objetivo general

El objetivo general de este trabajo es contribuir a la comprensión del efecto de la variación de la salinidad en la movilidad de larvas de *Concholepas concholepas* y *Loxechinus albus*, a través de modelamiento estadístico y estocástico.

1.2. Objetivos específicos

A partir de las preguntas de investigación se derivan los siguientes objetivos específicos:

- **OE1:** Definir observables que midan el impacto de un cambio de salinidad en la movilidad de larvas de *Concholepas concholepas* y *Loxechinus albus*.
- **OE2:** Proponer un modelo estocástico para la movilidad de larvas y estimar la perturbación inducida por el cambio en la salinidad. La propuesta tiene la forma general

$$\vec{r}_t^* = \vec{r}_0 + \int_0^t b(\vec{r}_s^*, \Delta S) ds + \phi(\vec{r}_s^*, \Delta S) + \sigma B_s \quad (1.1)$$

donde \vec{r}_t^* es la posición en el tiempo t , \vec{r}_0 la posición inicial, $\int_0^t b(\vec{r}_s^*) ds$ es el término de deriva que depende de la posición en el tiempo s , σ es la intensidad del ruido y B_s un proceso de Wiener, que introduce la aleatoriedad. En la [Ecuación 1.1](#) el término $\phi(\vec{r}_s^*, \Delta S)$, que se busca estimar, captura los efectos de la perturbación inducida ΔS .

Esta memoria está organizada en dos partes principales: la primera se centra en la identificación de la variable cinemática que presenta una mayor variación entre salinidades, mientras que en la segunda se propone y calibra un modelo estocástico para la movilidad larval.

El [Capítulo 2](#) presenta un breve marco teórico sobre la ecología de las especies estudiadas, junto con una descripción del origen de los datos experimentales utilizados. En el [Capítulo 3](#) se detalla el procesamiento y se realiza una exploración inicial de estos datos, describiendo las transformaciones realizadas para obtener observables cinemáticos adecuados.

El [Capítulo 4](#) está dedicado a la evaluación del efecto de la salinidad en el movimiento larval, aplicando modelos estadísticos clásicos y enfoques de análisis de datos funcionales para la identificación de diferencias significativas entre grupos experimentales.

Posteriormente, en el [Capítulo 5](#), se introduce un primer modelo estocástico para la movilidad larval, construido a partir de la variable más sensible a la variación de salinidad. Se discute la formulación del modelo, la estimación de sus parámetros y la comparación cualitativa de trayectorias simuladas con aquellas obtenidas experimentalmente.

1.2. Objetivos específicos

Finalmente, en el [Capítulo 6](#) se presentan las conclusiones del trabajo, destacando los principales hallazgos y su aporte a la comprensión del efecto de la salinidad sobre el movimiento larval, junto con las proyecciones para futuros desarrollos metodológicos y aplicaciones.

Capítulo 2

Preliminares

2.1. Antecedentes biológicos y ecológicos

2.1.1. *Concholepas concholepas*

Concholepas concholepas (Bruguière, 1789), conocido comúnmente como loco, es un gastrópodo murícido de gran importancia ecológica y pesquera a lo largo de la costa chilena. Esta especie representa uno de los recursos marinos más relevantes para la pesca artesanal y sustento económico de diversas comunidades costeras [4]. Se encuentra en hábitats intermareales y submareales —zonas del litoral que están parcial o totalmente sumergidas según el ciclo de mareas—, formando agregaciones reproductivas y depositando cápsulas ovígeras sobre sustratos rocosos hasta los 30 m de profundidad [4]. Las larvas recién eclosionadas se distribuyen principalmente en la columna de agua superficial (menor a 10 m), aunque también pueden desarrollarse a mayor profundidad dependiendo de la hidrodinámica local.

Frente a la zona centro–norte de Chile, la distribución, dispersión y abundancia larval de *C. concholepas* en el borde somero cercano a la costa (a menos de 5 km de la orilla) están influenciadas por las corrientes de marea y por la actividad de surgencia costera (Figura 2.1) presente durante todo el año [15].

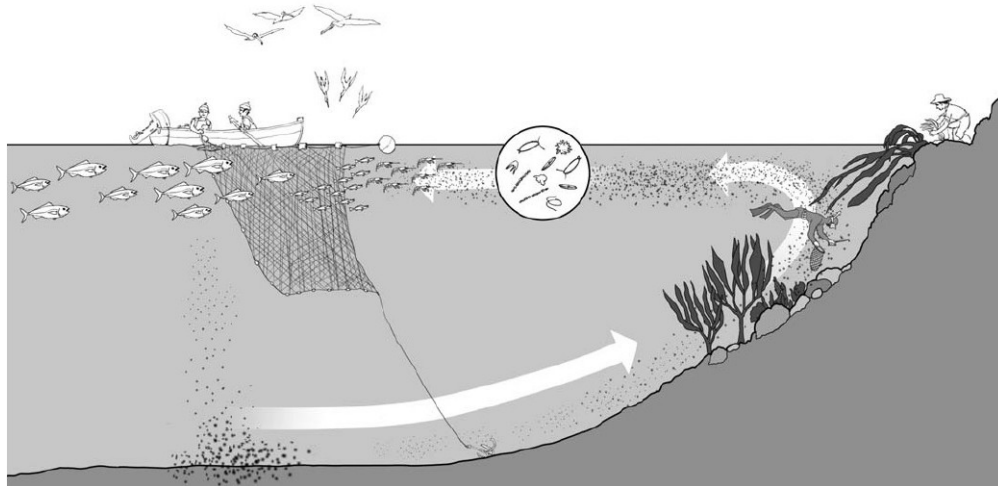


Figura 2.1: El movimiento de aguas en el fenómeno de surgencia se produce desde el fondo hacia mar adentro. Este flujo favorece el ascenso de nutrientes que fertilizan las zonas costeras y oceánicas y permiten el crecimiento de organismos más grandes, que son el sustento de las pesquerías [16].

Tras 1–2 meses de desarrollo intracapsular, las larvas de *C. concholepas* ingresan a una fase pelágica velígera¹ de aproximadamente tres meses, caracterizada por alimentación activa y natación planctónica [4]. Durante esta etapa, las larvas recién eclosionadas se encuentran comúnmente en la superficie o a profundidades menores a 10 m [17], lo que las expone potencialmente a vertidos de salmuera provenientes de plantas desalinizadoras ubicadas cerca del lecho marino. Al alcanzar competencia, estas larvas exhiben comportamiento epineustónico², encontrándose principalmente en la superficie [17], lo que facilita su retención costera durante eventos de surgencia.

Su capacidad de posicionarse verticalmente evita que sean arrastradas mar adentro por la corriente de Ekman, un fenómeno que ha sido asociado a la dispersión larval en otros sistemas marinos similares [18, 19]. Dado que estos patrones natatorios son esenciales para el asentamiento exitoso de la población bentónica, las variaciones de salinidad generadas por descargas de salmuera podrían alterar su desempeño larval y comprometer dicho proceso [4].

En ambientes marinos, muchos invertebrados son osmoconformadores, adaptando sus fluidos internos a la salinidad externa, aunque esto limita su tolerancia a fluctuaciones. Por contraste, los osmorreguladores mantienen una osmolaridad interna constante, lo que les permite habitar zonas con salinidades variables. Las larvas recién eclosionadas de *C. concholepas* en el sur de Chile, por ejemplo, se encuentran en aguas con salinidades bajas (5–30 PSU), y su escasa abundancia bajo 15–20 PSU sugiere que podrían haber desarrollado mecanismos fisiológicos para enfrentar la variabilidad salina [17]. Manríquez et al. indican que “aún no está claro si lo logran mediante osmoconformidad pasiva u osmorregulación activa y energéticamente costosa” [4].

Las larvas recién eclosionadas de *C. concholepas* corresponden al estadio velígero temprano (ver Figura 2.2), caracterizado por escasa capacidad de desplazamiento autónomo. En esta fase, su propulsión

¹Etapa planctónica del ciclo de vida, donde las larvas flotan y se dispersan en la columna de agua.

²Habitan o nadan en la capa superficial del agua.

se realiza mediante una vela ciliada, lo que les permite movimientos lentos y principalmente verticales, quedando expuestas directamente a impactos físicos como las plumas hipersalinas provenientes de plantas desalinizadoras. Manríquez et al. han observado una reducción significativa en su capacidad natatoria cuando son expuestas a salinidades superiores a 42 PSU, con efectos persistentes incluso tras la recuperación en salinidad normal si han sido expuestas a más de 48 PSU, así como un aumento del metabolismo (consumo de oxígeno) a partir de los 38 PSU, señalando un costo energético derivado de mantener la homeostasis osmótica [4]. Por otro lado, Gallardo (1994) [20] expone que salinidades de 20 PSU o menores son dañinas para los embriones de esta especie: se vuelven altamente inactivos y la mortalidad ocurre alrededor de los 30 o 40 días.

Este comportamiento y su limitada capacidad evasiva en estadios tempranos los convierte en bioindicadores sensibles para estudios de riesgo ambiental. Su prolongada fase pelágica, sensibilidad ambiental y valor pesquero hacen de *C. concholepas* una especie modelo en la evaluación de efectos subletales de descargas hipersalinas. Por ahora, se recomienda situar puntos de descarga en zonas altamente turbulentas o con difusores que reduzcan la salinidad a niveles menores a 48 PSU [4].

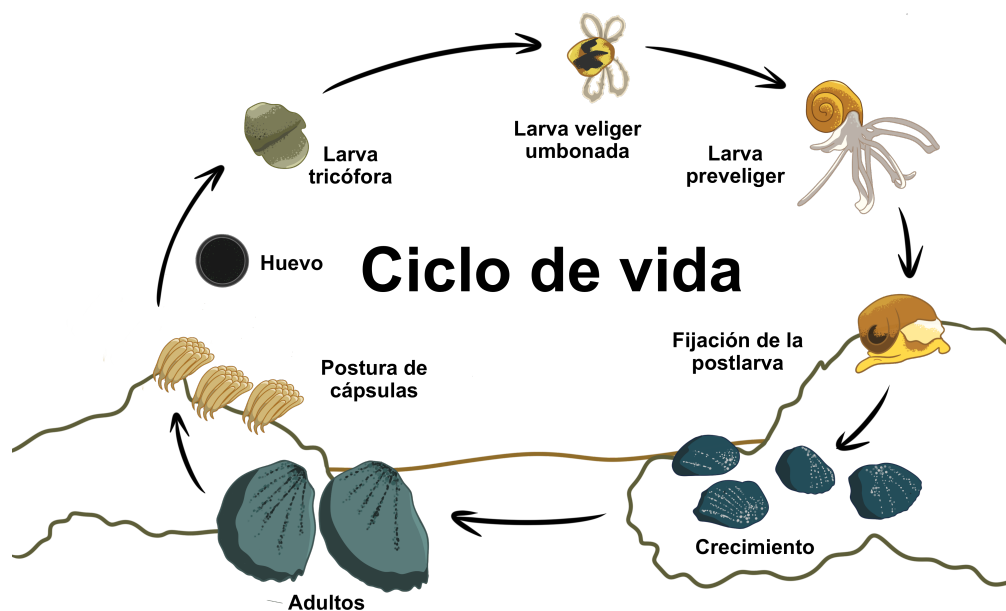


Figura 2.2: Ciclo de vida de *Concholepas concholepas*. Elaboración propia basada en Bustos y Navarrete (2001) [21].

2.1.2. *Loxechinus albus*

Loxechinus albus (Molina, 1782) es un equinoideo herbívoro esencial en ecosistemas marinos del Pacífico suroriental, y una especie fundamental en pesquerías artesanales y en la regulación de praderas algales [22]. Se distribuye desde Ecuador hasta el sur de Chile, incluyendo Juan Fernández y Tierra del

Fuego, habitando zonas intermareales bajas hasta los 40 m de profundidad, preferentemente en fondos rocosos con macroalgas como *Macrocystis pyrifera* (huiro flotador) [23, 24].

Como equinodermo, *L. albus* es un osmoconformador, lo que lo hace vulnerable a fluctuaciones salinas. Cambios fuera del rango óptimo de 28–38 PSU pueden afectar procesos como la fecundación, el desarrollo embrionario y la calcificación [25]. Además, esta especie presenta desoves sincrónicos, liberando larvas pluteus planctónicas que permanecen en la columna de agua entre 25 y 35 días. Esta etapa pelágica es especialmente sensible a las condiciones fisicoquímicas del entorno, incluyendo variaciones en la salinidad [25].

Como en otros equinodermos, las larvas de *L. albus* atraviesan etapas morfológicas diferenciadas que influyen en su comportamiento natatorio (ver Figura 2.3). Las larvas en estadio prisma presentan una movilidad limitada, basada en la acción de bandas ciliadas y sin estructuras esqueléticas, lo que restringe su capacidad de desplazamiento frente a gradientes ambientales abruptos. En etapas más avanzadas, como el equinopluteus de cuatro brazos, el desarrollo de elementos esqueléticos permite un mayor control sobre el posicionamiento vertical en la columna de agua. Aunque el desplazamiento sigue siendo mayormente pasivo, el batido ciliar coordinado les permite ajustar su flotabilidad y ubicación en microestratos más favorables [26]. Esta diferenciación morfológica influye en su susceptibilidad a perturbaciones como descargas hipersalinas, siendo las etapas tempranas especialmente vulnerables.

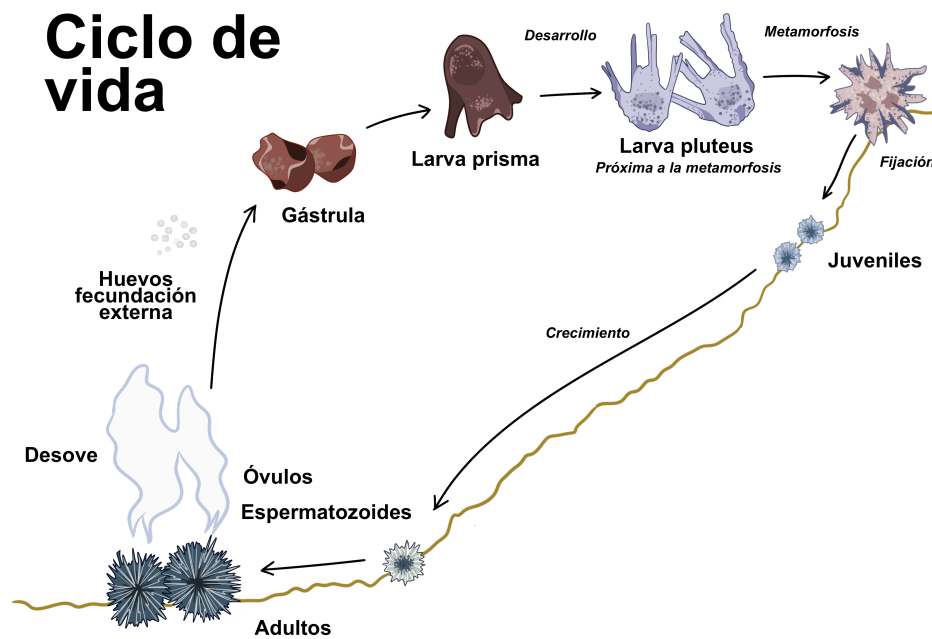


Figura 2.3: Ciclo de vida de *Loxechinus albus*. Elaboración propia basada en Bustos y Olave (2001) [27].

Aunque existen pocos estudios enfocados en *L. albus*, investigaciones sobre equinoideos similares sugieren que salinidades mayores a 42 PSU inhiben el desarrollo larval [26]. Por su valor ecológico,

económico y sensibilidad fisiológica, *L. albus* se considera un organismo bioindicador adecuado para estudios de impacto asociados a desalinización costera.

2.2. Proyecto IDRO y LabOcéano: Origen y caracterización de los datos experimentales

Los datos utilizados en este trabajo provienen de bioensayos desarrollados por el equipo de Biología Marina de la UNAB en el marco del proyecto FONDEF ID23I10336 "IDRO: Evaluación y modelo de impactos ambientales de plantas desalinizadoras por Osmosis Inversa (OI)". El estudio se centró en evaluar los efectos de una exposición aguda a la pluma hipersalina generada por una planta desaladora sobre larvas de *C. concholepas* y *L. albus*, simulando condiciones cercanas al punto de descarga.

Durante los experimentos, las larvas fueron expuestas durante seis horas a un rango de salinidades entre 5 y 60 PSU, abarcando variaciones de aproximadamente ± 30 PSU respecto del promedio ambiental de 35 PSU. Esta ventana de exposición representa una aproximación realista a las condiciones que organismos planctónicos podrían enfrentar en ambientes costeros influenciados por descargas de desalinizadoras. Posteriormente, se implementó un período de recuperación de 12 horas en condiciones normales de salinidad (35 PSU), con el fin de observar posibles efectos reversibles o compensatorios en la movilidad larval. Durante todo el procedimiento las larvas fueron mantenidas individualmente en placas de Petri, como se observa en la [Figura 2.4](#).

El comportamiento locomotor fue registrado mediante videograbaciones de alta resolución acopladas a un microscopio óptico, obteniéndose trayectorias bidimensionales individuales inmediatamente después de las fases de exposición (E) y recuperación (R). El procesamiento de los videos fue realizado por el Laboratorio de Procesos Costeros (LabOcéano) de la Escuela de Ingeniería Oceánica de la UV, utilizando herramientas de seguimiento digital para extraer métricas relevantes del movimiento larval. A partir de estas trayectorias se construyó la base de datos que constituye el insumo principal para los análisis desarrollados en esta memoria.

Si bien este estudio se centra principalmente en evaluar la exposición aguda a una pluma hipersalina generada por plantas desalinizadoras, se consideró pertinente incluir tratamientos con salinidades inferiores al promedio ambiental en los análisis. Esta decisión se basa en la comprensión de que la dinámica del entorno costero es altamente compleja y no siempre predecible. Procesos como la turbulencia, la es-correntía continental, la estratificación vertical o la interacción con masas de agua dulce pueden generar microzonas de dilución extrema, incluso en las cercanías de un punto de descarga hipersalina. Estudios como el de Soto et al. (2025) [10] han demostrado que, bajo ciertas condiciones físicas, pueden formarse acumulaciones localizadas de agua con salinidad reducida, lo que sugiere que escenarios hipotónicos no pueden descartarse a priori. Por ello, se consideró relevante explorar la sensibilidad larval en un rango más amplio de salinidades, reconociendo que la interacción entre el efluente y el ambiente receptor puede dar lugar a condiciones transitorias no intuitivas.



Figura 2.4: Aparato experimental en donde se realizaron los bioensayos. Fuente: Winckler et. al (2024) [12, 13].

Se dispone de los siguientes antecedentes experimentales que constituyen la base para el análisis presentado:

- La temperatura y el pH del agua se mantuvieron constantes durante todo el procedimiento.
- Las larvas no fueron alimentadas durante el experimento, por lo que no existen variaciones energéticas que influyan en su comportamiento locomotor.
- El espesor de la placa utilizada para aislar a cada larva fue de aproximadamente 2 mm, mientras que el tamaño de las larvas se sitúa en aproximadamente 230 μm (0,23 mm).
- El número de Reynolds (Re) del agua de cada placa, considerada como un fluido en estado estacionario, es del orden de $Re \approx 2 \cdot 10^{-2}$.
- No se registró movimiento en la dirección perpendicular a la placa, dada la baja magnitud relativa de dicho componente.
- Las observaciones corresponden a muestras independientes.

En cuanto a los posibles efectos del ayuno en el comportamiento larval, estudios previos en especies similares indican que las larvas recién eclosionadas pueden mantenerse activas durante varios días –e incluso semanas– sin alimentación, especialmente en etapas tempranas como la velígera o el prisma, donde el metabolismo basal es relativamente bajo y la movilidad depende principalmente del batido ciliar [26, 4]. Por lo tanto, se considera que esta condición no afecta significativamente la movilidad larval en este contexto, ni compromete la validez de las observaciones realizadas.

Por otro lado, el medio físico en el que se desarrolla cada observación corresponde a un entorno acuático altamente viscoso a escala microscópica, donde las larvas se desplazan a velocidades del orden de milímetros por segundo. Bajo estas condiciones, el número de Reynolds asociado al movimiento larval es considerablemente bajo, del orden de 10^{-2} a 10^0 , lo que caracteriza un régimen sobreamortiguado donde las fuerzas viscosas predominan sobre las fuerzas inerciales [28, 29]. En este régimen, los efectos de inercia son despreciables, y cualquier impulso externo se disipa casi instantáneamente debido a la

2.3. Naturaleza y estructura de los datos

resistencia del fluido. Este hecho es relevante desde el punto de vista del modelamiento matemático, ya que, en condiciones de bajo Re , el movimiento de organismos puede, bajo ciertas circunstancias, aproximarse como un proceso difusivo, en el que las fluctuaciones aleatorias del medio predominan sobre trayectorias inerciales persistentes. En este contexto, la dinámica larval podría explorarse mediante representaciones estocásticas, como caminatas aleatorias o ecuaciones diferenciales estocásticas de tipo Itô, en las que la posición se modela como una función continua del tiempo influida por perturbaciones aleatorias idealizadas como ruido blanco gaussiano [30].

2.3. Naturaleza y estructura de los datos

Los datos resultantes del procesamiento de videos se registraron siguiendo una jerarquía bien definida:

Ubicación > Especie > Tratamiento > Salinidad > Nombre de archivo

La descripción y el conjunto de valores posibles de cada nivel jerárquico se muestra en la [Tabla 2.1](#).

Categoría	Valores posibles	Descripción
Ubicación	{1, 2, 3}	1: Candelaria, 2: Quintay A, 3: Quintay B
Especie	{CC, LA}	CC: <i>Concholepas concholepas</i> , LA: <i>Loxechinus albus</i>
Tratamiento	{E, R}	E: Exposición, R: Recuperación
Salinidad	{5, 10, ..., 60}	Nivel de salinidad en PSU

Tabla 2.1: Estructura de categorización experimental.

Las condiciones experimentales de los ensayos por ubicación se detallan en la [Tabla 2.2](#). Algunos videos fueron descartados del análisis debido a interferencias externas, como vibraciones visibles en la imagen, ausencia de larvas en el campo visual o porque el individuo abandonaba rápidamente el encuadre, impidiendo un seguimiento adecuado.

Ubicación	Nombre del experimento	Aumento de microscopio	Obtención de muestras	Videos analizados	Videos descartados	Total
1	APR Norte	4x	Agosto 2023	129	15	144
2	APR Quintay	2x	Octubre 2023 (9 al 12)	116	4	120
3	APR Quintay	2x	Octubre 2023 (18 y 19)	44	16	60

Tabla 2.2: Información de los experimentos realizados según la ubicación de las muestras.

Las Tablas [2.3](#) y [2.4](#), reproducidas a partir del informe de Winckler et. al (2024) [12], presentan un resumen del número de videos analizados para *C. concholepas* y *L. albus*, respectivamente.

2.3. Naturaleza y estructura de los datos

Especie	Ubicación	Tratamiento	5	10	15	20	25	30	35	35-N	35-R	40	45	50	55	60	Se analizan	Se descartan	Total
CC Loco	1 Candelaria	E								■	■	■	■	■	■	■	37	5	42
		R								■	■	■	■	■	■	■	39	3	42
	2 Quintay A	E				■	■	■	■								23	1	24
		R	■	■	■	■	■	■	■								40	2	42
Total																	139	11	150
Total (%)																	93 %	7 %	100 %

Tabla 2.3: Distribución de videos analizados por ubicación, tratamiento y salinidad de *Concholepas concholepas*.

Especie	Ubicación	Tratamiento	5	10	15	20	25	30	35	35-N	35-R	40	45	50	55	60	Se analizan	Se descartan	Total
LA Erizo rojo	1 Candelaria	E								■	■	■	■	■			25	5	30
		R								■	■	■	■	■			28	2	30
	2 Quintay A	E				■	■	■	■								23	1	24
		R			■	■	■	■	■								30	0	30
	3 Quintay B	E				■	■	■	■								26	4	30
		R				■	■	■	■								18	12	30
Total																	150	24	174
Total (%)																	86 %	14 %	100 %

Tabla 2.4: Distribución de videos analizados por ubicación, tratamiento y salinidad de *Loxechinus albus*.

Una primera observación a partir de estas tablas es la diferencia en los rangos salinidad entre ambas especies. La literatura indica que los locos presentan un umbral de tolerancia a la salinidad más amplio que los erizos, posiblemente asociado a su mayor plasticidad fisiológica y a su distribución a lo largo de un extenso gradiente latitudinal en la costa chilena [31].

En los ensayos realizados con muestras provenientes de Candelaria se incluyen condiciones de salinidad etiquetadas como 35-N y 35-R. Esta diferenciación responde a la fuente del agua utilizada: mientras que 35-N corresponde a agua de mar natural, recolectada directamente del ambiente, 35-R representa agua reconstituida en laboratorio mediante la mezcla de sales puras en agua destilada. Aunque ambas comparten la misma salinidad nominal, su composición bioquímica puede diferir significativamente. Esta distinción permite evaluar si factores adicionales presentes en el agua natural —como materia orgánica, microorganismos o trazas de nutrientes— influyen en la respuesta de los organismos, a la vez que ofrece un mayor control experimental mediante el uso de agua reconstituida [32].

Capítulo 3

Procesamiento de datos y exploración inicial

3.1. Consolidación y preprocesamiento

Previo al análisis, se consolidó la información de *trackeo* desde un archivo de texto plano .txt por cada individuo hacia una base de datos estructurada en SQL. Cada registro corresponde al seguimiento bidimensional de una larva individual, proporcionando series temporales de posición $(x(t), y(t))$, velocidad $(v_x(t), v_y(t), v(t))$ y aceleración $(a_x(t), a_y(t), a(t))$. La frecuencia promedio de registro fue de aproximadamente 20 Hz, equivalente a un salto temporal de $\Delta t = 0,05$ s entre cuadros. Los factores principales son la ubicación de la muestra, la especie, el tratamiento y el nivel de salinidad.

A partir de los resúmenes experimentales oficiales, se incorporó información complementaria relativa al estadio larval, observaciones de vibración externa, número de ejemplares presentes, duración del video y un identificador único asignado a cada individuo. La presencia de vibraciones extrínsecas podría introducir ruido ambiental, mientras que la presencia de varios organismos en un mismo video podría comprometer el supuesto de independencia, ya que sus movimientos pueden verse mutuamente influidos por proximidad o interacción física.

Adicionalmente, se derivaron variables complementarias consideradas potencialmente sensibles a la salinidad y representativas del comportamiento locomotor larval. Para el suavizado y la estimación de valores faltantes de la coordenada vertical $y(t)$, se empleó interpolación PCHIP¹, seleccionada por su capacidad de preservar la forma y la monotonicidad de los datos originales, evitando oscilaciones artificiales que podrían distorsionar la interpretación del movimiento [33].

Con el fin de mantener la consistencia analítica y abordar la presencia de valores nulos en algunas variables, se recalcularon las magnitudes de velocidad y aceleración utilizando exclusivamente las coordenadas x, y y el tiempo t provenientes del conjunto de datos original. Estas, junto con el resto de variables calculadas, se definen a continuación.

¹*Piecewise Cubic Hermite Interpolating Polynomial*

Posición radial. La posición radial cuantifica la distancia de cada larva respecto a su posición inicial, la cual ha sido trasladada al origen del sistema centrado. Esta métrica resulta útil para analizar patrones de expansión o contracción del movimiento a lo largo del tiempo. Se define como la distancia euclidiana entre la posición actual y el punto de inicio de la trayectoria:

$$r(t) = \sqrt{x(t)^2 + y(t)^2}. \quad (3.1)$$

Rapidez. La magnitud de la velocidad escalar representa la magnitud del desplazamiento por unidad de tiempo, sin considerar la dirección. Se calcula como la norma del vector velocidad en el plano centrado:

$$v(t) = \sqrt{v_x(t)^2 + v_y(t)^2}, \quad (3.2)$$

donde $v_x(t)$ y $v_y(t)$ corresponden a las derivadas temporales de las coordenadas centradas $x(t)$ y $y(t)$, respectivamente.

Magnitud de la aceleración. La aceleración escalar describe la tasa de cambio de la velocidad en el tiempo, y permite identificar eventos de aceleración o desaceleración en la trayectoria. Se define como:

$$a(t) = \sqrt{a_x(t)^2 + a_y(t)^2}, \quad (3.3)$$

donde $a_x(t)$ y $a_y(t)$ son las derivadas temporales de $v_x(t)$ y $v_y(t)$, respectivamente.

Velocidad radial. Esta variable permite analizar la dinámica de expansión o retracción del movimiento, al cuantificar la velocidad con la que varía la distancia respecto al punto de origen:

$$\dot{r}(t) = \frac{dr(t)}{dt}. \quad (3.4)$$

Ángulo de la posición. El ángulo polar describe la orientación de la larva respecto al origen del sistema centrado. Se define como el ángulo entre el vector de posición centrado y el eje x :

$$\theta_r(t) = \arctan 2(y(t), x(t)). \quad (3.5)$$

El uso de $\arctan 2$ permite determinar correctamente el ángulo en todo el plano, considerando el signo de ambos argumentos y evitando ambigüedades en los cuadrantes. A diferencia de \arctan , que entrega resultados en el intervalo $(-\frac{\pi}{2}, \frac{\pi}{2})$ y puede fallar al distinguir entre direcciones opuestas, $\arctan 2$ retorna valores en el rango $(-\pi, \pi]$, lo que permite una representación continua y coherente de la orientación angular. Esta propiedad es especialmente relevante en trayectorias que cruzan ejes o presentan cambios de dirección abruptos.

Velocidad angular. La velocidad angular representa la rapidez con la que cambia la orientación de la larva en el plano. Se calcula como la derivada temporal del ángulo polar:

$$\omega(t) = \frac{d\theta_r(t)}{dt}. \quad (3.6)$$

Aceleración angular. La aceleración angular cuantifica los cambios en la velocidad angular, permitiendo detectar maniobras de giro o ajustes direccionales. Se define como la segunda derivada del ángulo polar respecto al tiempo:

$$\alpha(t) = \frac{d\omega(t)}{dt} = \frac{d^2\theta_r(t)}{dt^2}. \quad (3.7)$$

Ángulo de la velocidad. El ángulo de la velocidad indica la dirección instantánea del movimiento en el plano. Se calcula como el ángulo entre el vector velocidad y el eje x :

$$\theta_v(t) = \arctan 2(v_y(t), v_x(t)) \quad (3.8)$$

Curvatura local. La curvatura describe cuán pronunciadamente cambia la dirección de una trayectoria en cada punto. Para datos discretos, como los obtenidos a partir de imágenes o simulaciones, una opción robusta es el método del triángulo, basado en la fórmula de Herón. A diferencia de otros enfoques, este no requiere derivadas, lo que lo hace especialmente adecuado para trayectorias larvales digitalizadas. Este método fue utilizado para el cálculo de la curvatura por el equipo del LabOcéano [12].

Para cada punto interno de la trayectoria se forma un triángulo junto con los puntos anterior y posterior. Sean $A = \mathbf{r}(t_{i-1})$, $B = \mathbf{r}(t_i)$, $C = \mathbf{r}(t_{i+1})$, y se definen:

- Lados del triángulo: $a = \|C - B\|$, $b = \|C - A\|$, $c = \|B - A\|$.
- Semiperímetro: $s = \frac{a+b+c}{2}$.
- Área (Heron): $A_{\Delta} = \sqrt{s(s-a)(s-b)(s-c)}$.
- Radio del círculo circunscrito: $R = \frac{abc}{4A_{\Delta}}$.

La curvatura local en t_i se define entonces como:

$$k(t_i) = \frac{1}{R}. \quad (3.9)$$

En caso de que los tres puntos sean colineales (i.e., se viola la desigualdad triangular), se asigna $k(t_i) = 0$.

El radio R corresponde al de la circunferencia osculatriz, que representa la mejor aproximación circular a la trayectoria en ese punto. Esta comparte la dirección y curvatura locales, y permite interpretar de forma geométrica cuánto se desvía la trayectoria de una línea recta: a menor R , mayor curvatura.

Este método evita la necesidad de parametrizar o suavizar la curva, y se adapta naturalmente a datos discretos. En la implementación, el uso de `np.gradient` en Python evita saltos numéricos al manejar correctamente las diferencias finitas en primer y segundo orden.

Coordenadas centradas. Centrar las coordenadas respecto a un punto de referencia permite eliminar el efecto del desplazamiento absoluto y enfocarse en la dinámica relativa del movimiento, lo cual es útil para comparar trayectorias entre individuos o condiciones experimentales. Para cada trayectoria individual, las coordenadas se centran en el punto inicial:

$$x_{\text{centrada}}(t) = x(t) - x(t_0), \quad y_{\text{centrada}}(t) = y(t) - y(t_0). \quad (3.10)$$

3.2. Análisis de los tiempos de grabación

Se observó una considerable variabilidad en la duración de los videos registrados, fluctuando entre 3,54 y 33,57 s. Esta diferencia no se atribuye a un factor biológico intrínseco de las larvas, sino a decisiones durante el protocolo experimental de grabación. La [Figura 3.1](#) muestra, además, un número considerable de valores atípicos según el criterio del RIC². Sin embargo estos no parecen afectar de manera sustancial el valor de la media, ya que esta resulta muy similar a la mediana de las duraciones ([Figura 3.2](#)).

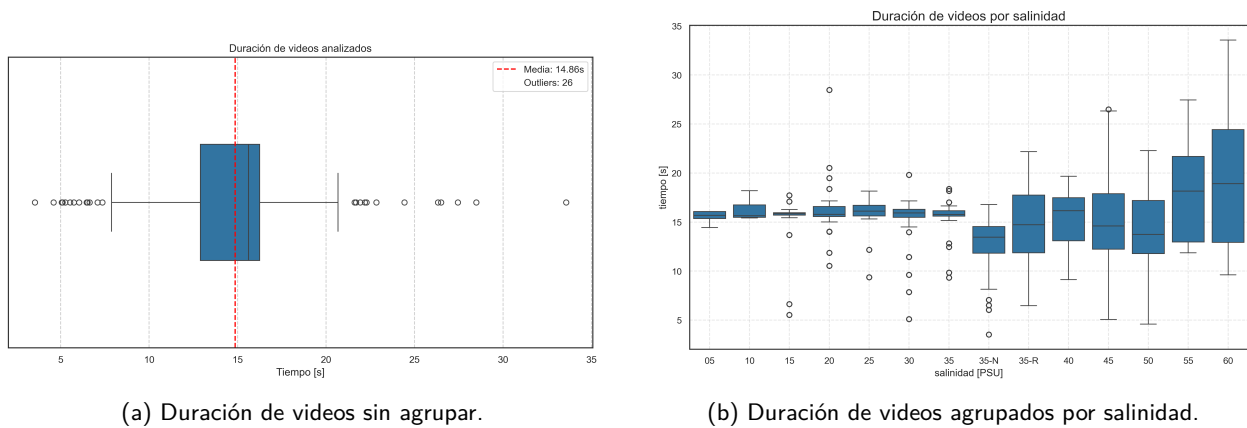


Figura 3.1: Diagramas de caja de la duración de los videos analizados: (a) sin agrupar y (b) agrupados por salinidad.

En la [Figura 3.1](#) no se distingue entre especies, dado que el objetivo inicial es definir un intervalo común y comparable, maximizando la retención de información frente a la alta variabilidad observada. El histograma de duraciones ([Figura 3.2](#)) muestra que los valores se concentran en torno a los 15s, con una media de $\bar{x} = 14,86$ s y una mediana de $M_e = 15,62$ s. Pese a la dispersión y presencia de valores atípicos, la cercanía entre ambas medidas sugiere una distribución aproximadamente simétrica.

²Rango intercuartil: $RIC = Q_3 - Q_1$, con Q_i el i -ésimo cuartil, $i \in \{1, 2, 3\}$.

3.2. Análisis de los tiempos de grabación

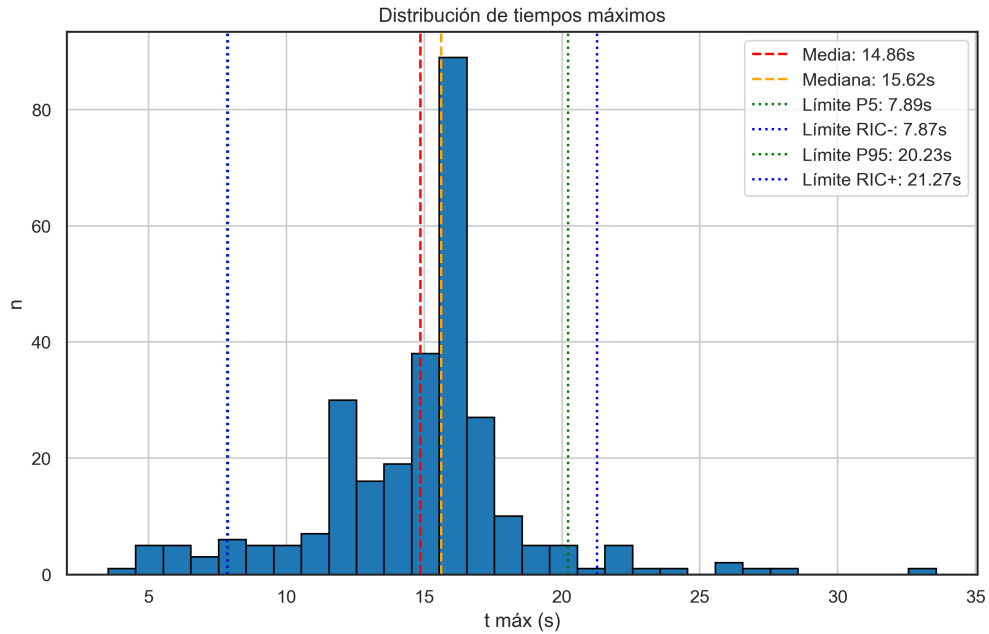


Figura 3.2: Histograma de duraciones de videos analizados. La media \bar{x} y la mediana M_e difieren por 0,86 s. Los valores P5 y P95 representan los percentiles 5 y 95, respectivamente. Los valores Límite RIC- y Límite RIC+ representan los extremos inferior y superior, respectivamente, del rango intercuartil.

En cuanto al paso temporal entre cuadros (*frames*), este no es homogéneo, con una media de 0,052 s y una desviación estándar de 0,012 s. La Figura 3.3 muestra la distribución de los valores de $\Delta t = t_i - t_{i-1}$, calculados para cada par consecutivo de tiempos de muestreo dentro de cada video. Sea I_k el conjunto de índices temporales del video k ; entonces, el conjunto total considerado es $\bigcup_{k=1}^{289} I_k$, es decir, la unión de todos los tiempos de muestreo, sin solapamiento entre videos. Esta distribución permite caracterizar el comportamiento general de los saltos temporales registrados, aspecto relevante para los análisis posteriores en el dominio temporal.

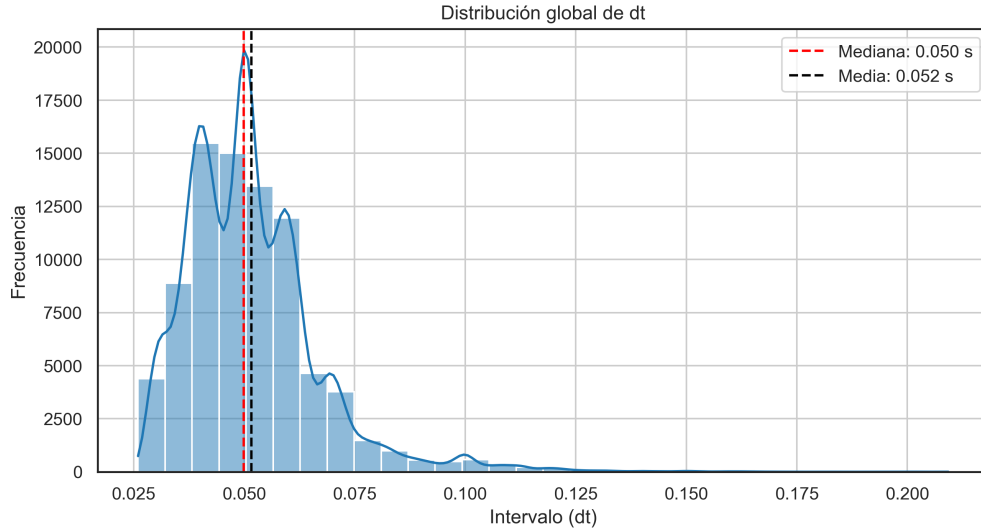


Figura 3.3: Distribución de los pasos temporales $\Delta t = t_i - t_{i-1}$, con $i \in I_k$ el conjunto de índices temporales del k -ésimo video, $k \in \{1, \dots, 289\}$. Se muestra el histograma junto con una estimación de densidad suavizada, cuyo ajuste no refleja con precisión la forma empírica observada.

En la figura anterior se observa prácticamente igualdad entre la media y la mediana de los intervalos de muestreo, lo que sugiere cierto grado de simetría y uniformidad en su distribución. Si bien existe una ligera cola hacia la derecha, esta se atenúa a poco más de una desviación estándar sobre la media. En general, los intervalos son consistentes entre trayectorias y no hay evidencia de valores atípicos extremos.

No obstante, pese a este leve grado de homogeneidad, una inspección directa de los instantes de muestreo t_k revela que no existe un conjunto común de tiempos compartido por todas las trayectorias. Cabe mencionar que cada t_k fue registrado con siete cifras significativas; por ende, al redondear y emplear menos decimales se obtienen algunas intersecciones, aunque a costa de una pérdida de precisión.

La falta de sincronización entre registros y la variabilidad en la duración de los videos hacen necesario homogeneizar el dominio temporal. Este paso es especialmente relevante para variables como la posición radial $r(t)$, cuya interpretación depende directamente del tiempo transcurrido. Sin este ajuste, los análisis podrían verse distorsionados por diferencias arbitrarias en los tiempos de observación, introduciendo sesgos que comprometen la comparabilidad entre trayectorias.

3.2.1. Criterio de optimización para una ventana temporal uniforme

Con el objetivo de asegurar la consistencia estadística y comparar elementos equivalentes, se llevó a cabo un procedimiento simple pero efectivo de homogeneización temporal. Este busca equilibrar la cantidad de información retenida con el número de trayectorias disponibles para el análisis.

Sea L la duración común a la que se desea truncar las trayectorias de forma uniforme, y sea T_i la duración de la trayectoria i -ésima, para $i = 1, \dots, N$, con $N = 289$. Para cada duración candidata L , se define la cantidad total de tiempo útil retenido como:

3.3. Caracterización inicial del comportamiento locomotor

$$I(L) := L \cdot \sum_{i=1}^N \mathbf{1}_{T_i \geq L}. \quad (3.11)$$

El valor óptimo se define como aquel que maximiza esta función:

$$L^* := \arg \max_L I(L). \quad (3.12)$$

Este criterio garantiza la selección del mayor volumen total de datos (en segundos) posible al truncar todas las trayectorias con duración al menos L . Para ello, se evaluaron únicamente los valores únicos de duración T_i observados en los datos.

El óptimo encontrado fue $L^* = 11,515$ s. Este valor se utilizó como duración común para los análisis subsecuentes, descartando aquellas trayectorias con duración menor a L^* .

La Figura 3.4 muestra la función de pérdida (Ecuación 3.13) asociada al problema de optimización definido en la Ecuación 3.12, la cual alcanza su mínimo global en L^* :

$$\text{Pérdida}(L) = \sum_{i=1}^N T_i - I(L) = \sum_{i=1}^N T_i - L \cdot \sum_{i=1}^N \mathbf{1}_{T_i \geq L}. \quad (3.13)$$

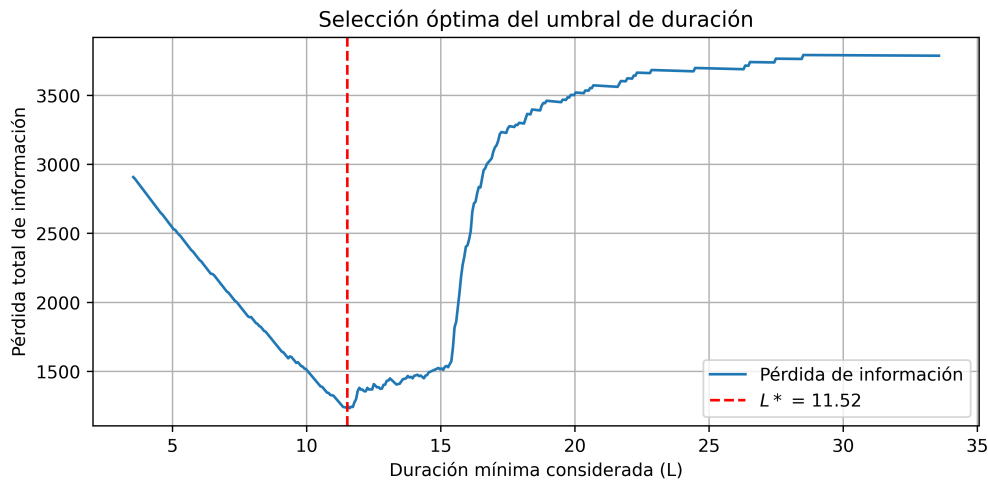


Figura 3.4: La función de pérdida total de información alcanza su mínimo global en $L^* = 11,515$. Se evaluaron únicamente las duraciones T_i observadas en los datos.

Con esta ventana temporal homogénea, junto con un Δt promedio que muestra baja variabilidad entre registros, se establecen las condiciones adecuadas para llevar a cabo los análisis planteados.

3.3. Caracterización inicial del comportamiento locomotor

Como primer paso, se obtuvieron medidas resumen que representan la dinámica del movimiento por individuo, facilitando la comparación entre condiciones experimentales. Con el objetivo de cuantificar la

3.3. Caracterización inicial del comportamiento locomotor

variabilidad en la respuesta en distintos escenarios, se trabajó con los cuatro grupos principales definidos por la combinación entre especie y tratamiento, presentados en la [Tabla 3.1](#). Esta estratificación se justifica en que las especies exhiben comportamientos diferenciados, por lo que analizarlas en conjunto podría introducir sesgos en los resultados. Por ejemplo, en condiciones de salinidad normal, el comportamiento esperado corresponde a un movimiento helicoidal o helicoidal rotatorio para las larvas de loco, y un movimiento ondulatorio hacia adelante en el caso de los erizos [13].

Código	Especie	Tratamiento
CC-E	<i>Concholepas concholepas</i>	Exposición
CC-R	<i>Concholepas concholepas</i>	Recuperación
LA-E	<i>Loxechinus albus</i>	Exposición
LA-R	<i>Loxechinus albus</i>	Recuperación

Tabla 3.1: Grupos experimentales considerados según especie y tratamiento.

En esta etapa se trabajó con datos agregados, donde cada larva queda representada por un conjunto de valores resumen. Para caracterizar la distribución de estas métricas dentro de cada grupo experimental se utilizaron dos medidas de tendencia central:

- **Media muestral** (\bar{x}): definida como $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$, representa el comportamiento promedio del grupo. Su uso está justificado por ser un estándar en análisis cuantitativo, y al normalizar por el tiempo, permite comparaciones entre trayectorias de distinta duración.
- **Mediana muestral** (\tilde{x}): corresponde al valor central de la muestra ordenada. Se emplea por su robustez frente a valores atípicos.

Estas fueron calculadas para todas las variables recién definidas, a saber: velocidad [mm s^{-1}], aceleración [mm s^{-2}], posición [mm], derivada de la posición (curvatura) [mm^{-1}], ángulo de la posición [rad], velocidad angular [rad s^{-1}], aceleración angular [rad s^{-2}], ángulo de la velocidad [rad] y curvatura local [mm^{-1}].

La comparación entre promedio y mediana permite evaluar la simetría de las distribuciones: una alta concordancia entre ambas sugiere simetría, mientras que discrepancias podrían indicar asimetrías o la presencia de valores atípicos. En esta etapa exploratoria, centrada en una agregación por grupo experimental sin considerar aún la salinidad, se observa una notable similitud entre ambas medidas ([Figura 3.5](#)).

3.3. Caracterización inicial del comportamiento locomotor

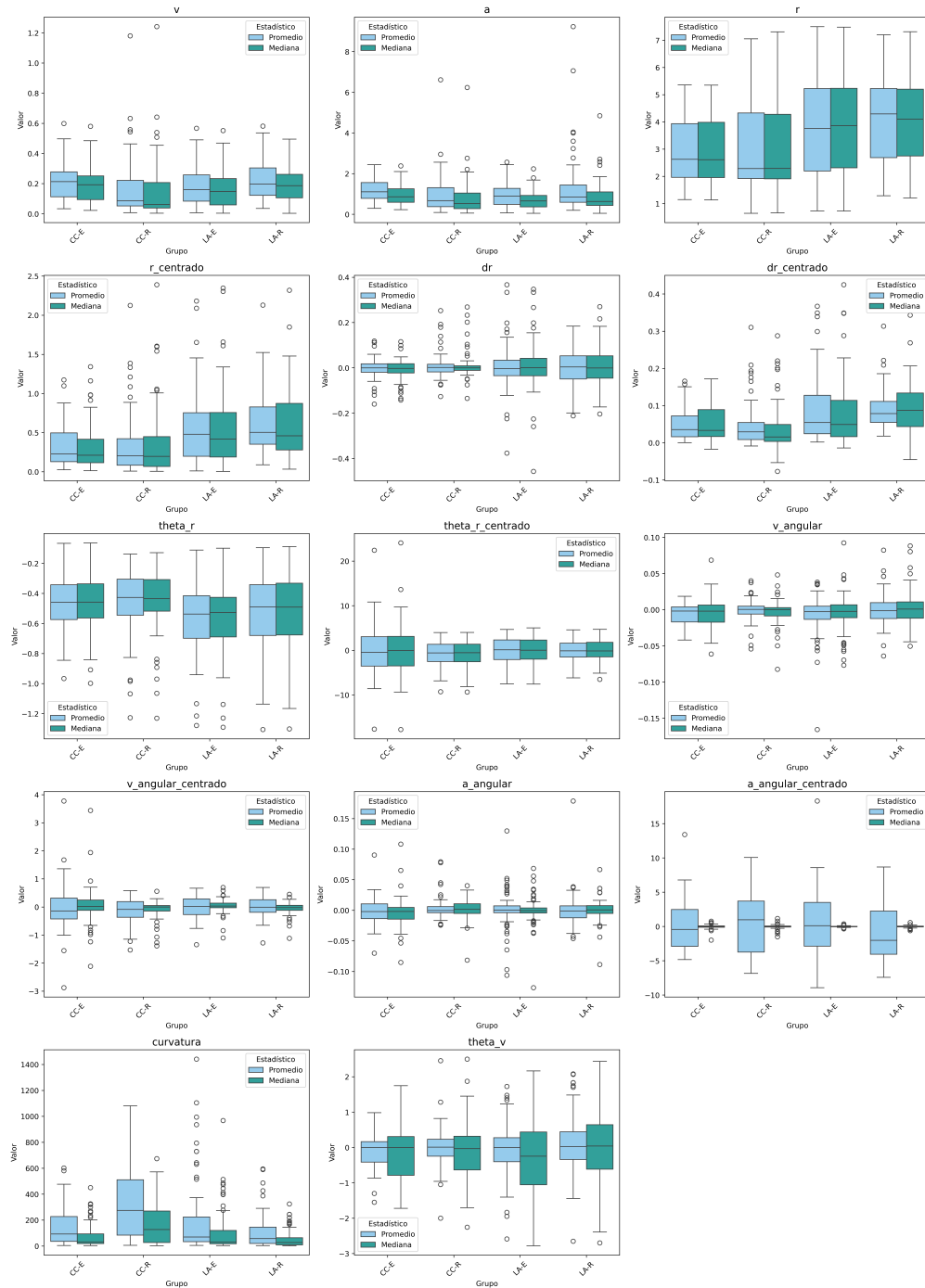


Figura 3.5: Diagramas de caja comparativos de las medidas de tendencia central (media y mediana) para cada variable analizada, estratificadas por grupo experimental.

En la mayoría de las variables se observa un comportamiento prácticamente idéntico entre la media y la mediana, lo que sugiere distribuciones simétricas y ausencia de sesgos marcados. Las diferencias presentes corresponden principalmente a la influencia de valores atípicos, los cuales afectan en mayor medida a la media que a la mediana. Esta mínima diferencia sugiere que ambas medidas capturan de

forma comparable el comportamiento central en esta etapa preliminar. Sin embargo, esta visualización no constituye un criterio suficiente para asumir su equivalencia en etapas posteriores. Al incorporar el efecto del nivel de salinidad podrían surgir diferencias relevantes entre promedio y mediana que justifiquen su análisis por separado.

La [Figura 3.5](#) también muestra comportamientos notoriamente distintos entre las variables calculadas a partir de coordenadas originales y centradas, lo que presenta la necesidad de evaluar el uso de un conjunto de variables consistente a lo largo de este estudio. Se detalla sobre esto en la [Subsección 3.4.2](#).

3.4. Análisis descriptivo

Una primera visualización relevante corresponde al conteo de registros por grupo. Para asegurar análisis comparables, se conservarán solo los datos con presencia en ambos tratamientos.

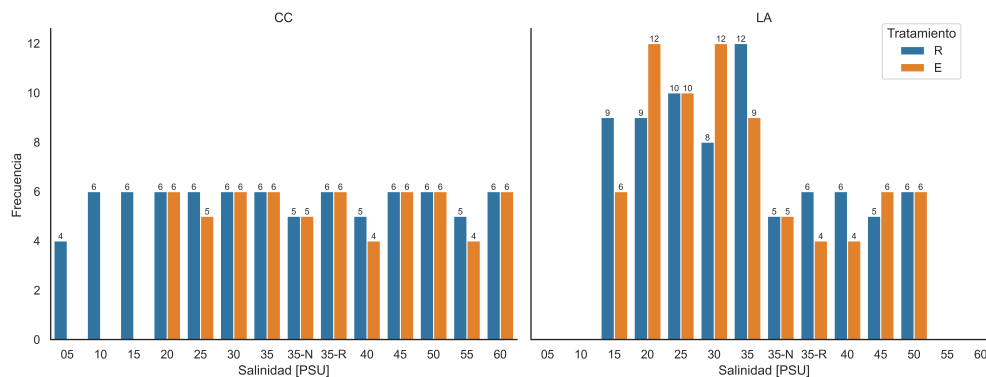


Figura 3.6: Número de observaciones por especie y tratamiento.

La [Figura 3.6](#) muestra que el número de observaciones por especie y tratamiento varía, y que en locos hay niveles de salinidad presentes únicamente en un tratamiento. Para permitir la comparación entre tratamientos, se descartaron del análisis los datos de esta especie con salinidades de 5, 10 y 15 [PSU]. En los demás grupos se observan cantidades dispares en ciertos niveles de salinidad, posiblemente debido a decisiones experimentales. Esta falta de equilibrio en el número de individuos por tratamiento refuerza la hipótesis de tratar las observaciones como muestras independientes.

3.4.1. La importancia de la ubicación

Durante la etapa exploratoria se evaluó la posibilidad de incluir la variable `id_ubicacion` como factor adicional junto con la salinidad. Sin embargo, al revisar la estructura del diseño experimental, se constató que cada nivel de salinidad está asociado de forma exclusiva a una ubicación determinada, sin solapamientos entre ellas. Esto implica que la ubicación no aporta información adicional respecto a la salinidad, ya que ambas variables están perfectamente correlacionadas en este contexto (ver [Tabla 2.3](#) y [Tabla 2.4](#)). Por lo tanto, se decidió excluir la ubicación como factor en los análisis, representando el tratamiento experimental únicamente mediante la salinidad, la cual encapsula adecuadamente la variabilidad geográfica.

Además, en los individuos de *L. albus*, la ubicación está estrechamente ligada al estadio larval observado, lo que añade una dimensión biológica relevante al filtrado. En larvas de equinodermos, los estadios iniciales como prisma exhiben una natación esencialmente pasiva, mientras que el estadio pluteus desarrolla un control activo del movimiento y mayor estabilidad estructural (brazos esqueléticos), generando diferencias locomotoras significativas [34]. Dado que en el experimento que dio origen a los datos analizados se observaron ambos estadios, en la [Subsección 4.5.1](#) se detalla el procedimiento metodológico aplicado previo al análisis.

Esta relación entre ubicación, salinidad y estadio larval queda reflejada en la [Figura 3.7](#), donde se muestra la distribución espacial de los estadios observados para cada especie. Esta visualización sirve como base para filtrar los datos antes de realizar comparaciones, asegurando consistencia tanto experimental como biológica. Se evaluará en cada caso si existen respaldos estadísticos para unificar los datos de *L. albus* según su estadio larval o, por el contrario, si son intrínsecamente distintos y no comparables.

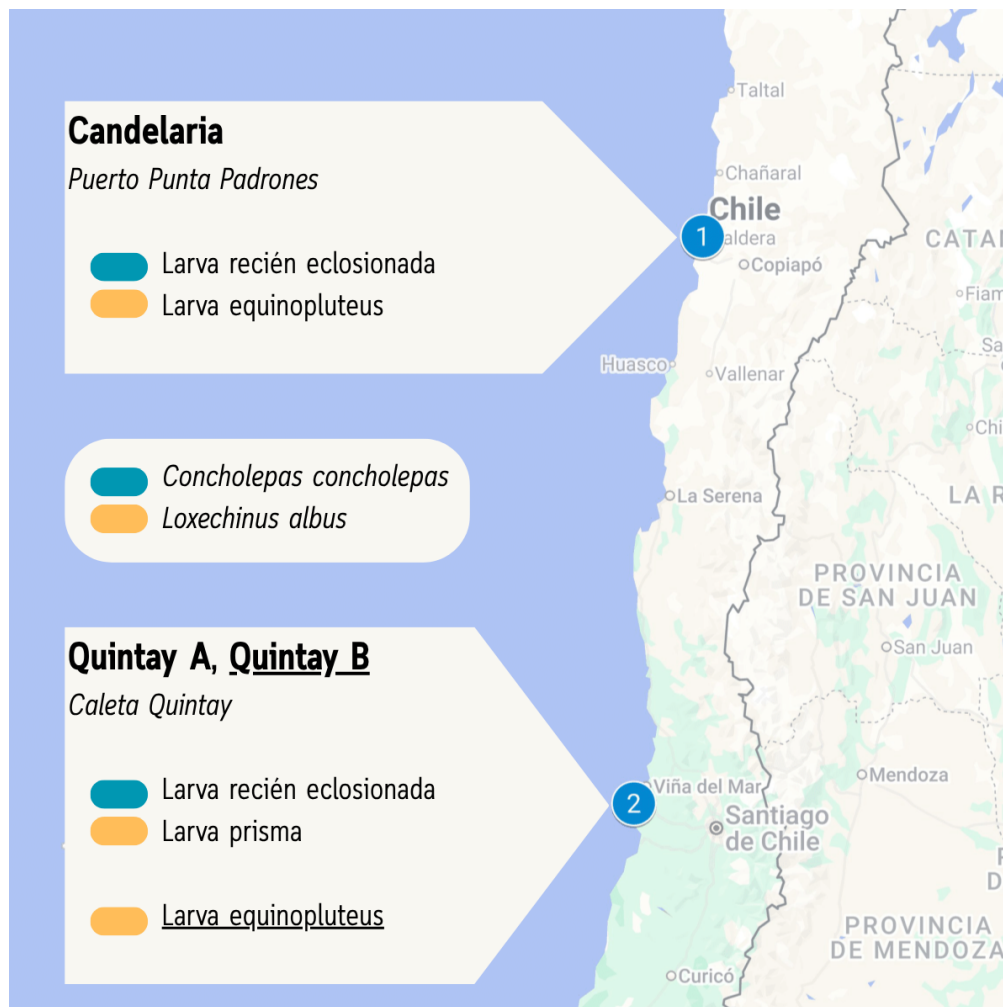


Figura 3.7: Mapa de recolección de muestras, con los estadios larvales por especie. Fuente: elaboración propia.

3.4.2. Análisis de la condición inicial

Con miras al análisis estadístico y el modelamiento estocástico, es necesario evaluar preliminarmente si las condiciones iniciales de las trayectorias generan sesgos en las comparaciones posteriores. Para este fin, resulta suficiente observar la distribución de la condición inicial en una variable globalmente descriptiva, como la posición radial $r(t)$.

Sea $r_0 = r(0)$ la posición inicial de cada larva. La [Figura 3.8](#) muestra su distribución por nivel de salinidad y grupo experimental. Se observan diferencias significativas entre grupos, atribuibles no a las condiciones experimentales, sino a una asignación espacial inicial no homogénea.

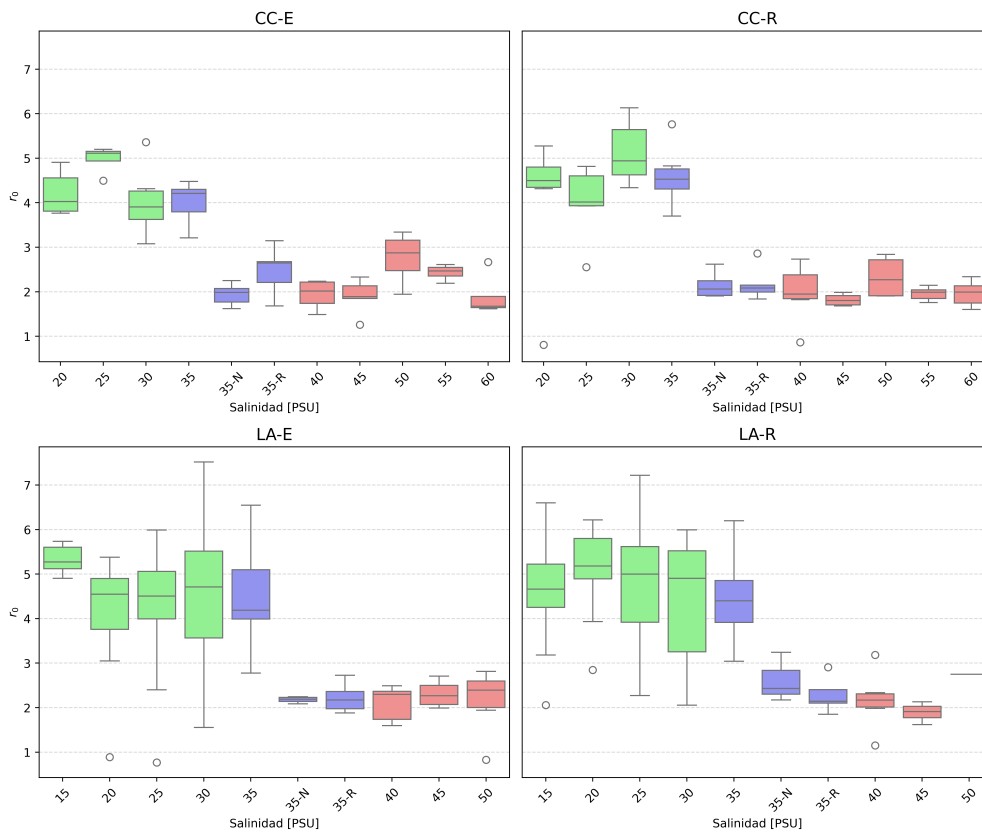


Figura 3.8: Diagramas de caja de la condición inicial r_0 por salinidad para cada grupo experimental. Se observan diferencias entre grupos y dentro de los grupos, lo que motiva el uso de la posición centrada $r_{\text{centrado}}(t)$ para asegurar consistencia.

Este resultado muestra que las trayectorias $r(t)$ no son directamente comparables entre condiciones, ya que combinan tanto la dinámica real como diferencias iniciales que no dependen del tratamiento; es decir, las trayectorias están sesgadas por construcción. En adelante, todos los análisis se realizan sobre las variables centradas, que permiten evaluar los efectos de forma consistente.

3.4.3. Métricas entre salinidades

La clasificación principal de las salinidades en baja, normal y alta resulta útil como punto de partida, pero puede ocultar respuestas específicas de los individuos frente a variaciones más finas del entorno. Esto lleva naturalmente a preguntarse si las salinidades agrupadas dentro de cada nivel representan efectivamente un mismo fenómeno, o si es necesario un análisis más detallado.

Para cada grupo experimental se visualizó el comportamiento de las variables consideradas a lo largo de los tres macroniveles de salinidad, desglosando además cada uno de sus subniveles. Cuando se aprecian diferencias cualitativas relevantes entre subniveles —más allá de la clasificación general—, se evalúa formalmente si dichas diferencias son estadísticamente significativas. Este procedimiento se aborda con mayor detalle en la [Subsección 4.5.1](#).

A continuación se presentan las estimaciones de densidad de las variables de interés, dado el componente rotatorio característico de los organismos estudiados [13]. Por simplicidad, se omiten gráficos análogos de variables que no presentaron diferencias notorias o dinámicas relevantes entre niveles de salinidad.

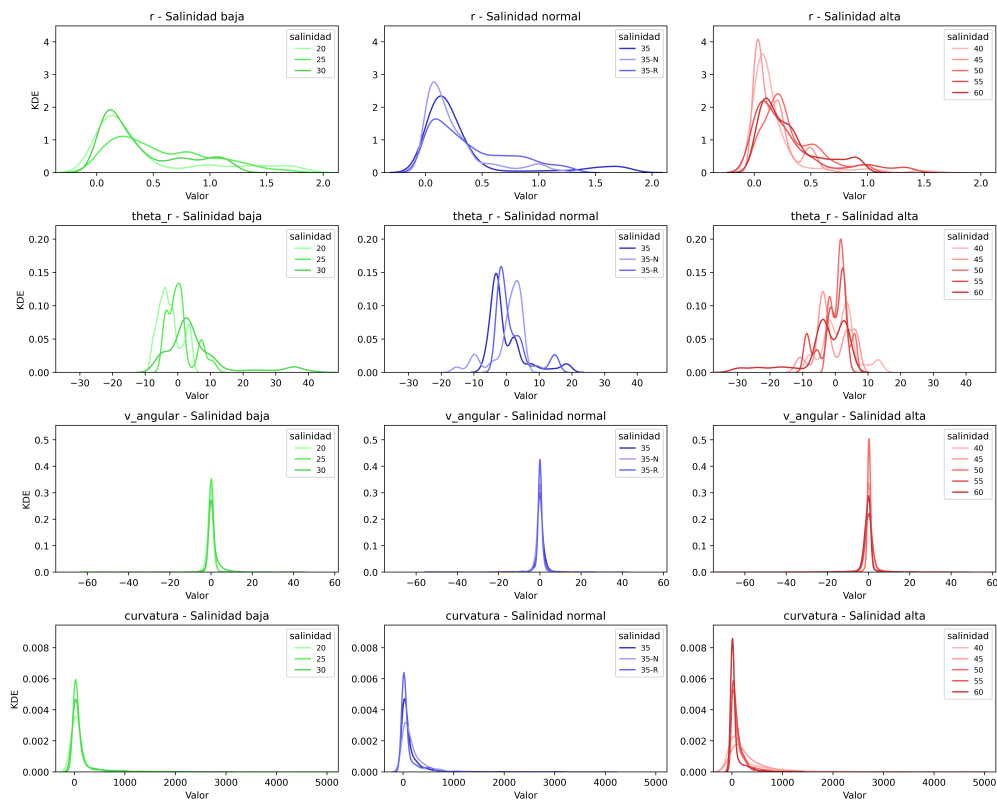


Figura 3.9: Estimación de densidad del *kernel* (KDE) para *C. Concholepas* en exposición en distintos macroniveles de salinidad

3.4. Análisis descriptivo

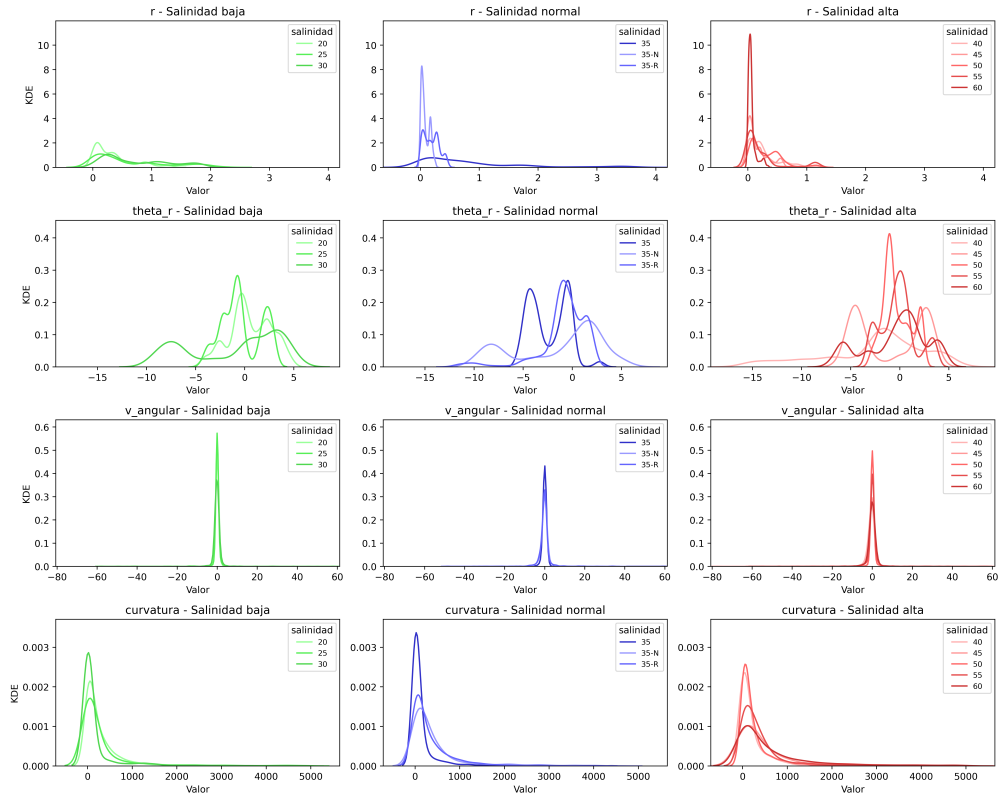


Figura 3.10: Estimación de densidad del *kernel* (KDE) para *C. Concholepas* en recuperación en distintos macroniveles de salinidad

3.4. Análisis descriptivo

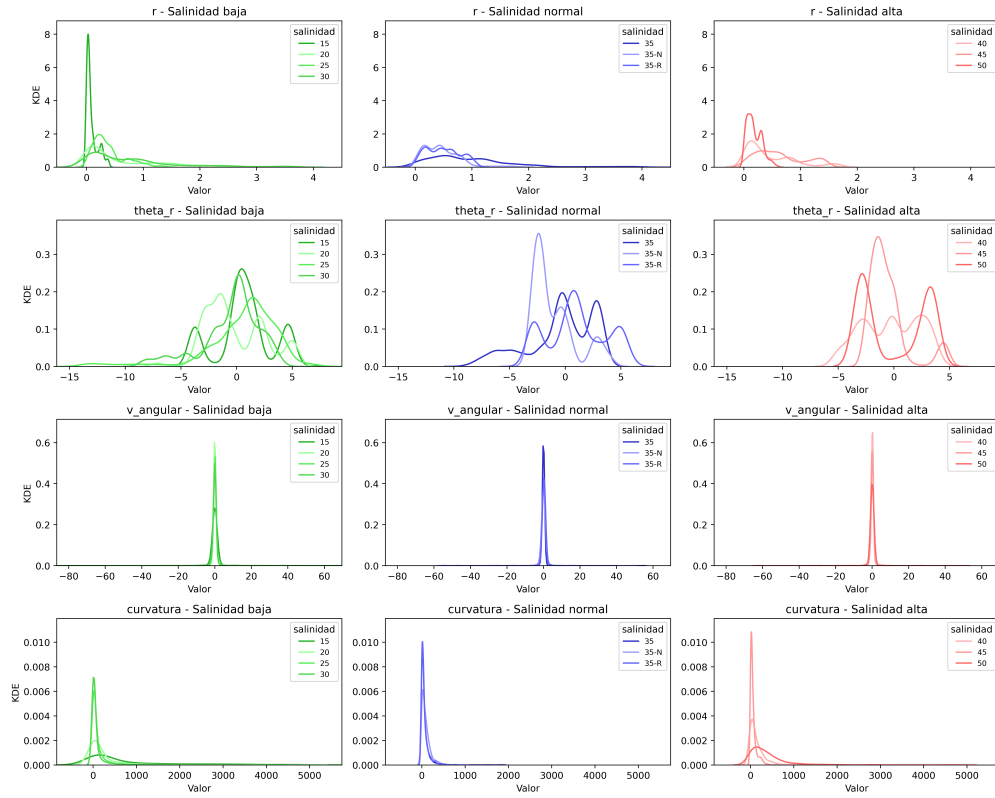


Figura 3.11: Estimación de densidad del *kernel* (KDE) para *L. albus* en exposición en distintos macro-niveles de salinidad

3.4. Análisis descriptivo

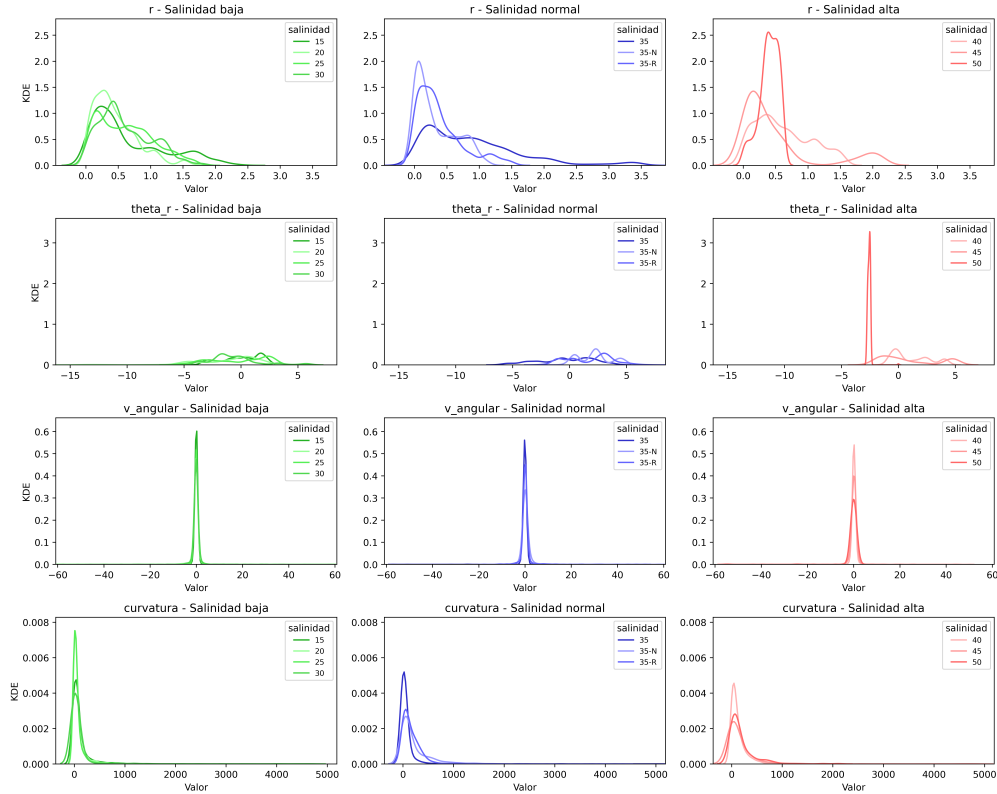


Figura 3.12: Estimación de densidad del *kernel* (KDE) para *L. albus* en recuperación en distintos macroniveles de salinidad

Las figuras 3.9, 3.10, 3.11 y 3.12 muestran que, de manera transversal entre grupos experimentales, las variables r_{centrado} y la curvatura exhiben distribuciones asimétricas con cola derecha, lo que sugiere una concentración de valores cercanos a cero acompañada de eventos extremos poco frecuentes. En contraste, la velocidad angular y el ángulo de la posición presentan perfiles más complejos: mientras la primera se mantiene estrechamente centrada en cero —indicando baja actividad rotacional—, el ángulo revela distribuciones multimodales, especialmente bajo salinidad normal y alta, lo que podría reflejar patrones de orientación divergentes entre individuos.

La velocidad angular y el ángulo presentan, en efecto, distribuciones más complejas. Las distribuciones de la velocidad angular son muy estrechas y centradas en cero, lo que implica que la mayoría del tiempo la velocidad es baja, con valores predominantemente entre -20 y 20 radian/s. Sin embargo, en el caso de θ_r , se observan distribuciones multimodales. Por ejemplo, en LA-E-normal, se aprecian picos claros alrededor de -5 , 0 y 5 , lo que podría indicar que los organismos se orientan en diferentes direcciones. Esto contrasta con las distribuciones de la velocidad angular, que son unimodales y más centradas.

En general, la salinidad parece ser un factor influyente en la forma de las distribuciones, especialmente en las variables de ángulo y posición radial, donde se observan cambios notorios en la forma y dispersión de los datos a medida que la salinidad varía. Por ejemplo, en los gráficos de salinidad alta se aprecia una mayor dispersión en los valores del ángulo, lo que podría reflejar una respuesta de estrés o una adaptación a la condición ambiental. Esto sugiere que las condiciones ambientales, como la salinidad,

3.4. Análisis descriptivo

alteran significativamente el comportamiento de movimiento de las especies estudiadas.

3.4.4. Correlaciones entre salinidades

Como complemento al análisis exploratorio, se calcularon matrices de correlación de Spearman para cada grupo experimental, con el objetivo de identificar patrones de asociación monotónica entre variables sin suponer normalidad ni relaciones lineales.

El coeficiente de Spearman, ρ_s , cuantifica la relación entre los rangos de dos variables, lo que le otorga mayor robustez frente a valores atípicos en comparación con la correlación de Pearson. Esta robustez se justifica formalmente por su función de influencia acotada, lo que implica que pequeñas alteraciones en los datos afectan de manera limitada su estimación. De hecho, como señalan Croux y Dehon (2010), esta propiedad hace de Spearman una herramienta especialmente adecuada en contextos con alta variabilidad experimental, como ocurre en estudios ecológicos [35]. Además, bajo condiciones de normalidad, mantiene una eficiencia comparable a la de Pearson, lo que refuerza su aplicabilidad como método preliminar ante incertidumbre en la distribución de los datos.

A continuación se presentan los mapas de calor de correlación de Spearman para los cuatro grupos principales, ilustrando la estructura relacional entre variables en cada caso.

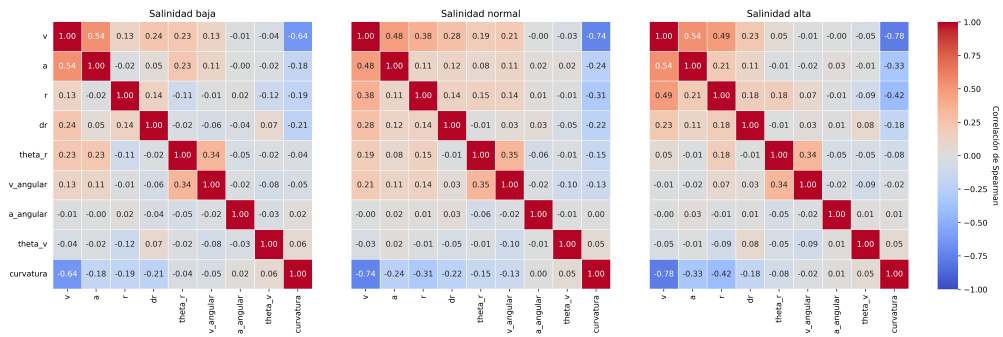


Figura 3.13: Correlaciones del mismo subconjunto de variables presentado en la Figura 3.9 para *Concholepas* en exposición.

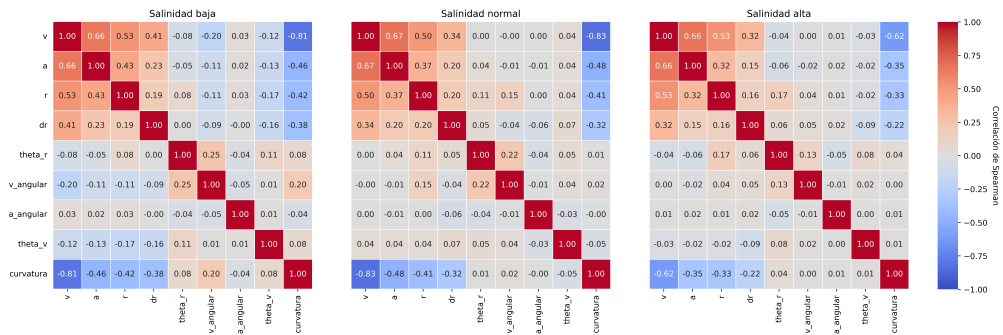


Figura 3.14: Correlaciones del mismo subconjunto de variables presentado en la Figura 3.10 para *Concholepas* en recuperación.

3.4. Análisis descriptivo

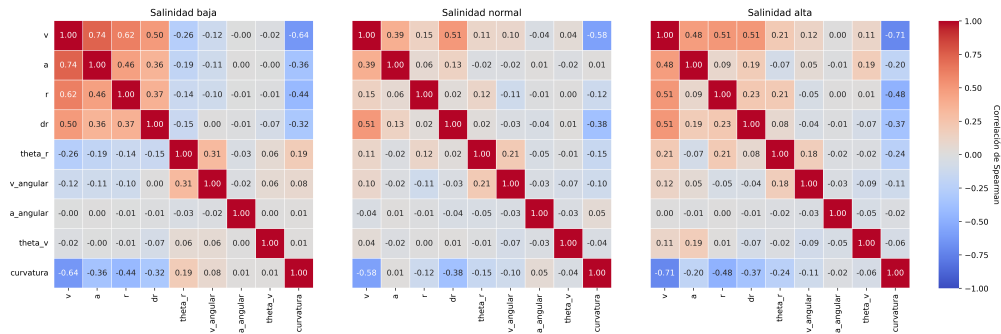


Figura 3.15: Correlaciones del mismo subconjunto de variables presentado en la Figura 3.11 para *L. albus* en exposición.

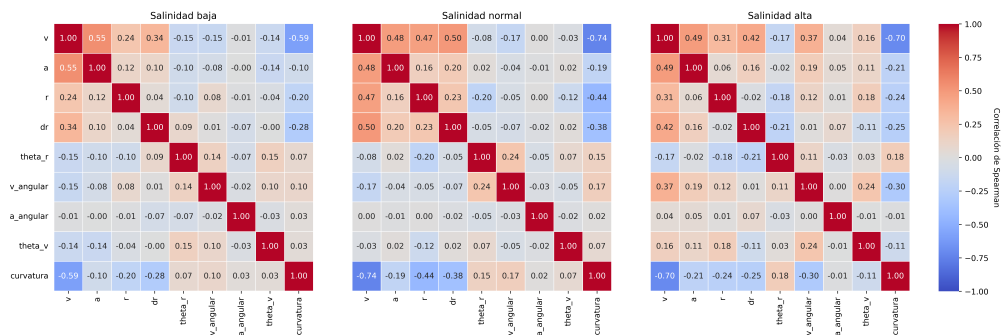


Figura 3.16: Correlaciones del mismo subconjunto de variables presentado en la Figura 3.12 para *L. albus* en recuperación.

De manera consistente en las figuras 3.13, 3.14, 3.15 y 3.16, la posición radial (r) exhibe una correlación positiva fuerte con la velocidad lineal (v) y la aceleración lineal (a) en casi todos los escenarios. Por ejemplo, en LA-R, salinidad baja, la correlación entre r y v es de 0,24, mientras que en salinidad alta aumenta a 0,31. Sin embargo, la correlación más consistente y significativa es entre r y dr (cambio en la posición radial), con valores consistentemente altos: 0,62 en LA-E, salinidad baja, y 0,49 en salinidad alta. Además, r muestra una fuerte correlación negativa con la curvatura, como se observa con un valor de $-0,64$ en LA-E, salinidad baja. Esto sugiere que a mayor distancia del centro, la trayectoria tiende a ser menos curva.

La velocidad angular (ω) generalmente presenta correlaciones débiles o nulas con las demás variables, excepto con la aceleración angular (α) y, en algunos casos, con el ángulo (θ_r). Por ejemplo, en LA-E, salinidad normal, la correlación entre ω y θ_r es débil pero positiva (0,21). La correlación más robusta es entre ω y α , que es consistentemente positiva, como lo demuestra el valor de 0,69 en CC-R, salinidad baja. Esto indica que a medida que la velocidad angular aumenta, la aceleración angular también lo hace.

El ángulo (θ_r) también muestra correlaciones generalmente débiles con la mayoría de las variables, aunque en algunos casos se observan asociaciones moderadas con la velocidad angular y aceleración angular. Por ejemplo, en CC-R, salinidad normal, se aprecia una correlación positiva de 0,22 entre θ_r y

ω , mientras que en LA-R, salinidad alta, esta correlación es de $-0,21$.

Por otro lado, la curvatura presenta correlaciones negativas consistentemente fuertes con la posición radial (r), como se mencionó, y con la velocidad lineal (v), con valores que a menudo superan $-0,40$. Por ejemplo, en CC-E, salinidad baja, la correlación entre curvatura y r es de $-0,64$, y entre curvatura y v , de $-0,46$. Esto sugiere que el movimiento más rápido y alejado del centro se asocia con trayectorias más rectas.

Finalmente, la velocidad y la aceleración lineal muestran correlaciones positivas muy fuertes y consistentes en todos los escenarios, como se observa con valores de $0,86$ en CC-R, salinidad baja, y $0,81$ en LA-E, salinidad alta. También se observa una fuerte correlación positiva entre v y dr , lo que indica que una mayor velocidad se asocia con un mayor desplazamiento. Por ejemplo, en LA-R, salinidad normal, la correlación alcanza $0,73$.

Capítulo 4

Evaluación del efecto de la salinidad

4.1. Modelos de comparación y variabilidad grupal

En el análisis estadístico de datos experimentales, una de las tareas centrales consiste en evaluar si existen diferencias significativas entre grupos definidos por uno o más factores. Para ello, se dispone de variadas técnicas que, dependiendo de los supuestos que cumpla la información, permiten realizar comparaciones con un respaldo estadístico robusto. Como se indica en [36, Cap. 10], utilizar pruebas estadísticas acorde al comportamiento de los datos reduce el riesgo de cometer errores de tipo I, que en este contexto significa detectar diferencias estadísticamente significativas entre los grupos cuando en realidad no existen.

En este trabajo se consideran tres procedimientos principales: el análisis de la varianza clásico (ANOVA¹), el ANOVA de Welch y la prueba no paramétrica de Kruskal–Wallis. Cada uno de estos métodos se basa en una formulación estadística distinta, asociada a condiciones específicas de aplicabilidad.

4.1.1. Análisis de la varianza

El ANOVA, introducido por R. Fisher en 1925 [37] como herramienta para evaluar diferencias entre medias grupales bajo un marco experimental, constituye uno de los pilares del análisis estadístico clásico. El ANOVA a un factor permite contrastar, para k grupos, la hipótesis nula

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_k \quad (4.1)$$

frente a la alternativa

$$H_1 : \exists i, j \text{ tales que } \mu_i \neq \mu_j, \quad (4.2)$$

donde μ_i representa la media poblacional del grupo i . Este contraste se formula mediante el modelo lineal:

$$y_{ij} = \mu + \tau_i + \epsilon_{ij}, \quad (4.3)$$

¹*Analysis of Variance*

con y_{ij} el valor de la j -ésima observación en el grupo i , μ la media global, τ_i el efecto del grupo i y ϵ_{ij} el término de error aleatorio, asumido como

$$\epsilon_{ij} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2). \quad (4.4)$$

En este análisis, la validez y fiabilidad de las conclusiones dependen en gran medida del cumplimiento de sus supuestos: independencia de las observaciones, normalidad de los residuos y homocedasticidad, entendida como la igualdad de varianzas poblacionales.

La variabilidad total se descompone en una suma de cuadrados *entre* grupos (SSB), atribuible al efecto del tratamiento o factor de agrupación, y una suma de cuadrados *dentro de* (o *intra*) grupos (SSW):

$$SSB = \sum_{i=1}^k n_i (\bar{y}_i - \bar{y})^2, \quad (4.5)$$

$$SSW = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2, \quad (4.6)$$

donde \bar{y}_i es la media del grupo i , \bar{y} la media global y n_i el tamaño del grupo i . El estadístico se prueba F se define como

$$F = \frac{MSB}{MSW} = \frac{SSB/(k-1)}{SSW/(N-k)}, \quad (4.7)$$

con N el número total de observaciones y MSB^2 y MSW^3 las medias cuadráticas entre e intra-grupos, respectivamente. Este se evalúa frente a la distribución de Fisher-Snedecor⁴ con $k-1$ grados de libertad entre grupos y $N-k$ grados de libertad intra grupos, i.e., $F_{k-1, N-k}$ [37, 38]. Si el valor observado de F supera el valor crítico de la distribución al nivel de significancia establecido –generalmente 5%–, se rechaza la hipótesis nula, concluyendo que al menos una de las medias grupales difiere significativamente del resto.

4.1.2. Análisis de la varianza de Welch

Cuando se cumplen las condiciones de normalidad e independencia, pero se viola el supuesto de homocedasticidad, el ANOVA de Welch se presenta como una alternativa robusta. Este método, propuesto por B. L. Welch en 1951 [39], no asume la igualdad de varianzas y ajusta el cálculo del estadístico de prueba para compensar las diferencias.

A diferencia del ANOVA clásico que utiliza un único estimador de la varianza del error, el método de Welch incorpora un peso $w_i = n_i/s_i^2$ para cada grupo i . Este peso es inversamente proporcional a la varianza muestral de cada grupo (s_i^2), lo que significa que los grupos con varianzas más pequeñas tienen

²Mean Square Between

³Mean Square Within

⁴También denominada distribución F, se define como la ley de la variable $\frac{U/m}{V/n} \sim F_{m,n}$, donde $U \sim \chi_m^2$ y $V \sim \chi_n^2$ son independientes y m, n son sus grados de libertad.

un mayor impacto en el cálculo de la media ponderada y del estadístico de prueba.

El estadístico F_W se calcula como una razón entre la varianza entre grupos y la varianza intra-grupos, análogamente al ANOVA clásico, pero ajustada por las varianzas individuales de cada uno de los k grupos:

$$F_W = \frac{\sum_{i=1}^k w_i (\bar{y}_i - \bar{y}_w)^2 / (k - 1)}{1 + \frac{2(k-2)}{k^2-1} \sum_{i=1}^k \frac{(1-w_i/n_i)^2}{n_i-1}}, \quad (4.8)$$

donde \bar{y}_i es la media ponderada del grupo i , $w_i = n_i/s_i^2$ es el peso del grupo i y \bar{y}_w es la media ponderada:

$$\bar{y}_w = \frac{\sum_{i=1}^k w_i \bar{y}_i}{\sum_{i=1}^k w_i}. \quad (4.9)$$

Un aspecto diferenciador del ANOVA de Welch es que su estadístico de prueba F_W no sigue una distribución F de Fisher-Snedecor con los grados de libertad estándar ($k - 1$ y $N - k$). En cambio, se evalúa contra una distribución F con grados de libertad ajustados, los cuales se calculan utilizando la aproximación de Welch-Satterthwaite⁵ [40]:

$$v_2 = \frac{\left(\sum_{i=1}^k \frac{s_i^2}{n_i} \right)^2}{\sum_{i=1}^k \frac{(s_i^2/n_i)^2}{n_i-1}}. \quad (4.10)$$

Esta fórmula produce un valor no entero para los grados de libertad, que es una característica distintiva de la aproximación de Satterthwaite. El valor de v_2 es generalmente menor que los grados de libertad del ANOVA clásico ($N - k$), lo que resulta en una distribución crítica más conservadora y un intervalo de confianza más amplio para la prueba. De acuerdo con Satterthwaite [40], esto reduce la probabilidad de cometer un error de Tipo I cuando las varianzas son desiguales.

4.1.3. Prueba de Kruskal–Wallis

Cuando los supuestos de normalidad y homocedasticidad no se cumplen, o si la escala de las mediciones es ordinal en lugar de continua, la prueba de Kruskal–Wallis –también conocida como prueba H de Kruskal–Wallis– se presenta como la alternativa no paramétrica principal para comparar tres o más grupos independientes. Este test, propuesto por W. H. Kruskal y W. A. Wallis en 1952 [41], es una extensión de la prueba U de Mann-Whitney⁶ y se basa en la comparación de las medianas de los grupos, aunque su hipótesis nula es más general.

A diferencia del ANOVA paramétrico, que trabaja con las medias de las variables, esta prueba utiliza los rangos de las observaciones. El procedimiento consiste en combinar todas las observaciones de todos los grupos y asignarles un rango, desde 1 hasta N , siendo N el tamaño total de la muestra. Luego, se suman los rangos de cada grupo.

⁵Esta aproximación, desarrollada inicialmente por F. E. Satterthwaite en 1946 y refinada por Welch, es un método para estimar los grados de libertad efectivos de una combinación lineal de varianzas muestrales.

⁶Prueba no paramétrica equivalente al t -test de Student para dos muestras independientes, que evalúa si hay una diferencia en la distribución o mediana entre los grupos, sin asumir normalidad ni homocedasticidad.

En este contexto, el *rango* de una observación se define como su posición relativa en el conjunto ordenado de todos los datos, sin considerar a qué grupo pertenece. Es decir, si se combinan todas las observaciones y se ordenan de menor a mayor, el valor más pequeño recibe el rango 1, el siguiente el rango 2, y así sucesivamente hasta el valor más grande, que recibe el rango N [41]. En caso de empates (valores idénticos), se asigna a cada observación el promedio de los rangos que ocuparían. Por ejemplo, si tres observaciones tienen el mismo valor y ocuparían los rangos 5, 6 y 7, cada una recibe el rango promedio:

$$\frac{5 + 6 + 7}{3} = 6.$$

Estos empates pueden ocurrir tanto dentro de un mismo grupo experimental como entre grupos distintos, ya que el ordenamiento se realiza sobre el conjunto total de observaciones. Por ejemplo, si dos larvas en salinidad baja y una en salinidad alta presentan exactamente el mismo valor en alguna variable observada, las tres compartirán el mismo rango promedio, independientemente de su grupo de origen. Este tratamiento de empates es fundamental para preservar la validez del estadístico H , y se discute en detalle en textos clásicos como el de Hollander et. al (2013) [42].

El estadístico de prueba, denotado como H , se calcula a partir de las sumas de los rangos de cada uno de los k grupos. Su fórmula es:

$$H = \frac{12}{N(N+1)} \sum_{i=1}^k n_i (\bar{R}_i - \bar{R})^2, \quad (4.11)$$

donde n_i es el tamaño de la muestra del grupo i , \bar{R}_i es el rango promedio del grupo i y $\bar{R} = \frac{N+1}{2}$ el rango promedio global.

La hipótesis nula de la prueba de Kruskal-Wallis establece que las distribuciones de todos los grupos son idénticas, lo que implica que sus medianas son iguales. La hipótesis alternativa indica que al menos una de las medianas de los grupos es diferente:

$$H_0 : \tilde{\mu}_1 = \tilde{\mu}_2 = \dots = \tilde{\mu}_k,$$

$$H_1 : \exists i, j \in \{1, \dots, k\} \text{ tal que } \tilde{\mu}_i \neq \tilde{\mu}_j$$

Para muestras con un tamaño de grupo suficientemente grande (generalmente $n_i \geq 5$), la distribución de H se aproxima a una distribución chi-cuadrado (χ^2) con $k - 1$ grados de libertad. Si el valor de H calculado es mayor que el valor crítico de la distribución χ^2 al nivel de significancia deseado, se rechaza la hipótesis nula, concluyendo que existe una diferencia estadísticamente significativa entre las distribuciones de los grupos.

4.2. Verificación de supuestos y selección del modelo

Antes de realizar el análisis, es fundamental verificar los supuestos de las pruebas estadísticas, con el fin de seleccionar el procedimiento más apropiado. Los tres supuestos principales para el ANOVA clásico son la independencia de las observaciones, la normalidad de los residuos y la homocedasticidad. Se tiene la independencia como hipótesis desde la [Sección 2.2](#); las otras dos se comprueban mediante las siguientes pruebas.

4.2.1. Prueba de Shapiro-Wilk para normalidad

Para evaluar la normalidad de los datos se utilizó la prueba de Shapiro-Wilk. Esta prueba, propuesta por S. Shapiro y M. Wilk en 1965 [43], es considerada una de las más potentes para detectar desviaciones de la normalidad. Contrasta las hipótesis:

H_0 : La muestra proviene de una población con distribución normal.

H_1 : La muestra no proviene de una población con distribución normal.

El estadístico de prueba W se calcula como:

$$W = \frac{\left(\sum_{i=1}^n a_i(x_{(i)})\right)^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

donde $x_{(i)}$ es el i -ésimo valor de los datos ordenados, \bar{x} es la media muestral y a_i son coeficientes calculados a partir de los valores esperados de los orden estadísticos de una muestra normal estándar y su matriz de covarianza [43]. El test se rechaza cuando el estadístico W es significativamente menor que 1, lo que indica que los datos se desvían de la normalidad. Formalmente, se rechaza la hipótesis nula cuando el valor- p asociado a W es menor que el nivel de significancia α (usualmente 0.05), sugiriendo que la muestra no proviene de una población normalmente distribuida [43].

4.2.2. Prueba de Bartlett para homocedasticidad

Esta prueba, propuesta por M. S. Bartlett en 1937 [44], es considerada potente, pero también muy susceptible a la violación del supuesto de normalidad. Si los datos no son normales, un resultado significativo en la prueba de Bartlett podría deberse a la no normalidad más que a la heterocedasticidad. Por esta razón, su uso se complementa con pruebas más robustas. Contrasta las hipótesis

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$$

$$H_1 : \exists i, j \text{ tal que } \sigma_i^2 \neq \sigma_j^2$$

El estadístico de prueba χ^2 se define como:

$$\chi^2 = \frac{(N - k) \log(S_p^2) - \sum_{i=1}^k (n_i - 1) \log(S_i^2)}{1 + \frac{1}{3(k-1)} \left(\sum_{i=1}^k \frac{1}{n_i - 1} - \frac{1}{N - k} \right)}$$

donde S_i^2 es la varianza muestral del grupo i , S_p^2 es la varianza combinada (*pooled*), n_i es el tamaño del grupo i y N es el total de observaciones [44]. Este estadístico sigue, bajo la hipótesis nula, una distribución aproximada χ^2 con $k - 1$ grados de libertad. El test se rechaza cuando el valor- p asociado es menor que el nivel de significancia α , lo que sugiere evidencia suficiente para rechazar la homocedasticidad entre los grupos [44].

4.2.3. Prueba de Levene para homocedasticidad

A diferencia de la prueba de Bartlett, la prueba de Levene es más robusta y menos sensible a desviaciones de la normalidad. Propuesta por H. Levene en 1960 [45], es un método estándar para verificar la homocedasticidad, especialmente cuando los datos no cumplen el supuesto de normalidad.

La prueba de Levene transforma el problema de la varianza en un ANOVA sobre las desviaciones absolutas de cada observación respecto a la mediana del grupo. Contrasta las mismas hipótesis que Bartlett, pero utilizando el estadístico

$$W = \frac{(N - k)}{(k - 1)} \cdot \frac{\sum_{i=1}^k N_i (Z_{i.} - Z_{..})^2}{\sum_{i=1}^k \sum_{j=1}^{N_i} (Z_{ij} - Z_{i.})^2}$$

donde:

- $Z_{ij} = |Y_{ij} - \tilde{Y}_i|$: diferencia absoluta respecto a la mediana del grupo,
- $Z_{i.}$: media de los Z_{ij} en el grupo i ,
- $Z_{..}$: media global de todos los Z_{ij} ,
- N_i : tamaño del grupo i , N : total de observaciones [45].

Este estadístico tiene distribución aproximada F de Snedecor con $k - 1$ y $N - k$ grados de libertad, bajo la hipótesis nula de igualdad de varianzas. El test se rechaza cuando el valor- p asociado a W es menor que el nivel de significancia α [46].

4.3. Análisis post hoc

Una vez que una prueba global (ANOVA clásico, ANOVA de Welch o Kruskal–Wallis) ha revelado una diferencia significativa entre los grupos, el análisis post hoc es indispensable para identificar cuáles son los pares de grupos específicos que difieren entre sí. La selección del procedimiento post hoc debe ser coherente con los supuestos de la prueba global y el tipo de datos.

En este trabajo se emplearon los siguientes procedimientos, según el resultado de la verificación de supuestos:

- Para el ANOVA clásico (normalidad y homocedasticidad): Se utilizó la prueba HSD de Tukey.
- Para el ANOVA de Welch (solo normalidad): Se empleó la prueba Games-Howell.
- Para la prueba de Kruskal-Wallis (ausencia de normalidad): Se aplicó la prueba de Dunn, con una corrección por comparaciones múltiples.

A continuación, se introduce brevemente cada procedimiento y sus fundamentos.

4.3.1. Prueba HSD de Tukey

La prueba HSD⁷ de Tukey es un método de comparaciones múltiples diseñado para realizar comparaciones pareadas entre todas las medias de grupos controlando la tasa de error familiar (FWER). Propuesta por J. Tukey en 1949 [47], parte del supuesto de homocedasticidad y normalidad de residuos. El estadístico se basa en el rango estudentizado (*studentized range*) q , y una comparación entre los grupos i y j se evalúa con el intervalo de confianza

$$\bar{y}_i - \bar{y}_j \pm q_{\alpha; k, N-k} \sqrt{\frac{MSW}{n}}$$

donde MSW es la media cuadrática intragrupos (error), k el número de grupos y n el tamaño muestral (para tamaños desiguales se usan sus fórmulas adaptadas). Tukey es apropiado cuando interesa controlar el FWER en comparaciones *todas contra todas* y los supuestos del ANOVA clásico son razonablemente válidos (Tukey, 1949) [47].

4.3.2. Prueba de Games-Howell

La prueba de Games-Howell es una alternativa recomendada cuando existen desigualdades de varianza y/o tamaños muestrales desiguales. Fue propuesta por J. R. Games y J. F. Howell en 1976, y se basa en una adaptación de la prueba t de Welch [48].

Para comparar los grupos i y j se utiliza un estadístico de tipo t corregido:

$$t_{ij} = \frac{\bar{y}_i - \bar{y}_j}{\sqrt{\frac{s_i^2}{n_i} + \frac{s_j^2}{n_j}}}, \quad (4.12)$$

Los grados de libertad efectivos para este estadístico se aproximan mediante la fórmula de Welch-Satterthwaite, lo que le permite mantener un control adecuado del error tipo I bajo condiciones de heterocedasticidad (Games & Howell, 1976) [48].

4.3.3. Prueba de Dunn y correcciones por múltiples comparaciones

Cuando la prueba global de Kruskal-Wallis resulta significativa, las comparaciones por pares se realizan comúnmente con la prueba de Dunn, propuesta por O. Dunn en 1964 [49]. Este procedimiento opera sobre las sumas de rangos y es la alternativa post hoc no paramétrica al método HSD de Tukey.

⁷Honestly Significant Difference

Dado que realizar múltiples comparaciones aumenta la probabilidad de cometer un error de tipo I, se aplica *a posteriori* una corrección para controlar la tasa de error familiar (FWER). Dos de las correcciones más comunes son:

- **Corrección de Bonferroni:** Divide el nivel de significancia α entre el número total de comparaciones por pares m . Es simple de aplicar pero muy conservadora [50], lo que reduce la potencia de la prueba y aumenta el riesgo de error tipo II.
- **Corrección de Holm:** Se ordenan los valores p de las comparaciones de menor a mayor y se comparan con $\alpha/(m - r + 1)$, donde r es el rango de la comparación. Si una comparación no es significativa, todas las que le siguen tampoco lo son.

La corrección de Holm, también conocida como procedimiento secuencialmente rechazante de Holm-Bonferroni [51], ha demostrado ser “uniformemente más potente” que la corrección clásica de Bonferroni, excepto en casos triviales, según el propio autor [51]. Esta superioridad se debe a que Holm ajusta los umbrales de significancia de forma progresiva, comenzando con el más exigente para el valor p más pequeño y relajando los criterios para los siguientes, lo que permite detectar más diferencias reales sin inflar el error tipo I.

Estudios posteriores como el de Giacalone et al. (2018) [50] compararon ambos métodos en un estudio sobre enfermedades inflamatorias intestinales y concluyeron que Holm permite identificar más comparaciones significativas que Bonferroni, manteniendo el control estricto de la FWER [50]. En su aplicación del procedimiento de Holm-Bonferroni, se detectaron nueve variables significativamente distintas entre grupos clínicos, mientras que Bonferroni habría pasado por alto varias de ellas, ilustrando su menor potencia en contextos reales.

4.4. Medidas de tamaño del efecto

Después de realizar una prueba de significancia como las de la [Sección 4.1](#), que indican si existe una diferencia entre grupos, es fundamental determinar la magnitud de dicha diferencia, conocida como *tamaño del efecto*. Estas medidas complementan el valor p , proporcionando una interpretación más completa de los resultados.

4.4.1. Pruebas paramétricas (ANOVA y ANOVA de Welch)

Las medidas de tamaño del efecto más comunes en ANOVA se basan en la proporción de la varianza total de la variable dependiente explicada por el factor de agrupación. Las más utilizadas son η^2 (eta cuadrado) y ω^2 (omega cuadrado).

- **Eta cuadrado (η^2):** Proporciona una estimación de la proporción de la variabilidad total que puede atribuirse a las diferencias entre grupos:

$$\eta^2 = \frac{SSB}{SST}$$

donde $SST = SSB + SSW$ es la suma total de cuadrados, y SSB es la suma de cuadrados del efecto entre grupos definida en la [Subsección 4.1.1](#). Es un estimador sesgado que tiende a sobrestimar el tamaño del efecto real [52, 53].

- **Omega cuadrado (ω^2):** corrige el sesgo de η^2 , especialmente en muestras pequeñas [54]:

$$\omega^2 = \frac{SSB - (k - 1)MSW}{SST + MSW},$$

siendo k el número de grupos y MSW la media cuadrática intra-grupos.

Debido a su menor sesgo, ω^2 es preferido sobre η^2 para la mayoría de los análisis, particularmente en estudios con muestras pequeñas.

4.4.2. Pruebas no paramétricas (Kruskal–Wallis)

En pruebas no paramétricas, las medidas se basan en rangos:

- **Épsilon cuadrado (ε^2):** Mide la proporción de la variabilidad explicada por el factor en términos de rangos y es menos sesgada que usar directamente η^2 calculado a partir de H en muestras pequeñas. Una forma estándar de calcularla para k grupos es la versión de Tomczak & Tomczak (2014) [55]:

$$\varepsilon^2 = \frac{H - k + 1}{N - k},$$

donde H es el estadístico de Kruskal–Wallis y N el tamaño total de la muestra. En casos raros, este valor puede ser negativo debido a fluctuaciones muestrales, interpretándose como 0.

- **Eta cuadrado basado en rangos (η_{H}^2):** Adaptación de η^2 para Kruskal–Wallis:

$$\eta_{H}^2 = \frac{H}{N - 1}.$$

En la literatura, ε^2 se considera superior porque corrige el sesgo de H y proporciona una estimación más precisa del tamaño del efecto, especialmente en muestras pequeñas [55].

En todos los casos, los valores cercanos a 0 indican un efecto pequeño, mientras que valores mayores reflejan una mayor proporción de varianza explicada por el factor. Los análisis post hoc se realizaron de manera coherente con la prueba utilizada, con el objetivo de comparar cada grupo frente al control (salinidad normal) y evaluar todas las comparaciones posibles cuando corresponda. Este procedimiento secuencial fue aplicado transversalmente en todos los análisis presentados en esta sección.

4.5. Análisis sobre indicadores agregados

En esta sección se presentan los resultados de la evaluación del efecto de la salinidad mediante métricas que condensan el comportamiento de las trayectorias en un solo valor escalar. Además, se analiza la posible influencia de otros factores.

4.5.1. Homogeneidad entre estadios larvales de *Loxechinus albus*

Una primera influencia importante a tener en cuenta es la de la etapa de desarrollo en las larvas de erizo rojo. Como se comentó en la [Subsección 3.4.1](#), este es un factor relevante en la respuesta motora de los organismos, por lo que evaluar efectos introducidos por ello es crucial para la obtención de conclusiones significativas.

Los resultados de la aplicación de pruebas de comparación entre grupos para el conjunto de datos correspondiente a *L. albus* arrojan las variables significativas presentadas en la [Tabla 4.1](#), por orden creciente del valor p . En esta, se aprecia casi total intersección de las variables significativas entre tratamientos.

Trat.	Variable	Test	Valor p	η^2 / η_H^2	ω^2 / ε^2
E	\tilde{k}^*	Kruskal-Wallis	0,000275	0,1813	0,1699
	\bar{v}^*	Kruskal-Wallis	0,002088	0,1297	0,1176
	\tilde{v}^*	Kruskal-Wallis	0,003845	0,1145	0,1022
	\bar{r}	Kruskal-Wallis	0,006857	0,1001	0,0876
	\bar{d}_r	Kruskal-Wallis	0,007102	0,0993	0,0868
	\bar{k}^*	Kruskal-Wallis	0,009684	0,0917	0,0791
	\bar{d}_r	Kruskal-Wallis	0,009684	0,0917	0,0791
	\tilde{r}	Kruskal-Wallis	0,016923	0,0781	0,0653
	\bar{a}^*	Kruskal-Wallis	0,018605	0,0759	0,0630
	$\bar{\omega}^*$	Kruskal-Wallis	0,025315	0,0685	0,0556
	$\bar{\theta}_r$	Welch's ANOVA	0,026253	0,0667	0,0531
	\tilde{a}^*	Kruskal-Wallis	0,028536	0,0657	0,0527
	$\tilde{\theta}_r$	Welch's ANOVA	0,035764	0,0598	0,0461
	R	\tilde{k}^*	Kruskal-Wallis	$2,15 \times 10^{-7}$	0,3586
\bar{k}^*		Kruskal-Wallis	$1,02 \times 10^{-6}$	0,3186	0,3094
\tilde{v}^*		Kruskal-Wallis	$4,65 \times 10^{-6}$	0,2797	0,2700
\bar{v}^*		Kruskal-Wallis	$5,14 \times 10^{-6}$	0,2771	0,2673
$\tilde{\omega}$		Kruskal-Wallis	$3,42 \times 10^{-4}$	0,1710	0,1598
\tilde{a}^*		Kruskal-Wallis	0,003134	0,1164	0,1044
\bar{a}^*		Kruskal-Wallis	0,003357	0,1147	0,1027
$\bar{\omega}^*$		Kruskal-Wallis	0,018316	0,0742	0,0617

Tabla 4.1: Variables con diferencias significativas entre larvas de erizo rojo en estadio prisma y equinopluteus bajo exposición (E) y recuperación (R). η^2 (ANOVA, ANOVA de Welch) o η_H^2 (Kruskal-Wallis): tamaño del efecto; ω^2 (ANOVA, ANOVA de Welch) o ε^2 (Kruskal-Wallis): estimadores corregidos del tamaño del efecto. *: Variable significativa en ambos tratamientos.

Los resultados de las comparaciones revelan diferencias significativas en varias variables de comportamiento entre las larvas de los estadios prisma y equinopluteus, tanto en el tratamiento de Exposición (E) como en el de Recuperación (R). En la [Tabla 4.1](#) se muestran solo las variables con $p < 0,05$; aún

así los tamaños de efecto (η^2 , ω^2/ε^2) evidencian diferencias marcadas entre tratamientos.

En Exposición, las diferencias se asocian a tamaños del efecto en un rango bajo a moderado, de 0,060 a 0,181. Esto sugiere que, si bien el estadio larval influye en la variabilidad del comportamiento, esta influencia es relativamente sutil en condiciones de exposición.

Por el contrario, en Recuperación, las diferencias entre ambos estadios son mucho más pronunciadas. Los tamaños del efecto para variables como la curvatura (\bar{k} y \tilde{k}) y la velocidad (\bar{v} y \tilde{v}) son notablemente altos, alcanzando hasta 0,359. Esto indica que, en este tratamiento, el estadio larval explica una porción mucho mayor de la variabilidad observada, actuando como un factor determinante en la respuesta de las larvas.

Esta divergencia en la respuesta entre ambos estadios larvales, que se hace particularmente evidente durante la recuperación, refuerza la idea de que representan fases de desarrollo funcionalmente distintas. La marcada diferencia del tamaño del efecto en el tratamiento R sugiere que las larvas de los distintos estadios no responden de la misma manera a los cambios ambientales, y su combinación en un único análisis podría hacer difusos patrones biológicos cruciales o introducir sesgos. Por lo tanto, con el objetivo de asegurar la consistencia metodológica y una interpretación precisa de los resultados, se optó por excluir las larvas prisma de los análisis posteriores. Aunque esta decisión implicó una reducción en el tamaño muestral, garantiza que los análisis se realicen sobre un grupo homogéneo y comparable.

4.5.2. Influencia de otros factores

En la base de datos se conservó información adicional relevante para la interpretación experimental: la presencia de vibraciones externas (obs) y la presencia de un único individuo en el video (larva_unica). Los casos no favorables de estos factores representan una fracción no despreciable de la muestra (aproximadamente 29 % y 13 %, respectivamente) y, por tanto, su posible efecto fue evaluado explícitamente para cada grupo experimental mediante el mismo flujo de trabajo de verificación de supuestos y test global, seguido de análisis post hoc.

Presencia de vibraciones externas

La variable obs codifica tres niveles de vibración en el entorno de registro: 0 corresponde a ausencia de vibraciones, 1 a vibración leve y 2 a vibración significativa. A continuación se muestran las variables que resultaron significativas para cada especie y condición experimental, ordenadas por valor p .

Variable	Test	Valor p	η^2	ω^2/ε^2
$\tilde{\alpha}$	Kruskal–Wallis	$8,00 \times 10^{-3}$	0,1328	0,1161
$\tilde{\theta}_r$	Kruskal–Wallis	$3,11 \times 10^{-2}$	0,0877	0,0701

Tabla 4.2: Variables con diferencias significativas según la variable obs en *C. concholepas* bajo exposición.

4.5. Análisis sobre indicadores agregados

Variable	Test	Valor p	η^2	ω^2/ε^2
\tilde{a}	ANOVA	0,00081	0,305969	0,285225
\bar{a}	Kruskal–Wallis	0,000242	0,306159	0,290024
$\tilde{\alpha}$	Kruskal–Wallis	0,16988	0,12949	0,10925
$\tilde{\theta}_r$	ANOVA	0,029833	0,10507	0,08254
$\tilde{\theta}_v$	ANOVA	0,032100	0,102422	0,079881
$\bar{\theta}_r$	ANOVA	$4,90 \times 10^{-2}$	0,08717	0,06457

Tabla 4.3: Variables con diferencias significativas según la variable obs en *L. albus* bajo exposición.

Variable	Test	Valor p	η_H^2	ε^2
\bar{a}	Kruskal–Wallis	0,003487	0,25098	0,228283
\tilde{a}	Kruskal–Wallis	0,007685	0,20898	0,185010

Tabla 4.4: Variables con diferencias significativas según la variable obs en *L. albus* bajo recuperación.

En *C. concholepas* las vibraciones afectan preferentemente a medidas angulares centradas y a la posición radial en recuperación (Tabla 4.2), con tamaños de efecto pequeños a moderados (η^2 en torno a 0,08–0,13) [56, 55]. En *L. albus* el efecto es más marcado: bajo exposición (Tabla 4.3) aparecen diferencias en medidas de aceleración (con η^2 próximas a 0,30), velocidades angulares y ángulos radiales; en recuperación (Tabla 4.4) las aceleraciones vuelven a aparecer con tamaños de efecto moderados.

En conjunto, estos resultados muestran que las vibraciones instrumentales pueden introducir una variabilidad no biológica considerable, especialmente en métricas como la aceleración y velocidad angular en erizos. Si bien no se incorporó directamente como covariable en este análisis, obs podría ser tomada en cuenta en estudios posteriores, ya sea mediante filtrado o como referencia contextual, para mejorar la interpretación de los efectos observados.

Número de individuos

La variable larva_unica indica si el video contiene un solo individuo (1) o más de uno (0). Los resultados presentados en las Tablas 4.5 a 4.7, muestran un efecto notoriamente más fuerte en *C. concholepas* que en *L. albus*.

Variable	Test	Valor p	η_H^2	ε^2
\bar{a}	Kruskal–Wallis	0,013922	0,114114	0,097077
\tilde{a}	Kruskal–Wallis	0,028421	0,0960611	0,073123

Tabla 4.5: Variables con diferencias significativas según la variable larva_unica en el grupo CC-E.

Variable	Test	Valor p	η_H^2	ϵ^2
\tilde{k}	Kruskal–Wallis	0,000078	0,273768	0,260799
\tilde{r}	Kruskal–Wallis	0,000141	0,254277	0,240961
\bar{v}	Kruskal–Wallis	0,000155	0,251099	0,237726
\tilde{v}	Kruskal–Wallis	0,000155	0,251099	0,237726
\bar{r}	Kruskal–Wallis	0,000206	0,241683	0,228142
\bar{k}	Kruskal–Wallis	0,000206	0,241683	0,228142
\tilde{a}	Kruskal–Wallis	0,000515	0,211598	0,197520
\bar{a}	Kruskal–Wallis	0,000515	0,211598	0,197520
$\bar{d}r$	Kruskal–Wallis	0,002161	0,165041	0,150131
$\tilde{d}r$	Kruskal–Wallis	0,007782	0,124261	0,108623

Tabla 4.6: Variables con diferencias significativas según la variable *larva_unica* en el grupo CC-R.

Variable	Test	Valor p	η_H^2	ϵ^2
\tilde{r}	Kruskal–Wallis	0,005421	0,175761	0,156593
\bar{r}	Kruskal–Wallis	0,017142	0,129130	0,108878
$\bar{d}r$	Kruskal–Wallis	0,022888	0,117658	0,097139
$\tilde{d}r$	Kruskal–Wallis	0,039533	0,096314	0,075298

Tabla 4.7: Variables con diferencias significativas según la variable *larva_unica* en el grupo LA-E.

En *C. concholepas*, la presencia de un único individuo en el video muestra efectos generalizados sobre variables radiales, de velocidad y curvatura, con tamaños de efecto elevados [56] (varios η^2 en torno a 0,12–0,25). Esto se observa durante la fase de recuperación. Por otro lado, en *L. albus* el efecto existe pero es más moderado: algunas medidas de r y sus derivadas aparecen afectadas con η^2 en torno a 0,09–0,17, pero solo durante la fase de exposición. Estos resultados sugieren que la presencia de más de un individuo, incluso sin contacto físico, podría modificar significativamente varios parámetros cinemáticos en locos, especialmente durante la fase de recuperación, mientras que en erizos este efecto es mucho más débil y no se encontró evidencia de su influencia en la fase de recuperación.

Consideración sobre el tamaño de muestra. Los niveles minoritarios de las variables *obs* ($\approx 29\%$) y *larva_unica* ($\approx 13\%$) generan desbalances apreciables entre grupos comparados. Esta asimetría puede comprometer la potencia estadística de las pruebas y la estabilidad de las estimaciones de efecto, incrementando el riesgo de que tamaños de efecto elevados reflejen, en parte, artefactos muestrales. En particular, en los casos donde se observan η^2 altos (como en locos para ambas variables), es posible que parte de esa magnitud esté influenciada por la menor representatividad de uno de los grupos. Por ello, aunque los estadísticos empleados toleran diferencias en el tamaño muestral y las diferencias detectadas son estadísticamente significativas, conviene interpretar estos resultados con cierta cautela. En las secciones que siguen, no se aplicó un filtrado adicional sobre estos dos factores, ya que ello habría

implicado una reducción aún mayor del volumen de datos, el cual ya es limitado. En estudios futuros, podría ser útil explorar estos patrones bajo condiciones de mayor equilibrio muestral o mediante análisis complementarios.

4.6. Resultados por grupo experimental

El resultado esencial de esta sección se desprende de la comparación de las medias de cada variable locomotora entre los tres niveles de salinidad principales: baja, normal y alta. A partir de este análisis se verifica la existencia de diferencias en el comportamiento larval en condiciones de salinidad normal (35, 35-N, 35- R PSU) frente a salinidad alterada (ya sea baja, de 20 a 30 PSU, o alta, mayor o igual a 40 PSU). Es sensato comparar bajo esta estratificación más general antes de hilar más fino en la evaluación de diferencias de valores intra-salinidad.

A continuación se detallan las variables que presentaron diferencias significativas en cada grupo experimental. Los contrastes globales fueron ajustados por comparaciones múltiples cuando correspondía.

Variable	Test aplicado	Valor p	η_H^2	ε^2
\bar{r}	Kruskal–Wallis	0,005765	0,180909	0,151124
\tilde{r}	Kruskal–Wallis	0,010880	0,158625	0,128029
$\bar{d}r$	Kruskal–Wallis	0,029930	0,123119	0,091233
$\tilde{d}r$	Kruskal–Wallis	0,044218	0,109425	0,077041

Tabla 4.8: Variables con diferencias significativas entre niveles de salinidad en *C. concholepas* bajo recuperación.

Variable	Test aplicado	Valor p	η_H^2	ε^2
\tilde{a}	Kruskal–Wallis	0,019555	0,178841	0,139738
$\tilde{d}r$	Kruskal–Wallis	0,020934	0,175745	0,136495
\bar{a}	Kruskal–Wallis	0,023505	0,170479	0,130978
$\tilde{\alpha}$	Kruskal–Wallis	0,028096	0,162370	0,122483
$\bar{d}r$	Kruskal–Wallis	0,028175	0,162242	0,122348
$\bar{\theta}_v$	Kruskal–Wallis	0,039928	0,146395	0,105747

Tabla 4.9: Variables con diferencias significativas entre niveles de salinidad en *L. albus* bajo exposición.

En *C. concholepas* en recuperación, la posición y velocidad radial, r y dr , son los indicadores más sensibles a cambios de salinidad (Tabla 4.8). El análisis post-hoc revela que las diferencias significativas en ambas variables se encuentran específicamente entre los grupos de salinidad baja y alta ($p < 0,05$). Esto sugiere que los individuos expuestos a estos extremos de salinidad se comportan de forma notablemente diferente entre sí, mientras que su comportamiento en relación con el grupo de salinidad normal no muestra una diferencia estadísticamente significativa.

En este sentido, una pregunta natural que surge es dónde, dentro del espectro de salinidad, se posiciona el umbral de cambio comportamental en la respuesta de los organismos de este grupo. Abordar esta pregunta, si bien escapa a los objetivos de este trabajo, se vincula directamente con los resultados de la [Sección 4.7](#) y con la propuesta de consideración de la salinidad como un factor de naturaleza continua en estudios futuros.

Por su parte, en *L. albus* bajo exposición, varios componentes locomotores presentan diferencias significativas, como se muestra en la [Tabla 4.9](#). Aunque el tamaño del efecto es moderado en todas estas variables (η_H^2 entre 0,1 y 0,18), su sensibilidad colectiva en ambos grupos destaca la importancia de monitorearlas.

El análisis post-hoc de Kruskal–Wallis revela detalles adicionales sobre las diferencias en LA-E:

- La mediana de la velocidad radial ($\tilde{d}r$) muestra una diferencia significativa entre los grupos de salinidad normal y baja ($p = 0,019879$).
- El promedio de la aceleración (\bar{a}) también presenta una diferencia significativa entre los grupos de salinidad normal y baja ($p = 0,021574$).
- La mediana de la aceleración angular ($\tilde{\alpha}$) presenta una diferencia significativa entre los grupos de salinidad alta y baja ($p = 0,028124$).
- Finalmente, el promedio de la velocidad angular ($\bar{\theta}_v$) también muestra una diferencia significativa entre los grupos de salinidad alta y baja ($p = 0,036318$).

Lo anterior indica que, en larvas de erizo rojo, la alteración del movimiento es más pronunciada en las variables relacionadas con la velocidad y la aceleración cuando se comparan los extremos de salinidad (alta vs. baja), así como la salinidad normal con la baja.

Una observación relevante es que en los grupos CC-E y LA-R, ninguna de las variables mostró diferencias estadísticamente significativas entre macroniveles de salinidad. Esta ausencia de efectos detectables podría interpretarse según los mecanismos de osmorregulación descritos previamente en el informe. En particular, se ha documentado que *C. concholepas* exhibe una mayor plasticidad fisiológica frente a variaciones salinas [57], mientras que *L. albus*, aunque más sensible en fase de exposición, muestra una recuperación conductual relativamente rápida bajo condiciones controladas, posiblemente asociada a una reestabilización osmótica post-exposición [58]. Esta interpretación refuerza la hipótesis de que la respuesta locomotora larval no depende únicamente de la salinidad ambiental, sino también del historial de exposición y de la capacidad intrínseca de cada especie para modular su homeostasis interna.

Dicho esto, resulta particularmente curiosa la inversión de las fases sensibles a variaciones salinas ambientales en cada especie. Las diferencias en la respuesta conductual entre *C. concholepas* y *L. albus* puede atribuirse a sus estrategias osmóticas. *C. concholepas* presenta una respuesta tardía, compatible con un comportamiento osmoconformador, en el que los efectos del estrés salino no se manifiestan durante la exposición sino en la fase de recuperación. Si bien, como indica [4], aun no se ha determinado con certeza su modalidad osmótica, estudios fisiológicos sugieren una baja activación de mecanismos excretores inmediatos y una relación entre locomoción y gasto energético en contextos de ajuste interno

(Navarro & Torrijos, 1995 [57]; Mangum, 1983) [59]. En contraste, *L. albus* muestra alteraciones conductuales durante la exposición, lo que podría reflejar una regulación osmótica más activa incluso en etapas larvales. Este patrón es coherente con lo descrito por Diehl (1986) [?], quien señala que algunos equinodermos movilizan osmólitos compatibles y ajustan su composición iónica interna en respuesta a variaciones salinas, lo que podría explicar la activación temprana de respuestas locomotoras en *L. albus*, aunque no se menciona esta especie en particular.

Las diferencias temporales observadas no implican mayor o menor sensibilidad, sino que responden a los momentos en que cada especie activa sus mecanismos compensatorios frente al estrés osmótico, incluso en estados tan tempranos como los considerados en este estudio. No obstante, es importante considerar que estos resultados se basan en promedios, lo que podría ocultar variaciones temporales transitorias en la respuesta de los organismos. En la Sección 4.8 se aplica un enfoque más integral que incorpora la dependencia temporal, y podría revelar hallazgos pasados por alto al comprimir todo el proceso en un solo valor resumen.

4.7. Evaluación intra-nivel de salinidad

El objetivo de este apartado es evaluar si, dentro de cada categoría amplia de salinidad, existen diferencias significativas entre los niveles específicos que la componen. Por ejemplo, se contrasta el comportamiento entre 20, 25 y 30 PSU, todos niveles de salinidad baja. Esta exploración permite indagar si las respuestas observadas obedecen a umbrales discretos o si, por el contrario, reflejan una sensibilidad más continua frente a variaciones graduales de la salinidad. Se evaluaron solamente las variables significativas en la comparación inicial entre categorías amplias de salinidad.

4.7.1. *C. concholepas*: Recuperación

Baja. No se observaron efectos intra-nivel detectables. La categoría *baja* se comporta de manera estable en esta fase, lo que refuerza su tratamiento como condición homogénea.

Normal. La Tabla 4.10) indica que la derivada radial promedio ($\bar{d}r$) es la única variable que muestra una diferencia estadísticamente significativa, con un valor $p = 0,043$. Los tamaños de efecto η_H^2 y ω^2 son altos (0.407 y 0.294, respectivamente), lo que sugiere que las variaciones internas en esta categoría afectan de manera importante la dinámica del desplazamiento radial de las larvas.

Variable	Test	Valor p	η_H^2	ω^2
$\bar{d}r$	Welch's ANOVA	0,043381	0,40723	0,29392

Tabla 4.10: Comparaciones intra-nivel dentro de *normal* para *C. concholepas* en recuperación.

Alta. Finalmente, se detecta una diferencia intra-nivel en la mediana de velocidad radial centrada, lo que sugiere que incluso en condiciones de alta salinidad, la dinámica del movimiento puede variar significativamente entre niveles específicos.

Variable	Test	Valor p	η^2_H	ω^2
$\tilde{d}r$	Kruskal–Wallis	0,030365	0,42733	0,31825

Tabla 4.11: Comparaciones intra-nivel dentro de *alta* para *C. concholepas* en recuperación.

En las larvas de *C. concholepas* en recuperación, la posición radial r destaca como una variable altamente sensible a las variaciones internas dentro de los macroniveles de salinidad *normal* (Tabla 4.10) y *alta* (Tabla 4.11). En ambos casos, los análisis post hoc indican que los individuos expuestos a las condiciones de salinidad natural y reconstituida del grupo de Candelaria –35-N y 35-R respectivamente– muestran un comportamiento significativamente diferente del grupo control de Quintay (35). Esto sugiere que la procedencia del agua, y por ende, el entorno experimental, introduce efectos sustantivos que trascienden la simple categorización por salinidad. Los resultados (ver ??) muestran que las condiciones 35-N y 35-R se comportan de manera homogénea entre sí, pero de forma distinta al grupo 35, lo que apunta a la influencia del entorno control, al menos en las variables aquí presentadas.

4.7.2. *L. albus*: Exposición

Baja. Se observan efectos intra-nivel fuertes sobre la aceleración y la velocidad radial:

Variable	Prueba	Valor- p	η^2	ω^2
\bar{a}	Welch's ANOVA	0,000023	0,730404	0,675356
\tilde{a}	Kruskal–Wallis	0,003057	0,661472	0,605051
$\bar{d}r$	Kruskal–Wallis	0,005436	0,602786	0,536583

Tabla 4.12: Comparaciones intra-nivel dentro de la categoría *baja* para *L. albus* en exposición.

Normal. No se detectaron cambios intra-nivel significativos.

Alta. Aparecen diferencias intra-nivel en aceleración:

Variable	Prueba	Valor- p	η^2	ω^2
\tilde{a}	ANOVA	0,009909	0,602616	0,503145
\bar{a}	ANOVA	0,019643	0,544338	0,433454

Tabla 4.13: Comparaciones intra-nivel dentro de la categoría *alta* para *L. albus* en exposición.

Las larvas de *L. albus* en exposición muestran una respuesta selectiva en la aceleración a , que se ve afectada en los límites inferior y superior del rango de salinidad. Dentro de la categoría *baja* (Tabla 4.12), se observaron diferencias significativas en la aceleración entre los subniveles 15, 25 y 30. De manera similar, en la categoría *alta*, la aceleración difiere entre los subniveles más extremos (40, 45 y 50). Este comportamiento puede interpretarse como una respuesta de evitación aguda ante el estrés salino, que se manifiesta con mayor fuerza en los umbrales de la tolerancia del organismo.

Como interpretación general de esta subsección, los resultados revelan que en muchos casos la categoría principal de salinidad oculta heterogeneidad interna que puede ser relevante tanto para la interpretación biológica y el modelamiento. En particular, a partir de lo expuesto se plantean dos decisiones metodológicas que se adoptarán en la modelación posterior:

1. Seleccionar $r(t)$ como observable principal para modelar la dinámica espacial de las larvas, dado que captura de forma directa la dispersión radial desde el origen y se relaciona estrechamente con la estructura del movimiento. Esta elección se fundamenta en los resultados obtenidos para el grupo CC-R, donde r se comportó como variable significativamente distinta entre condiciones salinas, por lo que el modelo considerará exclusivamente el conjunto de datos correspondiente a esta especie.
2. Implementar la modelación de forma individual a cada trayectoria larval, asignando parámetros específicos a cada ID sin agrupar previamente por salinidad. Esto permite tratar cada trayectoria como una unidad analítica independiente, conservando la variabilidad interna y la resolución temporal del movimiento. La salinidad se considerará como una variable contextual externa, utilizada solo en la fase de comparación de parámetros, para explorar posibles patrones vinculados a las condiciones experimentales. Aunque en este trabajo no se abordará la salinidad como variable continua por limitaciones de diseño y tamaño muestral, se plantea esta estrategia para investigaciones futuras que busquen caracterizar con mayor detalle las transiciones fisiológicas inducidas por gradientes salinos.

4.8. Análisis sobre estructuras funcionales

A modo de complemento al enfoque basado en medidas resumen, buscando compensar las limitaciones impuestas por su reducción de dimensionalidad inherente, se incorporó un enfoque de análisis de datos funcionales (FDA⁸) (Ramsay y Silverman (2005) [60]; Febrero-Bande y Oviedo de la Fuente (2012) [61]). Esta metodología permite representar las trayectorias de las variables estudiadas como funciones continuas en el tiempo, preservando tanto su forma como su dinámica. En contextos biológicos, donde las respuestas a estímulos pueden ser transitorias o no lineales, esta representación funcional resulta especialmente valiosa.

El FDA es una rama de la estadística que estudia datos cuya naturaleza subyacente es continua, es decir, que pueden considerarse como realizaciones de funciones suaves definidas sobre un dominio, generalmente temporal o espacial. A diferencia de los enfoques multivariantes tradicionales, que tratan observaciones como vectores discretos, el FDA modela la estructura funcional completa de los datos, preservando información sobre forma, derivadas y otras propiedades globales. Constituye una extensión natural de la estadística multivariada a espacios de dimensión infinita, donde cada observación es una función completa $x_i(t)$ definida en un dominio continuo $T \subset \mathbb{R}$ [60, 62].

⁸Functional Data Analysis

En este sentido, entendemos un dato funcional como una curva suave o función que representa una variable o unidad de interés en un conjunto de datos, y que puede combinarse con vectores o escalares [62]. Formalmente:

Definición (Datos funcionales). Un conjunto de datos funcionales consiste en una colección de N instancias

$$\mathcal{D} = \{(x_n(t), y_n); n = 1, \dots, N\} \quad (4.13)$$

donde $x_n(t)$, con $t \in \mathcal{T} \subset \mathbb{R}$, es la función de valores reales que caracteriza al n -ésimo ejemplo, y $y_n \in \{1, \dots, K\}$ es la etiqueta correspondiente a la clase. Típicamente, las funciones están disponibles como muestras de un proceso en una malla de M tiempos discretos $\mathbf{t} = (t_1, \dots, t_M)$ [63].

El objetivo del FDA es reconstruir la función subyacente que dio origen a estas observaciones y, a partir de allí, aplicar herramientas estadísticas adecuadas a este tipo de objetos.

4.8.1. Representación en bases de funciones

Para reconstruir funciones continuas a partir de observaciones discretas, se utiliza una expansión en bases funcionales, representando cada función $x(t)$ como combinación lineal de funciones base:

$$x(t) = \sum_{k=1}^K c_k \phi_k(t) + \epsilon(t) \quad (4.14)$$

donde $\{\phi_k(t)\}_{k=1}^K$ son funciones base conocidas, $\{c_k\}$ los coeficientes –a estimar– que definen la contribución de cada una y $\epsilon(t)$ es un término de error.

Elegir una base no es una decisión meramente técnica: cada tipo de base introduce propiedades matemáticas y supuestos sobre la señal. Aunque la suavidad no es obligatoria, es recomendable para evitar que el ruido, sobretudo al trabajar con datos experimentales, dominen la interpretación (Eilers y Marx, 1996) [64]. A continuación, se describen las bases más utilizadas y su relevancia en este estudio.

B-splines

Las B-splines ("*basic splines*") son funciones polinómicas por tramos, definidas de forma local en intervalos delimitados por puntos llamados nudos (*knots*). Su principal ventaja es la flexibilidad y el control local: aumentando el número de nudos, la función se adapta mejor a cambios rápidos; reduciéndolos, se obtiene un ajuste más suave. Al ser localmente definidas, los cambios en un tramo de la función afectan únicamente a un subconjunto de coeficientes, lo que mejora la estabilidad numérica [65]. Son ideales para representar trayectorias suaves con variabilidad localizada [66].

La i -ésima (de un total de K) función base B-spline de grado p , denotada $B_{i,p}(t)$, se define recursivamente sobre una secuencia no decreciente de nudos $\tau = \{t_0, t_1, \dots, t_{K+p+1}\}$. La definición de grado

0 es la siguiente:

$$B_{i,0}(t) = \begin{cases} 1 & \text{si } t_i \leq t < t_{i+1}, \\ 0 & \text{en otro caso.} \end{cases}$$

$B_{i,0}(t)$ es una función indicatriz en $[t_i, t_{i+1})$. Para grados superiores, la función se define como una combinación lineal de funciones de grado inferior:

$$B_{i,p}(t) = \frac{t - t_i}{t_{i+p} - t_i} B_{i,p-1}(t) + \frac{t_{i+p+1} - t}{t_{i+p+1} - t_{i+1}} B_{i+1,p-1}(t).$$

Esta fórmula recursiva se conoce como el algoritmo de Cox-De Boor (Cox, 1972; De Boor, 1972) [67, 68], y asegura la continuidad y suavidad de las funciones B-spline. Los coeficientes

$$\frac{t - t_i}{t_{i+p} - t_i} \text{ y } \frac{t_{i+p+1} - t}{t_{i+p+1} - t_{i+1}}$$

son factores de interpolación que aseguran que $B_{i,p}(t)$ sea no nula solo en el intervalo $[t_i, t_{i+p+1})$.

Esta base ofrece gran flexibilidad para aproximar formas funcionales, estabilidad numérica gracias a la ortonormalidad, y control local que permite modificar regiones específicas sin afectar el resto. Además, los B-splines brindan control sobre la suavidad y se adaptan bien a dominios irregulares. Sin embargo, requieren definir cuidadosamente la posición y número de nudos, lo que puede inducir sobreajuste si se emplean demasiados [67].

P-splines

Los P-splines ("*penalized B-splines*"), propuestos en 1996 por Eilers, P. y Marx, B. [64], son una extensión de los B-splines que añaden una penalización sobre los coeficientes para controlar la suavidad de la curva resultante, combinando una base de B-splines con un término de penalización que evita el sobreajuste [69]. Los P-splines simplifican el proceso de ajuste, ya que permiten usar un gran número de nudos equiespaciados, y la suavidad se controla de forma continua a través del parámetro de penalización, en lugar de depender de la cantidad de nudos [70].

El ajuste de un P-spline a un conjunto de datos se logra minimizando la función objetivo:

$$\text{RSS}(\mathbf{c}) = \sum_{j=1}^M \left[y_j - \sum_{k=1}^K c_k \phi_k(t_j) \right]^2 + \lambda \sum_{k=1}^{K-d} (\Delta^d c_k)^2. \quad (4.15)$$

En la expresión anterior, el primer término,

$$\sum_{j=1}^M \left(y_j - \sum_{k=1}^K c_k \phi_k(t_j) \right)^2,$$

es la suma de cuadrados de los residuos, que mide la discrepancia entre la curva ajustada y los datos

observados (t_j, y_j) . El segundo término,

$$\lambda \sum_{k=1}^{K-d} \left(\Delta^d c_k \right)^2,$$

es la penalización, que busca suavizar la curva. Esta penalización se basa en el operador de diferencias finitas Δ^d , que "castiga" las variaciones abruptas entre los coeficientes de los B-splines consecutivos.

El parámetro de suavizado λ controla el equilibrio entre el ajuste a los datos y la suavidad de la curva: un valor grande de λ impone una penalización fuerte, generando una curva más suave que ignora las pequeñas fluctuaciones en los datos; un λ pequeño reduce la penalización, permitiendo que la curva se ajuste más de cerca a los datos, lo que podría llevar al sobreajuste si hay ruido significativo.

Bases de Fourier

Las bases de Fourier representan señales como combinaciones de senos y cosenos. Son ideales para procesos periódicos o cuasiperiódicos, donde los componentes armónicos capturan bien la dinámica. Un ejemplo clásico es el análisis de señales de audio o vibraciones, donde los patrones repetitivos son de interés [71]. Además, su estructura ortogonal facilita la descomposición y reconstrucción de señales sin redundancia, lo que aporta estabilidad numérica y eficiencia computacional en contextos bien definidos [72].

Sin embargo, esta misma estructura implica una limitación importante: las funciones base de Fourier están definidas globalmente sobre todo el dominio, lo que reduce su capacidad para representar eventos localizados o transitorios. En consecuencia, su desempeño se ve comprometido en escenarios donde se requiere detectar discontinuidades, cambios abruptos o patrones no periódicos, como podría ser el caso al analizar trayectorias larvales con perturbaciones ambientales puntuales.

La k -ésima función base de Fourier, de período T , evaluada en el tiempo t se define como sigue:

$$\phi_k(t) = \begin{cases} 1 & \text{si } k = 0, \\ \sqrt{2} \cos\left(\frac{2\pi kt}{T}\right) & \text{si } k \text{ es impar,} \\ \sqrt{2} \sin\left(\frac{2\pi kt}{T}\right) & \text{si } k \text{ es par.} \end{cases}$$

En esta definición:

- $k = 0$ corresponde a la función constante $\phi_0(t) = 1$.
- Para k impar, $\phi_k(t)$ es una función coseno con un factor de escala $\sqrt{2}$.
- Para k par, $\phi_k(t)$ es una función seno con un factor de escala $\sqrt{2}$.

Wavelets

Las *wavelets* son funciones con una duración limitada y una amplitud que oscila. A diferencia de las bases de Fourier, que están definidas globalmente, las *wavelets* tienen una localización dual en tiempo y frecuencia, lo que las hace ideales para representar tanto componentes suaves y estables como eventos puntuales, abruptos o transitorios [73].

La familia de wavelets se genera a partir de una única función base $\psi(t)$, mediante operaciones de traslación (desplazamiento) y dilatación (escalado):

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right),$$

donde a es el factor de escala que controla la frecuencia (un a grande corresponde a baja frecuencia) y b es el factor de traslación que determina la posición temporal. La representación de una señal se logra evaluando su similitud con esta familia de funciones, lo que permite descomponerla en componentes locales de distinta resolución.

Las *wavelets* ofrecen una representación eficaz de señales con variabilidad localizada, combinando fases estables y transiciones abruptas, pero implica una mayor complejidad metodológica: la elección de la familia y parámetros puede influir en la interpretación, y su lectura resulta menos intuitiva que la de los splines, especialmente en análisis multiescala [74]. En este trabajo se aplicaron cuatro familias, todas con tres componentes: Daubechies (db4 y db8), caracterizadas por su buena capacidad de compresión y representación de regularidad [73]; Haar, la forma más simple, útil para detectar discontinuidades y cambios abruptos; y Coiflet (coif2), que asegura simultáneamente momentos nulos en la función y su transformada, lo que las hace apropiadas para el análisis de señales suaves con ciertas singularidades [75].

4.8.2. Regularización L_2 en P-splines

La regularización L_2 ⁹ consiste en añadir una penalización cuadrática sobre los parámetros del modelo con el fin de controlar su magnitud. Esta estrategia, ampliamente utilizada en estadística y aprendizaje automático [76, 77], atenúa la varianza de los estimadores y previene soluciones inestables al restringir la amplitud de los coeficientes.

Esta técnica resulta fundamental para controlar el sobreajuste en modelos de regresión y suavizado, al mantener un equilibrio entre flexibilidad y estabilidad sin sacrificar la capacidad de capturar la tendencia de los datos [64]. En el contexto de los P-splines, se implementa mediante un término de penalización que controla la suavidad de la curva ajustada, favoreciendo soluciones más estables [78].

Los coeficientes óptimos de las B-splines, \hat{c} , se obtienen minimizando una función de mínimos cuadrados penalizados:

$$\hat{c} = \arg \min_c \left(\|y - Bc\|_2^2 + \lambda c^\top P c \right). \quad (4.16)$$

La expresión anterior combina dos componentes fundamentales:

⁹También conocida como regresión Ridge o regularización de Tíjonov.

- **El término de ajuste** $\|y - Bc\|_2^2$ representa la suma de los errores al cuadrado, midiendo la discrepancia entre los datos observados (y) y la curva predicha (Bc). Su minimización sin restricciones podría resultar en una curva que se adapta excesivamente a las fluctuaciones del ruido, comprometiendo la generalización [64].
- **El término de penalización** $\lambda c^\top P c$ aplica una regularización L_2 sobre las diferencias finitas de los coeficientes de la base. La matriz de penalización P se construye a partir de un operador de diferencias (comúnmente de segundo orden), que actúa como una aproximación discreta de la segunda derivada de la curva. Penalizar este término disminuye los cambios bruscos y oscilaciones excesivas, resultando en una curva más suave [78].

El parámetro de suavizado λ regula el compromiso entre fidelidad y suavidad. Valores altos de λ generan curvas más lisas, al restringir la variabilidad de los coeficientes; valores bajos permiten mayor flexibilidad, acercándose a una interpolación directa de los datos [64].

En este trabajo, las P-splines se eligieron como base principal por su capacidad de regularización. Su implementación es eficiente y está bien soportada en librerías como `scikit-fda` en Python [79, 80], lo que las hace ideales para:

1. **Evitar el sobreajuste:** Al limitar la complejidad del modelo, se previene que la curva capture fluctuaciones artificiales en los datos. Esto es especialmente importante en entornos con ruido inherente, como los datos biológicos, donde la variabilidad puede ser alta [81].
2. **Mejorar la estabilidad numérica:** La penalización L_2 mitiga los problemas de un sistema mal condicionado que pueden surgir al utilizar una gran cantidad de funciones base, garantizando la eficiencia computacional y una solución única [82].

4.9. ANOVA funcional

El análisis de varianza para datos funcionales (fANOVA) extiende el ANOVA clásico a situaciones en que las observaciones corresponden a trayectorias continuas en el tiempo o en el espacio. En este caso, en lugar de comparar medias escalares, se contrasta la igualdad de las medias funcionales de distintos grupos. Este problema fue formalizado por Cuevas, Febrero y Fraiman (2004) [83], quienes propusieron un test basado en la norma L^2 de las diferencias entre funciones promedio de los grupos, constituyendo una versión asintótica del ANOVA de un factor.

Sea k el número de grupos y n_i el número de observaciones en el grupo i . Cada curva observada $X_{ij}(t)$ proviene de un proceso L^2 con media $m_i(t)$ y covarianza $K_i(s, t)$. El objetivo es contrastar:

$$H_0 : m_1(t) = m_2(t) = \dots = m_k(t), \quad \forall t \in [a, b]. \quad (4.17)$$

4.9.1. Estadístico de prueba

Un análogo funcional del estadístico de Fisher se construye mediante la variabilidad entre grupos, expresada a través de la norma L^2 . El estadístico propuesto por Cuevas et. al (2004) es:

$$V_n = \sum_{i < j} n_i \|\bar{X}_i - \bar{X}_j\|^2, \quad (4.18)$$

donde $\bar{X}_i(t)$ es la media funcional del grupo i , n_i el número de observaciones en el grupo i y $\|\cdot\|$ denota la norma L^2 :

$$\|f\| = \left(\int_a^b f(t)^2 dt \right)^{1/2}.$$

Este estadístico cuantifica la discrepancia global entre las curvas promedio de los grupos, en lugar de hacerlo punto a punto. Cuevas et al. demuestran en [83] que, bajo la hipótesis nula H_0 , la distribución asintótica de V_n se caracteriza mediante procesos gaussianos en espacios de Hilbert, lo que permite demostrar la consistencia del test.

A diferencia del ANOVA clásico, la distribución exacta de V_n bajo H_0 no es conocida en forma cerrada. Por ello, Cuevas et al. proponen un procedimiento de Monte Carlo asintótico: se simulan trayectorias gaussianas con covarianzas estimadas a partir de los datos y se generan replicaciones artificiales del estadístico límite. A partir de su distribución empírica se obtiene el valor crítico (cuantil) para un nivel de significancia α dado, o se calcula un valor- p aproximado. Se rechaza H_0 si el estadístico observado V_n excede este cuantil, lo cual es equivalente a tener un valor- p global menor que α [83].

Este procedimiento puede interpretarse como una forma de *bootstrap paramétrico asintótico*, válido bajo homocedasticidad y también en situaciones más generales, siempre que los tamaños muestrales sean suficientemente grandes [83].

4.9.2. Aplicación al conjunto de datos

En esta memoria se aplicó el test de Cuevas et al. para evaluar diferencias –bajo un nivel de significancia $\alpha = 5\%$ – en las trayectorias larvales entre macroniveles de salinidad, mediante la librería `scikit-fda` [79] de Python, que computa el estadístico global y su valor- p vía remuestreo/simulación bajo H_0 . Se realizaron 1000 permutaciones por cada combinación variable–grupo experimental. Cuando se detectaron diferencias significativas entre grupos, se aplicaron comparaciones múltiples *post hoc* con corrección de Bonferroni, con el fin de identificar qué niveles de salinidad contribuían a las discrepancias observadas.

Para la representación de las trayectorias funcionales se utilizaron distintas bases (P-splines, Fourier y *wavelets*), las cuales constituyen aproximaciones alternativas en L^2 y no alteran la validez del contraste, puesto que el test de Cuevas et al. (2004) se formula en términos de normas integrales. En la práctica, la elección de la base afecta la suavización y la capacidad de capturar estructuras locales, pero no modifica el fundamento teórico del ANOVA funcional.

4.10. Implementación de FDA

El análisis de los datos funcionales se llevó a cabo siguiendo un flujo de trabajo (*pipeline*) de tres etapas: la estandarización temporal de las trayectorias, la estimación funcional mediante P-splines y la optimización de sus hiperparámetros. La Figura 4.1 sintetiza el procedimiento completo.

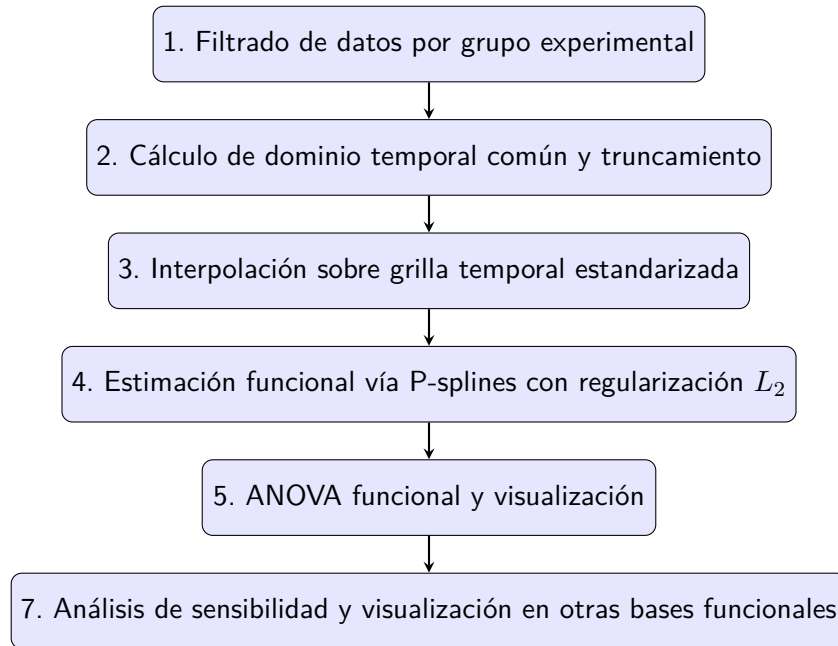


Figura 4.1: Flujo de trabajo para procesamiento de datos funcionales.

4.10.1. Estandarización temporal de trayectorias

El primer paso del análisis consistió en la estandarización de las trayectorias longitudinales para obtener un dominio temporal común y una grilla de evaluación uniforme. Para cada conjunto de trayectorias, se definió el dominio común $\mathcal{T}_c = [t_{\min}, t_{\max}]$ como el intervalo cerrado que abarca el tiempo máximo de inicio y el tiempo mínimo de finalización de todas las trayectorias:

$$t_{\min} = \max_i(t_{i,\text{inicio}}) \quad \text{y} \quad t_{\max} = \min_i(t_{i,\text{fin}}),$$

de esta manera, el análisis se limita estrictamente a períodos donde existen observaciones reales para todos los individuos, evitando extrapolaciones y garantizando una base de comparación homogénea.

Una vez definido el dominio común, se construyó una grilla de evaluación temporal con una resolución adaptativa. La cantidad de puntos en la grilla, $m_{\text{resolución}}$, se determinó en función de la densidad de observaciones más baja entre todas las trayectorias, acotada para prevenir el sobremuestreo o submuestreo:

$$m_{\text{resolución}} = \max \left(\min \left(\min_i \{n_{i,\text{dominio}}\}, 300 \right), 50 \right)$$

donde $n_{i,\text{dominio}}$ es el número de observaciones de la trayectoria i dentro del dominio común. Finalmente,

cada trayectoria observada fue recortada al intervalo \mathcal{T}_c e interpolada sobre esta grilla de puntos uniforme, creando una representación discreta pero estandarizada de las funciones. Las trayectorias mostraron una longitud de la serie de $\sim 225 - 235$, por lo que la pérdida de datos por truncamiento es mínima.

4.10.2. Representación mediante P-splines

El suavizado de las trayectorias estandarizadas se realizó usando P-splines cúbicas con regularización L_2 , siguiendo las recomendaciones de Eilers y Marx (1996) en [64]. Cada trayectoria suavizada, $\hat{f}_i(t)$, fue representada como una combinación lineal de K funciones B-spline:

$$\hat{f}_i(t) = \sum_{j=1}^K c_{ij} B_j(t), \quad t \in \mathcal{T}_c,$$

donde $B_j(t)$ son funciones base B-spline cúbicas, y c_{ij} los coeficientes estimados. La estimación de estos coeficientes se obtuvo minimizando la suma de cuadrados de los errores penalizada (PENSSE):

$$\text{PENSSE}(\mathbf{c}_i, \lambda) = \sum_{k:t_{ik} \in \mathcal{T}_c} \left(Y_i(t_{ik}) - \hat{f}_i(t_{ik}) \right)^2 + \lambda \|\mathbf{D}^{(2)} \mathbf{c}_i\|^2. \quad (4.19)$$

El primer término de la Ecuación 4.19 mide el ajuste a los datos, mientras que el segundo penaliza la rugosidad de la curva mediante las diferencias de segundo orden de los coeficientes. Esta formulación corresponde a una región penalizada cuya solución se obtiene resolviendo:

$$\hat{\mathbf{c}}_i = \arg \min_{\mathbf{c}} \|\mathbf{y}_i - \mathbf{Bc}\|^2 + \lambda \mathbf{c}^T \mathbf{Pc} \quad (4.20)$$

donde la matriz de penalización $\mathbf{P} = (\mathbf{D}^{(2)})^T \mathbf{D}^{(2)}$ se deriva del operador de diferencias discretas $\mathbf{D}^{(2)}$. El parámetro de suavizado λ controla el equilibrio entre la fidelidad del ajuste y la suavidad de la curva [70].

4.10.3. Optimización de hiperparámetros

La robustez del modelo depende críticamente de la selección de los hiperparámetros K (número de bases) y λ (suavizado). Para garantizar una implementación eficiente y consistente, se optó por fijar estos valores de forma manual, basándose en un análisis previo y en el buen rendimiento del modelo.

- **Número de bases (K):** Para asegurar que la complejidad del modelo fuera comparable entre grupos, se seleccionó un valor único y global para todo el conjunto de datos. Este valor fue optimizado mediante un análisis previo de validación cruzada y se fijó en $K = 25$.
- **Parámetro de suavizado (λ):** Aunque la práctica y la literatura sugieren optimizar λ de forma individual para cada grupo (ver [70, 82, 69]), en este trabajo se utilizó un valor fijo de $\lambda = 0,001$ para todas las trayectorias. Esta decisión responde a un equilibrio observado adecuado entre suavidad y ajuste a los datos, logrando curvas visualmente satisfactorias y estables, exceptuando,

claramente, los casos con demasiado ruido. Además, el uso de un valor constante entre grupos simplifica la implementación y asegura la comparabilidad de la suavidad entre las trayectorias, de manera que diferencias en la varianza de los datos de cada grupo no afecten la rigidez de la curva ajustada. Este enfoque es consistente con la filosofía de las P-splines de ofrecer una solución flexible pero controlada [70].

4.10.4. Análisis de sensibilidad

Para validar la robustez de las conclusiones y su independencia frente a los métodos empleados, se realizó un análisis de sensibilidad comparando los resultados del fANOVA sobre los P-splines con otras representaciones funcionales. El criterio de validación es el principio de consistencia inter-metodológica: si los resultados son coherentes a través de diferentes enfoques, las conclusiones son fiables, y por lo tanto no un artefacto del método de suavizado elegido. Las alternativas evaluadas fueron: datos sin suavizar, bases de Fourier y transformadas *wavelet*.

4.11. Resultados funcionales por grupo experimental

En esta sección se presentan los resultados del fANOVA para cada grupo experimental. Nos centramos principalmente en las variables que mostraron diferencias estadísticamente significativas entre los distintos niveles de salinidad, vale decir:

- **CC-R:** $r(t)$ y $\dot{r}(t)$.
- **LA-E:** $a(t)$, $\dot{r}(t)$, $\alpha(t)$ y $\theta_v(t)$.

También se evaluaron la velocidad angular ($\omega(t)$) y la posición radial ($r(t)$), con miras a su incorporación en un modelo formulado en coordenadas polares, a pesar de que sus indicadores resumen no indicaron diferencias significativas entre salinidades. Resulta relevante contrastar estos resultados con los obtenidos en la [Sección 4.6](#), lo que permite valorar si el enfoque funcional aporta sensibilidad adicional o revela partones que pasan desapercibidos para los métodos clásicos.

4.11.1. CC-R

Posición radial

Para las larvas de loco en recuperación, la posición radial fue el indicador con más diferencias entre macroniveles de salinidad. Esto se sostiene bajo un enfoque funcional, indicando de manera consistente que $r(t)$ varía entre condiciones salinas para este grupo experimental, independientemente de la base funcional utilizada en el fANOVA. La [Tabla 4.14](#) muestra los resultados de esta prueba para las bases consideradas, mientras que el ajuste por P-splines se muestra en la [Figura 4.2](#).

4.11. Resultados funcionales por grupo experimental

En las dos variables testeadas en este grupo experimental, la longitud del dominio temporal (original) discreto es 225, con volúmenes muestrales de 17, 15 y 26 observaciones en los macroniveles de salinidad baja, normal y alta, respectivamente.

Base	Valor- p	Estadístico V_n
P-spline	0,0125	73,0306
Ninguna	0,012	65,560
Fourier	0,012	65,0183
Daubechies (db4)	0,009	65,5558
Daubechies (db8)	0,009	65,5511
Haar	0,009	65,3763
Coiflet (coif2)	0,009	65,5623

Tabla 4.14: Resultados del fANOVA bajo diferentes bases funcionales; las últimas cuatro corresponden a familias de *wavelets*.

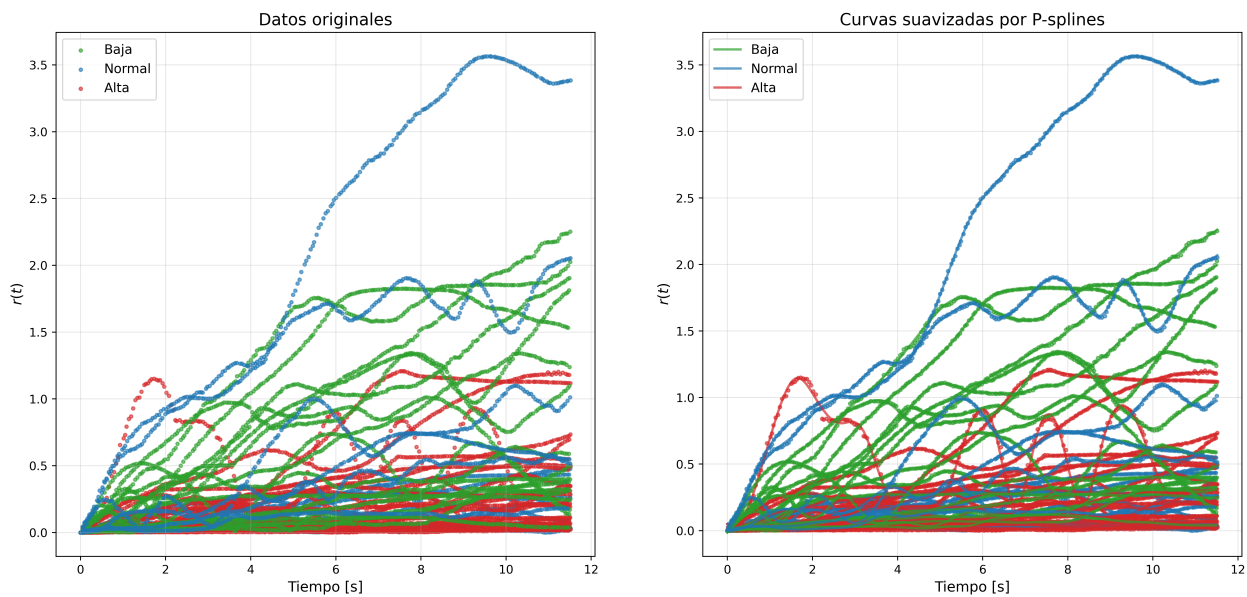


Figura 4.2: *Scatterplot* de los datos originales (izquierda) y curvas suavizadas (derecha, *scatterplot + P-splines*) para la variable $r(t)$ en el grupo CC-R.

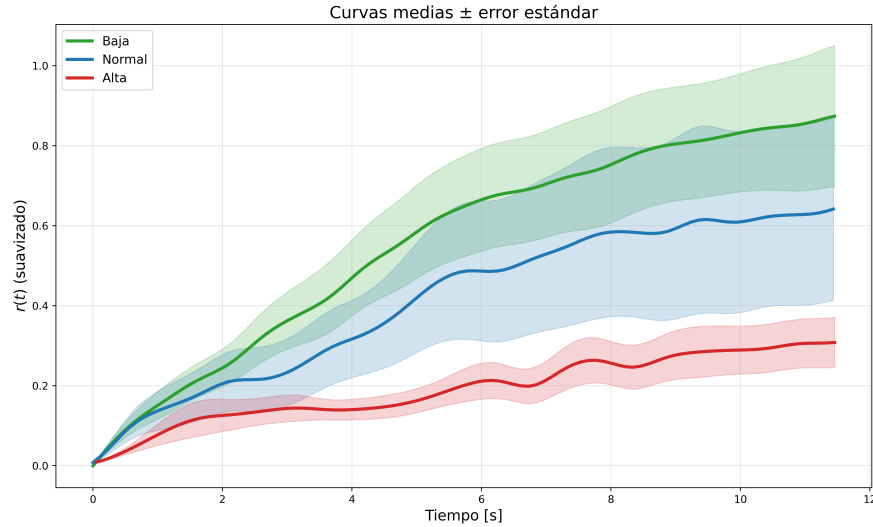


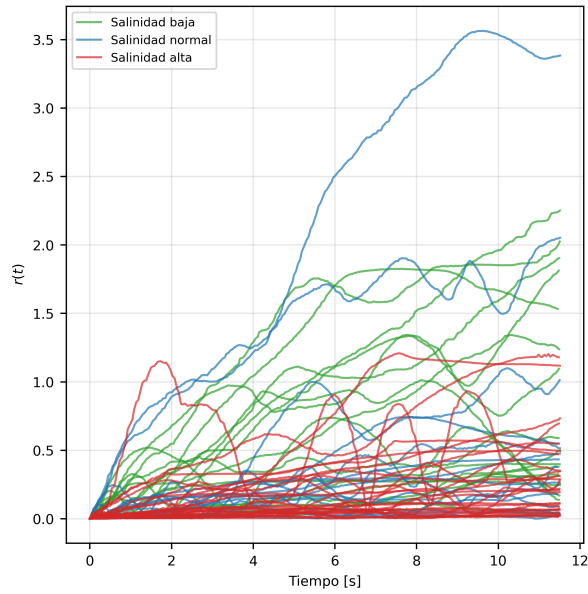
Figura 4.3: Curvas medias para la variable $r(t)$ en el grupo CC-R. Las bandas de error corresponden a los intervalos de confianza del 95 %.

El análisis *post hoc* con corrección de Bonferroni ($\alpha = 0,0167$) reveló una diferencia significativa en la posición radial $r(t)$ únicamente entre los grupos de salinidad alta y baja ($p = 0,0020$). Las comparaciones que involucraron el nivel de salinidad normal no alcanzaron significancia estadística: $p = 0,0840$ entre normal y alta; $p = 0,4560$ entre normal y baja. Esto se comprueba al observar la [Figura 4.3](#).

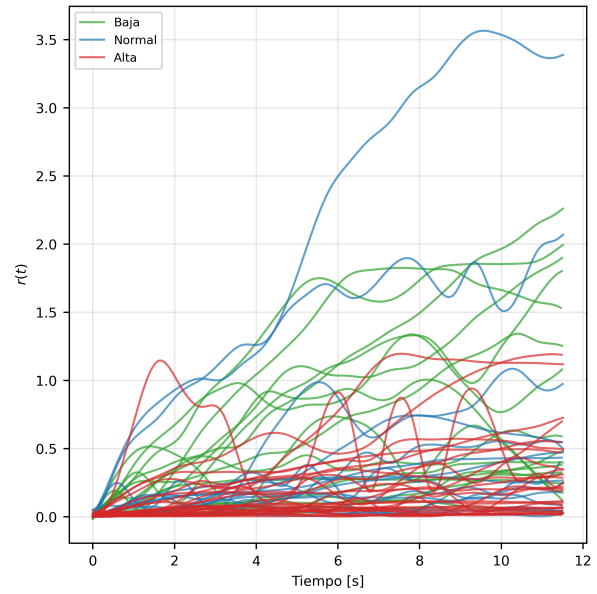
Si bien las pruebas *post hoc* se realizaron solo considerando la representación funcional principal, (P-splines) los resultados indican que el efecto de la salinidad sobre el comportamiento larval es más pronunciado en los extremos del gradiente evaluado. En particular, las trayectorias bajo salinidad normal se superponen parcialmente con las de ambos extremos, lo que dificulta detectar diferencias significativas en esas comparaciones, y se condice con las diferencias detectadas entre salinidades extremas en la [Sección 4.6](#).

Como parte del análisis de sensibilidad, la [Figura 4.4](#) muestra el ajuste de $r(t)$ usando las bases funcionales mencionadas anteriormente. En todos los casos se hallaron al menos dos grupos con medias significativamente diferentes, como indican los valores de la [Tabla 4.14](#).

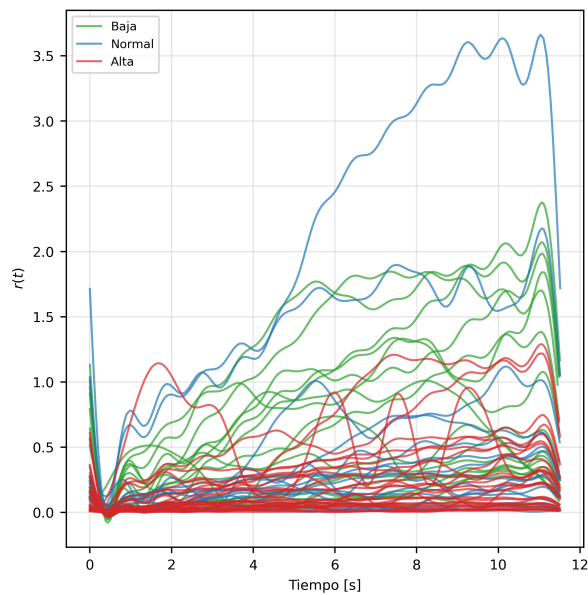
4.11. Resultados funcionales por grupo experimental



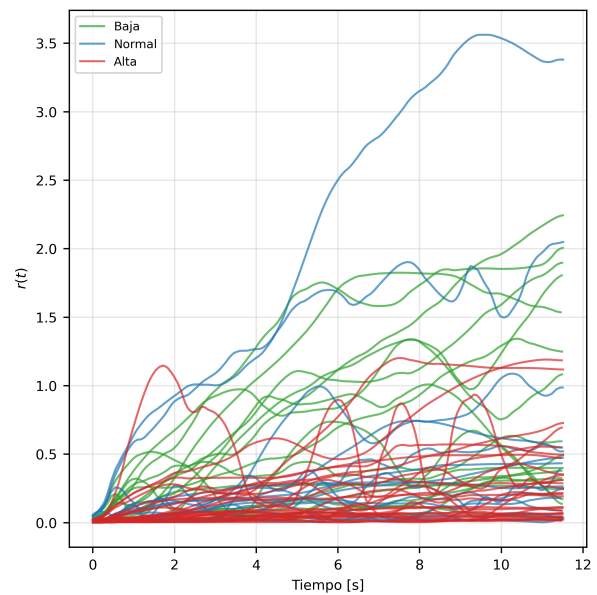
(a) Datos originales (interpolación lineal).



(b) P-splines.



(c) Fourier.



(d) Wavelet db4.

Figura 4.4: Trayectorias de $r(t)$ representadas en distintas bases funcionales, CC-R.

Velocidad radial

Por otro lado, el fANOVA para $\dot{r}(t)$ no mostró diferencias significativas entre los grupos de salinidad, sin importar la base funcional utilizada (véase la [Tabla 4.15](#)). Todos los valores- p resultaron ser altos, superando el umbral de significancia de $\alpha = 0,05$, mientras que los estadísticos V_n fueron bajos, lo que refuerza la conclusión de que no existen diferencias estadísticamente significativas en la tasa de cambio de la posición radial.

4.11. Resultados funcionales por grupo experimental

Este resultado es notable, ya que contrasta con el análisis por indicadores agregados de la [Sección 4.6](#), el cual sí había detectado diferencias significativas. Esta disparidad sugiere que, si bien la velocidad radial promedio de las larvas podría variar entre grupos, su patrón de cambio a lo largo del tiempo es muy similar.

Base	Valor- p	Estadístico V_n
P-spline	0,2365	2,9644
Ninguna	0,249	3,0012
Fourier	0,239	2,5716
Daubechies (db4)	0,22	2,6568
Daubechies (db8)	0,234	2,6469
Haar	0,216	2,5574
Coiflet (coif2)	0,218	2,65

Tabla 4.15: Resultados del fANOVA bajo diferentes bases funcionales; las últimas cuatro corresponden a familias de *wavelets*.

Como parte del análisis de sensibilidad, la [Figura 4.7](#) muestra los ajustes de las curvas de $\dot{r}(t)$ utilizando diferentes bases funcionales. En general, se observa que el suavizado de las curvas no es óptimo, ya que los datos originales presentan una alta variabilidad y las trayectorias no siguen un patrón coherente, lo que dificulta la detección de diferencias. Esto es particularmente evidente al comparar las curvas suavizadas con el *scatterplot* de los datos originales en la [Figura 4.5](#), donde la alta dispersión de los puntos sugiere que las larvas mantienen un comportamiento errático y que la tasa de cambio de su posición, o bien no es consistente a lo largo del tiempo, o es prácticamente nula. Las curvas medias, presentadas en la [Figura 4.6](#), confirman esta falta de patrón, mostrando una trayectoria fluctuante y superpuesta entre los grupos, sin un comportamiento claro que las diferencie.

4.11. Resultados funcionales por grupo experimental

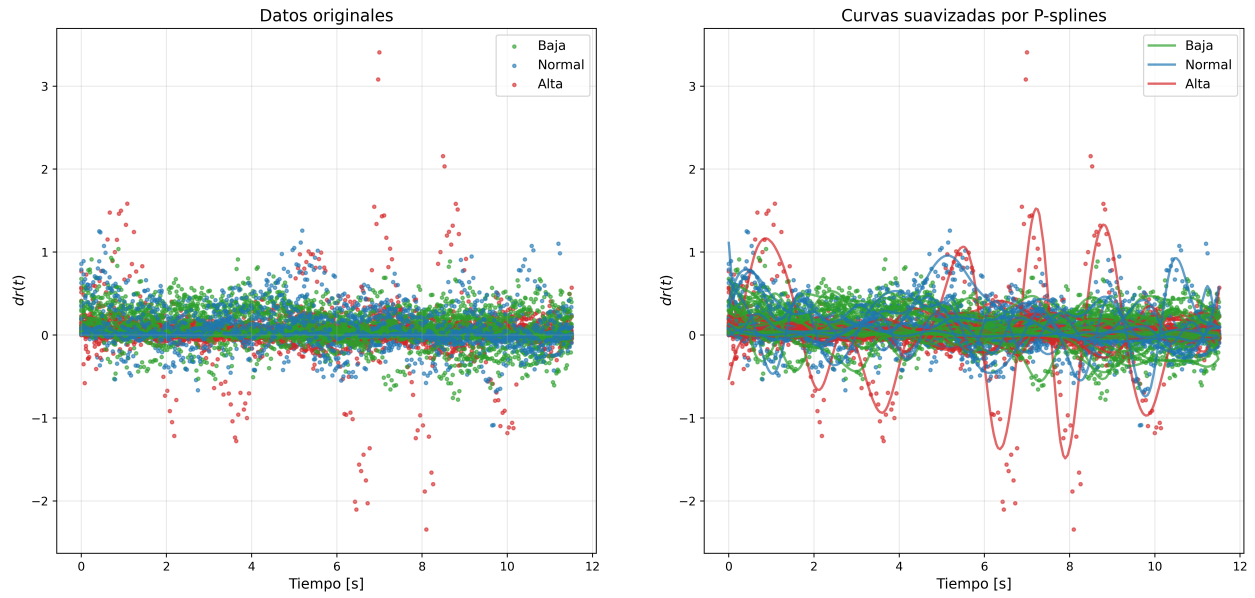


Figura 4.5: *Scatterplot* de los datos originales (izquierda) y curvas suavizadas (derecha, *scatterplot* + *P-splines*) para la variable $\dot{r}(t)$ en el grupo CC-R.

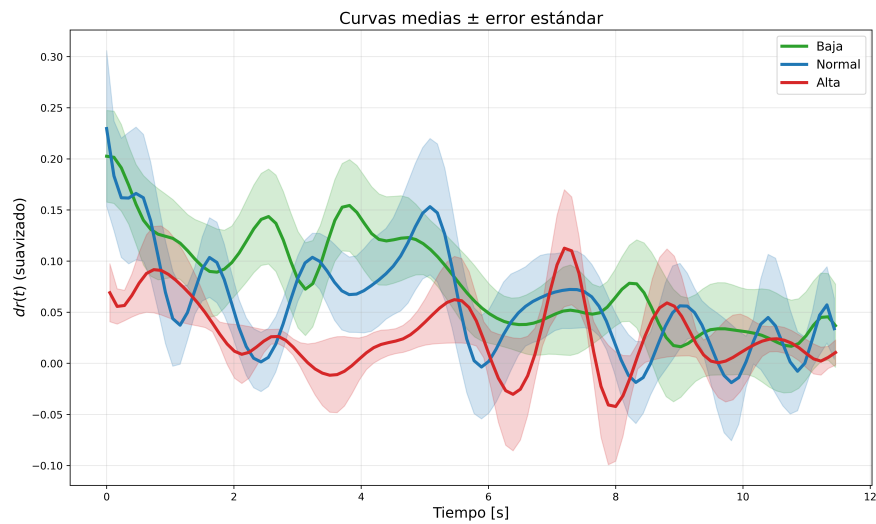


Figura 4.6: Curvas medias para la variable $\dot{r}(t)$ en el grupo CC-R. Las bandas de error corresponden a los intervalos de confianza del 95 %.

4.11. Resultados funcionales por grupo experimental

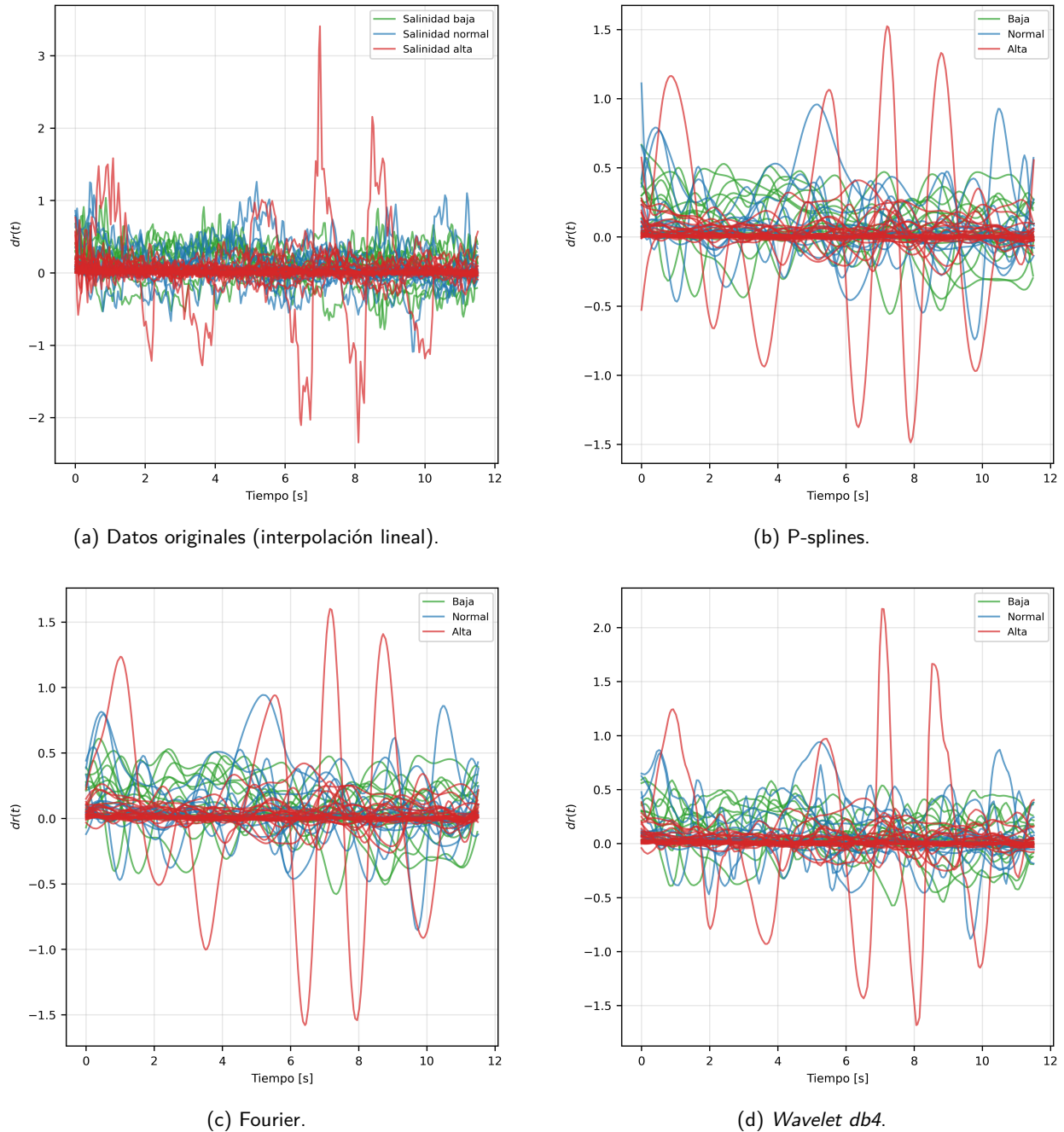


Figura 4.7: Trayectorias de $\dot{r}(t)$ representadas en distintas bases funcionales, CC-R.

4.11.2. LA-E

En las larvas de erizo rojo en exposición el catálogo de variables sensibles a la salinidad es, en principio, más amplio, presentando también en su mayoría variables angulares, pero también una prominente diferencia en la aceleración lineal. La longitud del dominio temporal (original) discreto en este grupo experimental es 227, con volúmenes muestrales de 22, 10 y 13 observaciones en los macroniveles de salinidad baja, normal y alta, respectivamente.

Aceleración lineal

Los resultados del fANOVA para $a(t)$ se muestran en la [Tabla 4.16](#). Se observa consistencia entre valores- p significativos en todas las bases funcionales, lo que refuerza la sensibilidad de esta variable frente a diferencias salinas del entorno. El análisis *post hoc* revela que la diferencia detectada ocurre entre los macroniveles normal y baja ($p = 0,008$). En los pares normal-alta ($p = 0,238$) y baja-alta ($p = 0,112$) no se detectaron diferencias significativas.

Base	Valor- p	Estadístico V_n
P-spline	0,0115	117,6782
Ninguna	0,0230	76,0577
Fourier	0,0270	63,4780
Daubechies (db4)	0,0210	66,6854
Daubechies (db8)	0,0150	71,2908
Haar	0,0180	69,5155
Coiflet (coif2)	0,0200	66,6503

Tabla 4.16: Resultados del fANOVA para $a(t)$ bajo diferentes bases funcionales en el grupo LA-E; las últimas cuatro corresponden a familias de *wavelets*.

En la [Figura 4.8](#) se observa la concentración de valores de $a(t)$ cerca de 0 en todo el dominio, junto con aparentes oscilaciones esporádicas principalmente en las salinidades baja y normal. Estos cambios en la magnitud de la aceleración pueden estar contribuyendo a la diferencia de medias, visualizada en la [Figura 4.9](#). Esta última muestra también que pareciera haber una estabilización, o más técnicamente, convergencia hacia un estado estacionario, en el "largo" plazo (recordemos que el horizonte temporal no pasa los 12 segundos). Las diferencias son más acentuadas en los primeros segundos, donde es probable que el test las detecte dada la intersección vacía de las bandas de confianza de las medias de salinidades normal y baja. A partir de este gráfico se puede concluir que, en los tres escenarios, los individuos de este grupo tienden a desplazarse con una rapidez baja y relativamente constante en el largo plazo.

En cuanto a la calidad del ajuste de las curvas, la consistencia entre los resultados indica un nivel suficiente para los fines aquí perseguidos, aun con alta presencia de ruido y oscilaciones agudas ([Figura 4.10](#)). Es particularmente notorio el comportamiento heterogéneo entre las trayectorias de salinidad baja (en verde) y alta (en rojo), observación que se corresponde con lo presentado en las [Tablas 4.12](#) y [4.13](#): la visualización de los datos funcionales confirma que solo la salinidad normal presenta curvas homogéneas para $a(t)$. La [Figura 4.10](#) revela también que, en este caso, la base de Fourier parece capturar mejor el comportamiento de las trayectorias que la P-spline, sin comprometer la suavidad de las curvas (aquí falla la *wavelet* aunque parece ajustarse mínimamente mejor que Fourier). Esto remarca las cualidades de las distintas bases y el rol del comportamiento intrínseco de la data a la hora de representar en bases funcionales.

4.11. Resultados funcionales por grupo experimental

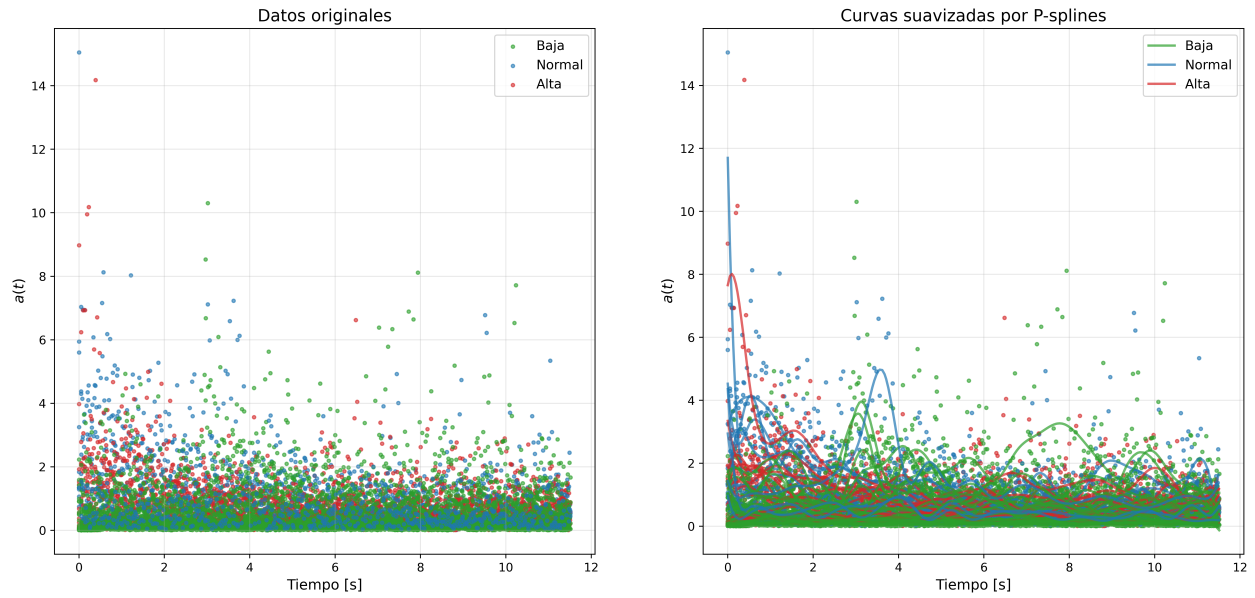


Figura 4.8: *Scatterplot* de los datos originales (izquierda) y curvas suavizadas (derecha, *scatterplot + P-splines*) para la variable $a(t)$ en el grupo LA-E.

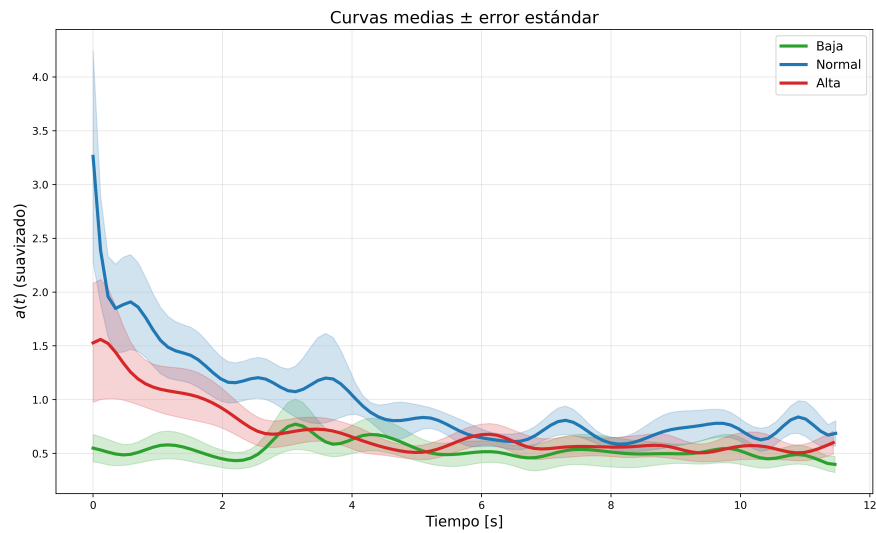


Figura 4.9: Curvas medias para la variable $a(t)$ en el grupo LA-E. Las bandas de error corresponden a los intervalos de confianza del 95 %.

4.11. Resultados funcionales por grupo experimental

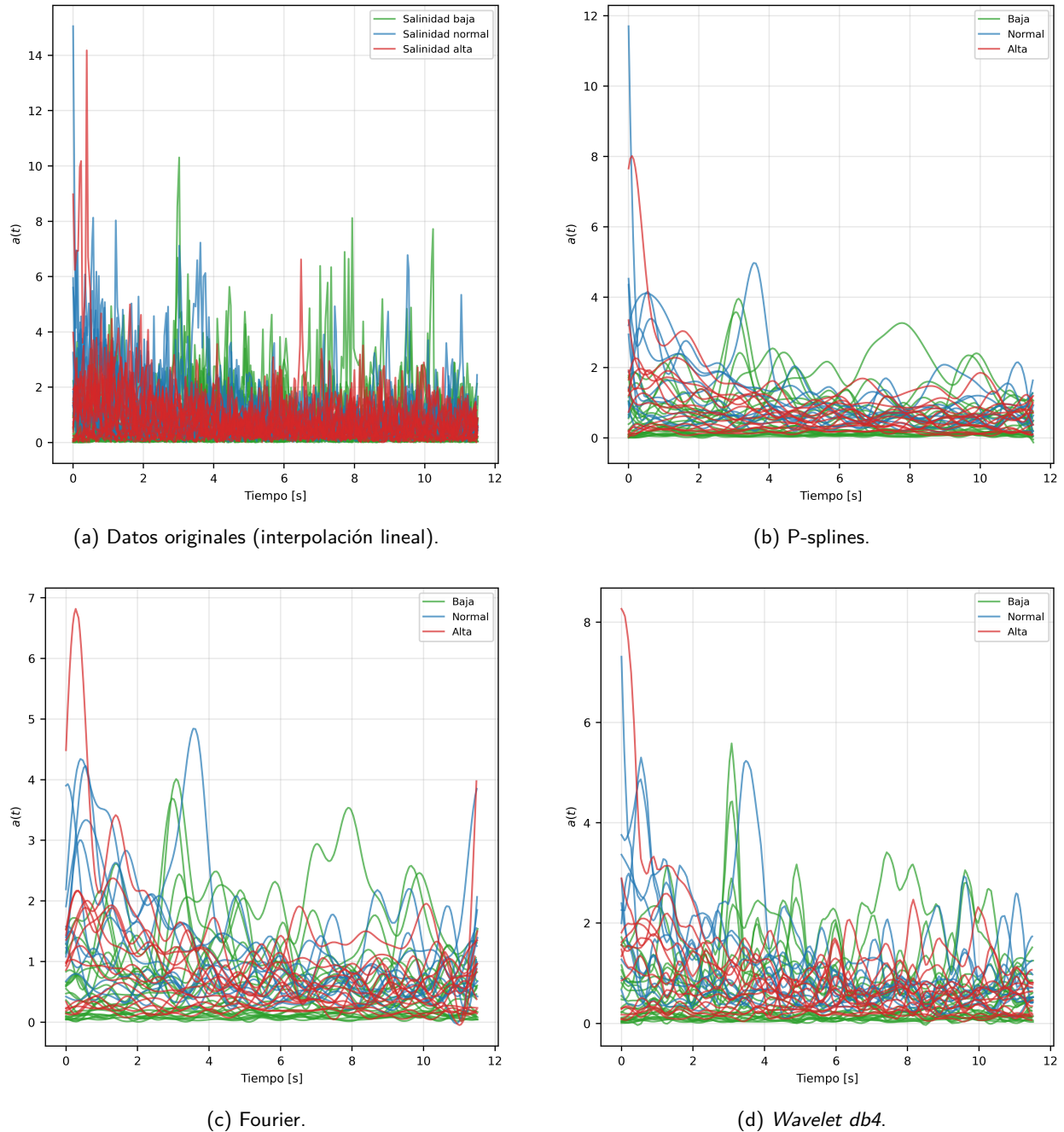


Figura 4.10: Trayectorias de $a(t)$ representadas en distintas bases funcionales, LA-E.

Velocidad radial

La velocidad radial $\dot{r}(t)$ es la única métrica en común entre los grupos de esta Sección. Sin embargo, en contraste con CC-R, en este caso el enfoque funcional refuta las diferencias entre macroniveles de salinidad detectadas mediante indicadores resumen.

4.11. Resultados funcionales por grupo experimental

Base	Valor- p	Estadístico V_n
P-spline	0,3975	1,2636
Ninguna	0,4460	0,9211
Fourier	0,4840	0,8060
Daubechies (db4)	0,4340	0,8457
Daubechies (db8)	0,4300	0,8488
Haar	0,4310	0,8366
Coiflet (coif2)	0,4320	0,8467

Tabla 4.17: Resultados del fANOVA para $\dot{r}(t)$ bajo diferentes bases funcionales en el grupo LA-E; las últimas cuatro corresponden a familias de *wavelets*.

La [Tabla 4.17](#) muestra valores- p no significativos para un nivel de significancia del 5% de manera transversal entre bases funcionales, mientras que la [Figura 4.11](#) respalda un efectivo comportamiento homogéneo –no se detectaron diferencias intra-niveles en la [Sección 4.7](#)– entre trayectorias de la misma salinidad a pesar de la alta presencia de ruido, sobre todo en salinidad baja. Por su parte, la [Figura 4.12](#) muestra que, controlando el ruido, las diferencias de medias funcionales entre condiciones experimentales no son detectables por este test.

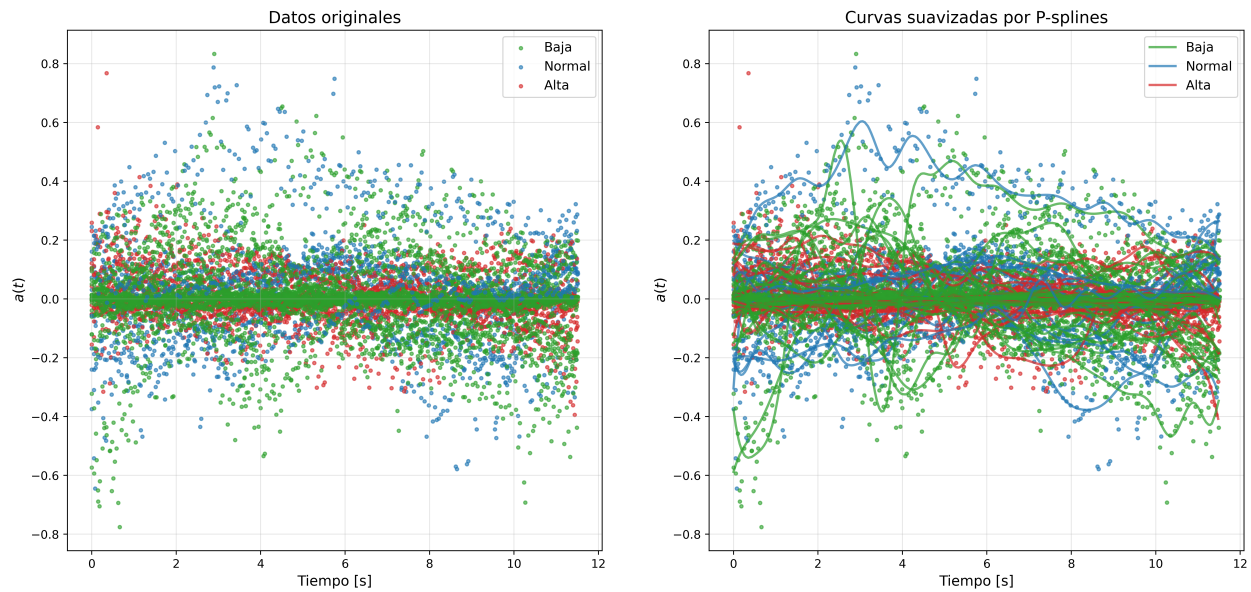


Figura 4.11: *Scatterplot* de los datos originales (izquierda) y curvas suavizadas (derecha, *scatterplot + P-splines*) para la variable $\dot{r}(t)$ en el grupo LA-E.

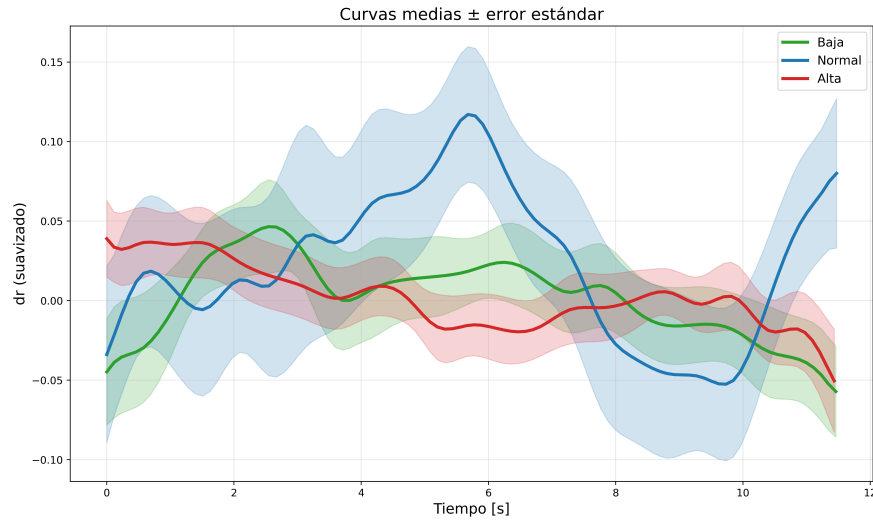


Figura 4.12: Curvas medias para la variable $\dot{r}(t)$ en el grupo LA-E. Las bandas de error corresponden a los intervalos de confianza del 95 %.

La Figura 4.13 muestra la semejanza de los valores promedio que alcanzan las trayectorias, en las tres bases funcionales adicionales consideradas, más la interpolación lineal. No obstante, las diferencias en la variabilidad y forma de cada grupo son considerables, observándose rangos de amplitud bien definidos, particularmente en la salinidad alta; si bien las bandas de confianza de la Figura 4.12 logran capturar parcialmente este comportamiento, se puede considerar este caso como un ejemplo de donde incluso el fANOVA, teniendo en cuenta la evolución temporal del proceso, puede verse limitado en la detección de diferencias de forma de las trayectorias. En este sentido, basar el criterio solo en la media funcional puede no ser lo más adecuado, especialmente cuando cuando las curvas de diferentes grupos tienen una media similar, pero su variabilidad o forma es distinta.

Existen en la literatura técnicas para profundizar en el análisis de forma y comparación de datos funcionales, ofreciendo caminos interesantes hacia un análisis más integral, pero que desviarían el flujo de esta memoria de ser implementados aquí. En estudios futuros, se podrían incorporar técnicas de comparación funcional más robustas como pruebas basadas en distancias L^2 , tests globalizados punto a punto (GPF) o enfoques de permutación, que permiten evaluar las diferencias en la forma completa de las trayectorias. Estos métodos son especialmente útiles cuando las curvas de distintos grupos oscilan en torno a un valor promedio similar, una situación en la que el ANOVA funcional clásico tiende a no detectar diferencias significativas en la media funcional. Su aplicación ha sido documentada en la literatura reciente, destacando el trabajo de Górecki & Smaga (2015) [84] y el de Zhang & Zhang (2014) [85]. El primero compara múltiples tests funcionales bajo distintos supuestos de homocedasticidad y normalidad, mientras que el segundo propone tests globales para datos funcionales con control del error de tipo I.

4.11. Resultados funcionales por grupo experimental

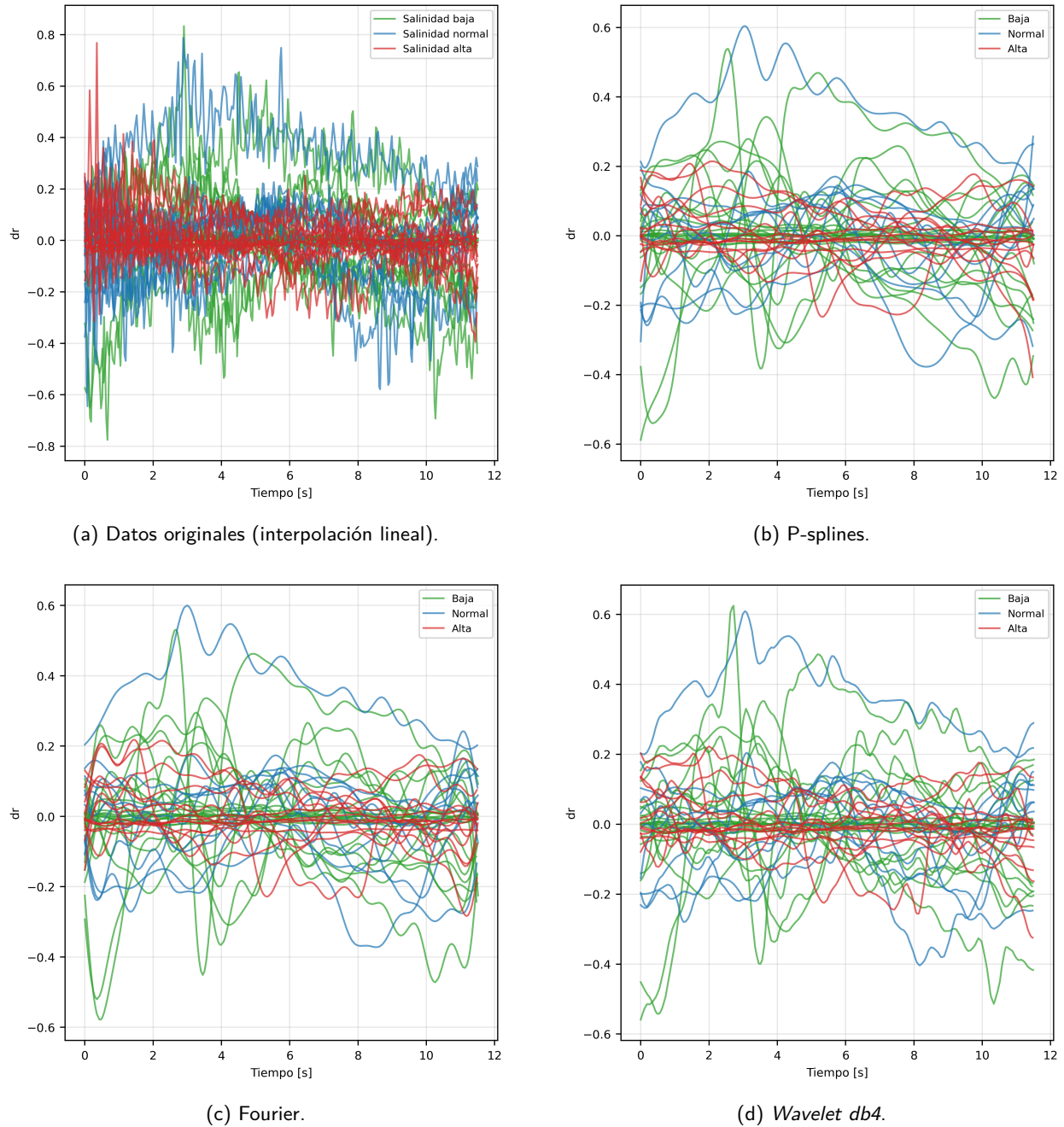


Figura 4.13: Trayectorias de $\dot{r}(t)$ representadas en distintas bases funcionales, LA-E.

Aceleración angular

La aceleración angular $\alpha(t)$ tampoco mostró diferencias significativas en su forma funcional, presentando unos estadísticos V_n particularmente altos en todas las bases (ver [Tabla 4.18](#) y [Figura 4.15](#)). Este comportamiento se explica por la escasa variabilidad intra e intergrupo en curvas prácticamente planas y cercanas a cero, donde pequeñas fluctuaciones relativas elevan el valor absoluto de V_n pese a no generar evidencia estadística suficiente frente a la distribución nula obtenida por permutación. La [Figura 4.14](#),

4.11. Resultados funcionales por grupo experimental

junto con el resultado basado en la *mediana* (y no la *media*) de α en la [Sección 4.6](#), sugieren que estas diferencias puntuales podrían explicarse por un retardo en la estabilización de las larvas en salinidad baja frente a las de salinidad alta, confirmando así lo indicado por la prueba *post hoc* clásica.

Base	Valor- p	Estadístico V_n
Ninguna	0,3880	97387,4011
Fourier	0,4760	17053,7656
P-spline	0,3005	75051,9467
Daubechies (db4)	0,4470	23429,6526
Daubechies (db8)	0,3260	59849,4008
Haar	0,3720	36032,3157
Coiflet (coif2)	0,4220	24895,1269

Tabla 4.18: Resultados del fANOVA para $\alpha(t)$ bajo diferentes bases funcionales en el grupo LA-E; las últimas cuatro corresponden a familias de *wavelets*.

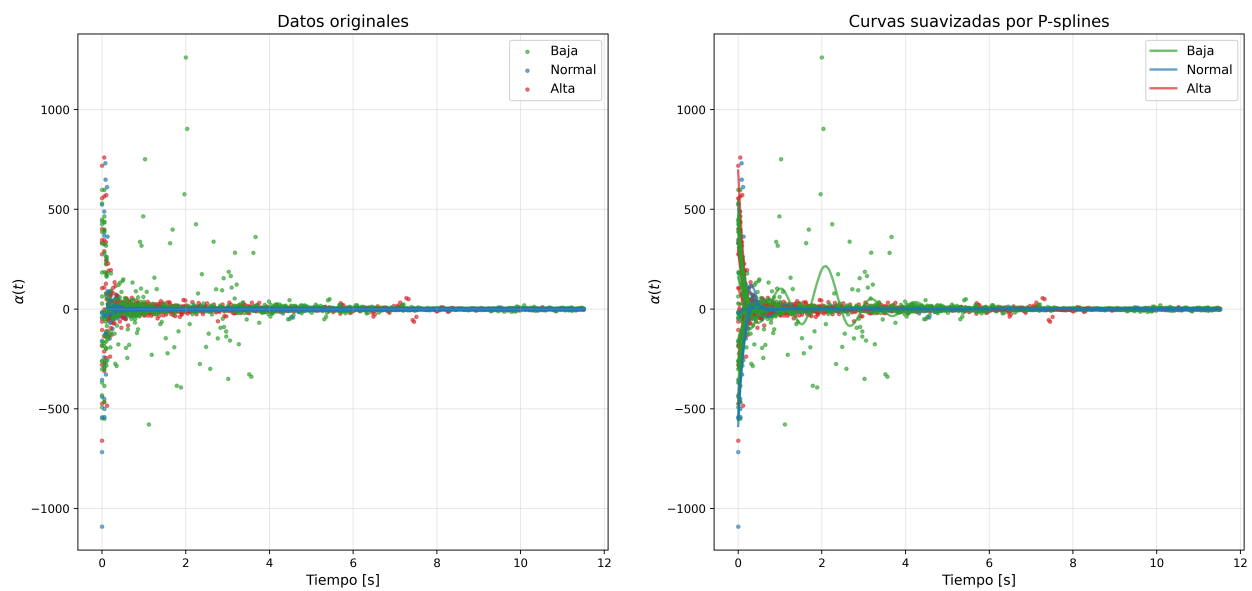


Figura 4.14: *Scatterplot* de los datos originales (izquierda) y curvas suavizadas (derecha, *scatterplot* + *P-splines*) para la variable $\alpha(t)$ en el grupo LA-E.

4.11. Resultados funcionales por grupo experimental

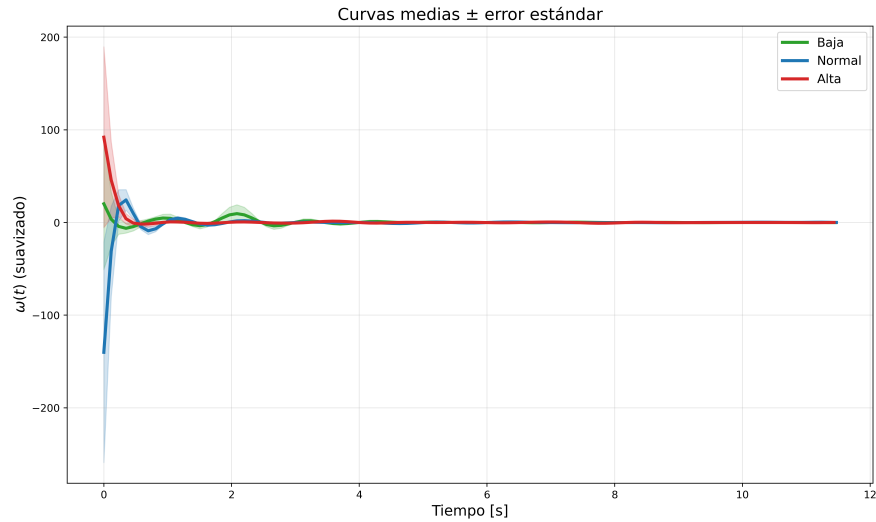


Figura 4.15: Curvas medias para la variable $\alpha(t)$ en el grupo LA-E. Las bandas de error corresponden a los intervalos de confianza del 95 %.

La Figura 4.16 muestra que el ajuste de la *wavelet* captura adecuadamente las oscilaciones, pero nuevamente a costa de una suavidad visible en los datos; Fourier, en cambio, genera curvas suaves que replican las fluctuaciones iniciales, pero introduce extrapolaciones artificiales en el resto del dominio. Todas fueron consistentes en la homogeneidad de las trayectorias medias entre salinidades.

4.11. Resultados funcionales por grupo experimental

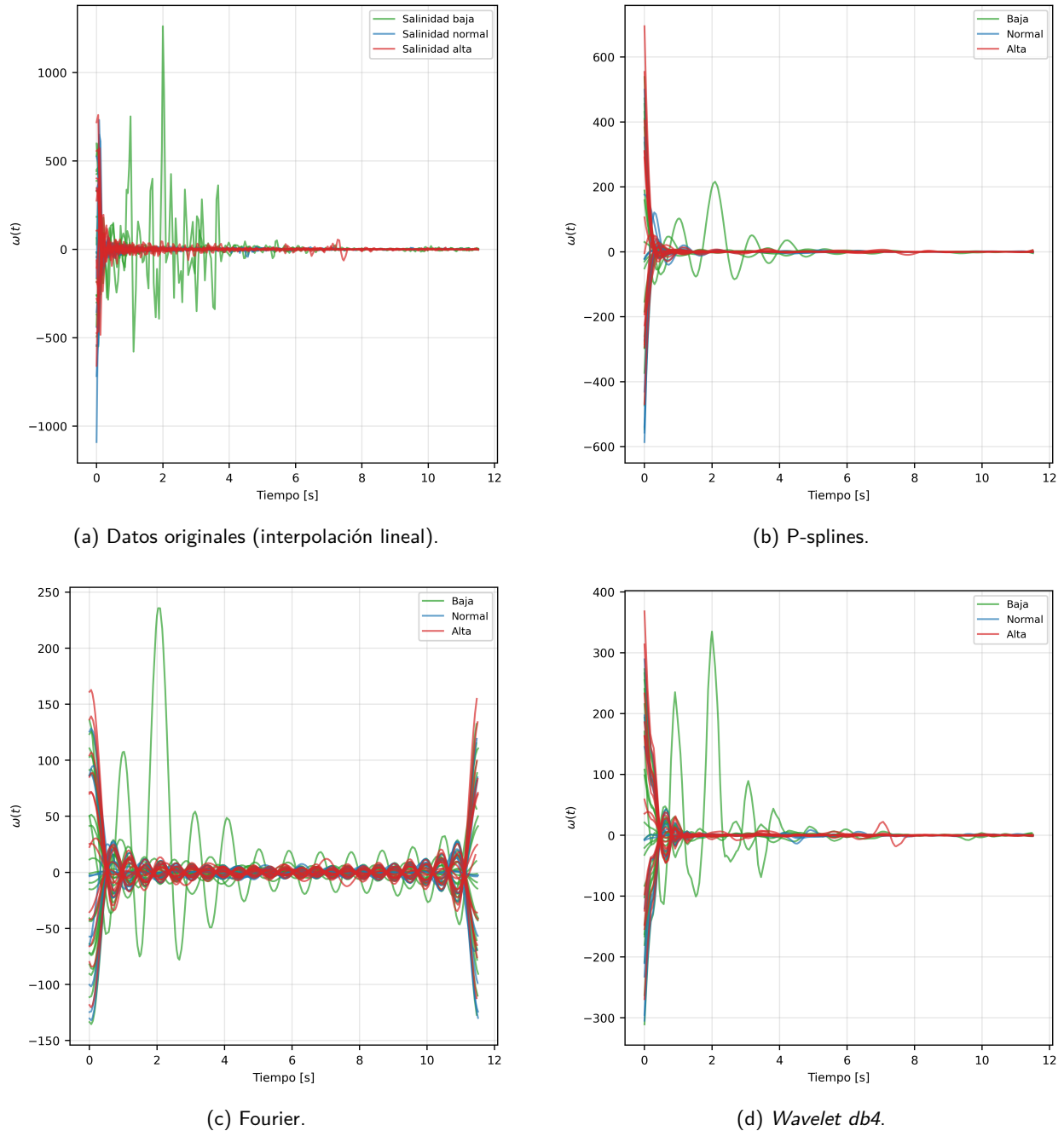


Figura 4.16: Trayectorias de $\alpha(t)$ representadas en distintas bases funcionales, LA-E.

Ángulo de la velocidad

Por último, el ángulo de la velocidad $\theta_v(t)$ tampoco presentó diferencias entre salinidades en su forma funcional ($p > 0,21$ en todos los casos; ver [Tabla 4.19](#) y [Figura 4.18](#)). Los estadísticos V_n muestran magnitudes relativamente altas ($\approx 150 - 193$) pero consistentes entre bases.

Base	Valor- p	Estadístico V_n
Ninguna	0,3100	192,9971
Fourier	0,2660	149,0424
P-spline	0,3835	185,9544
Daubechies (db4)	0,2220	161,0445
Daubechies (db8)	0,2250	161,0165
Haar	0,2220	158,0320
Coiflet (coif2)	0,2190	161,1279

Tabla 4.19: Resultados del fANOVA para $\theta_v(t)$ bajo diferentes bases funcionales en el grupo LA-E; las últimas cuatro corresponden a familias de *wavelets*.

La alta variabilidad y ruido en este conjunto de datos sugiere que las diferencias detectadas previamente sobre el promedio pueden deberse a fluctuaciones puntuales y no a un patrón sistemático entre grupos. En este sentido, y considerando el *scatterplot* de los datos originales de la Figura 4.17, la representación en bases funcionales presenta una escasa utilidad e interpretabilidad, puesto que el exceso de dispersión impide capturar información relevante y podría incluso distorsionar la lectura de la estructura real de los datos.

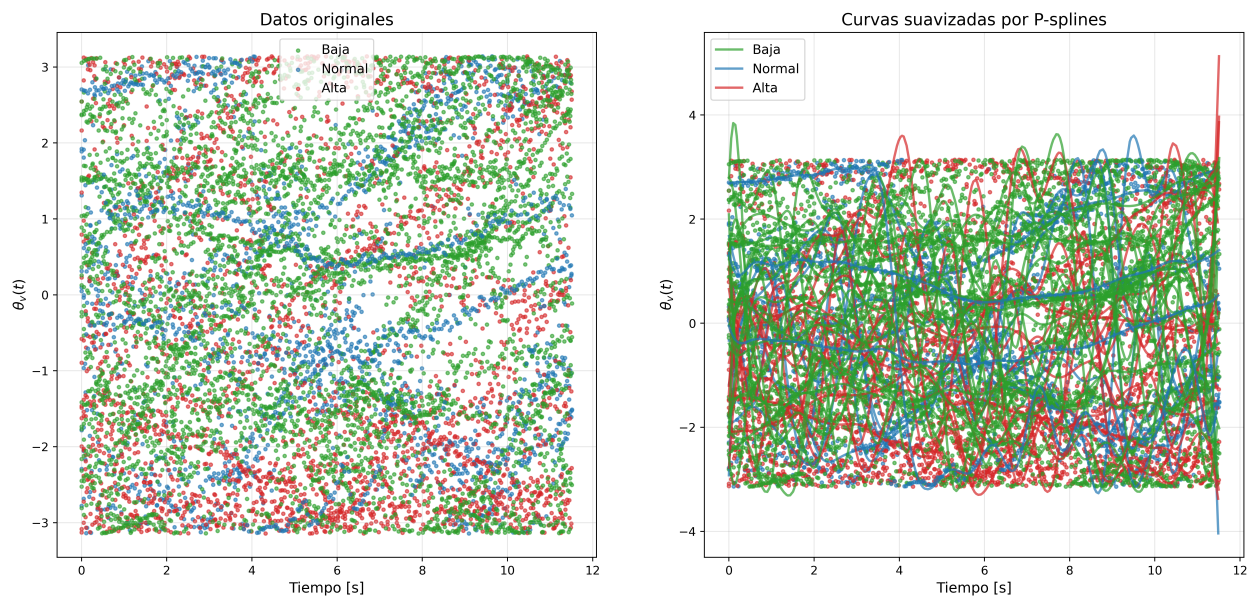


Figura 4.17: *Scatterplot* de los datos originales (izquierda) y curvas suavizadas (derecha, *scatterplot* + *P-splines*) para la variable $\theta_v(t)$ en el grupo LA-E.

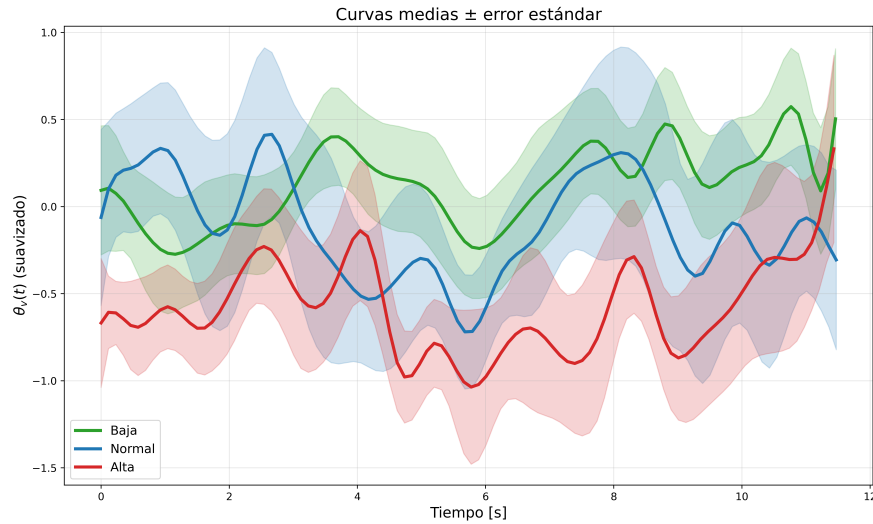


Figura 4.18: Curvas medias para la variable $\theta_v(t)$ en el grupo LA-E. Las bandas de error corresponden a los intervalos de confianza del 95 %.

Los resultados funcionales para ambos grupos experimentales contrastan con la significancia de las pruebas clásicas en, al menos, el 50 % de las variables testeadas. Esto sugiere que parte de los efectos detectados previamente podrían deberse a fluctuaciones locales o artefactos numéricos en lugar de patrones temporales sostenidos en el tiempo.

4.12. Variables angulares

Para cerrar este capítulo, se incluye un análisis funcional de aquellas variables que, aunque no mostraron una significancia clásica, son esenciales para el proceso de modelación estocástica. Esta sección final busca establecer los vínculos directos con el modelamiento que se presentará en el [Capítulo 5](#).

Como se verá más adelante, los modelos propuestos se sustentan en las métricas angulares de $r(t)$, $\omega(t)$ e, indirectamente, en $\theta_r(t)$. Al trabajar con datos longitudinales, el enfoque funcional se considera metodológicamente más robusto que el clásico, a pesar de sus propias limitaciones. Por ello, la evidencia sobre el efecto de los niveles de salinidad obtenida con este enfoque funcional se convierte en un insumo fundamental para evaluar el rendimiento de los modelos y, en particular, de los parámetros estimados en la siguiente etapa.

4.12.1. CC-R

Para las larvas de loco en recuperación, se cuenta ya con evidencia de significancia estadística para la posición radial $r(t)$, tanto bajo pruebas clásicas como en su forma funcional. De esta manera, la evaluación avanza hacia la velocidad angular $\omega(t)$.

La tabla 4.20 resume los valores obtenidos para V_n y sus respectivos valores- p . La representación en bases en ningún caso supera el umbral de significancia $\alpha = 0,05$, lo que proporciona evidencia sólida para asumir homogeneidad en las trayectorias de velocidades angulares de este grupo experimental. Esta se

comprueba en la Figura 4.19, que muestra un comportamiento similar a $\alpha(t)$ en LA-E (Figura 4.14), con valores altamente concentrados alrededor de 0 y presencia de fluctuaciones esporádicas, aunque en este caso con mayor presencia en salinidades normal y alta, en comparación con la salinidad baja que mostró este comportamiento en LA-E- $\alpha(t)$. De la misma manera, las curvas medias, con intervalos de confianza notoriamente estrechos, revelan la baja variabilidad en las curvas y una estabilización prácticamente instantánea, considerando además que probablemente los valores iniciales no nulos se deban a factores experimentales (*trackeo*, momento de grabación, cálculos iniciales) en lugar de un comportamiento diferenciado en los primeros instantes temporales.

Base	Valor- p	Estadístico V_n
P-spline	0,092	295,2044
Ninguna	0,581	441,583
Fourier	0,463	171,156
Daubechies (db4)	0,524	197,6482
Daubechies (db8)	0,706	237,11
Haar	0,553	218,7549
Coiflet (coif2)	0,563	185,7562

Tabla 4.20: Resultados del fANOVA para $\omega(t)$ bajo diferentes bases funcionales en el grupo CC-R; las últimas cuatro corresponden a familias de *wavelets*.

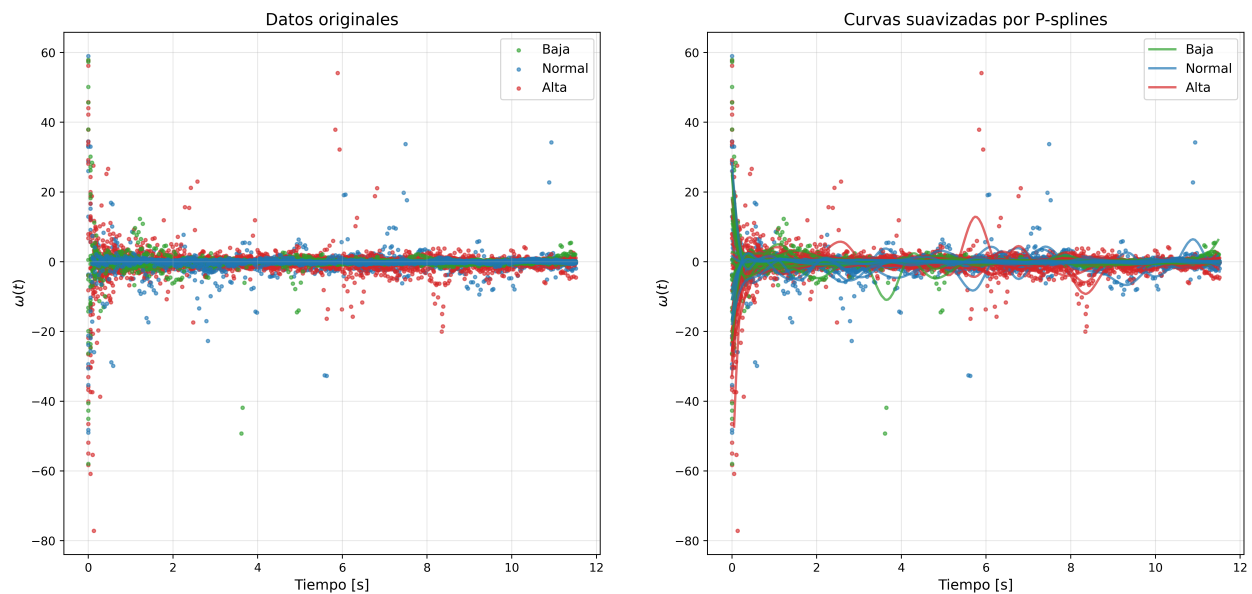


Figura 4.19: *Scatterplot* de los datos originales (izquierda) y curvas suavizadas (derecha, *scatterplot* + *P-splines*) para la variable $\omega(t)$ en el grupo CC-R.

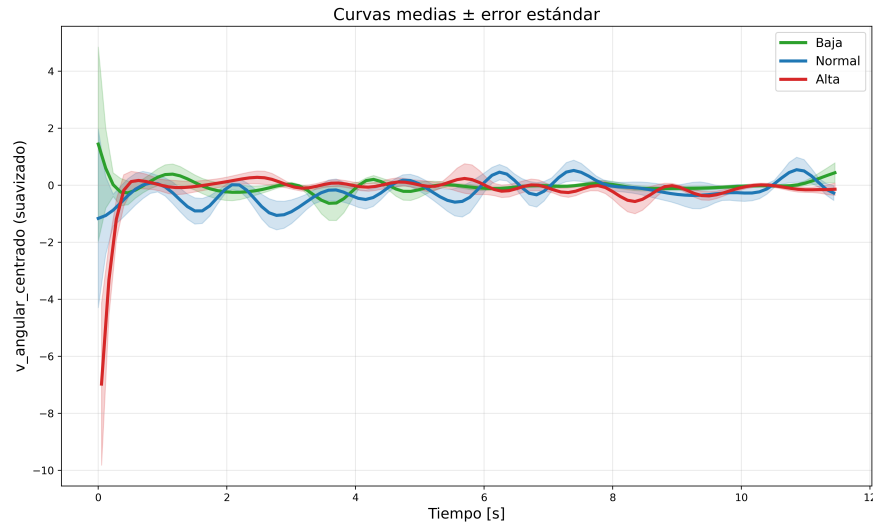


Figura 4.20: Curvas medias para la variable $\omega(t)$ en el grupo CC-R. Las bandas de error corresponden a los intervalos de confianza del 95 %.

4.12.2. LA-E

En el caso de las larvas de erizo rojo, los resultados obtenidos bajo el enfoque clásico no evidencian diferencias significativas en ninguna de las variables consideradas relevantes para el modelo propuesto en esta memoria. No obstante, esta ausencia de significancia no permite descartar completamente la posibilidad de que métodos funcionales detecten diferencias entre grupos.

Posición radial

Los resultados del fANOVA para $r(t)$ revelan una inconsistencia entre las bases funcionales utilizadas. La Tabla 4.21 muestra que P-spline indica significancia ($p = 0,045$), mientras que ninguna de las demás indica diferencias en las medias de $r(t)$ entre salinidades. Los valores de V_n son consistentes con la teoría, mostrando proporcionalidad inversa respecto al valor- p correspondiente.

No obstante, el análisis *post hoc* para la base P-spline retorna un valor de significancia ajustado mediante Bonferroni de $\alpha = 0,05/3 = 0,0167$. Bajo esta corrección, ninguna de las comparaciones entre pares presenta valores- p menores a α (0,07, 0,0820 y 0,9260 para baja vs. normal, normal vs. alta y baja vs. alta, respectivamente).

Si bien el valor- p obtenido mediante P-spline se ubicó marginalmente por debajo del umbral de significancia, esta señal no se sostiene tras aplicar correcciones por comparaciones múltiples como Bonferroni. La ausencia de significancia en el resto de las bases funcionales, junto con los resultados del análisis *post hoc*, permiten concluir que no existe evidencia estadística robusta que respalde diferencias funcionales en las curvas de $r(t)$ entre los niveles de salinidad. Aun así, el gráfico de medias de la Figura 4.22 muestra un comportamiento diferenciado en salinidad normal, el cual es muy probable que esté influenciado por la curva casi lineal observada en la Figura 4.21. Esta observación resalta la utilidad de otro aspecto del FDA: la detección de curvas atípicas. Se puede leer más sobre este tema en [62].

Base	Valor- p	Estadístico V_n
P-spline	0,045	55,0325
Ninguna	0,079	32,6677
Fourier	0,079	32,2800
Daubechies (db4)	0,069	32,6780
Daubechies (db8)	0,069	32,6760
Haar	0,069	32,5521
Coiflet (coif2)	0,069	32,6826

Tabla 4.21: Resultados del fANOVA para $r(t)$ bajo diferentes bases funcionales en el grupo LA-E; las últimas cuatro corresponden a familias de *wavelets*.

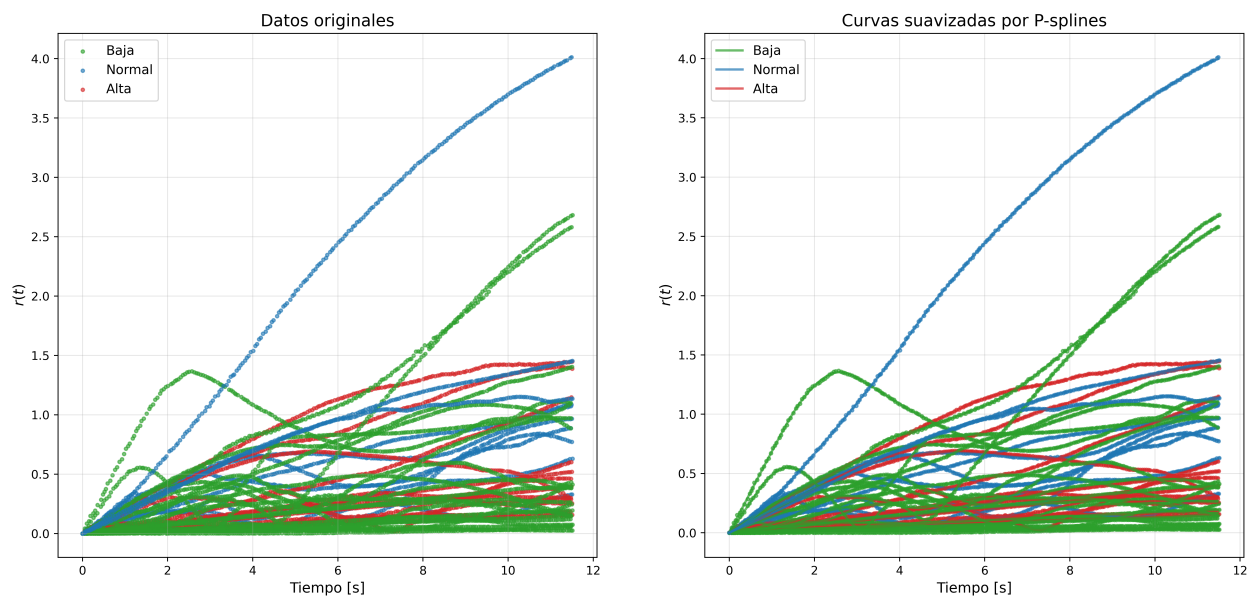


Figura 4.21: *Scatterplot* de los datos originales (izquierda) y curvas suavizadas (derecha, *scatterplot + P-splines*) para la variable $r(t)$ en el grupo LA-E.

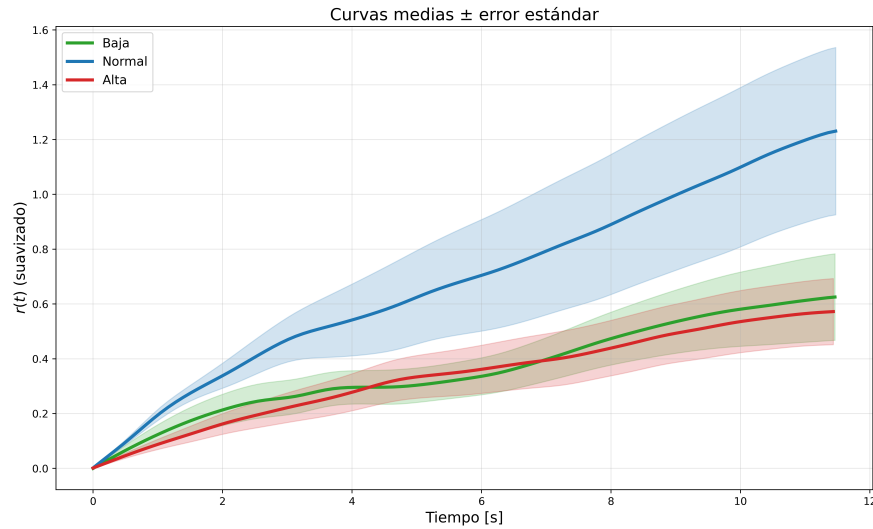


Figura 4.22: Curvas medias para la variable $r(t)$ en el grupo LA-E. Las bandas de error corresponden a los intervalos de confianza del 95 %.

Velocidad angular

La variable $\omega(t)$ tampoco presentó diferencias significativas en el fANOVA, como se evidencia en la [Tabla 4.22](#). P-spline muestra el valor más alto de V_n una vez más entre las bases consideradas. Las curvas ([Figuras 4.23](#) y [4.24](#)) exponen un comportamiento altamente similar al de la aceleración angular ([Figura 4.14](#), [Figura 4.15](#)), con una fuerte convergencia a 0 que se sostiene en el tiempo.

Base	Valor- p	Estadístico V_n
P-spline	0,571	204,9920
Ninguna	0,554	334,8255
Fourier	0,532	86,9054
Daubechies (db4)	0,578	113,1343
Daubechies (db8)	0,617	192,2646
Haar	0,527	166,9913
Coiflet (coif2)	0,578	115,3926

Tabla 4.22: Resultados del fANOVA para $\omega(t)$ bajo diferentes bases funcionales en el grupo LA-E; las últimas cuatro corresponden a familias de *wavelets*.

4.12. Variables angulares

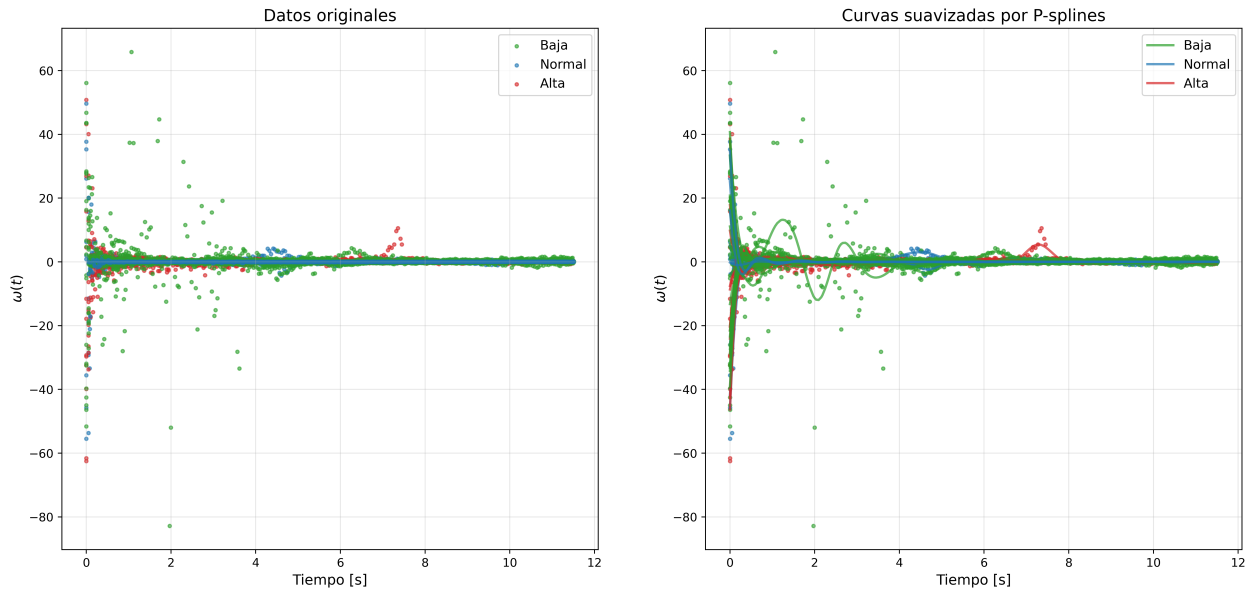


Figura 4.23: *Scatterplot* de los datos originales (izquierda) y curvas suavizadas (derecha, *scatterplot* + *P-splines*) para la variable $\omega(t)$ en el grupo LA-E.

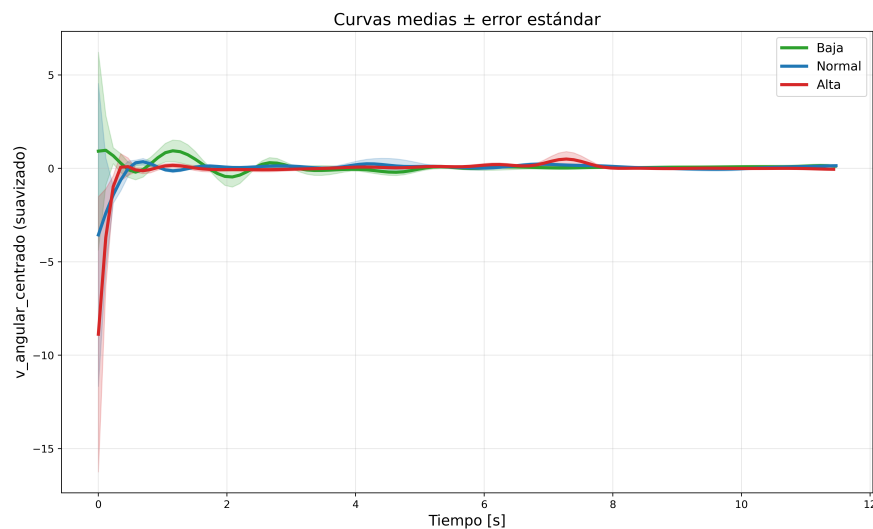


Figura 4.24: Curvas medias para la variable $\omega(t)$ en el grupo LA-E. Las bandas de error corresponden a los intervalos de confianza del 95 %.

Teniendo en cuenta lo anterior, no sería de extrañar un comportamiento subóptimo en la capacidad de los parámetros de los modelos del [Capítulo 5](#) para capturar el efecto de los niveles de salinidad. La homogeneidad intra-grupos de salinidad de $r(t)$ y $\omega(t)$ —ninguna presentó efectos intra-nivel detectables (CC-R en la [Subsección 4.7.1](#) y LA-E en la [Subsección 4.7.2](#))—, aunque no fue testada funcionalmente, junto con los resultados de esta última Sección, sugieren que modelos sobre la posición radial podrían capturar este efecto en las larvas de loco en recuperación y, tras fijar un nivel de confianza menor (como, por ejemplo, 90 %), en las de erizo rojo en exposición, de manera que sí haya evidencia de diferencias

entre las curvas de distinta salinidad.

Capítulo 5

Un primer modelo estocástico para la movilidad larval

En este capítulo se busca consolidar el aprendizaje estadístico del fenómeno estudiado, proponiendo un primer modelo estocástico para la movilidad larval. Como veremos en lo que sigue, la complejidad del fenómeno no logra ser completamente capturada, pero el trabajo realizado constituye un primer paso y entrega perspectivas de posibles mejoras. Se presentan los procedimientos y resultados del modelado, describiendo la estimación de parámetros y la metodología utilizada, junto con visualizaciones para ilustrar los procesos de optimización y simulación. La estructura del capítulo combina la teoría con los resultados, buscando generar un insumo útil para futuras investigaciones.

5.1. Ecuaciones diferenciales estocásticas

En diversos sistemas naturales, desde la dinámica de partículas en suspensión hasta el desplazamiento en organismos microscópicos o la actividad neuronal, el comportamiento observado presenta fluctuaciones abruptas que escapan a una descripción puramente determinista. Para capturar la variabilidad inherente a estos fenómenos, se recurre a las ecuaciones diferenciales estocásticas (EDEs), que extienden el formalismo clásico puramente determinista introduciendo términos aleatorios.

5.1.1. Movimiento browniano

El punto de partida para la formulación de una EDE es el proceso de Wiener $\{W_t\}_{t \geq 0}$, introducido formalmente por N. Wiener en 1923 como modelo matemático del movimiento de una partícula de polen en el agua, observado por R. Brown en 1827. El movimiento browniano describe el desplazamiento errático de partículas microscópicas suspendidas en un fluido, resultado de colisiones constantes con las moléculas del medio, y constituye uno de los fenómenos físicos más paradigmáticos de la aleatoriedad natural.

Desde una perspectiva matemática, el movimiento browniano se caracteriza por tener trayectorias continuas pero no diferenciables, cuya evolución temporal exhibe incrementos gaussianos independientes

y estacionarios. En este sentido, conviene recordar la definición de un proceso estocástico gaussiano.

Definición (Proceso estocástico gaussiano). Un proceso estocástico $\{X_t : t \in [t_0, T]\}$ es llamado *gaussiano* si cualquier combinación lineal de la forma $\sum_{i=1}^n a_i X_{t_i}$ es una variable aleatoria normal unidimensional.

Un proceso gaussiano está completamente caracterizado por su función de media, $\mathbb{E}[X_t]$, y su función de covarianza, $\text{Cov}(X_s, X_t)$. El caso del proceso de Wiener se aborda en Karatzas y Shreve (1991) [86]:

Definición (Proceso de Wiener). Un proceso de Wiener es un proceso estocástico gaussiano $\{W_t : t \geq 0\}$ cuyas distribuciones finito-dimensionales están determinadas por las siguientes condiciones:

1. $\mathbb{E}(W_t) = 0, \forall t \geq 0$.
2. $\text{Cov}(W_s, W_t) = \text{mín}(s, t)$.

La segunda propiedad en la definición anterior implica dos cosas importantes:

- Para $s = t$, la varianza es $\text{Var}(W_t) = \text{Cov}(W_t, W_t) = t$, mostrando que la variabilidad del proceso crece linealmente con el tiempo.
- Para $s < t$, la covarianza es $\text{Cov}(W_s, W_t) = s$. De esto se deriva la estructura de independencia de los incrementos del proceso.

Una EDE se define como una ecuación diferencial en la que al menos uno de sus términos está gobernado por un proceso estocástico. La formulación más común en la literatura, presentada en la [Ecuación 5.1](#), consiste en una ecuación diferencial ordinaria cuyo lado derecho está perturbado por una componente estocástica que representa las influencias aleatorias.

Definición (Ecuación diferencial estocástica). Una ecuación diferencial estocástica es una ecuación de la forma

$$dX_t = a(X_t, t)dt + \sigma(X_t, t)dW_t \quad (5.1)$$

$$X_{t_0} = x_0 \quad (5.2)$$

para $t_0 \leq t \leq T < \infty$. Aquí X_t es un proceso estocástico, $a(X_t, t)$ el coeficiente de deriva (*drift*), $\sigma(X_t, t)$ el coeficiente de difusión y dW_t un incremento de un movimiento browniano estándar.

La forma integral de la [Ecuación 5.1](#) está dada por

$$X_t = x_0 + \int_{t_0}^t a(X_s, s)ds + \int_{t_0}^t \sigma(X_s, s)dW_s \quad (5.3)$$

La interpretación rigurosa de una EDE se establece mediante la definición precisa de lo que se considera una *solución* a dicha ecuación. En este marco, se distinguen las nociones de *solución fuerte* y *solución débil*, ambas basadas en la existencia de un proceso X_t que satisface la versión integral de la

ecuación. La diferencia radica en la construcción del espacio de probabilidad: una solución débil permite definir conjuntamente el espacio y el proceso para que la ecuación se cumpla, mientras que una solución fuerte exige que el proceso esté adaptado a una fuente de aleatoriedad previamente especificada (ver [87, 88]).

La integración de una EDE requiere de la teoría de integración estocástica, un campo complejo que, para los fines de este trabajo, no se abordará aquí. Nos limitamos a señalar que los modelos a utilizar son bien conocidos en la literatura y se ha demostrado que poseen existencia y unicidad de soluciones fuertes.

5.2. Modelos de reversión a la media

Los modelos estocásticos de reversión a la media describen variables que tienden a regresar, con cierta velocidad, hacia un nivel de equilibrio en el largo plazo, exhibiendo fluctuaciones aleatorias alrededor de dicho valor. En matemática financiera, se han aplicado extensamente al análisis de precios de activos, tipos de cambio y tasas de interés [89].

Dentro de esta familia de modelos, destacan los procesos de Ornstein–Uhlenbeck (OU) y Cox–Ingersoll–Ross (CIR), ampliamente documentados y utilizados en contextos que van desde la dinámica molecular hasta la modelación de series temporales con estructura de dependencia estocástica.

5.2.1. Modelo de Ornstein–Uhlenbeck

El proceso de Ornstein–Uhlenbeck (OU) describe una reversión a la media aditiva y se aplica, entre otros campos, en física estadística, finanzas y neurociencia. Un proceso estocástico $\{X_t\}_{t \geq 0}$ es de tipo OU si satisface la EDE lineal homogénea:

$$dX(t) = \alpha X(t) dt + \sigma dW(t), \quad \omega(0) = X_0 \quad (5.4)$$

donde $\alpha > 0$ y $\sigma > 0$ son parámetros estrictamente positivos, y X_0 es una variable aleatoria independiente del movimiento browniano estándar $W(t)$. Como se muestra en Karatzas y Shreve (1991) [86], la ecuación Ecuación 5.4 admite una solución fuerte única dada por:

$$X(t) = e^{\alpha t} X_0 + \sigma \int_0^t e^{\alpha(t-s)} dW(s) \quad (5.5)$$

que además es un proceso de Markov con trayectorias continuas que satisface:

$$\mathbb{E}[X(t)] = e^{\alpha t} \mathbb{E}[X_0] \quad (5.6)$$

$$\text{Var}(X(t)) = \frac{\sigma^2}{2\alpha} + \left(\text{Var}(X_0) - \frac{\sigma^2}{2\alpha} \right) e^{2\alpha t} \quad (5.7)$$

$$\text{Cov}(X(s), X(s+t)) = \text{Var}(X_0) e^{\alpha t} + \frac{\sigma^2}{2\alpha} (e^{2\alpha s} - 1) e^{\alpha t} \quad (5.8)$$

Si $X_0 \sim \mathcal{N}\left(0, \frac{\sigma^2}{2\alpha}\right)$, entonces $X(t)$ es un proceso gaussiano estrictamente estacionario con función de covarianza:

$$c(t) = \text{Cov}(X(s), X(s+t)) = \frac{\sigma^2}{2\alpha} e^{-\alpha t} \quad (5.9)$$

En 1977, O. Vasicek adaptó la EDE de OU para crear su modelo de tasas de interés a corto plazo, conocido en finanzas como el modelo de Vasicek [90].

5.2.2. Modelo de Cox–Ingersoll–Ross

El proceso de Cox-Ingersoll-Ross (CIR) es ampliamente utilizado para modelar tasas de interés a corto plazo [91, 92]. Fue presentado inicialmente por Feller (1951) [93] para el estudio del crecimiento poblacional, y más tarde adaptado por Cox, Ingersoll y Ross (1985) [94]. Constituye una modificación del modelo de Vasicek, donde el coeficiente de difusión depende de la raíz cuadrada del estado del proceso, garantizando su positividad (o no negatividad) bajo ciertas condiciones [95]. Su EDE es:

$$dr(t) = \alpha(\mu - r_t)dt + \sqrt{r_t}\sigma dW_t, \quad (5.10)$$

donde r_t representa la variable de interés y α , μ y σ son los parámetros del modelo.

5.3. Un modelo para larvas de *C. concholepas* en recuperación

El proceso de modelación presentado en esta memoria busca capturar las diferencias observadas entre salinidades en alguna de las variables cinemáticas consideradas. Con base en los resultados del Capítulo 4, la posición radial $r(t)$ se posiciona como la candidata principal a modelar mediante una EDE. El proceso de modelación se restringe al conjunto de datos correspondiente al grupo CC-R.

Con miras en la simulación, la naturaleza de r como variable radial hace conveniente la transición de coordenadas cartesianas a coordenadas polares. En este sentido, la trayectoria bidimensional de los organismos se puede capturar mediante la transformación:

$$x_t = r_t \cos(\theta_t), \quad (5.11)$$

$$y_t = r_t \sin(\theta_t), \quad (5.12)$$

donde $x_t = x(t)$, $y_t = y(t)$ y $\theta_t = \theta_r(t)$ es el ángulo de la posición radial.

La significancia exclusiva e la posición radial —que interpretaremos en este contexto polar como el radio de la trayectoria semi-circunferencial de cada individuo— motiva la propuesta de un modelo sobre $r(t)$. En este marco, $\theta_r(t)$ se comporta como una variable auxiliar que adquiere relevancia exclusivamente en la simulación, más no en la captura de diferencias significativas entre salinidades.

5.3.1. Posición radial: proceso CIR

Para describir la posición radial de las larvas de *C. Concholepas*, se propone el modelo CIR, presentado en su forma integral:

$$r_t = r_0 + \int_0^t \alpha_r(r_\infty - r_s)ds + \int_0^t \sigma_r \sqrt{r_s} dB_s^{(r)} \quad (5.13)$$

donde $r_t = r(t)$, $r_0 = r(0)$ es la posición inicial de la larva, $\alpha_r = \alpha(r)$ la velocidad de reversión a la media, r_∞ el valor promedio a largo plazo y $\sigma_r = \sigma(r)$ la variabilidad del proceso.

El modelo CIR resulta naturalmente apropiado para representar una variable estrictamente no negativa, como lo es $r(t)$.

5.3.2. Velocidad angular: proceso OU

Por su parte, el ángulo radial $\theta_r(t)$ se modela indirectamente mediante un proceso de OU aplicado a la velocidad angular $\omega(t)$. La propuesta, en forma integral, es:

$$\omega_t = \omega_0 + \int_0^t \alpha_\omega(\omega_\infty - \omega_s) + \int_0^t \sigma_\omega dB_s^{(\omega)}, \quad (5.14)$$

donde $\omega_t = \omega(t)$, $\omega_0 = \omega(0)$ es la velocidad angular inicial, y los parámetros α_ω , ω_∞ y σ_ω son análogos al modelo CIR para $r(t)$.

A partir del proceso $(\omega_t : t \geq 0)$, se puede reconstruir el ángulo θ_t con un modelo cinético tipo Langevin:

$$\theta_t = \theta_0 + \int_0^t \omega_s ds \quad (5.15)$$

5.4. Calibración

En el contexto de modelamiento estocástico basado en datos, la calibración de modelos se refiere al proceso de estimación de los valores de los parámetros. Para el proceso de OU, el vector de parámetros es:

$$\Theta_{OU} = (\alpha, \omega_\infty, \sigma),$$

donde ω_∞ representa el nivel de reversión a la media. En el caso del proceso CIR, el vector de parámetros es:

$$\Theta_{CIR} = (\alpha, r_\infty, \sigma),$$

donde r_∞ cumple un rol análogo.

Para recuperar estos parámetros, existen diversos métodos de estimación. En este trabajo se emplea la estimación por variación cuadrática para σ y la estimación de máxima verosimilitud para α , r_∞ y

ω_∞ . Dado que los datos son discretos, el procedimiento requiere de un esquema de aproximación, como el método de Euler–Maruyama, así como de una etapa de optimización numérica para determinar los estimadores.

5.5. Esquema numérico de Euler-Maruyama

Los modelos formulados mediante EDEs se definen en tiempo continuo, mientras que, en la práctica, los datos experimentales disponibles suelen estar en tiempo discreto [96, 95]. Esta discrepancia dificulta el análisis estadístico y la estimación de parámetros, pues la distribución de transición del proceso es, en general, desconocida o carece de una forma cerrada [96, 97, 98]. Para superar este obstáculo, es necesario recurrir a esquemas de aproximación que discreticen el proceso continuo. El método de Euler–Maruyama es uno de los más utilizados por su simplicidad y eficiencia [95, 88].

Este método es una extensión natural del método de Euler para ecuaciones diferenciales ordinarias, adaptada para incorporar la componente estocástica introducida por el proceso de Wiener [95]. Dada una EDE general:

$$dX_t = m(X_t, \theta) dt + \sigma(X_t, \theta) dW_t, \quad (5.16)$$

donde X_t es el proceso estocástico, $m(X_t, \theta)$ es el coeficiente de deriva (*drift*), $\sigma(X_t, \theta)$ es el coeficiente de difusión (volatilidad) y dW_t es un incremento de un movimiento browniano estándar [95, 99], el esquema de Euler–Maruyama aproxima la trayectoria del proceso en puntos discretos del tiempo.

Para un intervalo de observación $[0, T]$ discretizado en n pasos de tiempo:

$$0 = t_0 < t_1 < \dots < t_n = T, \quad (5.17)$$

con incrementos constantes $\Delta = t_{i+1} - t_i$ ¹, la aproximación de Euler–Maruyama para el proceso X_t viene dada por:

$$X_{t_{i+1}} - X_{t_i} = m(X_{t_i}, \theta) \Delta + \sigma(X_{t_i}, \theta) \sqrt{\Delta} \varepsilon_{t_i}, \quad (5.18)$$

donde θ representa el vector de parámetros del modelo, y $\varepsilon_{t_i} \sim \mathcal{N}(0, 1)$ son variables aleatorias independientes e idénticamente distribuidas [95, 92]. Este esquema asume que las observaciones provienen de una distribución gaussiana local, donde la media es el *drift* y la desviación estándar es la función de difusión [95].

El método de Euler–Maruyama es fuertemente convergente de orden $1/2$ y débilmente convergente de orden 1^2 [95]. Esto indica que, aunque es una aproximación de Taylor simple, constituye una herramienta efectiva para la simulación y estimación de parámetros en procesos de difusión cuando no existen soluciones exactas o estas son inabordables.

¹Es decir, la [Ecuación 5.17](#) define una *partición* del dominio

²Bajo condiciones típicas de Lipschitz-continuidad y crecimiento lineal.

5.5.1. Aplicación al proceso CIR

La EDE del proceso CIR, dada en la [Ecuación 5.10](#), cumple la condición de Feller:

$$2\alpha\mu > \sigma^2, \quad (5.19)$$

la cual garantiza la positividad estricta del proceso [\[91, 92\]](#).

Aplicando el esquema de Euler–Maruyama, se identifican:

$$m(r_t, \theta) = \alpha(\mu - r_t), \quad \sigma(r_t, \theta) = \sigma\sqrt{r_t}. \quad (5.20)$$

Así, la forma discretizada del proceso es:

$$r_{t_{i+1}} - r_{t_i} = \alpha(\mu - r_{t_i}) \Delta + \sigma\sqrt{r_{t_i}} \sqrt{\Delta} \varepsilon_{t_i}, \quad (5.21)$$

donde $\varepsilon_{t_i} \sim \mathcal{N}(0, 1)$.

5.5.2. Aplicación al proceso OU

Para el proceso de OU, se tiene:

$$m(X_t, \theta) = \theta(\mu - X_t), \quad \sigma(X_t, \theta) = \sigma. \quad (5.22)$$

Por lo tanto, su forma discretizada es:

$$X_{t_{i+1}} - X_{t_i} = \theta(\mu - X_{t_i}) \Delta + \sigma\sqrt{\Delta} \varepsilon_{t_i}, \quad (5.23)$$

donde $\varepsilon_{t_i} \sim \mathcal{N}(0, 1)$.

Esta discretización es fundamental para la estimación de sus parámetros a partir de datos discretos [\[92, 95\]](#). Los métodos de estimación para el proceso de OU han sido ampliamente abordados en la literatura (ver son objeto de estudio en diversas investigaciones (ver [\[99, 100\]](#))).

Observaciones numéricas

1. El proceso de OU admite una forma exacta tipo AR(1) para pasos regulares, pero utilizar el esquema de Euler–Maruyama permite mantener coherencia metodológica si se aplica el mismo esquema a CIR y OU de manera conjunta.
2. La reconstrucción de ángulo a partir de ω se logra por integración discreta:

$$\theta_{n+1} = \theta_n + \omega_n \Delta t_n,$$

asegurando consistencia temporal con la discretización de ω .

5.6. Métodos de estimación

La estimación de los parámetros de modelos de difusión estocástica, como los procesos CIR y OU, se basa en métodos que consideran la naturaleza discreta de las observaciones. Entre los más utilizados destacan la máxima verosimilitud y la variación cuadrática.

5.6.1. Variación cuadrática para la estimación de σ

El coeficiente de difusión (σ), que representa la volatilidad o la magnitud de la aleatoriedad del proceso, puede estimarse de manera no paramétrica utilizando la variación cuadrática. Esta mide la “variación total” de una trayectoria a lo largo del tiempo.

Para un proceso de difusión

$$dX_t = m(X_t, t) dt + \sigma(X_t, t) dW_t, \quad (5.24)$$

la variación cuadrática está dada por:

$$[X]_t = \int_0^t \sigma^2(X_s, s) ds. \quad (5.25)$$

En observaciones discretas con un paso temporal pequeño, el estimador de la variación cuadrática puede aproximarse como la suma de los cuadrados de los incrementos del proceso [95].

Proceso OU En el proceso OU, donde σ es constante, se estima como:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (\Delta X_{t_i})^2}{\sum_{i=1}^n \Delta t_i}, \quad (5.26)$$

ignorando el término de deriva. Este estimador es consistente para el coeficiente de difusión cuando el tamaño de los pasos de tiempo tiende a cero [95].

Proceso de Cox-Ingersoll-Ross (CIR) En el proceso CIR, donde σ es proporcional a $\sigma\sqrt{r_t}$, la variación cuadrática depende de r_t . El estimador se define como:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (\Delta r_{t_i})^2}{\sum_{i=1}^n r_{t_{i-1}} \Delta t_i}. \quad (5.27)$$

Este método proporciona un estimador inicial robusto de σ , útil como punto de partida en procedimientos posteriores como la máxima verosimilitud [91].

5.6.2. Estimación de parámetros de reversión a la media

Antes de aplicar máxima verosimilitud, conviene obtener estimaciones iniciales de los parámetros de reversión a la media, puesto que facilitan la convergencia de los algoritmos de optimización y reducen

la probabilidad de caer en máximos locales [96]. Estas estimaciones iniciales se pueden obtener a partir de las propiedades estadísticas de las series de tiempo discretizadas.

Proceso OU

Para el proceso OU, α y μ pueden estimarse mediante mínimos cuadrados ordinarios (OLS³), ya que la versión discretizada del proceso se comporta como una regresión autorregresiva de primer orden (AR(1)), como indican Fergusson y Platen (2015) [92] y Holý y Tomanova (2022) [101]:

$$X_{t_{i+1}} = c + \phi X_{t_i} + \varepsilon_{t_{i+1}}, \quad (5.28)$$

donde $c = \mu(1 - e^{-\alpha\Delta})$ y $\phi = e^{-\alpha\Delta}$.

Mediante regresión lineal simple se obtienen \hat{c} y $\hat{\phi}$, y a partir de ellos, las estimaciones iniciales de α y μ .

Proceso CIR

En el proceso CIR, las estimaciones iniciales de α y μ se derivan de los momentos del proceso, aplicando OLS sobre su versión discretizada [91]:

$$r_{t+\Delta} - r_t = \alpha(\mu - r_t)\Delta + \sigma\sqrt{r_t}\varepsilon_t, \quad (5.29)$$

que puede reescribirse como:

$$\frac{r_{t+\Delta} - r_t}{\sqrt{r_t}} = \frac{\alpha\mu\Delta}{\sqrt{r_t}} - \alpha\sqrt{r_t}\Delta + \sigma\varepsilon_t. \quad (5.30)$$

A partir de esta expresión, Kladívko (2014) [91] deduce estimadores cerrados para α y μ mediante OLS. Sean N el número de observaciones y $\{r_{t_i}\}_{i=1}^N$ la serie observada de $r(t)$, entonces:

$$\hat{\alpha} = \frac{N^2 - 2N + 1 + \sum_{i=1}^{N-1} r_{t_{i+1}} \sum_{i=1}^{N-1} \frac{1}{r_{t_i}} - \left(\sum_{i=1}^{N-1} r_{t_i}\right) \left(\sum_{i=1}^{N-1} \frac{1}{r_{t_i}}\right) - (N-1) \sum_{i=1}^{N-1} \frac{r_{t_{i+1}}}{r_{t_i}}}{\left(N^2 - 2N + 1 - \sum_{i=1}^{N-1} r_{t_i} \sum_{i=1}^{N-1} \frac{1}{r_{t_i}}\right) \Delta}, \quad (5.31)$$

$$\hat{\mu} = \frac{(N-1) \sum_{i=1}^{N-1} r_{t_{i+1}} - \sum_{i=1}^{N-1} \frac{r_{t_{i+1}}}{r_{t_i}} \sum_{i=1}^{N-1} r_{t_i}}{N^2 - 2N + 1 + \sum_{i=1}^{N-1} r_{t_{i+1}} \sum_{i=1}^{N-1} \frac{1}{r_{t_i}} - \left(\sum_{i=1}^{N-1} r_{t_i}\right) \left(\sum_{i=1}^{N-1} \frac{1}{r_{t_i}}\right) - (N-1) \sum_{i=1}^{N-1} \frac{r_{t_{i+1}}}{r_{t_i}}}. \quad (5.32)$$

Kladívko también propone un estimador inicial de σ basado en la varianza de los residuos del ajuste OLS:

³Ordinary Least Squares.

$$\hat{\sigma} = \sqrt{\frac{\text{Var}(\text{residuos})}{\Delta}}, \quad (5.33)$$

aunque en este análisis se emplea solo el estimado por variación cuadrática, a fin de mantener un único criterio.

5.6.3. Estimación por máxima verosimilitud

El método de máxima verosimilitud (MLE⁴) es el estándar para la estimación de parámetros en modelos de difusión, ya que estima todos los parámetros del modelo de manera simultánea al maximizar la función de verosimilitud [96]. Esta función se construye a partir de la densidad de probabilidad de transición del proceso entre dos puntos de tiempo consecutivos. El objetivo es encontrar los valores de los parámetros que hacen que las observaciones sean lo más probables posible.

Proceso OU

El proceso OU cuenta con una densidad de transición conocida y expresable en forma cerrada, lo que permite construir una función de verosimilitud exacta. El proceso se puede caracterizar por una distribución normal condicionada al estado anterior.

1. Definición de la densidad de transición. La distribución de X_t dado X_s (con $s < t$) es normal, con una media y varianza que dependen de los parámetros del modelo (θ, μ, σ) y del paso de tiempo $\Delta t = t - s$ [99, 101]:

$$X_t | X_s \sim N\left(\mu + (X_s - \mu)e^{-\theta\Delta t}, \frac{\sigma^2}{2\theta}(1 - e^{-2\theta\Delta t})\right) \quad (5.34)$$

La función de densidad de probabilidad (PDF⁵) de esta transición, $p(X_t | X_s)$, es la de una distribución normal.

2. Formulación de la función de log-verosimilitud. Dada una serie de observaciones discretas en el tiempo, $\{X_{t_0}, X_{t_1}, \dots, X_{t_n}\}$, la función de densidad conjunta se puede expresar como el producto de las densidades de transición condicionales, asumiendo que las observaciones son independientes dado el estado anterior (propiedad de Markov) [99]:

$$f(X_{t_1}, \dots, X_{t_n} | X_{t_0}; \theta, \mu, \sigma) = \prod_{i=1}^n p(X_{t_i} | X_{t_{i-1}}; \theta, \mu, \sigma) \quad (5.35)$$

3. Maximización de la log-Verosimilitud Para simplificar el cálculo, se trabaja con el logaritmo de la función de verosimilitud (log-verosimilitud), que se convierte en una suma de logaritmos:

⁴Maximum Likelihood Estimation

⁵Probability Density Function

$$\log L(\theta, \mu, \sigma) = \sum_{i=1}^n \log p(X_{t_i} | X_{t_{i-1}}; \theta, \mu, \sigma) \quad (5.36)$$

La maximización de esta función de log-verosimilitud con respecto a los parámetros (θ, μ, σ) conduce a los estimadores de máxima verosimilitud. Este problema de optimización se resuelve numéricamente, y las estimaciones iniciales, obtenidas por ejemplo a través de un método de mínimos cuadrados, son cruciales para asegurar la convergencia.

La función de log-verosimilitud definida en la [Ecuación 5.36](#) es específica para cada trayectoria individual. De este modo, la estimación de los parámetros del modelo se realiza de forma separada para cada larva. Los resultados de esta estimación para el modelo de Ornstein-Uhlenbeck (OU) mostraron un comportamiento más homogéneo en comparación con el modelo CIR, presentando muy pocas trayectorias con un comportamiento abrupto o “agresivo” con saltos o picos. Esto es consistente con los análisis previos que no revelaron diferencias significativas en la velocidad angular entre los distintos niveles de salinidad.

Proceso de Cox-Ingersoll-Ross

En el proceso CIR, la densidad de transición no es una función simple como la normal, lo que hace que el cálculo de la MLE sea más complejo. La función de densidad de probabilidad (PDF) de transición del proceso CIR está relacionada con una distribución de chi-cuadrado no central [\[91\]](#).

1. Definición de la densidad de transición La densidad de probabilidad de transición para el proceso CIR, $p(r_t | r_s)$, corresponde a la de una distribución de chi-cuadrado no central. Para facilitar el cálculo, se utilizan transformaciones de la variable de estado. Una formulación estándar para la PDF de transición es [\[91\]](#):

$$p(r_t | r_s) = C_t e^{-U_t r_t} \left(\frac{r_t}{r_s} \right)^{q-1} I_q(2C_t U_t r_t r_s) e^{-U_t r_s} \quad (5.37)$$

donde:

$$I_q(\cdot) \text{ es la función de Bessel modificada de primera especie de orden } q, \quad (5.38)$$

$$q = \frac{2\alpha\mu}{\sigma^2} - 1, \quad (5.39)$$

$$C_t = \frac{2\alpha}{\sigma^2(1 - e^{-\alpha\Delta t})}, \quad U_t = \frac{2\alpha e^{-\alpha\Delta t}}{\sigma^2(1 - e^{-\alpha\Delta t})} \quad (5.40)$$

Esta densidad depende de los parámetros del modelo (α, μ, σ) y de las observaciones.

2. Formulación y maximización de la log-verosimilitud De manera análoga al proceso OU, la función de log-verosimilitud se construye como la suma de los logaritmos de las densidades de transición

para cada observación en el tiempo [91]:

$$\log L(\alpha, \mu, \sigma) = \sum_{i=1}^n \log p(r_{t_i} | r_{t_{i-1}}; \alpha, \mu, \sigma) \quad (5.41)$$

La maximización de esta función, aunque más compleja debido a la presencia de la función de Bessel, se realizó mediante métodos de optimización numérica. Los estimadores resultantes son los valores de los parámetros que mejor se ajustan a la trayectoria observada.

Es importante destacar que la función de log-verosimilitud es específica para cada trayectoria individual. Para el proceso CIR, la mayoría de las estimaciones mostraron un comportamiento esperado y bien ajustado. Sin embargo, un número limitado de trayectorias (no más de 3) presentaron comportamientos atípicos o abruptos, lo que probablemente se deba a errores numéricos o a una sensibilidad del método ante datos con alta variabilidad. Estas trayectorias no se consideraron para el análisis, ya que probablemente corresponden a aquellas que arrojaron parámetros negativos o inconsistentes, y que posteriormente fueron filtradas.

5.7. Resultados

Antes de presentar los resultados de la estimación, es importante considerar un aspecto particular del proceso CIR. A diferencia del proceso OU, la simulación del CIR presenta un comportamiento errático y volátil cuando los valores de la variable de estado se aproximan a cero [91]. Para mitigar este efecto y asegurar la fiabilidad de los resultados, se dejaron pasar aproximadamente 1,5 second de simulación, equivalentes a unas 30 observaciones dado el salto temporal promedio de 0,05 s. Este mismo procedimiento se aplicó al conjunto de datos utilizado para modelar la velocidad angular con OU, con el fin de mantener la consistencia en el preprocesamiento de ambas variables.

5.7.1. Prueba con datos simulados

La implementación de la estimación de parámetros se validó con una trayectoria simulada para ambos modelos. En cada caso, los valores estimados —tanto por el método inicial de OLS y variación cuadrática, como por la optimización— fueron cercanos a los parámetros originales, lo que indica que la implementación es correcta y robusta en este escenario de prueba.

CIR

Para el modelo CIR sobre $r(t)$, se simuló una trayectoria utilizando como condición inicial un muestreo aleatorio de la estimación de la densidad del núcleo (KDE⁶) de los valores iniciales r_0 en la data, como se muestra en la figura [Figura 5.1a](#). La trayectoria simulada se presenta en la [Figura 5.1b](#).

⁶Kernel Density Estimation

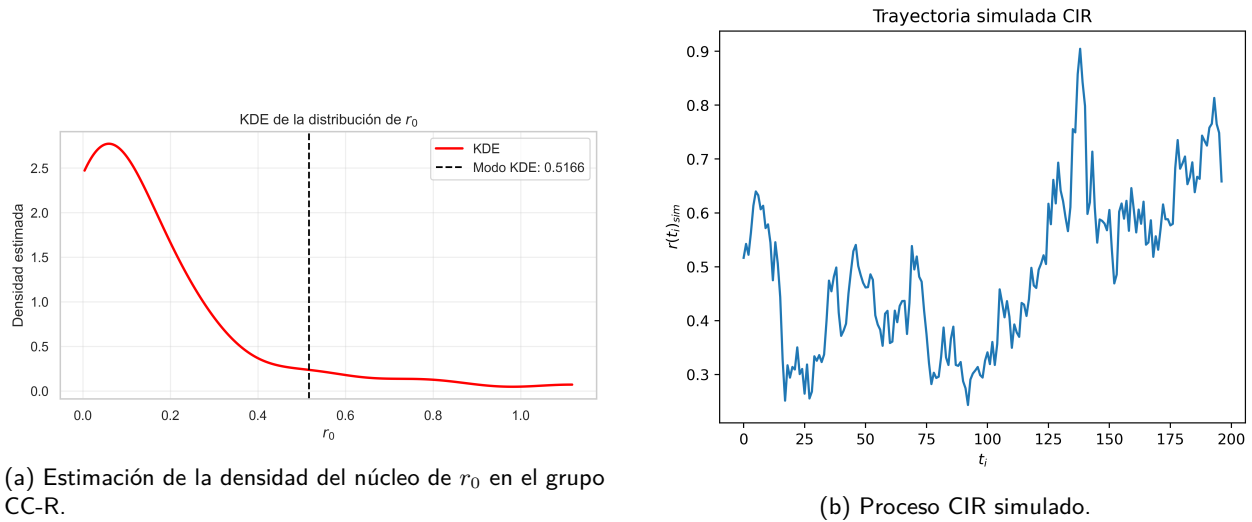


Figura 5.1: Simulación de una realización de un proceso CIR con parámetros $\Theta = (\alpha, r_\infty, \sigma) = (1, 2, 0, 5, 0, 3)$.

Tabla 5.1: Comparación de los valores de parámetros originales CIR frente a los estimados por diferentes métodos.

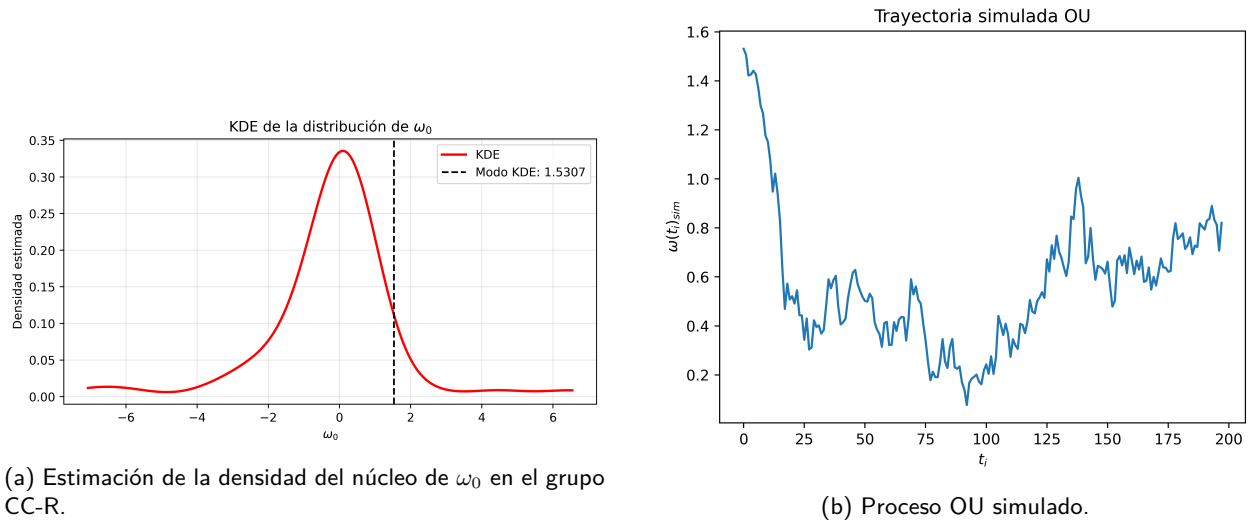
Parámetro	Original	OLS	Log-verosimilitud (Bessel/Chi-cuadrada)
α	1.2	1.0077556	1.0235860
r_∞	0.5	0.5069891	0.5071352
σ	0.3	0.3065225	No estimado

Notas sobre la optimización:

- El método de log-verosimilitud no estima el parámetro σ , el cual se mantiene fijo en 0,3.
- Ambos métodos de verosimilitud convergieron exitosamente, con un valor de la función objetivo de $-324,3959$.
- Durante la estimación de los parámetros iniciales mediante OLS y variación cuadrática se detectaron trayectorias con valores inválidos (negativos, no permitidos en el modelo CIR). De las 58 trayectorias de *C. concholepas* en recuperación (CC-R), 46 arrojaron parámetros iniciales válidos. Los casos problemáticos fueron penalizados con valores altos en la función objetivo para evitar la exploración de zonas no admisibles.

OU

Para el modelo OU sobre $\omega(t)$, se simuló una trayectoria bajo el mismo esquema, utilizando como condición inicial un muestreo aleatorio de la estimación KDE de los valores iniciales ω_0 , mostrada en la Figura 5.2a. La trayectoria resultante se presenta en la Figura 5.2b.

(a) Estimación de la densidad del núcleo de ω_0 en el grupo CC-R.

(b) Proceso OU simulado.

Figura 5.2: Simulación de una realización de un proceso OU con parámetros $\Theta = (\alpha, \omega_\infty, \sigma) = (1, 2, 0, 5, 0, 3)$.

Parámetro	Original	OLS	Log-verosimilitud
α	1.2	1.2218748	1.2081981
r_∞	0.5	0.5115432	0.4157006
σ	0.3	0.3127158	No estimado

Tabla 5.2: Comparación de los valores de parámetros originales OU frente a los estimados por diferentes métodos.

Notas sobre la optimización:

- El método de log-verosimilitud no estima el parámetro σ , que se mantuvo fijo en 0,3.
- El método convergió exitosamente, con un valor de la función objetivo de $-273,669$.
- De las 58 trayectorias de *C. concholepas* en recuperación (CC-R), 56 arrojaron parámetros iniciales válidos, lo que resultó en el descarte automático del 3.45% de los datos (2 trayectorias). Estos parámetros problemáticos fueron penalizados con valores altos durante el proceso de optimización para asegurar convergencia a soluciones admisibles.

Como se comentó anteriormente, para reducir la dimensionalidad de la optimización y minimizar errores numéricos, se mantuvo σ fijo en su estimación inicial por variación cuadrática. La función de log-verosimilitud, $\log L$, fue por tanto bidimensional.

5.7.2. Optimización de la log-verosimilitud

La estimación por máxima verosimilitud se implementó como una minimización de la log-verosimilitud negativa. Esta práctica es común en la optimización numérica, ya que la mayoría de los algoritmos están

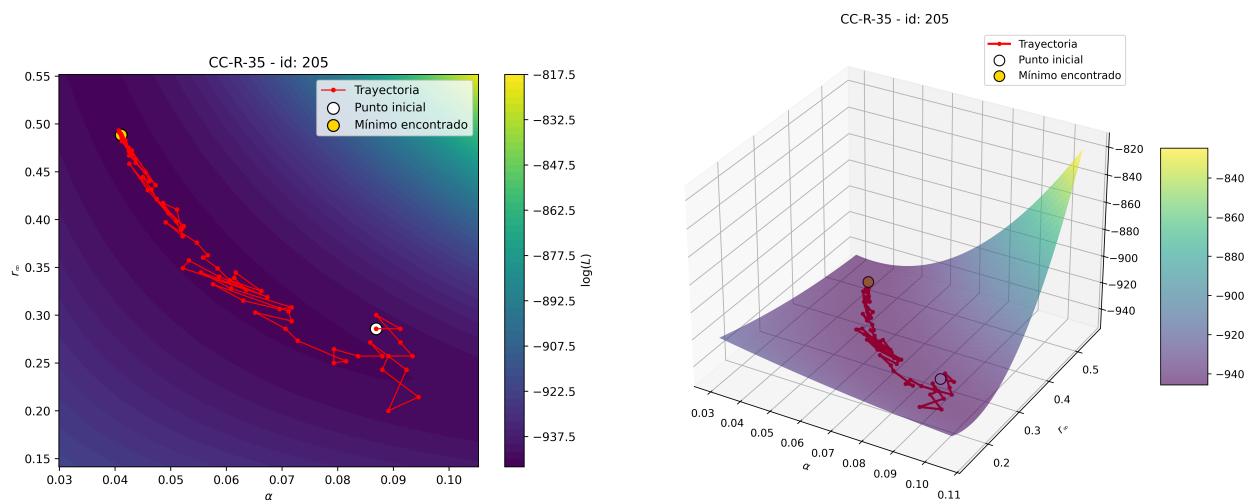
diseñados para resolver problemas de minimización. El procedimiento se aplicó de forma individual a cada trayectoria, resultando en un conjunto de estimadores $(\hat{\alpha}, \hat{\omega}_{\infty}, \hat{\sigma})$ para el modelo OU y $(\hat{\alpha}, \hat{r}_{\infty}, \hat{\sigma})$ para el modelo CIR.

Para la optimización, se empleó el algoritmo Nelder–Mead, el cual demostró una alta tasa de convergencia y una exploración eficiente del espacio de parámetros, como se observa en las figuras [Figura 5.3a](#) a [Figura 5.4b](#). Este método se eligió por su robustez para minimizar funciones multidimensionales, incluso sin el cálculo de gradientes. La optimización se realizó sobre una reparametrización que impone la positividad de los parámetros mediante transformaciones logarítmicas y penalizaciones para evitar zonas no admisibles en el dominio del modelo CIR.

La búsqueda del mínimo se inició desde semillas iniciales provistas por los métodos de OLS y variación cuadrática, lo que permitió una exploración eficiente de la superficie de la log-verosimilitud. El proceso se detuvo al cumplir las tolerancias predefinidas en el valor de la función objetivo. Como se ilustra en las figuras [Figura 5.3a](#) a [Figura 5.4b](#), el punto de partida OLS típicamente se ubicó en una cuenca cercana al mínimo, lo que llevó a una convergencia consistente hacia el valle principal.

Ejemplos del proceso de optimización

CIR Las figuras a continuación muestran un ejemplo representativo del proceso de optimización para el modelo CIR.

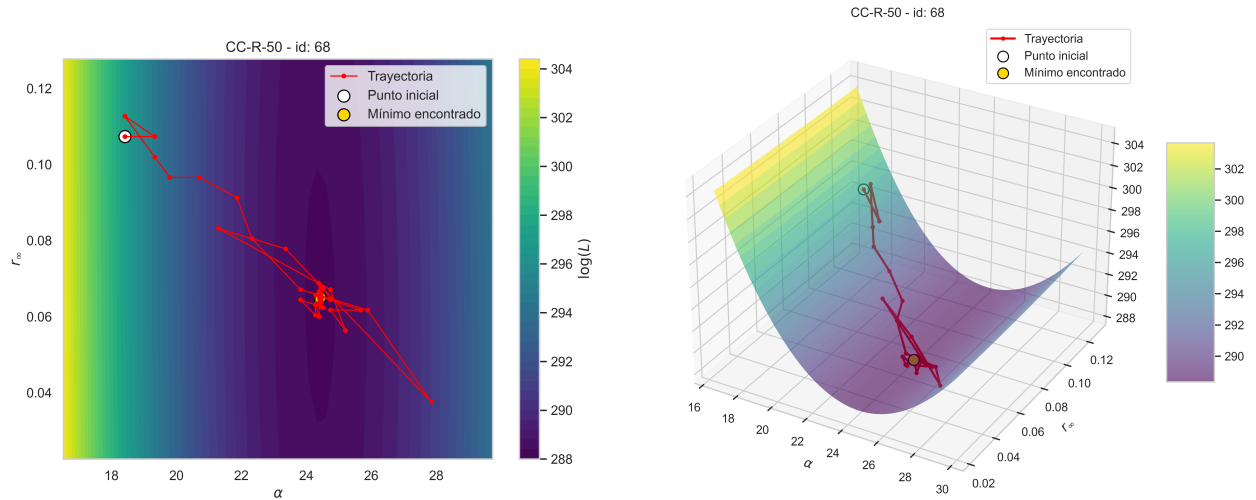


(a) Gráfico representativo del proceso de optimización de la función de log-verosimilitud del CIR (ID de trayectoria: 205).

(b) Gráfico tridimensional de la log-verosimilitud junto al punto de partida (estimadores iniciales por OLS) y el mínimo encontrado.

Figura 5.3: Optimización de la función de log-verosimilitud para el modelo CIR.

OU A continuación se ilustra un ejemplo del proceso de optimización para el modelo OU.



(a) Gráfico representativo del proceso de optimización de la función de log-verosimilitud del OU (ID de trayectoria: 68).

(b) Gráfico tridimensional de la log-verosimilitud junto al punto de partida (estimadores iniciales por OLS) y el mínimo encontrado.

Figura 5.4: Optimización de la función de log-verosimilitud para el modelo OU.

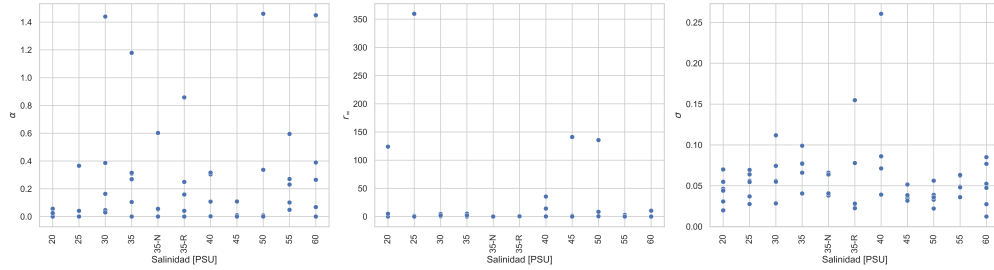
5.7.3. Visualización de parámetros estimados

Respondiendo a uno de los objetivos específicos de este trabajo, este apartado busca identificar patrones o relaciones en los parámetros estimados. Se plantea la hipótesis de que estos parámetros presenten diferencias entre trayectorias en salinidad baja y alta, donde el análisis post hoc del FDA detectó efectos significativos. En este sentido, el hecho de que al menos uno de los parámetros del modelo estocástico capture un comportamiento o dinámica asociado a la salinidad ambiental, puede ser un indicador importante de estrés subletal en las larvas, permitiendo caracterizar la respuesta de las mismas de manera mecanicista y no solo descriptiva.

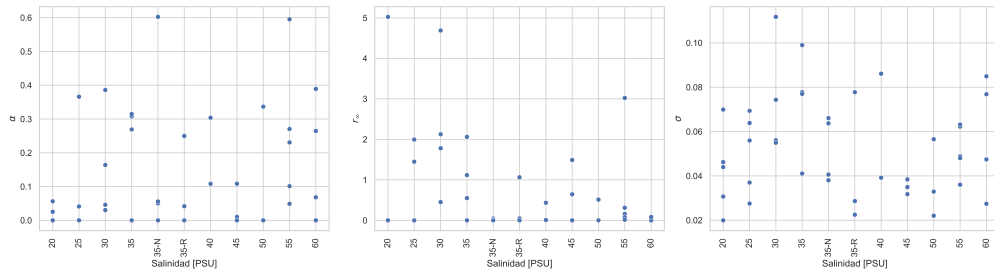
Modelo CIR para r

El gráfico superior de la [Figura 5.5](#) muestra los valores estimados del modelo CIR según la salinidad. El gráfico inferior presenta todos los datos, mientras que el inferior excluye valores atípicos mediante el criterio del RIC.

5.7. Resultados



(a) Diagrama de dispersión de los valores de los parámetros por salinidad del modelo OU.



(b) Diagrama de dispersión de los valores de los parámetros por salinidad filtrado para excluir valores atípicos.

Figura 5.5: Dispersión de los parámetros encontrados por salinidad del modelo CIR.

La visualización de los parámetros de forma unidimensional puede no ser la mejor manera de encontrar patrones o relaciones. Dada la naturaleza inherentemente correlacionada de los parámetros del modelo, al proyectarlos de manera individual se puede perder información crucial. En este sentido, un diagrama de dispersión de los parámetros entre sí es una visualización más apropiada para el objetivo de identificar grupos o clústeres. Por simplicidad, se presentan solo los datos de salinidad baja y alta.

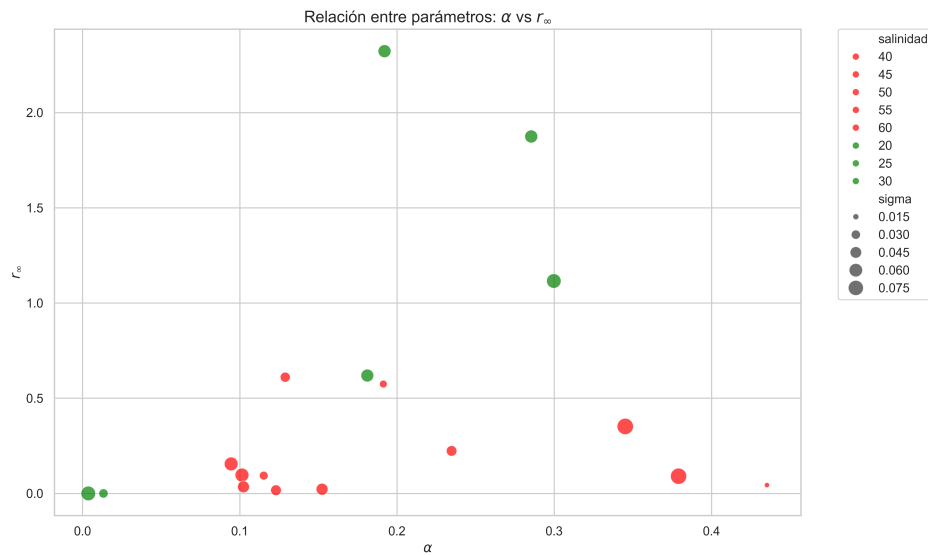


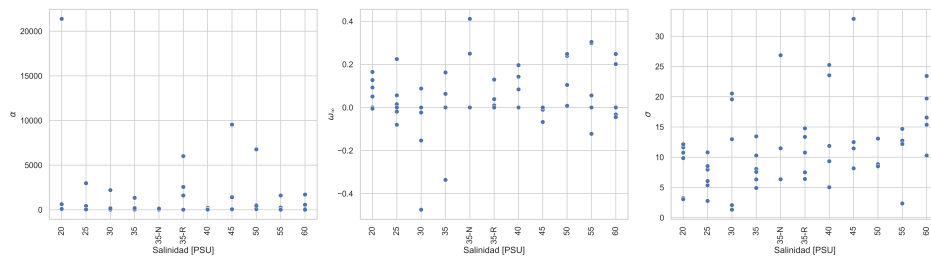
Figura 5.6: Diagrama de dispersión de α vs r_∞ . El color de cada punto representa su nivel de salinidad, y el tamaño la magnitud de la volatilidad de esa trayectoria (σ).

El diagrama de dispersión de α vs. r_∞ revela concentraciones diferenciadas por nivel de salinidad, aunque con cierto solapamiento entre grupos. Si bien no se observan clústeres perfectamente definidos, es posible identificar regiones del espacio de parámetros con mayor densidad de puntos para cada grupo, lo que sugiere que el modelo está capturando dinámicas de reversión a la media (α) y niveles de equilibrio (r_∞) que difieren en promedio según la salinidad.

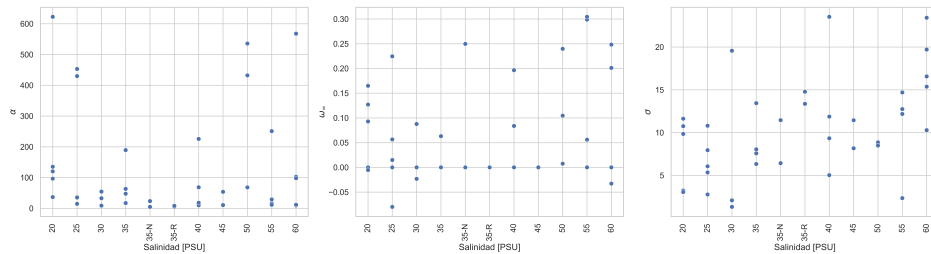
En particular, las trayectorias asociadas a salinidades bajas tienden a presentar valores menores de r_∞ , lo que implica que, en condiciones hipotéticas de equilibrio, el radio medio alcanzado por las larvas sería más reducido. En otras palabras, las larvas bajo salinidad baja tenderían a mantener trayectorias más confinadas espacialmente, posiblemente reflejando una disminución en la movilidad o en la amplitud de desplazamiento. Por el contrario, en salinidades altas se observan algunos casos con r_∞ más elevados, coherentes con trayectorias más amplias o dispersas. El tamaño de los puntos, que representa la volatilidad (σ), no muestra un patrón claro, lo que sugiere que las fluctuaciones locales de la posición radial no constituyen un factor determinante en la separación entre grupos.

Modelo OU para ω

Las Figuras 5.7 y 5.8 muestran los resultados obtenidos de manera idéntica a los de la subsección anterior.



(a) Diagrama de dispersión de los valores de los parámetros por salinidad.



(b) Diagrama de dispersión de los valores de los parámetros por salinidad filtrado para excluir valores atípicos.

Figura 5.7: Dispersión de los parámetros encontrados por salinidad.

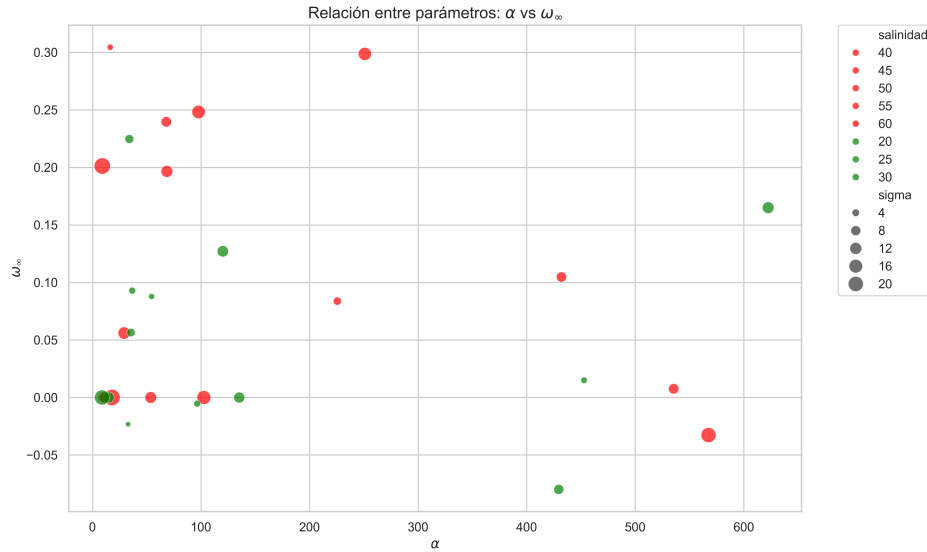


Figura 5.8: Diagrama de dispersión de α vs ω_∞ . El color de cada punto representa su nivel de salinidad, y el tamaño la magnitud de la volatilidad de esa trayectoria (σ).

A diferencia del CIR, el diagrama de dispersión de α vs ω_∞ muestra una superposición más marcada entre salinidades baja y alta. No se observan clústeres definidos, lo que concuerda con los resultados del capítulo anterior: la velocidad angular no constituye un buen bioindicador de estrés salino. La proyección bivariada sugiere un solapamiento amplio entre los niveles de salinidad, especialmente en los valores intermedios.

Una extensión natural sería evaluar la capacidad predictiva de los parámetros mediante clasificadores supervisados (por ejemplo, SVM con validación cruzada), usando como *features* los parámetros estimados y métricas como AUC o *balanced accuracy* [102].

5.8. Simulaciones

En esta última etapa, se utilizaron los parámetros estimados para los procesos que gobiernan la posición radial r (modelo CIR) y la velocidad angular $\omega = \dot{\theta}$ (modelo OU) para generar trayectorias simuladas y compararlas cualitativamente con las observadas.

Como el modelo para la componente angular está planteado sobre la derivada $\dot{\theta}$, se requiere recuperar el ángulo θ a través de una integración numérica a lo largo del tiempo. De este modo, se obtiene la evolución conjunta de (r_t, θ_t) , lo que permite reconstruir la trayectoria en el plano.

El procedimiento de simulación se implementó mediante un esquema de Euler–Maruyama, partiendo de condiciones iniciales observadas y actualizando en cada instante temporal tanto el radio como la velocidad angular. Una vez obtenida la secuencia de θ_t , se proyectó la trayectoria a coordenadas cartesianas mediante la transformación polar usual:

$$x_t = r_t \cos(\theta_t), \quad y_t = r_t \sin(\theta_t). \quad (5.42)$$

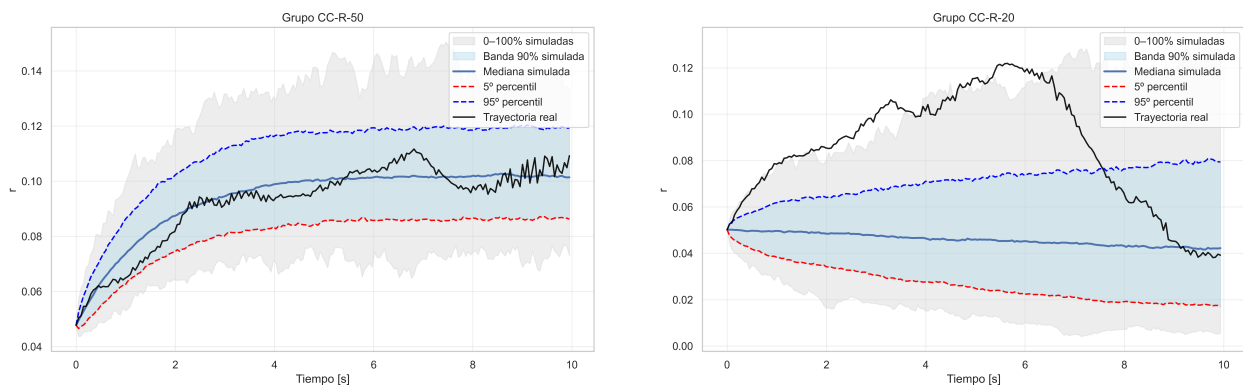
El enfoque es basado en datos: las EDE se ajustan para describir regularidades empíricas (reversión y volatilidad) sin apoyarse en primeros principios biológicos o ecológicos. En este sentido, el modelamiento se considera una aproximación informada descriptiva, pero no necesariamente explicativa. Por ello, las trayectorias simuladas deben interpretarse como realizaciones estocásticas plausibles dadas las estadísticas estimadas, no como reproducciones punto a punto.

5.8.1. Simulaciones unidimensionales

La calidad del ajuste del modelo se evaluó de manera individual, simulando 1000 trayectorias con los parámetros estimados para cada larva y visualizando la banda de confianza del 90 % (percentiles 5 y 95). Se consideró aceptable si la trayectoria observada permanecía dentro de la banda al menos el 80 % del tiempo.

Los resultados mostraron una notable heterogeneidad en el desempeño del modelo, sugiriendo que los tests estadísticos no capturan completamente la variabilidad individual o que el modelo posee capacidad limitada para reproducir la amplia gama de dinámicas exhibidas por las larvas. A continuación, se presentan ejemplos de simulaciones exitosas y deficientes para cada modelo.

CIR



(a) Trayectoria del ID 69 que se contiene el 100 % del tiempo dentro de las bandas de confianza simuladas al 90 %.

(b) Trayectoria del ID 184 que no se ajusta a las bandas de confianza simuladas.

Figura 5.9: Ejemplos de trayectorias observadas comparadas con simulaciones bajo el modelo CIR.

La Figura 5.12 presenta ejemplos de la calidad de ajuste del modelo CIR. El histograma de porcentajes de contención (Figura 5.10) concentra una alta densidad de probabilidad en valores elevados, lo que indica que el modelo logra un buen ajuste para un gran número de trayectorias. Sin embargo, se observa una cola de bajo ajuste (porcentajes de contención menores al 20 %), confirmando que el modelo no es efectivo para todos los individuos.

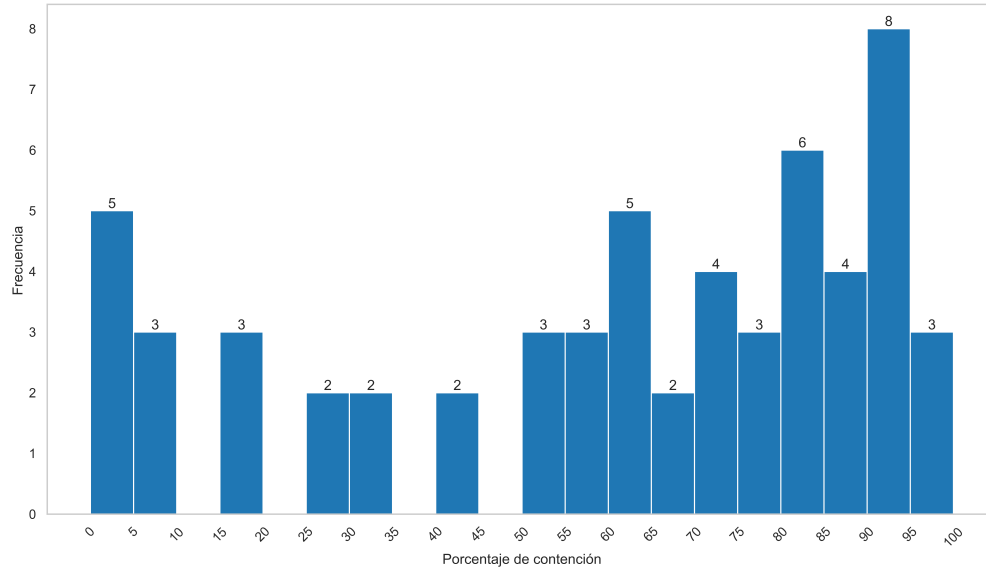


Figura 5.10: Histograma del porcentaje de tiempo en que cada trayectoria observada permanece dentro de las bandas de confianza (5–95 %) obtenidas por 1000 simulaciones CIR con sus parámetros estimados

Para tener una idea de cómo la salinidad influye en la calidad del ajuste, el diagrama de caja en la Figura 5.11 muestra que, en promedio, las trayectorias de salinidades más altas tienen un mejor ajuste. Este hallazgo es consistente con la menor heterogeneidad observada en los planos paramétricos para este grupo, lo que sugiere que el modelo estocástico captura de manera más homogénea las dinámicas de las larvas bajo estas condiciones ambientales.

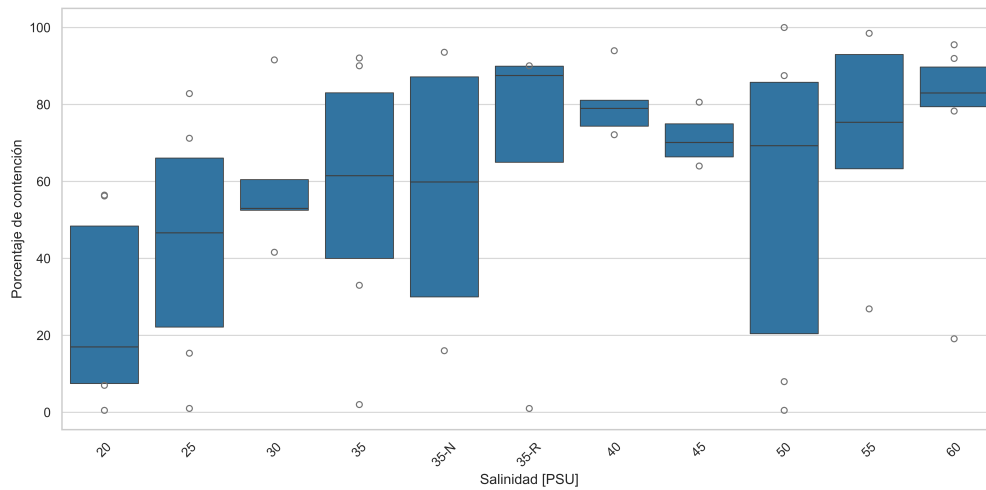
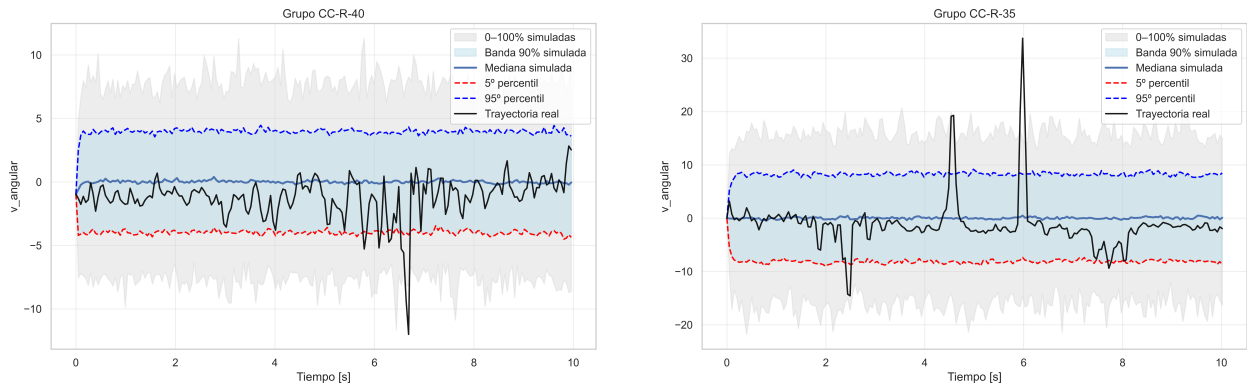


Figura 5.11: Diagrama de caja del porcentaje de contención por nivel de salinidad bajo el modelo CIR.

OU

En el modelo OU, el patrón de ajuste es dicotómico: una gran parte de las trayectorias exhibe un porcentaje de contención alto, mientras que un subconjunto muestra una contención muy baja. La

Figura 5.13 ilustra este comportamiento, mostrando una concentración de datos en los extremos del histograma. No se observa un gradiente por salinidad en los cuantiles superiores, por lo que el diagrama de caja en la Figura 5.14 se restringe a la fracción de contención alta para visualizar su estructura. Este resultado es consistente con la ausencia de diferencias funcionales por salinidad en la velocidad angular ($\omega(t)$) reportada en el Capítulo 4.



(a) Trayectoria del ID 57 que muestra un ajuste aceptable a las bandas de confianza simuladas al 90 % en los instantes temporales donde no hay saltos abruptos.

(b) Trayectoria del ID 48 que no se muestra un ajuste deficiente a las bandas de confianza simuladas en los intervalos donde $\omega(t)$ cambia abruptamente.

Figura 5.12: Ejemplos de trayectorias observadas comparadas con simulaciones bajo el modelo OU.

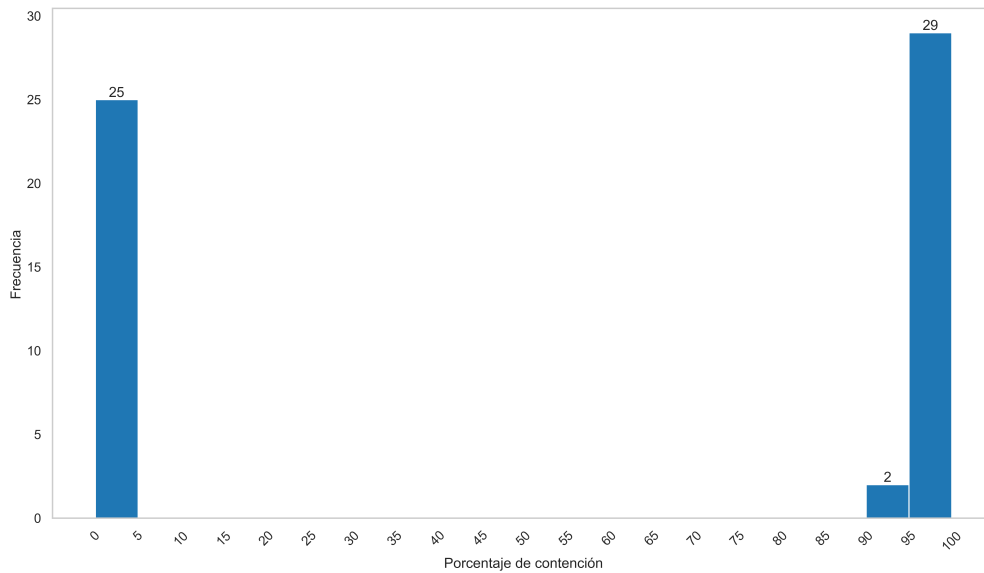


Figura 5.13: Histograma del porcentaje de tiempo en que cada trayectoria observada permanece dentro de las bandas de confianza (5-95 %) obtenidas por 1000 simulaciones OU con sus parámetros estimados.

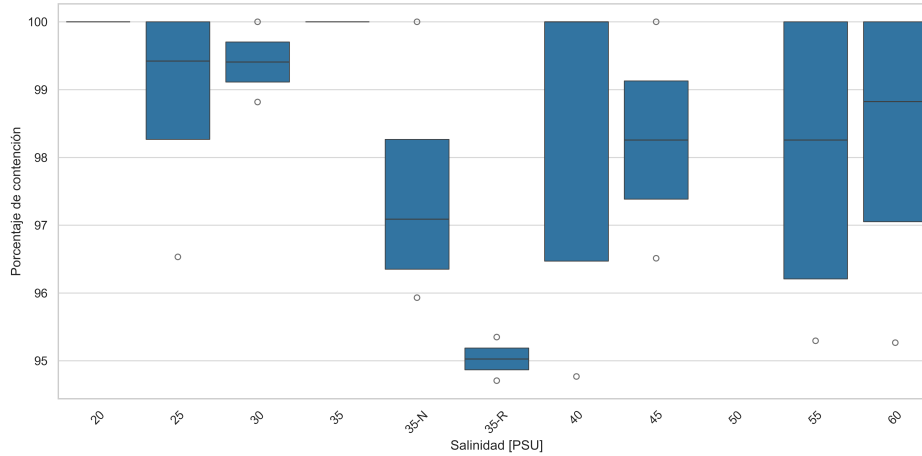


Figura 5.14: Diagrama de caja del porcentaje de contención por nivel de salinidad bajo el modelo OU. El gráfico solo considera los valores de contención superiores al 60 % para una mejor visualización de la estructura.

5.8.2. Simulaciones bidimensionales

Las simulaciones bidimensionales permiten comparar visualmente las trayectorias reales y las generadas. Las figuras siguientes presentan ejemplos representativos de estas simulaciones, mostrando cómo el modelo captura el comportamiento general de las larvas.

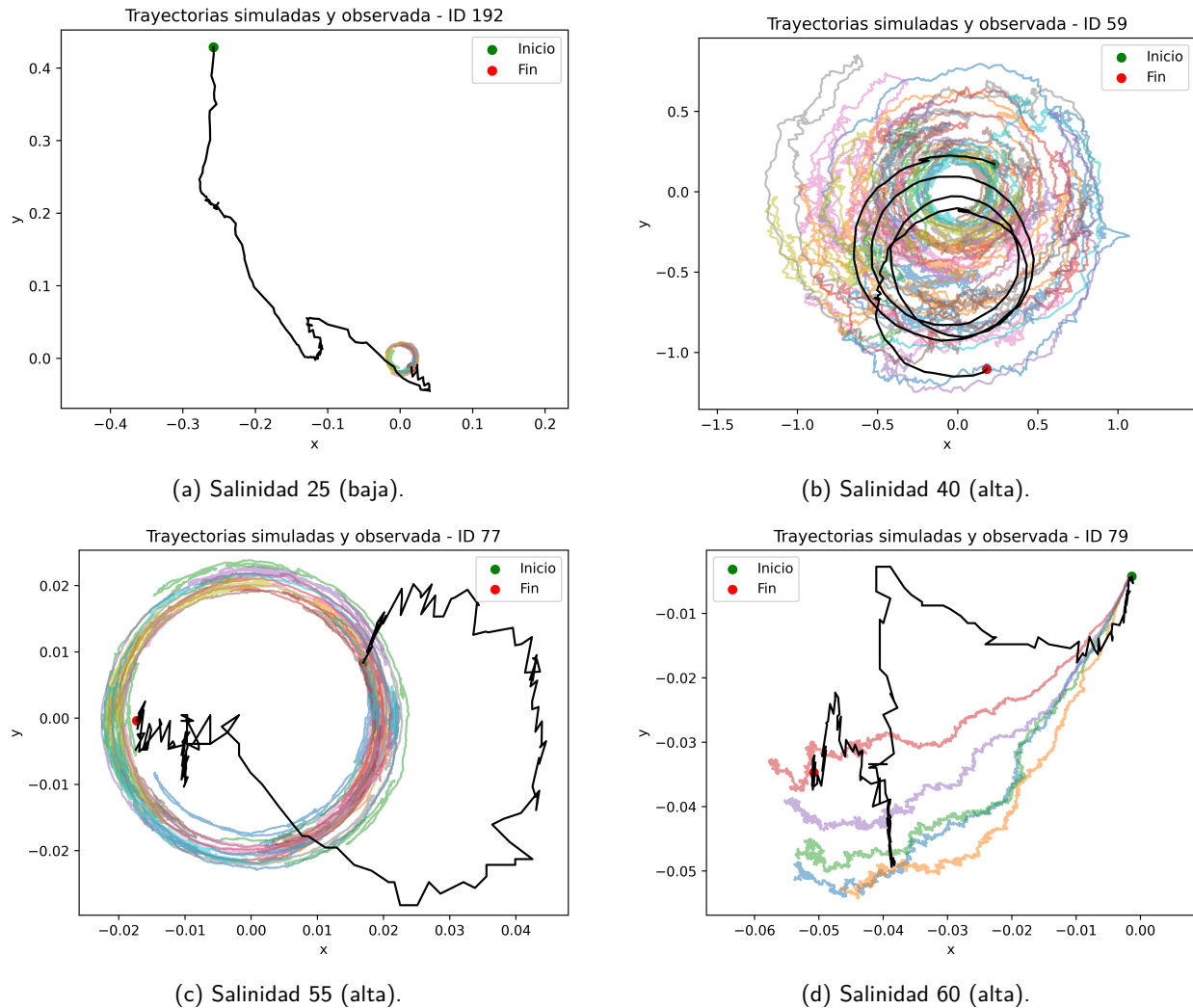


Figura 5.15: Ejemplos representativos de trayectorias simuladas en el plano a partir de $(\hat{r}_t, \hat{\theta}_t)$ obtenidos por Euler–Maruyama con parámetros estimados por ID; se superpone la trayectoria observada (en negro) para referencia.

En la Figura 5.15 se observa que el desempeño del modelo depende fuertemente de la salinidad. En salinidad baja (25 PSU) (panel Figura 5.15a) las simulaciones no logran replicar trayectorias erráticas y direccionales, generando en cambio patrones compactos y oscilatorios. En salinidad normal (40 PSU) (panel Figura 5.15b) y alta (55 PSU) (panel Figura 5.15c), el modelo muestra un desempeño moderado a bueno y deficiente, respectivamente, en el sentido de que captura bien las trayectorias circulares, pero no replica los comportamientos de exploración que se ven reflejados en cambios en la distancia al origen (r). Finalmente, en salinidad muy alta (60 PSU) (panel Figura 5.15d), las simulaciones vuelven a fallar, ya que la trayectoria real presenta un desplazamiento lineal y direccional que no es capturado por los movimientos oscilatorios generados por el modelo. Aun así cabe destacar la forma de las trayectorias simuladas, que parecen capturar la tendencia traslacional de la trayectoria, en lugar de exhibir un comportamiento puramente rotacional.

Estos resultados evidencian las limitaciones del modelo, tanto por su naturaleza empírica como por la variabilidad intrínseca de las trayectorias, que dificulta detectar puntos de cambio bien identificados en las variables de interés. Además, no es aplicable a *L. albus* ni a *C. concholepas* en exposición, lo que sin duda lo haría más robusto. En este sentido, modelos sobre alguna(s) de las variables significativas para larvas de erizo podrían capturar otro aspecto de la dinámica, como la aceleración y la velocidad. Un análisis de estas variables podría revelar patrones en la toma de decisiones direccionales o en la respuesta a cambios en la salinidad ambiental. Asimismo, sería interesante explorar si otras variables derivadas, como la curvatura de la trayectoria o la frecuencia de giros, son significativas para describir la transición entre movimientos lineales y circulares. Finalmente, también podría considerarse una dinámica acoplada entre componentes —por ejemplo, entre la posición radial y la velocidad angular— que permita representar interacciones más complejas entre los mecanismos de desplazamiento y orientación.

Capítulo 6

Conclusiones

Esta memoria de título ha buscado contribuir a la comprensión de los efectos de la variación de salinidad en la movilidad de larvas de *C. concholepas* y *L. albus* mediante un enfoque de modelamiento estadístico y estocástico. El trabajo cumplió con los objetivos planteados al examinar el efecto de la salinidad, proponer y calibrar un modelo estocástico para la dinámica de la movilidad larval, y evaluar su capacidad de replicar el comportamiento observado.

6.1. Análisis de resultados

El análisis descriptivo y de indicadores agregados confirmó que la salinidad es un factor determinante en la movilidad de ambas especies. Específicamente, las larvas de *C. concholepas* en fase de recuperación mostraron ser sensibles a salinidades extremas en su posición y velocidad radial. Del mismo modo, las larvas de *L. albus* exhibieron diferencias significativas en sus patrones de movimiento bajo distintas condiciones salinas. La variable de posición radial, $r(t)$, se estableció como un indicador principal de estrés ambiental en *C. concholepas*.

El FDA, y particularmente el fANOVA, complementaron estos hallazgos, evidenciando que $r(t)$ presenta diferencias funcionales significativas, principalmente entre las salinidades alta y baja, en el grupo experimental CC-R. Por su parte, la velocidad angular, $\omega(t)$, no mostró diferencias, lo que sugiere que el efecto de la salinidad se concentra en la dispersión más que en la orientación. Si bien se observaron contrastes con el análisis de métricas agregadas —como la ausencia de diferencias funcionales significativas en la velocidad radial, $\dot{r}(t)$, a pesar de haberlas en el análisis agregado— esto resalta las limitaciones inherentes a los métodos estadísticos y la importancia de complementar las distintas perspectivas de análisis.

El modelamiento estocástico, utilizando un proceso de Cox–Ingersoll–Ross (CIR) para la posición radial y un proceso de Ornstein–Uhlenbeck (OU) para la velocidad angular, permitió caracterizar la dinámica larval desde un punto de vista. Las simulaciones unidimensionales y bidimensionales, calibradas con los parámetros estimados, validaron cualitativamente la capacidad de este enfoque para reproducir los patrones de movimiento observados, como el confinamiento radial y las oscilaciones angulares. Se identificó que el modelo CIR mostró un mejor ajuste para las trayectorias bajo salinidades más altas en

C. concholepas, mientras que el modelo OU no mostró un patrón de ajuste claro, lo cual es coherente con los análisis previos que indicaban una menor sensibilidad de la velocidad angular a los cambios de salinidad. Es importante destacar que el modelo planteado es de naturaleza descriptiva y basada en datos, sin una explicación causal desde primeros principios biológicos, lo que lo convierte en una aproximación informada pero sujeta a las dinámicas observadas.

En base a estos resultados, se concluye que la variación de salinidad afecta directamente la capacidad de las larvas de *C. concholepas* para dispersarse desde un punto de origen, lo que podría tener implicaciones directas en su supervivencia y dinámica poblacional. El modelamiento estocástico se revela como una herramienta poderosa, aunque claramente no definitiva, para cuantificar la perturbación inducida por la salinidad, y puede servir como base para futuras investigaciones que busquen una representación más detallada de la compleja dinámica larval frente a perturbaciones ambientales.

Esta memoria aporta elementos modestos, pero relevantes, para la comprensión de los posibles efectos de las descargas de salmuera de plantas desaladoras en los ecosistemas costeros del país. Las especies estudiadas poseen un alto valor ecológico y económico, y la alteración de sus patrones natatorios en etapas tempranas debido a cambios en la salinidad ambiental podría tener consecuencias significativas en su supervivencia, dispersión, asentamiento y, en última instancia, en la estabilidad de sus poblaciones naturales.

6.2. Trabajo futuro

A pesar de los resultados obtenidos, el estudio presentó ciertas limitaciones. Se identificaron desequilibrios en la muestra asociados a las vibraciones externas y el número de individuos, lo que podría haber comprometido la potencia estadística de las estimaciones. Además, el tamaño muestral se volvió cada vez más restringido a lo largo de las etapas de preprocesamiento y limpieza de datos.

Se reconoció que los modelos propuestos tienen un carácter esencialmente descriptivo, lo que limita su capacidad para reproducir con precisión los comportamientos individuales observados y restringe su conexión con principios biológicos fundamentales. Además, se observó que el fANOVA basado en la media funcional puede resultar insuficiente cuando las diferencias entre grupos se manifiestan en la forma o variabilidad de las trayectorias, más que en sus medias. En tales casos, se recomienda emplear métodos de comparación funcional más robustos, como pruebas basadas en distancias L_2 , tests globalizados punto a punto (GPF) o enfoques de permutación (ver [103, 85, 84]).

Para consolidar y expandir los resultados presentados, se proponen varias líneas de trabajo a futuro. Un área fundamental es la exploración de otros tipos de ruido —por ejemplo, movimiento browniano fraccionario o Lévy— y el análisis de autocorrelación y memoria. Esto permitiría una representación más precisa de la persistencia o anti-persistencia en el movimiento larval. Estas ideas fueron exploradas de manera preliminar en esta memoria, pero no se desarrollaron en mayor profundidad.

Dado que las larvas son individuos propulsados, se sugiere investigar modelos estocásticos para nadadores, como los de partículas inerciales, que pueden capturar de manera más fiel la dinámica de un organismo que se mueve por sí mismo. Para una caracterización más detallada de las trayectorias, se

podrían aplicar técnicas de detección de puntos de cambio (*change points detection*) o modelos de tipo PPM¹, así como modelos autorregresivos continuos.

Adicionalmente, se propone un estudio más profundo de la densidad espectral de las trayectorias para analizar la distribución de la energía en las series de tiempo del movimiento. La incorporación de un componente periódico en el tiempo en los modelos permitiría representar las trayectorias oscilatorias, que se presentaron con una proporción relevante dentro de la muestra total. Finalmente, para una mejor clasificación de los patrones de los parámetros estimados, se sugiere implementar algoritmos de clasificación, como las máquinas de vectores de soporte (SVM).

Esta memoria de ingeniería ha mostrado cómo un enfoque cuantitativo, apoyado en el análisis funcional de datos y en el modelamiento estocástico, puede contribuir a explorar los efectos de la salinidad en la movilidad larval de especies relevantes para el país. Si bien los modelos propuestos presentan limitaciones y solo reproducen parcialmente los patrones observados, entregan una primera aproximación que abre la puerta a futuros desarrollos más precisos, y a un diálogo más estrecho entre la matemática y la biología en el estudio de los ecosistemas marinos.

¹*Prediction by Partial Matching*

Apéndice A

Nota sobre las unidades prácticas de salinidad (PSU, *Practical Salinity Units*)

PSU es una medida comúnmente utilizada en oceanografía para expresar la salinidad del agua de mar. Sin embargo, es importante tener en cuenta que no es una unidad formalmente definida en el Sistema Internacional de Unidades (SI), sino una escala práctica basada en la conductividad eléctrica del agua.

1 PSU es aproximadamente equivalente a 1 g de sal por kg de agua. La *Practical Salinity Scale 1978* (PSS-78) introdujo la unidad PSU como una medida práctica de la salinidad basada en la conductividad eléctrica del agua de mar. No tiene unidades físicas.

Desde principios de la década de 1980, los oceanógrafos han utilizado un valor calculado denominado formalmente *Salinidad Práctica* (representado como S_P) como un proxy de la salinidad verdadera. Esta salinidad se define como una función de la temperatura, la presión y la conductividad eléctrica, según el estándar conocido como la *Practical Salinity Scale 1978* (PSS-78) [?]. En la práctica oceanográfica, el término “salinidad” suele referirse a la salinidad práctica, aunque es preferible utilizar el nombre completo para evitar ambigüedades terminológicas.

Es importante destacar que la salinidad práctica no posee unidades físicas. Esta característica, que puede resultar confusa para no especialistas, se debe a limitaciones técnicas que impidieron una definición absoluta durante la formulación de la *Practical Salinity Scale 1978* (PSS-78) [104]. En ocasiones, esta ausencia de unidades se maneja de forma incorrecta mediante la adición del acrónimo PSU (*Practical Salinity Units*) al valor numérico, lo cual es formalmente erróneo y desaconsejado por organismos oceanográficos [105].

Desde el punto de vista cuantitativo, los valores de salinidad práctica son aproximadamente 0,5 unidades menores que la fracción másica de materia disuelta, cuando esta se expresa en gramos de soluto por kilogramo de agua de mar. No obstante, la PSS-78 fue diseñada para mantener continuidad histórica con los valores numéricos basados en clorinidad, asegurando así una comparabilidad razonable entre escalas [104].

- 1 PSU \approx 1 ppt (partes por mil)

Bibliografía

- [1] Marc A. Rosen and Aida Farsi. *Sustainable energy technologies for seawater desalination*. Academic Press, 2022. [13](#)
- [2] Amaya Alvez, Douglas Aitken, Diego Rivera, Marcelo Vergara, Neil McIntyre, and Fernando Concha. At the crossroads: can desalination be a suitable public policy solution to address water scarcity in Chile's mining zones? *Journal of Environmental Management*, 258:110039, 2020. [13](#)
- [3] David A. Roberts, Emma L. Johnston, and Nathan A. Knott. Impacts of desalination plant discharges on the marine environment: A critical review of published studies. *Water Research*, 44(18):5117–5128, Enero 2010. [13](#)
- [4] Patricio H. Manríquez, Claudio P. González, Alejandro Abarca, Katherine Jenó, Viviana Jofré, Orlando Astudillo, and Victor M. Aguilera. Effects of brine discharges on newly hatched larvae of the rocky-shore keystone gastropod *Concholepas concholepas*. *Desalination*, 581:117579, 2024. [13](#), [14](#), [17](#), [18](#), [19](#), [22](#), [58](#)
- [5] Mauricio Cerda, Bastiaan Knoppers, Jorge Valdés, Abdel Fettah Siffedine, Luc Ortlieb, and Elisamara Sabadini-Santos. Variación espacial y temporal de las masas de agua, nutrientes y sedimentación de la materia orgánica e inorgánica en la bahía Mejillones del sur (23° S), Chile. *Revista Chilena de Historia Natural*, 83(3):409–420, 2010. [13](#)
- [6] Instituto de Fomento Pesquero (IFOP). Evaluación hidroacústica del recurso jurel (*trachurus murphyi*) en la zona económica exclusiva de Chile. Technical Report FIPA 1998-11, IFOP, Fondo de Investigación Pesquera y de Acuicultura (FIPA), Subsecretaría de Pesca y Acuicultura, Valparaíso, Chile, 1998. Jefe de proyecto: José Córdova. [13](#)
- [7] Gary Shaffer, Samuel E. Hormazabal, Oscar Pizarro, and Sergio Salinas. Dynamics of seasonal and interannual variability of the Peru-Chile Undercurrent. *Geophysical Research Letters*, 29(24):10–1, 2002. [13](#)
- [8] Fabio Blanco-Murillo. A risk assessment on *Zostera chilensis*, the last relict of marine angiosperms in the South-East Pacific Ocean, due to the development of the desalination industry in Chile. In *Environmental Footprint of Water Desalination: Addressing Desalination Environmental Impact Amidst the Water Crisis*, pages 307–319. Springer, 2024. [13](#)

- [9] Patricio H. Manríquez, Claudio P. González, Alejandro Abarca, Viviana Jofré, Orlando Astudillo, and Victor M. Aguilera. Behavioural and physiological responses of the surf clam *Mesodesma donacium* to brine discharges. *Science of The Total Environment*, 922:171233, 2024. 13
- [10] Elizabeth Soto Muñoz, Pablo Madrid Arancibia, and Hernán Ramírez Rueda. Herramientas para el debate sobre desalinización: potenciales impactos y escenario en Chile. Technical report, Fundación Terram, Santiago, Chile, 2025. 14, 21
- [11] Escuela de Ingeniería Civil Oceánica. Laboratorio de procesos costeros (labocéano). Technical report, Facultad de Ingeniería, Universidad de Valparaíso, 2021. Documento institucional sobre infraestructura y capacidades técnicas del LabOcéano. 14
- [12] Patricio Winckler. Procesamiento de imágenes para bioensayos con larvas de *Concholepas concholepas* y *Loxechinus albus* expuestas a distintos niveles de salinidad. Proyecto IDRO: Evaluación y modelación de impactos ambientales de plantas desalinizadoras por ósmosis inversa (RO), 2024. 14, 22, 23, 27
- [13] Adrián Ortiz, Cristian Duarte, and Patricio Winckler. Evaluación y modelo de impactos ambientales de plantas desalinizadoras por osmosis inversa. <https://idrochile.cl/>, 2024. Proyecto financiado por ANID IDEA I+D (ID23110336). Instituciones participantes: USM, UV, UNAB. 14, 22, 32, 37
- [14] Proyecto SWAM. Sea, waves, and ecosystems: Stochastic models for perturbed marine environments. Proyecto institucional, 2024. Proyecto asociado entre Inria, Universidad de Valparaíso, Universidad de Concepción y otras instituciones. Coordinadores: Mireille Bossy y Kerlyns Martínez. 14
- [15] Elie Poulin, Álvaro T. Palma, Germán Leiva, Eduardo Hernández, Pilar Martínez, Sergio A. Navarrete, and Juan C. Castilla. Temporal and spatial variation in the distribution of epineustonic competent larvae of *concholepas concholepas* along the central coast of Chile. *Marine Ecology Progress Series*, 229:95–104, 2002. 17
- [16] Facultad de Ciencias Biológicas. Surgencia: Un fenómeno oceanográfico. <https://ecim.bio.puc.cl/boletin-remam/surgencia-un-fenomeno-oceanografico/>, n.d. Estación Costera de Investigaciones Marinas. Recuperado el 19 de julio de 2025. 18
- [17] Carlos Molinet, Alejandra Arévalo, María Teresa González, Carlos A. Moreno, Javier Arata, and Edwin Niklitschek. Patterns of larval distribution and settlement of *concholepas concholepas* (Bruguere, 1789) (Gastropoda, Muricidae) in fjords and channels of southern Chile. *Revista Chilena de Historia Natural*, 78(3):409–423, 2005. 18
- [18] Sebastián Flores Valiente. Modelación del transporte larval de *Argopecten purpuratus* en un sistema fuente-sumidero: Bahía Independencia, Perú. Tesis de magíster en ciencias con mención en oceanografía, Universidad de Concepción, 2016. 18

- [19] H. Reyes and C. A. Moreno. Asentamiento y crecimiento de los primeros estadios bentónicos de concholepas concholepas (bruguière, 1789) en mehuín, Chile. *Estudios Oceanológicos*, 8:67–76, 1989. [18](#)
- [20] Carlos Gallardo. Efecto de diferentes salinidades en la sobrevivencia embrionaria y desarrollo intracapsular del gastrópodo *Concholepas concholepas* (bruguière, 1789) bajo condiciones de laboratorio. *Revista de Biología Marina*, 29(2):263–279, 1994. [19](#)
- [21] Eduardo Bustos and Florencia Navarrete. *Manual: El cultivo del loco (Concholepas concholepas)*. Chile, 2001. Proyecto FONDEF D96 I 1101. [19](#)
- [22] R. H. Bustamante and J. C. Castilla. Impacto ecológico de la extracción de erizos. *Revista Chilena de Historia Natural*, 63:73–82, 1990. [19](#)
- [23] Juan C. Castilla and María Varas. Impacto de manejo de recursos bentónicos en Chile. *Revista Chilena de Historia Natural*, 71(2):155–167, 1998. [20](#)
- [24] Carlos Molinet, Carlos A. Moreno, Edwin J. Niklitschek, Manira Matamala, Mónica Neculman, Alejandra Arévalo, José Codjambassis, Patricio Díaz, and Manuel Díaz. Reproduction of the sea urchin *Loxechinus albus* across a bathymetric gradient in the Chilean inland sea. *Revista de Biología Marina y Oceanografía*, 47(2):257–272, August 2012. [20](#)
- [25] Pedro F. Cárcamo. Massive production of larvae and seeds of the sea urchin *Loxechinus albus*. In J. M. Lawrence and O. Guzmán, editors, *Sea Urchins: Fisheries and Ecology*, pages 299–306. DEStech Publications Inc., 2004. [20](#)
- [26] Anne Mercier, Zhen Sun, and Jean-François Hamel. Salinity-induced developmental shifts in echinoderms. *Marine Ecology Progress Series*, 374:209–218, 2009. [20](#), [22](#)
- [27] Eduardo Bustos and Sergio Olave. *Manual: El cultivo del erizo (Loxechinus albus)*. Chile, 2001. Proyecto FONDEF D96 I 1101. [20](#)
- [28] E. M. Purcell. Life at low Reynolds number. *American Journal of Physics*, 45(1):3–11, 1977. [22](#)
- [29] Howard C. Berg. *Random Walks in Biology*. Princeton University Press, Princeton, NJ, expanded edition, 1993. [22](#)
- [30] J. Happel and H. Brenner. *Low Reynolds Number Hydrodynamics*. Springer Netherlands, 1983. [23](#)
- [31] Jorge Alvarado, Laura Huaquín, Hugo Carrillo, Patricio Manríquez, and otros. Comportamiento y parámetros reproductivos de concholepas concholepas en la VIII y X regiones. Informe técnico FIP 2006-24, Universidad Austral de Chile, 2009. [24](#)
- [32] F. J. Millero. Sea water as an electrolyte. In A. Gianguzza, E. Pelizzetti, and S. Sammartano, editors, *Chemistry of Marine Water and Sediments*, pages 3–34. Springer-Verlag Berlin Heidelberg, 2002. [24](#)

- [33] F. N. Fritsch and R. E. Carlson. Monotone piecewise cubic interpolation. *SIAM Journal on Numerical Analysis*, 17(2):238–246, 1980. 25
- [34] Jeanette D. Wheeler, Kit Yu Karen Chan, Erik J. Anderson, and Lauren S. Mullineaux. Ontogenetic changes in larval swimming and orientation of pre-competent sea urchin *arbacia punctulata* in turbulence. *Journal of Experimental Biology*, 219(9):1303–1310, 2016. 35
- [35] Christophe Croux and Catherine Dehon. Influence functions of the spearman and kendall correlation measures. *Statistical Methods & Applications*, 19(4):497–515, 2010. 41
- [36] Larry Wasserman. *All of Statistics: A Concise Course in Statistical Inference*. Springer, New York, 2004. 44
- [37] Ronald A. Fisher. *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburgh, 1925. 44, 45
- [38] Douglas C. Montgomery. *Design and Analysis of Experiments*. John Wiley & Sons, Hoboken, NJ, 9th edition, 2017. 45
- [39] Bernard L. Welch. On the comparison of several mean values: An alternative approach. *Biometrika*, 38(3/4):330–336, 1951. 45
- [40] F. E. Satterthwaite. An approximate distribution of estimates of variance components. *Biometrics Bulletin*, 2(6):110–114, 1946. 46
- [41] William H. Kruskal and W. Allen Wallis. Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, 47(260):583–621, 1952. 46, 47
- [42] Myles Hollander, Douglas A. Wolfe, and Eric Chicken. *Nonparametric Statistical Methods*. John Wiley & Sons, 3rd edition, 2013. 47
- [43] Samuel S Shapiro and Martin B Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3-4):591–611, 1965. 48
- [44] Maurice Stevenson Bartlett. Properties of sufficiency and statistical tests. *Proceedings of the Royal Statistical Society, Series A*, 160:268–282, 1937. 48, 49
- [45] Howard Levene. Robust tests for equality of variances. In Ingram Olkin, S. G. Ghurye, Wassily Hoeffding, C. Reid, and John W. Tukey, editors, *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, pages 278–292. Stanford University Press, 1960. 49
- [46] George W. Snedecor and William G. Cochran. *Statistical Methods*. Iowa State University Press, 8th edition, 1989. 49
- [47] John W. Tukey. Comparing individual means in the analysis of variance. *Biometrics*, 5(2):99–114, 1949. 50

- [48] Paul A. Games and John F. Howell. Pairwise multiple comparison procedures with unequal n's and/or variances: A monte carlo study. *Journal of Educational Statistics*, 1(2):113–125, 1976. [50](#)
- [49] Olive J. Dunn. Multiple comparisons using rank sums. *Technometrics*, 6(3):241–252, 1964. [50](#)
- [50] Massimiliano Giacalone, Agata Zirilli, Paolo C. Cozzucoli, and Angela Alibrandi. Bonferroni-holm and permutation tests to compare health data: methodological and applicative issues. *BMC Medical Research Methodology*, 18(81), 2018. [51](#)
- [51] Sture Holm. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6:65–70, 1979. [51](#)
- [52] Kazuo Okada. Is omega squared less biased than eta squared? reexamining the issue from a simulation study. *Journal of Research in Psychology*, 47(3):253–261, 2013. [52](#)
- [53] H. J. Keselman. The problem of overestimating the effect size. *Journal of Educational and Psychological Measurement*, 35(1):321–324, 1975. [52](#)
- [54] D. Lakens. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and anovas. *Frontiers in Psychology*, 4:863, 2013. [52](#)
- [55] Maciej Tomczak and Ewa Tomczak. The need to report effect size estimates revisited: An overview of some recommended measures of effect size. *Trends in Sport Sciences*, 21(1):19–25, 2014. [52](#), [55](#)
- [56] Jacob Cohen. *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Erlbaum Associates, Hillsdale, NJ, 2nd edition, 1988. [55](#), [56](#)
- [57] JM Navarro and RA Torrijos. Fisiología energética de concholepas concholepas (bruguière, 1789)(gastropoda: Muricidae) en la bahía de yaldad, sur de chile. *Revista Chilena de Historia Natural*, 68:61–77, 1995. [58](#), [59](#)
- [58] William B. Stickle and William J. Diehl. Effects of salinity on echinoderms. *Advances in Marine Biology*, 23:197–255, 1986. [58](#)
- [59] Charlotte P. Mangum. Osmoregulation in marine and estuarine animals: Its influence on respiratory gas exchange and transport. *Italian Journal of Zoology*, 53(1):1–7, 1986. [59](#)
- [60] J.O. Ramsay and B.W. Silverman. *Functional Data Analysis*. Springer, New York, 2005. [61](#)
- [61] M. Febrero-Bande and M. Oviedo de la Fuente. Statistical computing in functional data analysis: The r package fda.usc. *Journal of Statistical Software*, 51(4):1–28, 2012. [61](#)
- [62] Piotr Kokoszka and Matthew Reimherr. *Introduction to Functional Data Analysis*. CRC Press, 2017. [61](#), [62](#), [89](#)

- [63] Carlos Ramos-Carreño, José Luis Torrecilla, and Alberto Suárez. Classification of functional data: A comparative study. *Proceedings of the 21st IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 866–871, 2022. [62](#)
- [64] Paul H. C. Eilers and Brian D. Marx. Flexible smoothing with b-splines and penalties. *Statistical Science*, 11(2):89–121, 1996. [62](#), [63](#), [65](#), [66](#), [69](#)
- [65] Paul Dierckx. *Curve and Surface Fitting with Splines*. Oxford University Press, Oxford, 1993. [62](#)
- [66] Larry L. Schumaker. *Spline Functions: Computational Methods*. Society for Industrial and Applied Mathematics, Philadelphia, 2015. [62](#)
- [67] M. G. Cox. The numerical evaluation of b-splines. *Journal of the Institute of Mathematics and its Applications*, 10(2):134–149, 1972. [63](#)
- [68] Carl de Boor. On calculating with b-splines. *Journal of Approximation Theory*, 6(1):50–62, 1972. [63](#)
- [69] Paul H. C. Eilers, Brian D. Marx, and María Durbán. P-splines: a tool for functional regression. *Journal of Computational and Graphical Statistics*, 24(1):1–22, 2015. [63](#), [69](#)
- [70] Paul H. C. Eilers and Brian D. Marx. Why p-splines?, 2021. Documento que acompaña al libro "Practical Smoothing: The Joys of P-splines", disponible online. [63](#), [69](#), [70](#)
- [71] Elias M Stein and Rami Shakarchi. *Fourier analysis: an introduction*, volume 1. Princeton University Press, 2003. [64](#)
- [72] George A. Papakostas and Dimitrios A. Karras. Fast numerically stable computation of orthogonal fourier–mellin moments. *IET Computer Vision*, 1(1):1–16, 2007. [64](#)
- [73] Ingrid Daubechies. *Ten Lectures on Wavelets*, volume 61 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. SIAM, 1992. [65](#)
- [74] A. Graps. An introduction to wavelets. *IEEE Computational Science and Engineering*, 2(2):50–61, 1995. [65](#)
- [75] A. Cohen. *Wavelets and their Applications*. Jones and Bartlett Publishers, Boston, 1992. [65](#)
- [76] Arthur E. Hoerl and Robert W. Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970. [65](#)
- [77] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York, 2nd edition, 2009. [65](#)
- [78] O. Sysoev and O. Burdakov. A smoothed monotonic regression via l2 regularization. *Knowledge and Information Systems*, 59:1–20, 2019. [65](#), [66](#)

- [79] Carlos Ramos-Carreño, José L. Torrecilla, Miguel Carbajo Berrocal, Pablo Marcos Manchón, and Alberto Suárez. scikit-fda: A python package for functional data analysis. *Journal of Statistical Software*, 109(2):1–37, May 2024. [66](#), [67](#)
- [80] The scikit-fda developers. scikit-fda: Functional data analysis in python, Febrero 2024. [66](#)
- [81] David Ruppert, M. P. Wand, and R. J. Carroll. *Semiparametric Regression*. Cambridge University Press, Cambridge, 2003. [66](#)
- [82] Simon N. Wood. *Generalized Additive Models: An Introduction with R*. Chapman & Hall/CRC, Boca Raton, 2017. [66](#), [69](#)
- [83] Antonio Cuevas, Manuel Febrero, and Ricardo Fraiman. An anova test for functional data. *Computational Statistics & Data Analysis*, 47(1):111–122, 2004. [66](#), [67](#)
- [84] Tomasz Górecki and Łukasz Smaga. A comparison of tests for the one-way anova problem for functional data. *Computational Statistics*, 30(4):931–957, 2015. [81](#), [120](#)
- [85] Xiaohui Zhang and Jin-Ting Zhang. One-way anova for functional data via globalizing the point-wise f-test. *Scandinavian Journal of Statistics*, 41(1):51–71, 2014. [81](#), [120](#)
- [86] Ioannis Karatzas and Steven E. Shreve. *Brownian Motion and Stochastic Calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer, 2 edition, 1991. [95](#), [96](#)
- [87] Bernt Øksendal. *Stochastic Differential Equations: An Introduction with Applications*. Universitext. Springer, 6 edition, 2003. [96](#)
- [88] Peter E. Kloeden and Eckhard Platen. *Numerical Solution of Stochastic Differential Equations*, volume 23 of *Applications of Mathematics*. Springer, 1992. [96](#), [99](#)
- [89] Alejandra María López Pérez. Modelos de tipo de interés: revisión y comparativa de métodos de estimación. Trabajo de fin de máster en técnicas estadísticas, Universidade de Santiago de Compostela, 2018. [96](#)
- [90] Oldrich A. Vasicek. An equilibrium characterisation of the term structure. *Journal of Financial Economics*, 5:177–188, 1977. [97](#)
- [91] Kamil Kládívko. Maximum likelihood estimation of the cox-ingersoll-ross process: The MATLAB implementation. *Journal of Computational Finance*, 18:1–19, 2014. [97](#), [100](#), [101](#), [102](#), [104](#), [105](#)
- [92] K. Fergusson and E. Platen. Application of maximum likelihood estimation to stochastic short rate models. *Annals of Financial Economics*, 10:1–26, 2015. [97](#), [99](#), [100](#), [102](#)
- [93] William Feller. Two singular diffusion problems. *Annals of Mathematics*, 54(1):173–182, 1951. [97](#)
- [94] John C. Cox, Jonathan E. Ingersoll, and Stephen A. Ross. A theory of the term structure of interest rates. *Econometrica*, 53(2):385–407, 1985. [97](#)

- [95] Eduardo Selim Martínez Mayorga. *Simulación y estimación en tres procesos de difusión fundamentales: Ornstein-Uhlenbeck, Cox-Ingersoll-Ross y Wright Fisher*. Tesis de licenciatura, Universidad Nacional Autónoma de México (UNAM), 2017. [97](#), [99](#), [100](#), [101](#)
- [96] Yacine Aït-Sahalia. Maximum likelihood estimation of discretely sampled diffusions: A closed-form approximation approach. *Econometrica*, 70:223–262, 2002. [99](#), [102](#), [103](#)
- [97] H. Sørensen. Parametric inference for diffusion processes observed at discrete points in time: a survey. *International Statistical Review*, 72:337–354, 2004. [99](#)
- [98] C. Y. Tang and S. X. Chen. Parameter estimation and bias correction for diffusion processes. *Journal of Econometrics*, 149:65–81, 2009. [99](#)
- [99] Paul Mullenney and Satish Iyengar. Maximum likelihood estimation and computation for the ornstein-uhlenbeck process. Technical report, Department of Statistics, University of Pittsburgh, 2008. [99](#), [100](#), [103](#)
- [100] Luis Valdivieso, Wim Schoutens, and Francis Tuerlinckx. Maximum likelihood estimation in processes of ornstein-uhlenbeck type. *Statistical Inference for Stochastic Processes*, 12(1):1–19, 2009. [100](#)
- [101] Vladimír Holý and Petra Tomanová. Estimation of ornstein-uhlenbeck process using ultra-high-frequency data with application to intraday pairs trading strategy. *Submitted to a journal or conference*, 2022. [102](#), [103](#)
- [102] Junwen Yao, Jonas Mueller, and Jane-Ling Wang. Deep learning for functional data analysis with adaptive basis layers. *Proceedings of the 38th International Conference on Machine Learning*, 139:11898–11908, 2021. [112](#)
- [103] J.T. Zhang. *Analysis of Variance for Functional Data*. CRC Press, Boca Raton, 2013. [120](#)
- [104] UNESCO. *Background Papers and Supporting Data on the Practical Salinity Scale 1978*, volume 37 of *UNESCO Technical Papers in Marine Science*. UNESCO, Paris, 1981. [122](#)
- [105] SCOR/IAPSO Working Group 127. *The International Thermodynamic Equation of Seawater – 2010: Abridged Version*, volume 56 of *IOC Manuals and Guides*. IOC-UNESCO, Paris, 2010. [122](#)