

UNIVERSITAT POLITÈCNICA DE VALENCIA
Departamento de Estadística e Investigación Operativa Aplicadas y
Calidad



PROGRAMA DE DOCTORADO EN ESTADÍSTICA Y
OPTIMIZACIÓN

**Monitorización y Predicción de Procesos Multivariantes
y Autocorrelacionados. Aplicación a la producción de
Tilapia en Mozambique.**

Autor: Sérgio Afonso Mulema
Dirigido por: Andrés Carrión García

VALENCIA, 2018

A mi familia y amigos.

Agradecimientos

Agradezco a Dios por la vida, la salud y fuerzas para realizar esta tesis. A mi Director, Profesor Andrés Carrión García, por su orientación y apoyo en todas las etapas del trabajo.

A toda mi familia: esposa, hija, madre, hermanos, que me dieron el apoyo moral para seguir con esta investigación.

En el ámbito profesional, agradezco al Profesor Carlos Lauchande, por su apoyo administrativo y financiero. A los Dirigentes de la Universidad Pedagógica de Mozambique: Profesor Manuel Morais, Mestre Dulce Passades, Stela Pinto Novo, Luck Njange, Fauzia Ramos, Acácio Muluta y Teodosio de Jesus Cosme por la concesión de la beca con la cual he podido realizar esta tesis. A los colegas del *Departamento de Ciências Naturais e Matemática da Universidade Pedagógica-Quelimane*, especialmente a los del curso de Matemática.

A mis amigos: Tanleque, Jackson, Ginet, Ulrich, Mendel y Acácio Don Luis con quien compartía las dificultades enfrentadas a lo largo de la tesis.

A todos los profesores y colegas del Máster en Ingeniería de Análisis de Datos, Mejora de Procesos y Toma de Decisiones, impartido en la Universidad Politécnica de Valencia, y a todos los profesores y colegas del Máster en Estadística impartido en la Universidad Pedagógica de Mozambique en colaboración con la Universidad Complutense de Madrid.

A la dirección de la empresa Aquapesca.Lda de Mozambique, especialmente al Mestre Vicente por la practica concedida.

A todos, muchas gracias.

Resumen

La presente tesis aborda los métodos estadísticos de monitorización y predicción de procesos multivariantes y autocorrelacionados en la industria. La investigación se orienta, por un lado, al estudio de los métodos de monitorización que culmina con el desarrollo de un gráfico de control de procesos; y por otro lado, al análisis de los métodos de predicción, donde se elabora el modelo dinámico sobre estructuras latentes (MDEL). En los dos casos, los modelos de series temporales constituyen una herramienta fundamental para modelar la estructura de autocorrelación en los datos.

El estudio del modelo de monitorización sigue un orden lógico de complejidad de los métodos de control de procesos usados en la industria. Se empieza por analizar los gráficos de control univariante; los gráficos multivariantes, y sus limitaciones en el control de procesos autocorrelacionados, con lo cual se introducen los gráficos multivariantes para procesos autocorrelacionados, donde se desarrolla el gráfico de control *multivariate autocorrelated and adapted EWMA chart* (MAAEWMA).

El desempeño del gráfico MAAEWMA fue comparado con el gráfico MEWMA ajustado a ruidos blancos. Los resultados mostraron que el gráfico MAAEWMA es más rápido en detectar cambios en la media del proceso, y es más eficiente para procesos con media y alta autocorrelación.

Por otro lado, el método predictivo fue elaborado para modelar los procesos multivariantes y autocorrelacionados. Para ello, se estudiaron los métodos sobre estructuras latentes PCA y PLS, y sus respectivas versiones dinámicas: dynamic PCA (DPCA) y dynamic PLS (DPLS).

El desarrollo del modelo dinámico sobre estructuras latentes (MDEL) consistió en la combinación del modelo DPCA y de la función de transferencia (FT). El modelo DPCA fue aplicado a los datos de entrada y de salida del proceso de manera separada para determinar las variables latentes que recogen la máxima variabilidad dinámica en los datos. La estructura dinámica en las variables latentes resultantes fue modelada mediante una función de transferencia para analizar la influencia de los datos de entrada sobre los datos de salida.

Los modelos MDEL y DPLS fueron ajustados en condiciones idénticas a un conjunto de datos simulados. La calidad de ajuste de los modelos fue evaluada mediante la variabilidad explicada de las variables latentes (R^2), y la capacidad predictiva fue testada mediante validación cruzada, donde fue calculado el mean absolute error (MAE) de los dos modelos. El MDEL presentó mejores resultados comparado al DPLS, con mayor variabilidad explicada por

las variables latentes, y con una capacidad predictiva robusta y estable.

Los modelos propuestos fueron aplicados para monitorizar y predecir el proceso de cría de tilapia en Mozambique. El gráfico MAAEWMA señaló cambios significativos del ambiente acuático en el proceso, y con el modelo MDEL se analizó el impacto de dichos cambios en el crecimiento de la tilapia.

Abstract

This thesis is based on statistical methods to monitor and predict the multivariate and autocorrelated industrial processes. The research is oriented on the monitoring methods that culminate with the development of a process control chart and analysis of the prediction methods, where the dynamic model on latent structures (MDEL) was developed. In both cases, the time series models were used to model the autocorrelation structure in the data.

The analysis of the monitoring models follows a logical order of complexity of the process control methods used in the industry. The univariate and multivariate charts were analysed, and their limitations in the control of autocorrelated processes were considered. Therefore, the multivariate charts for autocorrelated processes were introduced and the multivariate, autocorrelated and adapted EWMA chart (MAAEWMA) was developed. The performance of the MAAEWMA chart was compared with the MEWMA chart adjusted to white noise. The results showed that the MAAEWMA chart was faster in detecting changes in the process mean and was more efficient for processes with medium and high autocorrelation.

On the other hand, the predictive method was developed to model the multivariate and autocorrelated processes. To this end, the methods on latent structures PCA and PLS and their respective dynamic versions: dynamic PCA (DPCA) and dynamic PLS (DPLS) were studied.

The dynamic model on latent structures (MDEL) consists of the combination of DPCA model and the transfer function (FT). The DPCA model was applied separately to the input and output data of the process. This strategy was applied to guarantee the maximum dynamic variability in the latent variables. The dynamic structure in the latent variables was modelled by a transfer function to analyse the influence of the input on the output data.

The MDEL and DPLS models were adjusted in identical conditions to a set of simulated data. The goodness of fit was evaluated by the explained variability of the latent variables (R^2) and the predictive capacity of the models was tested through cross validation, where the mean absolute error (MAE) was calculated. The MDEL presented better results than DPLS model, with greater variability explained by the latent variables and a robust and stable predictive capacity.

The proposed models were applied to monitor and predict the tilapia crop process in Mozambique. The MAAEWMA chart indicated significant changes in the aquatic environment and with the MDEL was analysed the impact of these changes on the growth of tilapia.

Resum

La present tesi aborda els mètodes estadístics de monitoratge i predicció de processos multivariants i autocorrelacionados en la indústria. La investigació s'orienta, d'una banda, a l'estudi dels mètodes de monitorització que culmina amb el desenvolupament d'un gràfic de control de processos; i d'altra banda, a l'anàlisi dels mètodes de predicció, on s'elabora el model dinàmic sobre estructures latents (mdel). En els dos casos, els models de sèries temporals constitueixen una eina fonamental per a modelar l'estructura d'autocorrelació en les dades.

L'estudi del model de monitoratge segueix un ordre lògic de complexitat dels mètodes de control de processos usats en la indústria. Es comença per analitzar els gràfics de control univariants; els gràfics multivariants, i les seves limitacions en el control de processos autocorrelacionados, amb la qual cosa s'introdueixen els gràfics multivariants per a processos autocorrelacionados, on es desenvolupa el gràfic de control multivariate autocorrelated and adapted EWMA chart (MAAEWMA).

L'acompliment del gràfic MAAEWMA va ser comparat amb el gràfic MEWMA ajustat a sorolls blancs. Els resultats van mostrar que el gràfic MAAEWMA és més ràpid en detectar canvis en la mitjana del procés, i és més eficient per a processos amb mitja i alta autocorrelació.

D'altra banda, el mètode predictiu va ser elaborat per modelar els processos multivariants i autocorrelacionados. Per a això, es van estudiar els mètodes sobre estructures latents PCA i PLS, i les seves respectives versions dinàmiques: dynamic PCA (DPCA) i dynamic PLS (DPLS).

El desenvolupament del model dinàmic sobre estructures latents (mdel) va consistir en la combinació del model DPCA i de la funció de transferència (FT). El model DPCA va ser aplicat a les dades d'entrada i de sortida del procés de manera separada per a determinar les variables latents que recullen la màxima variabilitat dinàmica en les dades. L'estructura dinàmica en les variables latents resultants va ser modelada mitjançant una funció de transferència per analitzar la influència de les dades d'entrada sobre les dades de sortida.

Els models mdel i DPLS van ser ajustats en condicions idèntiques a un conjunt de dades simulades. La qualitat d'ajust dels models va ser avaluada mitjançant la variabilitat explicada de les variables latents (R^2), i la capacitat predictiva va ser testada mitjançant validació creuada, on va ser calculat el pïxen absolute error (MAE) de els dos models. L'mdel presentar millors resultats comparat al DPLS, amb major variabilitat explicada per les variables latents, i amb una capacitat predictiva robusta i estable.

Els models proposats van ser aplicats per monitoritzar i predir el procés de cria de tilàpia a Moçambic. El gràfic MAAEWMA va assenyalar canvis significatius de l'ambient aquàtic en el procés, i amb el model mdel es va analitzar l'impacte d'aquests canvis en el creixement de la tilàpia.

Índice general

Agradecimientos	IV
Resumen	V
Abstract	VII
Resum	IX
Lista de figuras	XV
Lista de tablas	XIX
1. INTRODUCCIÓN Y JUSTIFICACIÓN	1
1.1. Introducción	1
1.2. Justificación	4
1.3. Estructura de la Memoria	5
2. OBJETIVOS	7
3. ANTECEDENTES	11
3.1. Monitorización de Procesos de Producción	11
3.1.1. Introducción	11
3.1.2. SPC univariado	12
3.1.3. SPC Multivariante	16
3.2. Métodos de proyección sobre estructuras latentes	19
3.2.1. Análisis de Componentes Principales (PCA)	19
3.2.2. Análisis de componentes principales para procesos dinámicos (Dynamic PCA-DPCA)	27
3.2.3. Partial Least-Square Regression (PLS)	29
3.2.4. PLS para procesos dinámicos (Dynamic PLS-DPLS)	32
4. METODOLOGÍA	37
4.1. Introducción	37
4.2. Gráfico MEWMA para procesos autocorrelacionados	39
4.3. Modelo dinámico sobre estructuras latentes (MDEL)	41

5. PROPUESTA DE MÉTODOS DE MONITORIZACIÓN Y PREDICCIÓN DE PROCESOS MULTIVARIANTES Y AUTOCORRELACIONADOS	43
5.1. Introducción	43
5.2. Gráficos MEWMA para procesos autocorrelacionados	44
5.2.1. Análisis del desempeño del gráfico MAAEWMA	51
5.2.2. Conclusión	58
5.3. Modelo dinámico sobre estructuras latentes	60
5.3.1. Construcción del modelo propuesto	63
5.3.2. Simulación de procesos multivariantes y autocorrelacionados	69
5.3.3. Resultados	71
5.3.4. Conclusión	87
6. APLICACIÓN DE LOS MÉTODOS ESTADÍSTICOS PROPUESTOS AL PROCESO DE PRODUCCIÓN EN ACUICULTURA	89
6.1. Introducción	89
6.2. La Acuicultura	90
6.2.1. La Acuicultura en Mozambique	92
6.3. Métodos estadísticos en acuicultura	95
6.4. Parámetros que caracterizan el proceso en acuicultura	98
6.4.1. Parámetros físico-químicos	99
6.4.2. Parámetros Biometricos	102
6.4.3. Alternativas en la alimentación	102
6.5. Estructura y toma de datos	103
6.5.1. Muestreo	105
6.6. Análisis e Interpretación de los datos	106
6.6.1. Análisis del efecto del tipo de tratamiento en el crecimiento de la tilapia	107
6.6.2. Monitorización del ambiente acuático en Acuicultura	110
6.6.3. Predicción del índice de crecimiento de la tilapia mediante el modelo dinámico sobre estructuras latentes	129
6.7. Conclusión del capítulo	142
7. CONCLUSIONES GENERALES Y FUTURAS LINEAS DE INVESTIGACIÓN	145
7.1. Conclusiones	145
7.2. Futuras líneas de investigaciones	148
A. APÉNDICES: Figuras	157
B. CÓDIGO R DESARROLLADO	177
B.1. Código R para el ajuste del gráfico MAAEWMA	177

B.2. Código R para el ajuste del modelo dinámico sobre estructuras latentes (MDEL)	180
B.3. Código R para el gráfico MAAEWMA: Caso acuicultura	187
B.4. Código R para MDEL: Caso acuicultura	190

Índice de figuras

2.1. Esquema de los objetivos	8
3.1. Descomposición del modelo PCA	20
4.1. Esquema de las etapas del estudio	40
5.1. Construcción del modelo de monitorización	45
5.2. Efecto del coeficiente autoregresivo de la variable p_i en el RL (Run of Length) para los tres cambios considerados	53
5.3. Efecto del coeficiente autoregresivo de la variable p_j en el RL (Run of Length) para los tres cambios considerados	54
5.4. Comparación del RL (Run of Length) para los tres cambios considerados	54
5.5. Interacciones de procesos con diferentes coeficientes autoregresivos	55
5.6. Monitorización de un proceso autocorrelacionado bivalente con coeficientes $p_i(\phi = 0,9)$ y $p_j(\phi = 0,9)$	56
5.7. Monitorización de un proceso autocorrelacionado bivalente con coeficientes $p_i(\phi = 0,7)$ y $p_j(\phi = 0,5)$	56
5.8. Monitorización de un proceso autocorrelacionado bivalente con coeficientes $p_i(\phi = 0,5)$ y $p_j(\phi = 0,3)$	57
5.9. Monitorización de un proceso autocorrelacionado bivalente con coeficientes $p_i(\phi = 0,3)$ y $p_j(\phi = 0,3)$	57
5.10. Esquema de construcción del modelo dinámico propuesto	64
5.11. Representación de las variables latentes t_1 y t_2	72
5.12. Representación de las variables latentes u_1 y u_1	73
5.13. Correlaciones cruzadas entre los residuos a_1 y a_1^* y a_2 y a_2^*	73
5.14. Representación de las variables latentes t_1 y t_2	76
5.15. Representación de las variables latentes u_1 y u_2	76
5.16. Comparación del \bar{R}^2 de las variables latentes del bloque X obtenidas en el MDEL y DPLS	81
5.17. Comparación del \bar{R}^2 de las variables latentes del bloque Y obtenidas en el MDEL y DPLS	81
5.18. Comparación del MAE de MDEL y DPLS	82
5.19. Efecto de la autocorrelación al $\bar{R}^2(X)$	83
5.20. Efecto de la autocorrelación al $\bar{R}^2(Y)$	83

5.21. Ajuste del modelo dinámico y del DPLS para procesos con coeficientes autoregresivos igual a 0.3	84
5.22. Ajuste del modelo dinámico y del DPLS para procesos con coeficientes autoregresivos igual a 0.5	85
5.23. Ajuste del modelo dinámico y del DPLS para procesos con coeficientes autoregresivos igual a 0.9	86
6.1. Localización de Mzambique en el globo	92
6.2. parámetros biometricos	107
6.3. Comparación del peso (a la izquierda) y de la longitud (a la derecha) en los cuatro tratamientos (i-j, marca la comparación entre tratamientos C_i y C_j)	108
6.4. Clasificación de los individuos de los cuatro tratamientos	109
6.5. Variabilidad de los parámetros ambientales en el tratamiento C1	112
6.6. Variabilidad de los parámetros ambientales en el tratamiento C2	113
6.7. Variabilidad de los parámetros ambientales en el tratamiento C6	114
6.8. Variabilidad de los parámetros ambientales en el tratamiento C7	115
6.9. Monitorización del ambiente acuático en el tratamiento C1 mediante el gráfico MAAEWMA	127
6.10. Monitorización del ambiente acuático en el tratamiento C2 mediante el gráfico MAAEWMA	128
6.11. Monitorización del ambiente acuático en el tratamiento C6 mediante el gráfico MAAEWMA	128
6.12. Monitorización del ambiente acuático en el tratamiento C7 mediante el gráfico MAAEWMA	129
6.13. Función de correlación cruzada entre peso y longitud en los cuatro tratamientos	131
6.14. Representación de las componentes principales	133
6.15. Representación de las variables latentes en el tratamiento C1	134
6.16. Representación de las variables latentes en el tratamiento C2	134
6.17. Representación de las variables latentes en el tratamiento C6	135
6.18. Representación de las variables latentes en el tratamiento C7	135
6.19. Correlaciones simple y parciales de las variables latentes en el tratamiento C7	136
6.20. Predicción del índice del crecimiento de la tilapia en los cuatro tratamientos	142
A.1. Funciones de correlaciones simple y parciales del peso en los cuatro tratamientos	158
A.2. Funciones de correlaciones simple y parciales de la longitud en los cuatro tratamientos	159
A.3. Funciones de correlaciones simple y parciales de la temperatura en los cuatro tratamientos	160
A.4. Funciones de correlaciones simple y parciales de la oxígeno en los cuatro tratamientos	161

A.5. Funciones de correlaciones simple y parciales de la pH en los cuatro tratamientos	162
A.6. Funciones de correlaciones simple y parciales de la transparencia en los cuatro tratamientos	163
A.7. Función de correlación cruzada entre temperatura y oxígeno en los cuatro tratamientos	164
A.8. Función de correlación cruzada entre temperatura y pH en los cuatro tratamientos	165
A.9. Función de correlación cruzada entre temperatura y transparencia en los cuatro tratamientos	166
A.10. Función de correlación cruzada entre oxígeno y pH en los cuatro tratamientos	167
A.11. Función de correlación cruzada entre oxígeno y transparencia en los cuatro tratamientos	168
A.12. Función de correlación cruzada entre pH y transparencia en los cuatro tratamientos	169
A.13. Correlaciones simple y parciales de las variables latentes en el tratamiento C1	170
A.14. Correlaciones simple y parciales de las variables latentes en el tratamiento C2	171
A.15. Correlaciones simple y parciales de las variables latentes en el tratamiento C6	172
A.16. Función de correlación cruzada entre el índice del crecimiento de la tilapia (u) y las componentes t_1 y t_2) en los tratamientos C1 y C2.	173
A.17. Función de correlación cruzada entre el índice del crecimiento de la tilapia (u) y las componentes t_1 y t_2) en los tratamientos C6 y C7.	174
A.18. Función de autocorrelación simples y parcial para los residuos del modelo dinámico sobre estructuras latentes ajustados en los cuatro tratamientos	175

Índice de tablas

5.1. Comparación de los ARL de gráfico MAAEWMA (ARL_B) respecto al gráfico MEWMA (ARL_A)	59
5.2. Estimación de los parámetros del modelo de función de transferencia entre u_1 y t_1	74
5.3. Estimación de los parámetros del modelo de función de transferencia entre u_2 y t_2	74
5.4. Estimación de los coeficientes del modelo DPLS	77
5.5. Comparación del ajuste y predicción de los modelos MDEL y DPLS	80
6.1. efectos de cada intervalo de fluctuación de oxígeno	100
6.2. significación de la funciones discriminante	110
6.3. Límites de control para los parámetros ambientales	111
6.4. Resumen del estado del proceso	118
6.5. Resultados Numéricos del proceso de monitorización mediante gráfico MEWMA en C1	121
6.6. Resultados Numéricos del proceso de monitorización mediante gráfico MEWMA en C2	122
6.7. Resultados Numéricos del proceso de monitorización mediante gráfico MEWMA en C6	123
6.8. Resultados Numéricos del proceso de monitorización mediante gráfico MEWMA en C7	124
6.9. Límites de control y observaciones fuera de control	126
6.10. Resultados Numéricos de las estimación del Modelo dinámico sobre estructuras latentes en los tratamientos C1 y C2	138
6.11. Resultados Numéricos de las estimación del Modelo dinámico sobre estructuras latentes en los tratamientos C6 y C7	139
6.12. Expresiones analíticas de los modelos dinámicos en todos los tratamientos	141

C apitulo 1

INTRODUCCI ON Y JUSTIFICACI ON

1.1. Introducci on

La mayor parte de los procesos de producci on en la industria suelen caracterizarse por m as de una variable del producto, la calidad del producto suele definirse por un conjunto de par ametros espec ificos que han de cumplir con los requisitos establecidos en la manufactura, por otro lado, las condiciones en la que se realiza la producci on depende de muchos par ametros (clima, presi on, etc) que, aunque no sean registrados directamente en el producto, ejercen una influencia sobre su calidad.

El registro continuo de cada caracter istica de dichos procesos que se traducen en una base de datos, forman un sistema multivariante, con variables de entrada y salida (multi-input, multi-output (MIMO)) (Barcel o *et al.*, 2011). El  xito de dichos procesos dependen del cumplimiento simultaneo de los requisitos espec ificos en cada caracter istica en causa, de modo que un fallo en el producto puede estar asociado a deficiencias en el proceso de producci on. Por tanto, el reto que se presenta a los profesionales del campo es c omo controlar a la vez dichas caracter isticas de modo que se puedan se alar los cambios ocurridos en el proceso para que sean reparados y repuestos en marcha.

El control estad stico de procesos SPC cl sico (el que se basa en controlar solamente una caracter istica del producto) es inadecuado para monitorizar la estructura compleja presentada por estos procesos.

Los gr ficos de control de Shewart (Shewhart, 1931), EWMA (Roberts, 1959) y CUSUM (Page, 1954) fueron desarrollados para monitorizar solamente una caracter istica del producto. Una estrategia consiste en monitorizar todas las caracter isticas de un proceso multivariante aplicando estos gr ficos de manera independiente para cada variable del proceso. Es decir, monitorizar las n caracter isticas de un proceso multivariante mediante n gr ficos independientes, lo cual se llama control estad stico m ltiple. Este m todo no modela la estructura de correlaci on que define la caracter istica com n entre las variables del proceso, y es muy susceptible a falsas alarmas. Para subsanar este inconveniente

niente, los gráficos T^2 de Hotelling (Hotelling, 1947), MEWMA (Lowry *et al.*, 1992) y MCUSUM (Crosier, 1988) fueron desarrollados, y son aplicados para controlar en simultaneo las variables de un proceso multivariante.

La idea de monitorizar a la vez las variables del proceso permite controlar, además del efecto principal de una característica, el efecto simultáneo entre ellas. Los datos de un proceso multivariante suelen estar correlacionadas, de hecho, son necesarios métodos que exploten la relevante información que contiene dichos datos. Los gráficos multivariantes presentados anteriormente incluyen en sus estadísticos la matriz de varianza-covarianza, que permite incluir en el modelo la varianza de cada variable y de cada par de ellas.

Los parámetros de los gráficos multivariantes presentados anteriormente son calculados bajo las hipótesis de normalidad e independencia entre observaciones. Sin embargo, la dependencia del tiempo en las variables suele ser muy marcada en muchos procesos industriales, siendo que, además de la correlación entre las variables, cada una de ellas presenta una autocorrelación significativa.

En los procesos donde además de la correlación, las variables presentan una estructura de autocorrelación, los gráficos de control basados en las hipótesis de normalidad e independencia entre observaciones se torna inadecuado. Para monitorizar procesos con dichas características, otros enfoques en los gráficos de control fueron considerados (Claro *et al.*, 2007)(Testik, 2004)(Patel y Divecha, 2013).

El método de control de procesos autocorrelacionados se basa principalmente en dos enfoques: (1) ajustar un modelo adecuado de series temporales para cada característica y construir los gráficos de control sobre los ruidos blancos obtenidos (Montgomery, 2001)(Superville y Adams, 1994); (2) incluir la estructura de autocorrelación en el cálculo de los parámetros del gráfico y reemplazarlos en los gráficos multivariantes convencionales (Vasilopoulos y Stamboulis, 1978)(Patel y Divecha, 2013).

Nuestra línea de investigación a estos gráficos de control se orienta al enfoque (2), y consiste en adaptar el gráfico MEWMA (Multivariate Exponentially Weighted Moving Average) a los procesos donde cada característica se ajusta a un modelo autoregresivo.

La estructura de autocorrelación se incluye en el modelo a través de las autocovarianzas y covarianzas cruzadas calculadas usando las formulas correspondientes al modelo autoregresivo, y adaptadas según los objetivos del planteamiento del modelo. A continuación, estos parámetros son usados en el cálculo del estadístico y límite de control del gráfico, reemplazando las varianzas y covarianzas instantaneas usadas en los gráficos convencionales.

El error que se puede cometer al no incluir la estructura de autocorrelación en los parámetros del gráfico de control fue analizado por Claro et al. (Claro *et al.*, 2007), entre otros aspectos, se destaca el hecho de que los gráficos de control convencionales ajustados a los procesos autocorrelacionados son muy susceptibles a falsas alarmas, ya que los límites de control suelen ser más apretados de lo normal y las características comunes del proceso causadas por la autocorrelación entre las observaciones son señaladas como características fuera de control.

Por tanto, la presente propuesta busca incluir la estructura de autocorrelación de las variables en el cálculo de los parámetros del gráfico de control para minimizar las falsas alarmas y, por otro lado, garantizar que el gráfico sea rápido en detectar cambios en la media del proceso.

Los procesos industriales a que nos referimos en esta tesis presentan dos bloques de variables: las variables del proceso (input) y las variables del producto (output). Se considera que la calidad del producto depende de las características del proceso, y no al revés. En un proceso donde se busca medir la influencia de las características del proceso sobre el producto, se hace necesario un modelo predictivo. Por tanto, además de la propuesta de un gráfico de monitorización se presenta una metodología estadística para modelar estos dos grupos de variables.

La estructura de correlación y autocorrelación existente en cada bloque y, la dependencia de las variables del producto en las variables del proceso, nos sugiere el planteamiento de un modelo predictivo dinámico sobre estructuras latentes, donde se puede predecir la calidad del producto a través de las características del proceso.

El modelo de predicción es desarrollado combinando los modelos de análisis de componentes principales para procesos dinámicos (DPCA) y la función de transferencia (FT). El DPCA es aplicado en cada bloque de manera independiente con vista a explotar la estructura de correlación dinámica en cada bloque de variables, y sobre las estructuras latentes resultantes es ajustada una función de transferencia.

Por tanto, el método estadístico propuesto en esta tesis es constituido por un modelo MEWMA adaptado a los procesos autocorrelacionados para monitorizar las características del proceso (o del producto), y un modelo dinámico ajustado sobre estructuras latentes para predecir la calidad del producto a través de las características del proceso.

1.2. Justificación

La complejidad de la estructura de los datos en los procesos industriales originada por la correlación entre variables y la autocorrelación entre observaciones hace que los gráficos de control convencionales no sean adecuados para monitorizar este tipo de procesos. Las hipótesis de normalidad e independencia entre observaciones que son la base en el cálculo de los estadísticos de dichos gráficos de control no se verifican en los casos referidos, siendo necesarios otros enfoques que permitan lidiar con dicha estructura correlacionada y dinámica del proceso.

Como comentamos anteriormente, algunos métodos han sido desarrollados para monitorizar procesos multivariantes y autocorrelacionados. Sin embargo, el método que consiste en monitorizar el ruido blanco resultante de un modelo de series temporales no es muy sensible a pequeños cambios del proceso, además, el cambio registrado en la media del ruido blanco modelado no siempre es igual al cambio ocurrido en la media del proceso, lo que torna el gráfico difícil de interpretar (Montgomery, 2001).

Por otro lado, los gráficos desarrollados con la estrategia de incluir la estructura de autocorrelación en los parámetros del modelo presentan cálculos laboriosos (Ong, 2009), y de cierto modo, se pierde la sencillez del gráfico. Nuestra investigación está orientada a este enfoque, donde buscarse estrategias alternativas de inclusión de la autocorrelación en los gráficos de control sin comprometer la sencillez y que sean relativamente rápidos en detectar cambios del procesos.

Los métodos estadísticos propuestos en esta tesis están orientados al reto de poder modelar la estructura compleja de los datos registrados en los procesos de producción en la industria, con vista a proporcionar mejores resultados en la monitorización y predicción de dichos procesos. Como afirma Ferrer (Ferrer, 2003) asociado a los pensamientos de Box, Fisher, Pearson, Tukey y Youden:

“... es necesario trabajar junto con científicos e ingenieros para dar respuesta como estadísticos a los nuevos retos que la revolución tecnológica plantea, abandonando viejos paradigmas que nada tienen que ver con los verdaderos problemas que acucian a los procesos industriales del siglo XXI, y contribuyendo al desarrollo de nuevos métodos estadísticos para abordar los nuevos retos asociados con los grandes volúmenes de datos de naturaleza multivariante que nos inundan”.

De un modo general, los modelos propuestos fueron desarrollados con vista a modelar los procesos multivariante y autocorrelacionados en la industria. El

desempeño de los modelos desarrollados fue testado mediante simulación de datos. Como ejemplo ilustrativo, los modelos fueron aplicados al proceso de cría de especies acuáticas en acuicultura.

El proceso de producción en acuicultura es un caso particular de los procesos MIMO con estructuras dinámicas referidos en este trabajo. Las variables en acuicultura son formadas por dos grupos de parámetros: (1) Parámetros de la calidad del agua (parámetros ambientales) y (2) Parámetros de la calidad del producto cultivado (parámetros biométricos). La calidad del producto es caracterizado por los parámetros biométricos, que de cierto modo, indican el nivel de aceptación del producto final. Dicha calidad del producto depende de los parámetros del proceso, es decir, de la calidad del agua caracterizada por los parámetros físico-químicos y biológicos que deben ser regularmente controlados.

Los parámetros ambientales están correlacionados entre si, de manera que el cambio de uno afecta al otro, e influyen al desarrollo de la especie en el proceso de cría. Los métodos estadísticos propuestos son aplicados para monitorizar el ambiente acuático, y analizar su influencia a la calidad del producto.

1.3. Estructura de la Memoria

Esta memoria está estructurada en 7 capítulos. El primer capítulo está constituido por la introducción general y la justificación, donde los conceptos básicos que orientaron la investigación son presentados. En la introducción se discute la problemática de la incompatibilidad de las técnicas estadísticas convencionales en el control y predicción de procesos multivariantes con estructura de autocorrelación significativa en las variables, y en el apartado de la justificación se presenta la motivación, la relevancia de la investigación y el aporte del tema al campo en cuestión.

El segundo capítulo corresponde a los objetivos de la investigación, donde el objetivo general y los específicos son presentados de manera detallada a fin de clarificar lo que se pretende alcanzar en el trabajo.

Los métodos estadísticos uni y multivariantes de monitorización y predicción, son presentados en el tercer capítulo como los antecedentes. En este capítulo se realiza un enfoque sobre la evolución de las técnicas estadísticas fundamentales usadas en la monitorización y predicción de los procesos de producción en la industria. Las diversas técnicas han sido presentadas según un orden lógico basado en la creciente complejidad de las mismas y en su creciente capacidad para modelizar procesos complejos.

En el cuarto capítulo se presenta la metodología del trabajo, en la que se basa en presentar los procedimientos usados para desarrollar la investigación y lograr los objetivos planteados.

Los métodos de monitorización y predicción de los procesos multivariantes y autocorrelacionados propuestos en esta tesis son desarrollados en el capítulo cinco. Por tanto, en vista a la falta de adecuación de los métodos convencionales de control en monitorizar los procesos autocorrelacionados, los gráficos de control MEWMA fueron estudiados y adaptados para controlar procesos con estas características. Por otro lado, el modelo dinámico sobre estructuras latentes fue desarrollado como metodología estadística para predecir la calidad del producto a través de las características del proceso.

Una aplicación de los métodos propuestos es ilustrada en el capítulo seis, donde los modelos son aplicados en la monitorización y predicción de proceso de cría de tilapias en Mozambique. Para dar una visión del proceso, el capítulo empieza presentando algunos conceptos básicos respecto a los métodos aplicados en acuicultura y termina con la presentación y discusión de los resultados.

Finalmente, en el capítulo siete se presentan las conclusiones generales y las futuras líneas de investigación. Las conclusiones presentadas en este capítulo corresponden al resumen de las conclusiones presentadas en los capítulos del trabajo. Debido a ciertas limitaciones respecto al alcance de la investigación a aspectos peculiares de cada proceso en la industria, se presentan en este capítulo, otras líneas que se pueden considerar pertinentes en futuras investigaciones.

C pulo 2

OBJETIVOS

A partir de los aspectos comentados anteriormente, en el presente trabajo se pretende lograr el siguiente objetivo principal: **Desarrollar una metodolog a estad stica para modelar y monitorizar procesos multivariantes y autocorrelacionados en la industria.**

El desarrollo de dicha metodolog a implica un estudio estad stico que incluye dos aspectos importantes: La monitorizaci n y la predicci n de los procesos. Nuestra visi n es que al final de esta tesis hayamos podido lograr mecanismos eficientes capaces de detectar fallos en los procesos multivariantes con caracter sticas que dependen del tiempo, y poder predecir la calidad del producto manufacturado a trav s de las variables de su entorno de producci n. Esto comienza por un estudio de los m todos estad sticos aplicados en la industria dirigidos a estos procesos, de cara a identificar sus limitaciones y poder sugerir nuevas aportaciones.

La estructura general que se va a seguir para lograr el objetivo planteado se representa en la figura 2.1.

Empezaremos por un estudio bibliogr fico sobre los m todos de control y predicci n de procesos en la industria. Se busca por un lado, comprender el funcionamiento de los m todos estad sticos convencionales aplicados en los procesos de producci n en la industria en general, y por otro lado, identificar sus limitaciones en la modelaci n de los procesos multivariantes y autocorrelacionados considerados en este trabajo, para adaptarlos a la monitorizaci n y predicci n en estos entornos.

El estudio de las caracter sticas del proceso permitir  comprender la naturaleza de las variables en cuesti n, y a trav s de sus propiedades se podr  ajustar a un modelo espec fico con lo cual se podr  calcular los par metros relevantes del modelo.

De acuerdo con las caracter sticas del proceso de producci n considerado en este trabajo, y por la necesidad de obtener cada vez m s mejores resultados en la industria, el estudio sigue dos direcciones:

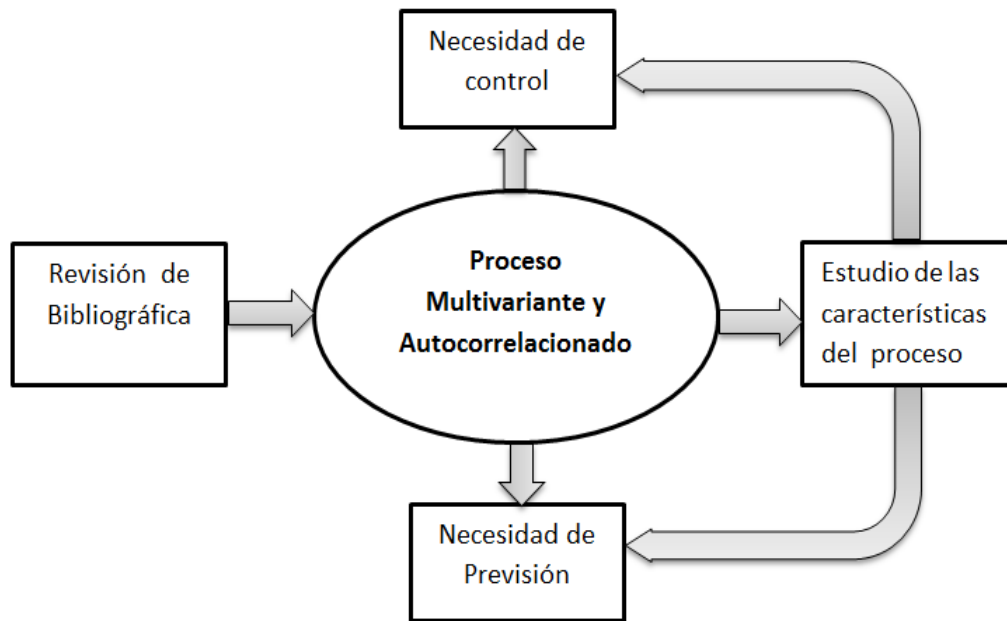


Figura 2.1: Esquema de los objetivos

1. Monitorización de las características del proceso y del producto;
2. Predicción de la calidad del producto a través de las características del proceso.

En la monitorización se estudian los métodos alternativos de control de procesos, su evolución y adecuación a la estructura de los datos en cada caso. De cara a la estructura de correlación de las características del proceso en cuestión, los gráficos MEWMA serán estudiados y adaptados a la estructura dinámica de dichos procesos. Por lo tanto, el gráfico de monitorización propuesto será desarrollado para lograr dos aspectos importantes: (1) modelar la estructura compleja de los datos, y (2) dotarlo de capacidad en detectar rápidamente cambios en el proceso.

La relación entre los dos bloques de variables (variables de proceso y del producto, o datos de entrada y de salida) será estudiado mediante un modelo de predicción. Considerando la dependencia del producto a las características de su entorno de manufactura, y a la estructura de correlación y autocorrelación en los datos, el planteamiento del modelo de predicción será realizado mediante la aplicación de un modelo DPCA para explorar la estructura de correlación dinámica entre las variables de cada bloque y, sobre las nuevas variables latentes dinámicas resultantes se ajusta un modelo de función de transferencia para analizar la influencia dinámica de las características del proceso en la calidad del producto. El modelo de predicción resultante se denomina *modelo*

dinámico sobre estructuras latentes.

De modo sintético, los objetivos específicos que se buscan lograr en el trabajo son los siguientes:

1. Lograr un conocimiento profundo sobre los métodos estadísticos de monitorización y predicción aplicados en los procesos multivariantes en la industria;
2. Analizar los métodos alternativos de monitorización de procesos autocorrelacionados;
3. Proponer un modelo de control que se adecue a los procesos multivariantes y autocorrelacionados;
4. Identificar estructuras latentes en el proceso;
5. Proponer un Modelo estadístico multivariante para predecir la calidad del producto a través de las características del proceso;

A estos objetivos se podría añadir uno adicional:

6. Aplicar las metodologías desarrolladas a un caso real.

Cápítulo 3

ANTECEDENTES

3.1. Monitorización de Procesos de Producción

3.1.1. Introducción

La capacidad de manejar de manera eficiente los métodos de monitorización de procesos de producción de cara a lograr la calidad deseada del producto final, es un tema discutido por muchos investigadores. Un aspecto importante y que fue muy discutido es que el control estadístico de procesos (en inglés, Statistical Process Control (SPC)) clásico, o sea, el que se basa en el control de solo una variable, no es adecuado para enfrentar los problemas de los procesos de producción caracterizados por un conjunto de variables que generan una estructura compleja de datos.

El SPC clásico, basado en el control de una sola característica del producto, ha traído buenos resultados en la industria. Sin embargo, la complejidad en la estructura de los datos de los procesos multivariantes sugieren nuevos métodos, capaces de explotar la variabilidad conjunta presentado por las características en cuestión. Por este motivo el SPC fue adaptado al entorno multivariante para superar las limitaciones encontradas en el método clásico (Macgregor, 1996).

Con el desarrollo de la industria y de la tecnología, los procesos son más exigentes, siendo que el reto que enfrenta actualmente la motorización de procesos es cómo manejar la gran cantidad de información correlacionada que puede llegar a registrarse en tiempo real en un proceso (Ferrer, 2003).

La monitorización de procesos mediante SPC se basa fundamentalmente en los gráficos de control, que suelen ser constituidos por un estadístico y los límites de control. Los estudios indican que Walter A. Shewhart (Shewhart, 1931), Page (Page, 1954), Roberts (Roberts, 1959), Holgate (Holgate, 1964) fueron los pioneros a usar esta metodología, cada uno con un enfoque particular según la naturaleza de la característica a controlar.

A continuación, presentamos los diferentes métodos de monitorización de procesos y su evolución de acuerdo con la naturaleza y el número la característica controla.

3.1.2. SPC univariado

3.1.2.1. Gráficos de Shewhart

El matemático Walter A. Shewhart es considerado por muchos como el padre del SPC, en 1924 introdujo el control estadístico de calidad, creando la base de los gráficos de control con objeto de detectar causas que podrían afectar a la calidad del producto en proceso y corregirlas para que no perjudique la calidad final (Shewhart, 1931).

La idea de Shewhart era la reducción de la variabilidad en el proceso para asegurar la calidad. Para ello, consideró dos fuentes de variabilidad: (1) la originada por causas aleatorias y (2) la originada por causas especiales. Shewhart llevaba a cabo el estudio de SPC con gráficos de control, que constituye una herramienta para identificar causas aleatorias y especiales de la variabilidad del proceso.

El gráfico de control de Shewhart consiste en monitorizar una característica de calidad de un proceso (cualitativa o cuantitativa) a través de un estadístico muestral y límites de control. En cada etapa de control, si el valor del estadístico cae dentro de los límites es que el proceso está bajo control, caso contrario, el proceso podría estar fuera de control, siendo necesario estudiarse si el desajuste es creado por causas aleatorias o especiales. Las causas aleatorias no se pueden controlar, es parte de la naturaleza del producto, pero las especiales pueden estar asociadas a un cambio en el proceso, errores de medición, etc, que cuando detectado se puede corregir y volver al estado normal.

Para monitorizar característica cuantitativas de un producto, Shewhart utilizó habitualmente los gráficos de la media \bar{x} del rango R y de la desviación típica s . Shewhart también estudió los gráficos c , u , p y np para controlar características cualitativas en un proceso para monitorizar el número de defectos por lote, la proporción de defectos por lotes, la proporción de unidades no conformes por lotes y el número de unidades defectuosas por lotes, respectivamente.

De lo expuesto anteriormente, se puede afirmar que los gráficos de control de Shewhart consisten básicamente en un estadístico de la característica del producto en control y sus respectivos límites, en términos estadísticos, diríamos que se trata de una operación de estimación puntual seguida por una prueba de hipótesis sobre la estabilidad del estadístico. Dicha característica puede ser continua o discreta, para cada uno de los casos es necesario conocer su

distribución de probabilidad, para facilitar el cálculo de los estadísticos y sus respectivos límites.

Se espera que una característica continua se ajuste a una distribución normal, ya que la función de probabilidades de varios estadísticos está basada en este supuesto. Sin embargo, en las características no mensurables, por ejemplo, cuando asume solo dos estados (defectuoso o no), la distribución del estadístico suele ser binomial y para casos en que se pretende estudiar número de defectos en un lote la distribución utilizada es poisson.

Si en una ocasión el estimador de la característica en estudio se presenta fuera de los límites de control, implicaría que el proceso podría estar fuera de control. Sin embargo, hay una probabilidad de que se generen falsas alarmas, es decir, que el proceso esté bajo control y el gráfico indique lo contrario, o que en realidad el proceso esté fuera de control y el gráfico no sea capaz de detectarlo. A estos errores los llamamos por Error Tipo I y Error Tipo II, respectivamente, y pueden ser cuantificados a través de las probabilidades $\alpha = P(\text{Error Tipo I})$ y $\beta = P(\text{Error Tipo II})$.

Por tanto, lo ideal en el control de calidad es obtener gráficos con mínimos valores de α y β , algo difícil de lograr a la vez. Una medida de rendimiento de los gráficos de control muy utilizada es el ARL (Average Run Length), que es el número promedio de muestras monitorizadas hasta que el gráfico presente una señal.

Necesitaríamos un valor de ARL muy grande cuando el proceso esté bajo control, indicando que el gráfico es menos susceptible a errores del tipo I, como indica la fórmula $ARL = \frac{1}{\alpha}$. Por otro lado, necesitaríamos valores pequeños de ARL cuando el proceso está fuera de control, indicando que el gráfico es rápido en detectar anomalías, y la expresión para este caso es $ARL = \frac{1}{1-\beta}$.

Se podría cuestionar, ¿porqué no atribuir valores muy pequeños a α y a β ?, así tendríamos un ARL grande para el proceso bajo control y un ARL pequeño para el proceso fuera de control. El precio de hacerlo de tal manera sería recurrir a tamaños de muestra demasiado grandes, lo cual habitualmente no es factible. Una forma de lograr la minimización de α y β a la vez es a través de optimización (García-Díaz y Aparisi, 2003).

En muchos estudios de control de calidad suele considerarse $\alpha = 0,0027$, lo que corresponde a un $ARL \approx 370$ para el proceso bajo control, indicando que serian necesarias tomar un promedio de 370 muestras para detectar una falsa alarma cuando el proceso está bajo control.

Otros métodos para el cálculo del ARL pueden ser encontrados en el estu-

dio de Roberts (Roberts, 1959) que para ello, usó por primera vez el método de Monte Carlo; Brook y Evans (Brook y Evans, 1972) fueron los primeros a utilizar las cadenas de Markov para el cálculo del ARL; y un método de sistema de acuación integral para el cálculo de ARL fue planteado por Crowder (Crowder, 1972).

3.1.2.2. Gráficos EWMA

Los gráficos de Shewhart fueron muy utilizados en la monitorización de productos en la industrial. Este método consiste en analizar solamente la información actual del proceso, considerando que los posibles cambios que pueden ocurrir en un determinado instante no dependen de los ocurridos en los instantes anteriores. Este aspecto contribuye a que los gráficos de Shewart sean insensibles a cambios de pequeña magnitud. En los casos donde además de la información del instante actual, es relevante incluir la información del pasado, se pueden usar los gráficos de Medias móviles ponderadas exponencialmente, representado por su sigla en inglés EWMA (Exponentially Weighted Moving Average) (Montgomery, 2001) .

El gráfico EWMA ha sido utilizado por primera vez por Roberts (Roberts, 1959), y el método consiste en controlar un proceso mediante el siguiente estadístico:

$$z_i = r\bar{x}_i + (1 - r)z_{i-1} \quad (3.1)$$

Donde $0 < r \leq 1$

$z_0 = \mu_0$ y $E(z_i) = \mu_0$ cuando el proceso está bajo control.

El estadístico z_i depende de los instantes anteriores, y a los valores altos de λ implica mayor peso a la observación actual, asimismo a los valores bajos de λ indica mayor peso a las observaciones anteriores.

En general, los límites superior (UCL) e inferior (LCL) de control del gráfico EWMA son representados por la siguiente expresión:

$$\begin{aligned} UCL &= \mu_0 + L\sigma_{z_i} \\ CL &= \mu_0 \\ LCL &= \mu_0 - L\sigma_{z_i} \end{aligned} \quad (3.2)$$

Donde L es el ancho de los límites de control.

La varianza de z_i es calculado a través de la expresión:

$$\frac{s}{\sqrt{n}} \sqrt{\frac{r}{2-r}} [1 - (1 - \lambda)^{2i}]$$

que de manera asintótica se reduce a $\frac{s}{\sqrt{n}}\sqrt{\frac{r}{2-r}}$. Por lo tanto, el estadístico z_i se distribuye según una normal $N\left(m_0, \frac{s}{\sqrt{n}}\sqrt{\frac{r}{2-r}}\right)$ considerando que la variable \bar{x} sigue una distribución normal $N(m_0, s_0)$ cuando el proceso está bajo control y las observaciones o medias muestrales son independientes (Montgomery, 2001). Los límites de control pasan a calcularse usando la expresión:

$$m_0 \pm L \frac{s_0}{\sqrt{n}} \sqrt{\frac{r}{2-r}} \quad (3.3)$$

3.1.2.3. El gráfico EWMA con variable autocorrelacionada

En muchos procesos industriales la dependencia del tiempo en las variables del producto no se puede evitar, y para estos casos, las condiciones generalmente aceptadas para los gráficos EWMA no se verifican. Tal es el caso de las características biométricas de una especie en acuicultura, los valores del peso y de la longitud registradas a lo largo del tiempo no suelen seguir una distribución normal, tampoco sus observaciones pueden ser consideradas independientes, puesto que los registros de esas características en un instante dependen del instante anterior. Por lo tanto, estos procesos están sujetos a una estructura de autocorrelación.

Claro et al. (Claro *et al.*, 2007) presentan en su trabajo una alternativa de monitorización de procesos con características autocorrelacionadas, explicando que si la autocorrelación en la variable es derivada por una causa especial, esta debe ser eliminada del proceso, sin embargo, si es parte inherente de la variabilidad resultante de causas comunes debe ser considerada en el planeamiento de las cartas de control, evitándose estimaciones incorrectas de los parámetros que causan una elevada tasa de falsas alarmas (Wiel, 1996)(Reynolds-Jr. y Lu, 1997).

La estructura de autocorrelación presente en la característica de calidad puede ser modelada mediante modelos de series temporales. Una estrategia considerada por varios autores consiste en monitorizar los residuos independientes e idénticamente distribuidos resultantes de estos modelos (Montgomery y Mastrangelo, 1991)(Superville y Adams, 1994). Sin embargo, estos métodos presentan algunas limitaciones, de las cuales se puede destacar su difícil interpretación. Más detalles sobre este tema puede ser visto en Wardell et al.(Wardell *et al.*, 1994), Harris y Ross (Harris y Ross, 1991), y Faltin et al.(Faltin *et al.*, 1997).

Otro método usado en la monitorización de procesos autocorrelacionados consiste en diseñar un sistema de control para las variables originales, incluyendo la estructura de autocorrelación en los parámetros del gráfico (Van-Brackle-III y Reynolds-Jr., 1997).

El control de procesos autocorrelacionados basado en este último criterio implica la determinación de un modelo que se ajusta a la característica de calidad a controlar. Para ello, Morettin y Toloi (Morettin y Toloi, 2004) consideraron un modelo autoregresivo de orden uno estacionario para las variables X_t del proceso, dada por la siguiente expresión:

$$X_t - \mu = \phi(X_{t-1} - \mu) + \epsilon_t, \quad t = 1, 2, 3, \dots, n \quad (3.4)$$

donde μ es la media del proceso, ϕ es el coeficiente autoregresivo y ϵ_t es el error aleatorio, i.i.d, $\epsilon_t \sim N(0, \sigma_t)$. Con la varianza:

$$\sigma_X^2 = \frac{\sigma_\epsilon^2}{1 - \phi^2} \quad (3.5)$$

y la función de autocovarianza es dada por:

$$\gamma_k = \phi^k \sigma_X^2 \quad (3.6)$$

Con $k=1,2,\dots$

Considerando un proceso con estas características, el estadístico Z_i de EWMA es calculado por la expresión 3.1 y los límites de control son obtenidos mediante la expresión:

$$LC = \mu_0 \pm L\sigma_{\bar{X}} \sqrt{\frac{\lambda}{2 - \lambda}} \quad (3.7)$$

donde λ es equivalente a r en el caso anterior, y $\sigma_{\bar{X}}$ es la varianza de la media de las muestras observadas en cada instante t (presentaremos los detalles en el capítulo 5).

Por tanto, la diferencia de estos límites con respecto al EWMA presentado en el apartado anterior, reside en la varianza utilizada $\sigma_{\bar{X}}$ que es determinada a través del modelo autoregresivo planteado, lo que garantiza la inclusión del efecto de autocorrelación en el modelo (fórmula 3.7).

3.1.3. SPC Multivariante

3.1.3.1. Gráficos T^2 Hotelling

La metodología de Hotelling ha sido desarrollada para controlar en simultaneo un conjunto de variables que se pueden registrar en un proceso. El hecho de que el gráfico considere simplemente la información presente del proceso es considerado una extensión del gráfico de control de Shewart.

Hotelling (Hotelling, 1947) presentó en 1947, un procedimiento para controlar el vector de medias de una población normal multivariante, mediante el estadístico T^2 que mide la distancia (distancia de Mahalanobis) entre el vector de medias esperado (parámetro del proceso) y el vector de medias observado (medias aritméticas de la muestra) tomando en cuenta su matriz de varianzas-covarianzas. El criterio consiste en comparar dicha distancia con los límites de control establecidos. Si la distancia resulta mayor que el límite de control establecido, el proceso podría estar fuera de control.

El estadístico T^2 de Hotelling toma la siguiente expresión:

$$T^2 = (\bar{x}_i - \mu)S^{-1}(\bar{x}_i - \mu)' \quad (3.8)$$

Donde \bar{x}_i es el vector de las medias de las variables tomadas en el instante i , μ es el vector de medias mientras el proceso este bajo control y S es una estimación de la matriz de varianzas-covarianzas de las variables x , para el proceso bajo control. Además, se asume que $x \sim N(\mu, S)$.

En el control de procesos mediante el gráfico T^2 Hotelling se supone que la estimación de la matriz de varianza-covarianzas (S) es no singular y que la matriz de los datos X es completa, es decir, sin datos faltantes (Ferrer, 2007). Sin embargo, hay casos donde no se puede registrar todos los datos del proceso y que la estructura de correlación entre las variables es muy fuerte, lo que hace que estos supuestos no se cumplan. En estos casos, los métodos de proyecciones sobre estructuras latentes suelen ser aplicados, ya que con un PCA se logra reducir la dimensión de los datos en unas pocas variables latentes (con observaciones completas) ortogonales entre si (MacGregor y T.Kourti, 1995).

3.1.3.2. El gráfico MEWMA

El gráfico de control MEWMA es considerado la versión multivariante del gráfico EWMA y, la primera referencia sobre estos gráficos se atribuye a Lowry et al. (Lowry *et al.*, 1992). El método consiste en controlar a la vez m variables correlacionadas entre si a través de un gráfico de control.

Sean $\vec{x}_1, \vec{x}_2, \dots, \vec{x}_m$ los vectores que recogen las medias muestrales tomadas del proceso m -variante. El gráfico MEWMA fue desarrollado considerando que los vectores aleatorios \vec{x}_i son independientes e idénticamente distribuidos según una normal m -variante de media \vec{m} y la matriz de covarianza Σ . En estas condiciones el estadístico MEWMA \vec{z}_i se define como:

$$\vec{z}_i = \Lambda \vec{x}_i + (1 - \Lambda) \vec{z}_{i-1} \quad (3.9)$$

donde

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ 0 & 0 & \lambda_j & \dots \\ 0 & 0 & \dots & \lambda_m \end{pmatrix}$$

$$\vec{z}_0 = \vec{m}_0.$$

y $0 < \lambda_j < 1$ ($j = 1, \dots, m$), \vec{x}_i es el vector de medias en cada instante i y tiene m elemento, \vec{m}_0 el vector de medias del proceso bajo control.

Se considera que el proceso está fuera de control si:

$$T_{1i}^2 = \vec{z}_i' \Sigma_{\vec{z}_i}^{-1} \vec{z}_i > h_1 \quad (3.10)$$

donde $h_1 > 0$ es el limite superior y se calcula de modo a obtener un especifico ARL bajo control. La matriz de covarianza de \vec{z}_i es calculado de la siguiente forma:

$$\Sigma_{\vec{z}_i} = \left(\frac{\lambda}{2 - \lambda} \right) [1 - (1 - \lambda)^{2i}] \Sigma$$

que asintoticamente puede ser calculado como:

$$\Sigma_{\vec{z}_i} = \left(\frac{\lambda}{2 - \lambda} \right) \Sigma$$

asumiendo la igualdad de los pesos de las observaciones pasadas para las m variables ($\lambda_1 = \dots = \lambda_m = \lambda$).

3.1.3.3. Gráfico MMOEWMA (Multivariate Modified EWMA chart)

El gráfico MMOEWMA fue desarrollado por Patel y Devicha (Patel y Divecha, 2013) como una extensión natural del gráfico EWMA modificado propuesto por ellos en 2011. El gráfico fue desarrollado para monitorizar la media de un proceso ajustado al vector autoregresivo de orden uno VAR(1). Además de introducir el impacto de las observaciones anteriores en el gráfico, el MMOEWMA incluye el cambio anterior del proceso, y el estadístico del gráfico es definido por:

$$\vec{z}_i = \Lambda \vec{x}_i + (1 - \Lambda) \vec{z}_{i-1} + (\vec{x}_i - \vec{x}_{i-1}), i \geq 1 \quad (3.11)$$

donde \vec{z}_0 es un vector de dimensión m y sus elementos corresponden a las medias del proceso bajo control para cada variable j , Λ es definida de manera análoga al caso anterior, $0 < \lambda_j < 1$, ($j = 1, \dots, m$).

Se asume que el proceso está fuera de control si:

$$T_{2i}^2 = \vec{z}_i' \Sigma_{\vec{z}_i}^{-1} \vec{z}_i > h_2 \quad (3.12)$$

donde $h_2 > 0$ es el límite superior que es calculado para obtener un específico ARL bajo control; en este caso, la matriz de covarianza de \vec{z}_i puede ser calculada de manera asintótica por

$$\Sigma_{\vec{z}_i} = \left(\frac{\lambda}{2 - \lambda} + \frac{2\lambda(1 - \lambda)}{2 - \lambda} \right) \Sigma$$

bajo la hipótesis de igualdad de pesos de las observaciones anteriores λ para las m variables ($\lambda_1 = \dots = \lambda_m = \lambda$).

3.2. Métodos de proyección sobre estructuras latentes

3.2.1. Análisis de Componentes Principales (PCA)

El Análisis de Componentes Principales, representado muchas veces por su símbolo en inglés PCA (Principal Componente Analysis) es la técnica multivariante que analiza una matriz con observaciones que presentan una estructura alta de correlación entre si. Es considerada una de las más antiguas técnicas multivariante (Abdi y Williams, 2010). El PCA en muchos sentidos constituye la base para el análisis multivariante de datos (Esbensen y Geladi, 1987), y es considerada uno de los más valiosos resultados de la aplicación del Álgebra lineal (Shlens, 2005)

El PCA se basa fundamentalmente en dar una aproximación a la matriz de datos $X = [x_1, x_2, \dots, x_m]$ en términos de producto de dos pequeñas matrices $T = [t_1, t_2, \dots, t_m]$ (de las puntuaciones o scores) y $P = [p_1, p_2, \dots, p_m]$ (de los pesos o loadings). Estas dos matrices (T y P) capturan el patrón esencial de los datos de X . La matriz T expresa el patrón de los individuos de X en las nuevas coordenadas. La matriz P recoge los pesos de cada variable x_j en las componentes principales. La matriz T se puede expresar en función de X y P de la siguiente manera:

$$T = X' * P \tag{3.13}$$

o sea, las componentes principales son las nuevas variables definidas por:

$$t_j = X' * p_j \quad j = 1, 2, \dots, m \tag{3.14}$$

La primera componente principal t_1 es calculada usando el vector p que maximiza la varianza de la matriz T , con la condición $p'p = 1$. Esto se traduce en calcular los autovalores y autovectores de la matriz de covarianza de X .

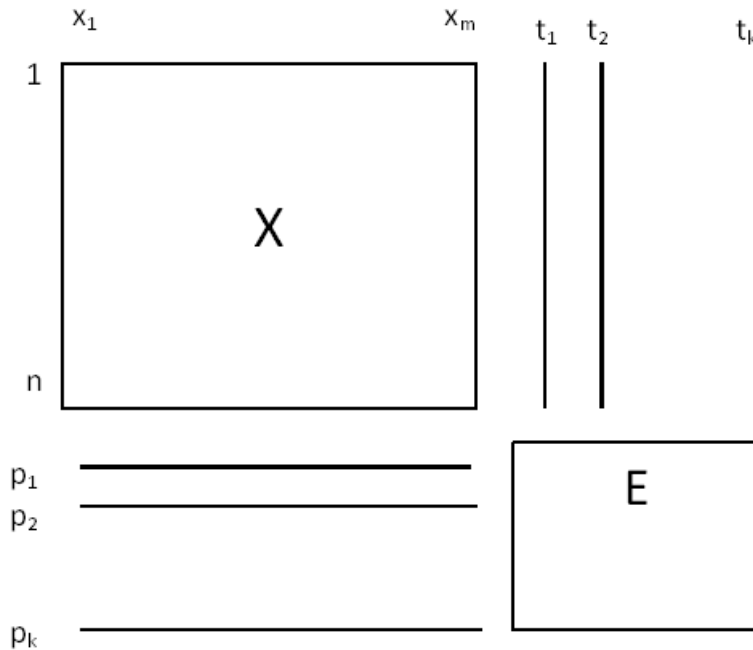


Figura 3.1: Descomposición del modelo PCA

Sea $S = \text{var}(X)$ la matriz de covarianzas de X . Puesto que $S \geq 0$ y simétrica, la diagonalización de S viene dada por:

$$S = PAP' \quad (3.15)$$

donde $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_m)$ es una matriz diagonal de misma dimensión que S y recoge los autovalores de S , y P es ortogonal, es decir $PP' = P'P = I$ cuyas columnas son autovectores de S , $\lambda_1 > \lambda_2 > \dots > \lambda_m$.

Para cada j , la nueva variable latente t_j se construye del j -ésimo autovector de S , y $T = [t_1, t_2, \dots, t_m]$.

Para las k primeras componentes, la matriz de datos X queda descompuesta como muestra la figura 3.1.

Se supone que la matriz X es centrada, es decir, que las variables fueron restadas por su respectiva media. Para las k primeras componentes, la matriz X puede ser descompuesta como se muestra a continuación:

$$X = \sum_{j=1}^k t_j * p_j' + E = TP' + E \quad (3.16)$$

Donde E es la matriz que contiene la variabilidad no explicada por el modelo PCA, o sea, la parte de varianza correspondiente a las $T_{m-k} = (t_{k+1}, \dots, t_m)$

componentes principales descartadas del modelo.

El algoritmo del PCA consiste en la descomposición de la matriz X en valores singulares (SVD), o mediante el NIPALS (Non-Linear Iterative Partial Least Squares) (Wold *et al.*, 1987).

3.2.1.1. Descomposición en valores singulares de la matriz X (SVD)

La descomposición de la matriz de datos X es uno de los métodos más usados para calcular los scores y loadings del modelo PCA.

El método consiste en descomponer la matriz X en tres elementos U, V y S , de modo que:

$$\text{SVD}(X) = USV' \quad (3.17)$$

donde V es la matriz que recoge los autovectores de X , es decir, la matriz de los loadings, S es la matriz diagonal que contiene las raíces cuadradas de los autovalores de la matriz X , y U es tal que los scores son las columnas de US (Wold *et al.*, 1987).

3.2.1.2. Algoritmo NIPALS

El algoritmo NIPALS es un método alternativo de cálculo de los scores en el análisis de componentes principales. El método consiste en un proceso iterativo lo cual se busca determinar los loadings que maximizan la varianza de los scores correspondientes. De modo sintético, el algoritmo sigue los siguientes pasos:

1. tomar $t_{\text{inicial}} = x_j$
2. $p' = t'X/t't$
3. $p = p/\|p\|$ (normalizar)
4. $t = Xp/p'p$
5. Se asume convergencia si el valor de t no cambió significativamente respecto al valor calculado en la iteración anterior. Si se obtiene la convergencia, parar. Caso contrario, volver al paso 2.

3.2.1.3. Propiedad de las componentes Principales

1. Las componentes principales tienen varianza decreciente

$$\begin{aligned} \text{var}(t_1) &= \text{var}(Xp_1) = p_1' S p_1 = \lambda_1 p_1' p_1 = \lambda_1 \\ \text{var}(t_2) &= \text{var}(Xp_2) = p_2' S p_2 = \lambda_2 p_2' p_2 = \lambda_2 \\ &\vdots \\ \text{var}(t_m) &= \text{var}(Xp_m) = p_m' S p_m = \lambda_m p_m' p_m = \lambda_m \end{aligned}$$

donde $\lambda_1 > \lambda_2 > \dots > \lambda_m$

2. La correlación entre ellas es nula:

$\text{cov}(t_i, t_j) = \text{cov}(Xp_i, Xp_j) = p_i' S p_j = \lambda_j p_i' p_j = 0$, para $i \neq j$, ya que P es una matriz ortogonal.

3. Las covarianzas entre cada componente principal y las variables originales X_i son:

$$\text{cov}(t_j, [x_1, \dots, x_m]) = \lambda_j t_j', \quad j = 1, \dots, p.$$

Utilizando las expresiones (3.13) y (3.15) tenemos:

$$\text{cov}(T, X) = \frac{1}{n} T' X' X = T' S = T' (T \Lambda T') = \Lambda T'$$

La fila de esta matriz proporciona las covarianzas entre t_j y las variables originales x_1, x_2, \dots, x_m .

En las nuevas coordenadas representadas por las componentes principales, el individuo i -ésimo, es decir, la fila $x_i' = (x_{i1}, x_{i2}, \dots, x_{im})$ de la matriz de datos X , se expresa como:

$$t_i' = x_i' P = (x_i' p_1, \dots, x_i' p_m).$$

La matriz de datos transformados es $T=XP$, que representa las observaciones de las nuevas variables (las variables latentes) sobre los n individuos de la muestra.

La variación total de X se define como $\text{tr}(S) = \sum_{i=1}^m \lambda_i$.

La variación total de $T=XP$ es igual a la variación total de X para las m componentes principales determinadas:

$$\text{tr}(\text{var}(T)) = \text{tr}\left(\frac{1}{n} T' X' T\right) = \text{tr}(T' S T) = \text{tr}(T' T \Lambda T' T) = \sum_{i=1}^m \lambda_i$$

puesto que, $S = T\Lambda T'$, donde T es una matriz ortogonal.

Cuando el cociente (porcentaje de la variabilidad explicada)

$$\text{VE \%} = \frac{\sum_{i=1}^k \lambda_i}{\text{tr}S} * 100, k < m,$$

es aproximadamente a 100 %, entonces las variables t_1, t_2, \dots, t_k pueden reemplazar a x_1, x_2, \dots, x_m sin gran pérdida de información, con respecto a la varianza total.

3.2.1.4. Número de componentes principales

El PCA reduce la dimensión de los datos correlacionados sin gran pérdida de información. Es decir, las variables del proceso se reducen en unas 'pocas' variables latentes que se obtienen mediante la explotación de la estructura de correlación en los datos, conservando la mayor parte de su variabilidad.

En términos matemáticos, en la matriz de datos X de dimensión $n * m$ se pueden determinar m componentes principales que recogen el 100 % de variabilidad de X . Sin embargo, en la practica se busca reducir la matriz X en una matriz T (de scores) de dimensión $n * k$, con $k \ll m$, que recoge la mayor información de X . Se asume que las $m - k$ componentes excluidas del modelo no presentan información relevante y constituyen la matriz de los residuos E . Para lograr este equilibrio (número de componentes principales que explican lo esencial de la matriz X), varios criterios de determinación de componentes principales fueron estudiados.

Baillo y Grané (Baillo y Grané, 2007) presentan 3 criterios de selección de componentes principales:

1. Porcentaje Explicado: Es el método más sencillo. Consiste en fijar un porcentaje de variabilidad explicado, por ejemplo el 90 %, y considera las sucesivas componentes principales hasta superar el porcentaje prefijado.
2. Criterio de Kaiser: Se incluyen aquellas componentes cuyos autovalores sean mayores que $\bar{\lambda} = \frac{\text{tr}(S)}{p}$, o bien mayores que 1 si se han calculado las componentes a partir de la matriz de correlación R .
3. Modificación de Jolliffe: Se ha comprobado que cuando $m \leq 20$ el criterio de Kaiser tiende a incluir pocas componentes. La modificación de Jolliffe excluye aquellas componentes cuyos autovalores sean menores que $0,7\bar{\lambda} = 0,7\frac{\text{tr}(S)}{p}$, o bien menores que 0.7 si se han calculado las componentes a partir de la matriz de correlación R .

Dependiendo del caso, uno u otro criterio puede ser usado, otros métodos pueden ser vistos en Otoole (Otoole *et al.*, 1993) lo cual advierte que la selección de número de componentes depende del estudio, y los criterios usados no deben limitar el estudio con riesgo que se pierda información importante en los datos. Más detalles sobre este tema se puede ver en Jolliffe (Jolliffe, 2002), Jackson (Jackson, 1991) y Peres (Peres-Neto *et al.*, 2005).

3.2.1.5. Inferencia sobre el PCA (Evaluación de la calidad del modelo)

Los conceptos presentados referentes al modelo PCA corresponden a los efectos fijos (Abdi y Williams, 2010), o sea, las conclusiones del modelo se limitan a las observaciones de la muestra y no se aplican a la población en estudio.

Consideremos la matriz X_k que representa la estimación de la matriz de datos X con las primeras k componentes principales obtenidos en el modelo PCA de los efectos fijos. A partir de la expresión 3.16, se deduce que

$$X_k = TP' \quad (3.18)$$

y

$$X = X_k + E \quad (3.19)$$

o sea, $E = X - X_k$

Para evaluar la calidad de la reconstrucción de la matriz X con k componentes, se evalúa la similitud entre las matrices X y X_k , para ello, el coeficiente de determinación puede ser utilizado (Abdi, 2007). El método más popular para ello es la suma de los cuadrados residuales (RESS) que se obtiene por la expresión:

$$\begin{aligned} RESS_k &= \| X - X_k \|^2 \\ &= \text{traza} E'E \\ &= I - \sum_{i=1}^k \lambda_i \end{aligned} \quad (3.20)$$

donde $\| \cdot \|$ es la norma y *traza* es la suma de los elementos de la diagonal de la matriz $E'E$. Por tanto, cuanto menor es el valor de RESS, el modelo PCA correspondiente presenta mejor ajuste.

Cuando se pretende estimar el valor de una nueva observación de la población se debe garantizar que el modelo PCA ajustado no cambia mucho con la introducción de nuevas observaciones. Para estimar la capacidad generalizada del modelo PCA, no se debe usar los procedimientos estándar, sino mediante técnicas de remuestreo, como el bootstrap y validación cruzada. En estas

técnicas los datos son separados en dos grupos: de aprendizaje y de test. La popular técnica de validación cruzada es 'Jackknife' ('leave one out' procedure), que significa el procedimiento de dejar uno fuera (Quenouille, 1956). Cada observación es dejada fuera del conjunto a su vez y las demás constituyen el conjunto de aprendizaje (learning set)(Efron, 1982).

El conjunto de aprendizaje es usado después para estimar la observación dejada fuera que constituye el conjunto de pruebas (test set). Usando este procedimiento, se busca garantizar que el modelo sea robusto en cuanto a la introducción de nuevas observaciones. Los valores predichos son recogidos en una matriz que denotaremos por \hat{X}_k .

La calidad general del modelo PCA con efectos aleatorios usando k componentes es evaluada a través de la similitud entre X y \hat{X}_k .

Mientras en el PCA de efectos fijos se utiliza el RESS para evaluar la calidad en la construcción del modelo, en este caso el PRESS (Predicted Residual Sum of Squares) es utilizado para evaluar la calidad predictiva del modelo,

$$\text{PRESS}_k = \| X - \hat{X}_k \|^2$$

Cuanto menor es el PRESS mejor es la calidad del modelo PCA. Podría observarse un decrecimiento de la calidad de predicción de un modelo ajustado al aumentar el número de componentes. Esto significaría que el modelo está sobrestimado, o sea, la información del conjunto del entrenamiento no es útil para ajustar el conjunto test. En muchos casos se usa este estadístico para determinar el número óptimo de componentes que mejor representan al modelo (Abdi y Williams, 2010).

Un estadístico de bondad de ajuste que evalúa en simultaneo el número óptimo de componentes y la robustez del modelo es usado por Abdi1 (Abdi, 2007) y se define como:

$$Q_k^2 = 1 - \frac{\text{PRESS}_k}{\text{RESS}_{k-1}} \tag{3.21}$$

donde k es el número de componentes.

En general se mantienen en el modelo las componentes con un valor de Q_k^2 mayor o igual a $1-0.952=0.048$.

3.2.1.6. SPC Multivariante mediante PCA

La mayor limitación de los gráficos univariantes en la monitorización de procesos multivariantes corresponde a la incapacidad de modelar la estructura de correlación entre las variables del proceso. De acuerdo con Macgregor

(Macgregor, 1996), aplicar un SPC univariante a un conjunto de variables correlacionadas es tan ineficiente como la presencia de interacciones entre los factores en un diseño de experimentos desarrollados bajo la inoperante técnica de ir modificando cada vez los niveles de un factor, dejando fijos los demás (Ferrer, 2003).

Un método muy usado en el control de procesos consiste en adaptar el SPC en el entorno de variables latentes. El gráfico T^2 Hotelling para controlar k variables latentes de un proceso es representada por:

$$T_k^2 = \sum_{j=1}^k \frac{t_j^2}{s_{t_j}^2} \quad (3.22)$$

El estadístico T^2 Hotelling para este caso es la suma de orden k de los cocientes entre el cuadrado de cada variable latente y sus respectivas varianzas. Estas varianzas son la diagonal de la matriz de varianza-covarianza (Σ) y son iguales a los autovalores (λ_j) de dicha matriz. Así, la expresión 3.22 es equivalente a:

$$T_k^2 = \sum_{j=1}^k \frac{t_j^2}{\lambda_j} \quad (3.23)$$

La división de cada t_j^2 por la respectiva varianza, permite que cada componente principal tenga el mismo peso en el cálculo del estadístico T^2 . Es decir, las $k + 1, \dots, m$ componentes descartadas del modelo contribuirían de forma similar en el cálculo del estadístico T^2 .

No siempre la matriz X está bien estructurada para que las k primeras componentes principales expliquen su mayor parte de variabilidad. Al dividir los t_j^2 por sus pequeñas varianzas, ligeras desviaciones de estas t_j^2 que no tienen casi ningún efecto sobre X daría lugar a una señal fuera de control en T^2 (MacGregor y T.Kourti, 1995).

Por tanto, el estadístico T_k^2 permite monitorizar el proceso basado en la información de las k componentes principales determinadas. Con ello, el proceso sólo señala cambios si la variación en las características del proceso se registra en el entorno de las primeras k componentes. Sin embargo, si sucede como indica MacGregor (MacGregor y T.Kourti, 1995), un evento especial, totalmente nuevo que no estaba presente en los datos de referencia utilizados para desarrollar el PCA, entonces nuevas componentes aparecerán y la nueva observación estará fuera de alcance. Para ello, Kresta (Kresta *et al.*, 1995) sugiere el gráfico del error cuadrático de predicción (SPE) para estudiar tales casos:

$$SPE_x = \sum_{j=1}^m (x_{\text{nuevo},i} - \hat{x}_{\text{nuevo},i})^2 \quad (3.24)$$

Mientras el proceso esté bajo control, el valor de SPE_x debe ser muy pequeño. Jackson (Jackson, 1991) indica que el límite superior de este estadístico puede ser calculado a través de los datos históricos, utilizando los resultados aproximados para la distribución de formas cuadráticas.

Cuando el modelo está bajo control, SPE representa ruido, la parte de la variabilidad no explicada por el modelo. No obstante, un alto valor de SPE_x indica que el modelo de proyección no es válido para dicha observación. Por tanto, la monitorización debe ser llevada a cabo aplicando estos dos gráficos: T^2 y SPE.

3.2.2. Análisis de componentes principales para procesos dinámicos (Dynamic PCA-DPCA)

El análisis de componentes principales presentado en el apartado anterior fue desarrollado para modelar un conjunto de datos con variables correlacionadas y observaciones independientes (Jolliffe, 2002)(Vanhatalo y KulaHCI, 2015). Muchos procesos industriales presentan, además de una estructura de correlación entre las variables, una fuerte dependencia de las observaciones con el tiempo. En estos casos, los algoritmos usados en el cálculo de la matriz T de los scores y P de los loadings no logran captar la relevante información en la matriz de los datos X. Vanhatalo y KulaHCI (Vanhatalo y KulaHCI, 2015) demuestran el impacto negativo de la autocorrelación en un modelo PCA. Uno de los aspectos mencionados por estos autores se refiere a la influencia de los coeficientes autoregresivos en la varianza explicada por las componentes determinadas.

Este trabajo se orienta a la investigación de procesos multivariantes con una estructura de correlación y autocorrelación en los datos. Cuando los datos contienen informaciones dinámicas, el PCA no revela las relaciones exactas entre las variables, sino una aproximación estática lineal (Ku *et al.*, 1995).

Para modelar los procesos dinámicos, Ku *et al.* (Ku *et al.*, 1995) desarrollaron el modelo Dynamic PCA (DPCA) lo cual fue usado y mejorado por muchos autores, entre ellos, destacamos el estudio de Vanhatalo *et al.* (Vanhatalo *et al.*, 2017).

El método DPCA consiste en incluir en la matriz de los datos las variables decaladas en el tiempo (lag), y ajustar un PCA sobre dicha nueva matriz extendida.

3.2.2.1. Definición del DPCA

Sea X la matriz de dimensión $t * m$ que recoge las variables correlacionadas y autocorrelacionadas de un proceso, y $\tilde{X} = [X_0, X_1, X_2, \dots, X_s]$ la matriz extendida, de dimensión $t * [m(s + 1)]$, que recoge las variables de X y las respectivas variables decaladas de orden $0, 1, 2, \dots, s$. Consideremos: $X_0 = X$; X_1 la matriz que recoge las variables de X decaladas con lag=1; y en general X_s recoge las variables de X decaladas con lag=s. Las componentes principales son calculadas como:

$$PC_{j,t} = P'_j \tilde{X} \quad (3.25)$$

donde $PC_{j,t}$ es la componente principal j y P'_j su correspondiente vector de los loadings.

Consideremos un caso particular donde X tiene solo dos variables correlacionadas y autocorrelacionadas. Por ejemplo, con $s=1$ (lag=1), la matriz \tilde{X} se reduce en $\tilde{X} = [X_0, X_1]$ y las componentes principales pueden ser calculadas como:

$$\begin{aligned} PC_{1,t} &= p_{11}x_{1,t} + p_{12}x_{2,t} + p_{13}x_{1,t-1} + p_{14}x_{2,t-1} \\ PC_{2,t} &= p_{21}x_{1,t} + p_{22}x_{2,t} + p_{23}x_{1,t-1} + p_{24}x_{2,t-1} \\ PC_{3,t} &= p_{31}x_{1,t} + p_{32}x_{2,t} + p_{33}x_{1,t-1} + p_{34}x_{2,t-1} \\ PC_{4,t} &= p_{41}x_{1,t} + p_{42}x_{2,t} + p_{43}x_{1,t-1} + p_{44}x_{2,t-1} \end{aligned}$$

donde p_{ij} son los elementos de la matriz P asociado al loading i y a la variable j .

A diferencia del modelo PCA, las componentes principales en el modelo DPCA son combinaciones lineales de las variables originales y sus respectivos retardos. En el caso particular considerado, además de las variables $x_{1,t}$ y $x_{2,t}$ que forman la matriz X , las variables $x_{1,t-1}$ y $x_{2,t-1}$ son incluidas en el modelo con vista a explotar la estructura de autocorrelación de orden uno en dichas variables. Por tanto, las componentes principales en este modelo DPCA explotan más información en los datos comparado con el correspondiente modelo PCA ajustado a la matriz X .

3.2.2.2. Lags y número de componentes en DPCA

La determinación de número de lags necesarios para formar la matriz \tilde{X} fue discutido por varios autores. En este trabajo destacamos tres métodos:

(1) El método KSG-95: propuesto en 1995 por Ku *et al.* (Ku *et al.*, 1995), consiste en identificar el número de lags a introducir en el modelo, analizando la relación lineal entre las componentes principales de la matriz extendida \tilde{X} mediante análisis de los gráficos de autocorrelaciones y correlaciones cruzadas de los scores;

(2) En 2013 Rato y Reis (Rato y Reis, 2013) propusieron un método (RR-13) que consiste en identificar los lags buscando minimizar en simultaneo el principal valor singular y el principal ratio del valor singular de la matriz \tilde{X} ;

(3) En 2017 Vanhatalo *et al.* (Vanhatalo *et al.*, 2017) propusieron un método que consiste en identificar el número máximo de lags analizando los autovalores de las matrices de correlaciones simples y parciales calculadas a partir de la matriz \tilde{X} . Si el autovalor entre X y X_s es muy bajo (prójimo a cero), indica que la variable correspondiente al $\text{lag}=s$ no presenta información relevante al modelo, por lo que, puede ese lag ser descartado.

Por otro lado, las componentes principales en el modelo DPCA pueden ser seleccionadas usando diferentes criterios, como el de establecer un límite para la varianza total explicada, establecer un mínimo de autovalor para la componente, y la validación cruzada (Vanhatalo *et al.*, 2017). Ku *et al.* (Ku *et al.*, 1995) recomiendan el uso paralelo o combinado de varios métodos para una sección final de las componentes principales más coherente.

3.2.3. Partial Least-Square Regression (PLS)

Partial Least-Square (PLS) es un método que se aplica a los procesos multi-input, multi-output (MIMO) y consiste en el ajuste de un modelo de regresión entre las variables explicativas X y respuestas Y , en el entorno de variables latentes. Este método fue propuesto por Herman Wold (Wold, 1982) a finales de los años sesenta para estudios econométricos. A posterior fue aplicado en la química por Wold *et al.* (Wold *et al.*, 1983) a finales de los años setenta, después del estudio inicial de Kowalski *et al.* (Kowalsk *et al.*, 1982).

El PLS es considerado la generalización del modelo de regresión lineal múltiple (MLR), lo cual subsana los problemas de multicolinealidad, y presenta la ventaja de incluir en el modelo más de una variable respuesta (Wold *et al.*, 1984)(Wold *et al.*, 1993) (Tenenhaus, 1998)(Hoskuldsson, 1996)(Hoskuldsson, 1988)(Wold, 1982).

La calidad predictiva de los modelos de regresión suele depender del número de variables explicativas en el modelo. En este sentido, el modelo PLS presenta ventaja respecto a otros modelos de regresión debido al número reducido de

variables latentes que suele tener.

3.2.3.1. La metodología PLS

La base de esta metodología consiste en proyectar la gran cantidad de datos correlacionados en estructuras de variables latentes. A diferencia del PCA, en este caso las variables latentes son construidas en la dirección de máxima covarianza entre la matriz explicativa X y respuestas Y (Altaf, 2013).

Considere que la matriz de variables explicativas X tiene dimensión $n * m$, y la matriz de variables respuestas Y de dimensión $n * p$. El modelo PLS aplica una relación externa, la que modela X e Y por separado (expresiones 3.29 y 3.30), y otra interna que relaciona las dos matrices X e Y (expresión 3.28).

Sea t_j la variable latente de la matriz X , y w_j su respectivo peso; u_j la variable latente de la matriz Y y su respectivo peso q_j , $j = 1, 2, \dots, k$. Las expresiones que definen t_j y u_j vienen dadas por:

$$t_j = Xw_j \quad (3.26)$$

$$u_j = Yq_j \quad (3.27)$$

El algoritmo PLS busca la dirección de mayor covarianza entre las componentes de X con las de Y . Por tanto, los vectores w y q tienen norma unitaria y son calculados de modo a maximizar la covarianza entre t_j y u_j , es decir: maximizar $\text{cov}(t, u)$ para $\|w\| = \|q\| = 1$.

La relación entre escores t_j y u_j (relación interna) es definida por:

$$u_k = b_k t_k \quad (3.28)$$

donde $b_k = u'_k t_k / t'_k t_k$ y está bajo las mismas reglas que el coeficiente de regresión en un modelo MRL o PCR.

El modelo PLS define las relaciones externas en las matrices X e Y como se muestra a continuación:

$$X = T_k W'_k + E \quad (3.29)$$

$$Y = U_k Q'_k + F \quad (3.30)$$

donde:

T, W, E son los scores, loadings y residuos de los bloques X ;
 U, Q, F son los scores, loadings y residuos de los bloques Y .

3.2.3.2. Algoritmo NIPALS para PLS

El NIPALS (Non-linear Iterative Partial Least Squares) es el algoritmo más popular usado en la construcción de los scores del modelo PLS. El algoritmo es aplicado en la dirección de máxima covarianza entre los scores. A continuación presentamos el algoritmo NIPALS según la estructura presentada por Altaf (Altaf, 2013):

1. Considere las matrices $E_{j+1} = E_j$, $F_{j+1} = F_j$, con $j = 1, \dots, k$, $E_1 = X$ y $F_1 = Y$, tomar u_j igual a una columna de Y ;
2. $w'_j = u'_j E_j / u'_j u_j$
3. $w_j = w'_j / \|w_j\|$ (normalizar w_j)
4. $t_j = E_j w_j / w'_j w_j$
5. Use la matriz F_j para calcular el weight w_j , $q'_j = t'_j F_j / t'_j t_j$
6. $q_j = q'_j / \|q_j\|$ (normalizar q_j)
7. Calcular nuevo output score $u_{\text{nuevo}} = F_j q_j / q'_j q_j$
8. Verifique se hubo convergencia (Se asume convergencia si el valor de u_{nuevo} no cambió significativamente respecto al valor de u_j en el paso anterior). Si se obtiene la convergencia, seguir al paso 9. Caso contrario, volver al paso 2 y sustituir u_j por u_{nuevo} ;
9. El loading p_j de las variables explicativas viene dado por: $p'_j = t'_j E_{j-1} / t'_j t_j$
10. Se usan las matrices U y T de los scores determinados para calcular el coeficiente b_j de regresión interna para las variables latentes:

$$b_j = t'_j u_j / t'_j t_j$$
11. Las matrices de los residuos de las matrices X y Y serán dadas por:

$$E_{j+1} = E_j - t'_j p_j,$$

$$F_{j+1} = F_j - t_j b_j p'_j$$

3.2.3.3. Selección de componentes PLS

El número óptimo de componentes en un modelo PLS garantiza la captura de la variabilidad esencial de los datos originales. De acuerdo con Altaf (Altaf, 2013) la inclusión de un gran número de variables latentes puede causar un ajuste excesivo debido a la inclusión de variables latentes que explican el ruido del proceso. Por otro lado, cuando se considera un número muy reducido de variables latentes no se logra modelar el comportamiento principal del proceso, y resulta en predicción muy pobre.

Wold (Wold, 1978) presenta un métodos de selección de variables latentes designado por R ajustado de Wold, basado en la validación cruzada y es presentado de modo sintético por Altaf (Altaf, 2013) como mostramos a continuación:

1. La base de datos es dividida en s conjuntos, donde $s-1$ conjuntos son usados para entrenamiento y la parte excluida para testar el modelo (grupo de pruebas);
2. Una variable latente es determinada a partir del conjunto del entrenamiento y se aplica al grupo de pruebas. Se calcula el Predicted Error Sum of Squares (PRESS). El proceso se repite para cada uno de los subconjuntos excluidos, donde para cada caso el PRESS es calculado y al final, un total de PRESS es determinado;
3. El procedimiento es repetido para la variable latente 2, 3, hasta $\min(n,m)$ y los correspondientes PRESS total son calculados.

El estadístico R de Wold es calculado como:

$$R = \frac{\text{PRESS}(k+1)}{\text{PRESS}(k)} \geq \alpha \quad (3.31)$$

donde k es el número de variables latentes y el umbral $\alpha = 0,95$.

El estadístico R evalúa la contribución de cada variable latente al modelo PLS. Recordemos que $\text{PRESS}_k = \|X - \hat{X}_k\|$ y $\text{PRESS}_{k+1} = \|X - \hat{X}_{k+1}\|$. Si k es el número óptimo de variables latentes, PRESS_{k+1} no será significativamente inferior que PRESS_k ($R \geq 0,95$) de modo que la nueva variable latente no aportará información significativa al modelo.

3.2.4. PLS para procesos dinámicos (Dynamic PLS-DPLS)

El modelo PLS fue originalmente desarrollado para modelar procesos con datos de entrada (X) y de salida (Y) con alta correlación (Barceló *et al.*, 2011). Como explicamos anteriormente, el modelo consiste en reducir la dimensión de X e Y formando variables latentes, donde se ajusta un modelo de regresión. En los casos donde además de la estructura de correlación entre las variables, existen relaciones dinámicas entre X e Y , el PLS tradicional dejará una gran cantidad de covarianza sin modelar, lo que hace el PLS inadecuado para procesos dinámicos (Dong y Qin, 2015).

Muchos estudios fueron desarrollados con vista a adaptar el modelo PLS a sistemas dinámicos. El modelo partial least square in time series (PLS-TS) fue desarrollado para lograr dicho objetivo, y sus detalles pueden ser vistos en

Eriksson et al. (Eriksson *et al.*, 1984), Dayal y MacGregor (Dayal y MacGregor, 1996), Barceló et al. (Barceló *et al.*, 2011).

En este trabajo destacamos los diferentes enfoques del modelo PLS dinámico (DPLS) presentados por Wold et al. (Wold *et al.*, 1984), Qin y McAvoy (Qin y McAvoy, 1996), Kaspar y Ray (Kaspar y Ray, 1993), Lakshminarayanan et al. (Lakshminarayanan *et al.*, 1997), y nos extenderemos un poco más al trabajo de Dong y Qin (Dong y Qin, 2015).

El método propuesto por Wold et al. (Wold *et al.*, 1984) consiste en incluir los decalajes de las variables en la matriz de las variables explicativas y ajustar el modelo PLS sobre la nueva matriz construida. Qin y McAvoy (Qin y McAvoy, 1996) presentan una propuesta similar a la de Wold et al. (Wold *et al.*, 1984) que consiste en extender la matriz de entrada X incluyendo un cierto número de variables explicativas y respuestas decaladas.

La propuesta de Kaspar y Ray (Kaspar y Ray, 1993) consiste en diseñar filtros mediante un conocimiento dinámico previo de modo que se eliminen la estructura dinámica en la matriz de entrada X. A continuación, la relación dinámica entre los scores de X e Y se ajusta mediante un modelo dinámico, después de ajustar un modelo estático entre las variables explicativas y las respuestas filtradas. Un método similar fue propuesto por Lakshminarayanan et al. (Lakshminarayanan *et al.*, 1997) que consiste en ajustar la relación interna mediante un modelo dinámico entre los scores de X e Y.

Según Dong y Qin (Dong y Qin, 2015) las propuestas anteriormente mencionadas presentan una inconsistencia entre el modelo dinámico ajustado a los scores, que modela la relación interna entre las variables, y el modelo externo estático, ya que los scores extraídos de manera estática son forzados a tener una relación dinámica en el modelo interno.

Para subsanar tal inconveniente, Dong y Qin (Dong y Qin, 2015) proponen el modelo dynamic-inner PLS (DiPLS) lo cual presentamos sus detalles a continuación.

Sea x_k e y_k las variables explicativa y respuesta en el instante k ($k=0,1,\dots,N+1$), y u_k y t_k sus respectivos scores. Según Dong y Qin (Dong y Qin, 2015) un modelo PLS dinámico con consistencia en la relación interna y externa debe verificar lo siguiente:

$$u_k = \beta_0 t_k + \beta_1 t_{k-1} + \dots + \beta_s t_{k-s} + r_k \quad (3.32)$$

y $u_k = y'_k q$ y $t_k = x'_k w$. Donde w es el peso de x_k , q el peso de y_k , y r_k el ruido.

Para cada factor, la relación interna será definida por:

$$\begin{aligned}\hat{u}_k &= x'_k w \beta_0 + x'_{k-1} w \beta_1 + \cdots + x'_{k-s} w \beta_s \\ &= [x'_k \ x'_{k-1} \ \cdots \ x'_{k-s}] (\beta * w)\end{aligned}\quad (3.33)$$

donde $\beta = (\beta_0, \beta_1, \dots, \beta_s)'$

El modelo externo consistente con el modelo interno anterior presentado debe maximizar la covarianza entre u_k y \hat{u}_k como se muestra a continuación:

$$\text{cov}(u_k, \hat{u}_k) = \frac{1}{N} \sum_{k=s}^{N+s} q' y_k [x'_k \ x'_{k-1} \ \cdots \ x'_{k-s}] (\beta * w) \quad (3.34)$$

Dong y Qin (Dong y Qin, 2015) definen las matrices X e Y de la siguiente forma:

$$X = [x_0 \ x_1 \ \cdots \ x_{s+N}]'$$

$$Y = [y_0 \ y_1 \ \cdots \ y_{s+N}]'$$

y definen las matrices Y_s y Z_s como:

$$Y_s = [y_s \ y_{s+1} \ \cdots \ y_{s+N}]'$$

$$Z_s = [X_s \ X_{s-1} \ \cdots \ X_0]'$$

y el modelo es ajustado resolviendo el siguiente problema de maximización:

$$\max [q' Y'_s Z_s (\beta * w)] \quad (3.35)$$

con las restricciones $\|w\| = 1$, $\|q\| = 1$, $\|\beta\| = 1$.
donde s es el orden dinámico del modelo.

Para solucionar este problema de optimización, Dong y Qin (Dong y Qin, 2015) usaron multiplicadores de Lagrange, y los vectores q , w y β que satisface dicha optimización fueron determinando mediante un proceso iterativo que inicia por seleccionar los vectores unitarios q , w y β , y a continuación son calculados como:

$$q = Y'_s Z_s (\beta * w); \quad q := q / \|q\|$$

$$w = (\beta * I)' Z'_s Y_s q; \quad w := w / \|w\|$$

$$\beta = (\mathbf{I} * \mathbf{w})' \mathbf{Z}'_s \mathbf{Y}_s \mathbf{q}; \quad \beta = \beta / \|\beta\|$$

y los scores \mathbf{t} y \mathbf{u} de las variables explicativas y respuestas respectivamente son representados por:

$$\mathbf{t} = [\mathbf{t}_0 \quad \mathbf{t}_1 \dots \mathbf{t}_{s+N}] = \mathbf{X}\mathbf{w}$$

$$\mathbf{u} = [\mathbf{u}_0 \quad \mathbf{u}_1 \dots \mathbf{u}_{s+N}] = \mathbf{Y}\mathbf{q}$$

y la relación dinámica del modelo es definida por:

$$\mathbf{u}_s = \alpha_0 \mathbf{t}_s + \alpha_1 \mathbf{t}_{s-1} + \dots + \alpha_s \mathbf{t}_0 + \mathbf{r}_s \quad (3.36)$$

donde \mathbf{r}_s es el residuo del modelo.

En los casos donde se pretende modelar la estructura autoregresiva de las variables respuestas, el modelo DiPLS pasa a ser representado como sigue:

$$\mathbf{u}_s = \varphi_0 \mathbf{u}_{s-1} + \dots + \varphi_1 \mathbf{u}_{s-f} + \alpha_0 \mathbf{t}_s + \alpha_1 \mathbf{t}_{s-1} + \dots + \alpha_s \mathbf{t}_0 + \mathbf{r}_s \quad (3.37)$$

resultante del siguiente problema de maximización:

$$\begin{aligned} \max \quad & \mathbf{q}'(\gamma_0 \mathbf{Y}'_s + \gamma_1 \mathbf{Y}'_{s-1} + \dots + \gamma_f \mathbf{Y}'_{s-f}) * \\ & * (\beta_0 \mathbf{X}_s + \beta_1 \mathbf{X}_{s-1} + \dots + \beta_s \mathbf{X}_0) \mathbf{w} \end{aligned} \quad (3.38)$$

con las restricciones $\|\mathbf{w}\| = 1$, $\|\mathbf{q}\| = 1$, $\|\beta\| = 1$, $\|\gamma\| = 1$.

Cápitulo 4

METODOLOGÍA

4.1. Introducción

En este capítulo se presentan los procedimientos utilizados en cada etapa del trabajo donde se buscan lograr los objetivos planteados en el capítulo 2. El estado del arte presentado en el capítulo anterior pudo dar una visión general sobre los métodos estadísticos de monitorización y predicción de los procesos de producción. Los métodos fueron presentados en orden del nivel de complejidad de los mismos y del tipo de procesos que modelan. Para cada caso, fueron identificadas las limitaciones que cada uno presenta de cara a la modelación de los procesos en la industria.

Nuestra investigación parte de la hipótesis de que los métodos estadísticos de monitorización y predicción convencionales (que se basan solamente en una variable y cuyos estadísticos se calculan con base en las hipótesis de normalidad e independencia entre observaciones) no son adecuados para explotar la información relevante en los procesos multivariantes y autocorrelacionados. Con base en ello, la investigación se orienta a desarrollar métodos capaces de subsanar tal inconveniente. De acuerdo con los objetivos de este trabajo, la investigación se orienta en dos direcciones:

1. Monitorización de procesos: se elabora un modelo para monitorizar los procesos multivariantes autocorrelacionados y;
2. Predicción de procesos: se propone un modelo dinámico ajustado sobre estructuras latentes para predecir la calidad del producto a través de las características del proceso.

Respecto a la monitorización, el estudio se orienta a la investigación de los gráficos EWMA multivariantes para procesos autocorrelacionados. El uso de este método se justifica por la ventaja que tiene al detectar pequeños cambios en el proceso, y la posibilidad de poder adaptarlo a las variables autocorrelacionadas.

La autocorrelación de las variables del proceso fue modelada mediante modelos autoregresivos, como lo hicieron Morettin y Toloí (Morettin y Toloí, 2004). Fueron simulados procesos multivariantes con variables autoregresivas bajo control, donde los parámetros del gráfico MEWMA propuesto fueron calculados. Cambios de diferentes magnitudes fueron simulados en las series determinadas y el desempeño del gráfico fue testado a través de los estadísticos RL (Run of Length) y ARL (Average Run of Length).

A los ruidos blancos de los procesos anteriores simulados fueron ajustados los gráficos MEWMA convencionales, y a través de AR y ARL su desempeño fue comparado con el gráfico MEWMA propuesto.

Respecto al modelo de predicción, se parte de la hipótesis de que existe influencia de las características del proceso en la calidad del producto. En vista a los casos donde eso ocurre (casos donde además de la estructura multivariante y autocorrelacionada, hay dos bloques de variables: X-del proceso e Y-del producto) se ajustó un modelo de regresión dinámica sobre los scores obtenidos en los dos bloques de variables, al aplicar un modelo DPCA.

El método consiste básicamente en la combinación del modelo DPCA y la función de transferencia (FT). El modelo DPCA se aplica en los dos bloques de variables con vista a explotar la estructura de correlación dinámica entre las variables. A continuación, se ajusta un modelo de función de transferencia sobre las variables latentes determinadas para estudiar la influencia dinámica de las características del proceso sobre la calidad del producto.

De modo sintético el modelo de predicción fue desarrollado siguiendo principalmente estas dos etapas:

1. Aplicar un análisis de componentes principales dinámico en cada bloque de variables (X-proceso e Y-producto);
2. Sobre las variables latentes obtenidas ajustar un modelo de función de transferencia.

La función de transferencia fue ajustada mediante la metodología de Box-Jenkins, la cual será presentada con detalle en el próximo capítulo.

Los resultados del modelo dinámico ajustado sobre estructuras latentes fueron comparados con el modelo dynamic partial least squares (DPLS) ajustado a los mismos datos y en condiciones idénticas. Para ello, la variabilidad explicada por las variables latentes y la calidad predictiva de los modelos ajustados fueron considerados para comparar las dos metodologías.

Por tanto, el método estadístico propuesto en esta tesis corresponde al gráfico MEWMA adaptado a las variables autocorrelacionadas para controlar

las características del proceso (o del producto); y el modelo dinámico ajustado sobre estructuras latentes, para predecir la calidad del producto a través de las características del proceso.

Una visión integral de la metodología estadística referida se presenta en el esquema de la figura 4.1. A continuación presentamos sus detalles.

4.2. Gráfico MEWMA para procesos autocorrelacionados

Este método consiste en adaptar el gráfico MEWMA convencional propuesto por Lowry et al. (Lowry *et al.*, 1992) a los procesos multivariantes autocorrelacionados. Con la estrategia usada buscarse incluir la estructura de autocorrelación en el gráfico para evitar falsas alarmas, y dotarlo de capacidad en detectar rápidamente cambios en la media del proceso.

El control de procesos multivariantes y autocorrelacionados mediante el gráfico MEWMA construido bajo las hipótesis de normalidad e independencia excluye información relevante del proceso causada por la correlación entre observaciones. De acuerdo con Claro et al. (Claro *et al.*, 2007) esto resulta en estimaciones erróneas en los parámetros del gráfico. Estos autores realizaron un estudio de control de un proceso autocorrelacionado univariando mediante el gráfico EWMA, donde sus parámetros fueron calculados con base en un modelo autoregresivo de primer orden. Los resultados fueron comparados con un gráfico EWMA basado en las hipótesis de normalidad e independencia, lo cual resultó ser muy susceptibles a falsas alarmas.

Nos basamos a esta metodología para extender el gráfico EWMA a los procesos multivariantes autocorrelacionados. Para ello, los modelos de series temporales (en particular los modelos autoregresivos de orden uno) constituyen una herramienta fundamental para el desarrollo de nuestra metodología.

De manera general, el ajuste del modelo MEWMA consiste en:

1. Ajustar un modelo autoregresivo a cada variable del proceso y determinar las respectivas medias y varianzas;
2. De forma bi-variante, se calcula las covarianzas cruzadas;
3. Con los resultados de 1 y 2 se construye un vector de medias y una matriz de varianzas-covarianzas cruzadas.
4. Con los parámetros anteriores, se construye el gráfico MEWMA

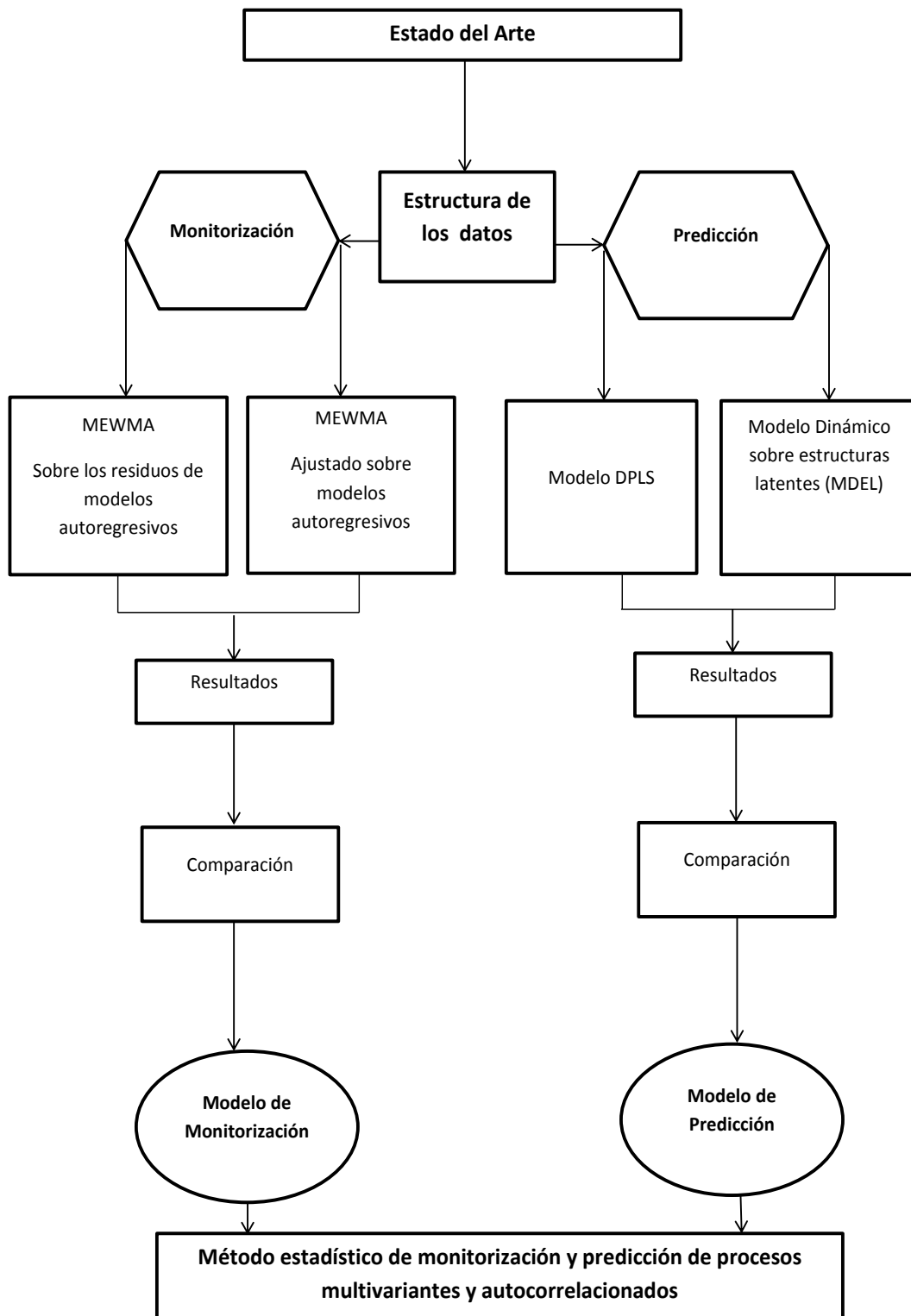


Figura 4.1: Esquema de las etapas del estudio

Para una definición formal de la función de covarianza cruzada se puede recurrir a Box y Jenkins (Box y Jenkins, 1976) que consideran dos procesos conjuntamente estacionarios x_t e y_t para $t = 1, 2, \dots$ y establecen que la covarianza cruzada de orden k se puede calcular mediante la siguiente expresión:

$$\gamma_{xy}(k) = E((X_t - \mu_x)(Y_{t+k} - \mu_y)) \quad (4.1)$$

para $k = \pm 1, \pm 2, \dots$. Así, en función de k , $\gamma_{xy}(k)$ es la llamada función de covarianza cruzada entre x_t e y_t .

Designamos por Σ_γ la matriz de varianzas-covarianzas cruzadas que recoge todas las varianzas y covarianzas cruzadas de las variables del proceso, como veremos adelante, no depende de k . Las varianzas de esta matriz fueron calculadas con base en la fórmula usada por Vasilopoulos y Stamboulis (Vasilopoulos y Stamboulis, 1978) y por Costa et al. (Costa *et al.*, 2005), con la cual se extiende para el cálculo de las covarianzas cruzadas. Esta matriz representa, en cierto modo, el factor de inclusión de la estructura de autocorrección en el gráfico.

4.3. Modelo dinámico sobre estructuras latentes (MDEL)

El modelo dinámico sobre estructuras latentes corresponde a nuestra propuesta metodológica para la predicción de los procesos multivariantes y autocorrelacionados. El planteamiento de esta metodología se fundamenta en la estructura compleja de los datos de muchos procesos en la industria, donde además de la alta correlación entre las variables, autocorrelación entre observaciones, la calidad del producto depende de las condiciones del entorno de fabricación.

Sean X e Y las matrices que recogen las variables del proceso y del producto, respectivamente. El modelo DPCA se aplica en las dos matrices de manera independiente.

Las variables latentes son calculadas explorando la estructura de correlación dinámica entre las variables en los dos bloques X e Y . Dichas variables conservan la estructura dinámica de las variables originales, por lo que son series temporales y una metodología adecuada fue considerada para su modelación. Por tanto, el estudio de la influencia de las variables latentes del proceso sobre las variables del producto fue llevado a cabo mediante una función de transferencia.

C apitulo 5

PROPUESTA DE M ETODOS DE MONITORIZACI ON Y PREDICCI ON DE PROCESOS MULTIVARIANTES Y AUTOCORRELACIONADOS

5.1. Introducci on

Los m etodos estad isticos presentados en el capitulo anterior conforman la base de nuestra investigaci on. De hecho, hemos realizado esa descripci on para dar una visi on general sobre los aspectos relevantes que corresponden al fundamento de su desarrollo. En base a los resultados obtenidos en el an alisis bibliogr afico realizado y a los objetivos del trabajo, pretendemos en este capitulo presentar los algoritmos usados en nuestra propuesta metodol ogica basada en los modelos estad isticos de monitorizaci on y predicci on para los procesos multivariantes y autocorrelacionados.

El gr afico MEWMA adaptado a las variables autocorrelacionadas y el modelo dinámico ajustado sobre las estructuras latentes son el m etodo estad istico que proponemos para afrontar el reto en la monitorizaci on y predicci on de los procesos multivariantes y autocorrelacionados. El cap itulo est a dividido en dos partes, donde en cada una se desarrolla cada uno de los modelos mencionados.

La primera parte corresponde a la construcci on del gr afico MEWMA propuesto. El m etodo consiste en adaptar el gr afico MEWMA propuesto por Lowry et al. (Lowry *et al.*, 1992) a los procesos autocorrelacionados, donde se busc o incluir la estructura de autocorrelaci on en el c alculo de los par ametros del gr afico.

Los modelos de series temporales constituyen una herramienta muy relevante en nuestro estudio. Sus detalles matem aticos y funcionales son presentados en la segunda parte del capitulo (predicci on), en la que su algoritmo tiene un papel fundamental. Los modelos y conceptos relacionados con series temporales son mencionados en la primera parte (monitorizaci on) como estrategia de c alculo del vector de medias y de la matriz de varianzas-covarianzas cruzada que se usa en la construcci on del gr afico MEWMA propuesto.

Como comentamos anteriormente, el modelo de predicción consiste en la combinación del modelo DPCA y de la función de transferencia. En la segunda parte del capítulo se muestra el modo en que estos dos métodos son implementados en el modelo para lograr un buen ajuste.

5.2. Gráficos MEWMA para procesos autocorrelacionados

Considere un conjunto de variables correlacionadas entre si y cada una dependiente del tiempo. El estudio consiste, en primer lugar, en ajustar modelos autoregresivos sobre las series temporales, donde se calculan las medias, las varianzas, autocovarianzas y las covarianzas cruzadas entre cada par de variables. Como resultado se obtiene un vector de medias y una matriz de varianzas-covarianzas cruzadas, con los cuales se calculan los parámetros del gráfico MEWMA.

En la figura 5.1 se presenta un esquema que visualiza los procedimientos usados en la construcción del modelo y su posterior comparación con el gráfico MEWMA ajustado sobre los ruidos blancos resultantes del modelo autoregresivo ajustado a las series del proceso. Pasamos a explicar cada uno de los aspectos considerados:

(1) Variables del proceso

Consideremos las siguientes variables que caracterizan un proceso (o producto) y que sus observaciones dependen del tiempo t :

$$p_{1t}, p_{2t}, p_{3t}, \dots, p_{mt}$$

Supongamos que las m variables están correlacionadas entre si, y que la dependencia de sus observaciones en el tiempo causa una estructura de autocorrelación significativa en cada variable, entonces se puede afirmar que cada variable p_{it} ($i = 1, \dots, m$ y $t = 1, \dots, n$) es una serie temporal con lo cual se puede aplicar la metodología correspondiente.

En cada instante, el valor de p_{it} será o bien la única observación realizada en ese instante t , o bien el promedio de las observaciones en t . En el caso de la construcción de un gráfico MEWMA, habitualmente p_{it} será la media de las observaciones de la variable p_i tomadas en el instantes t .

(2) Cálculo de medias y varianzas del proceso

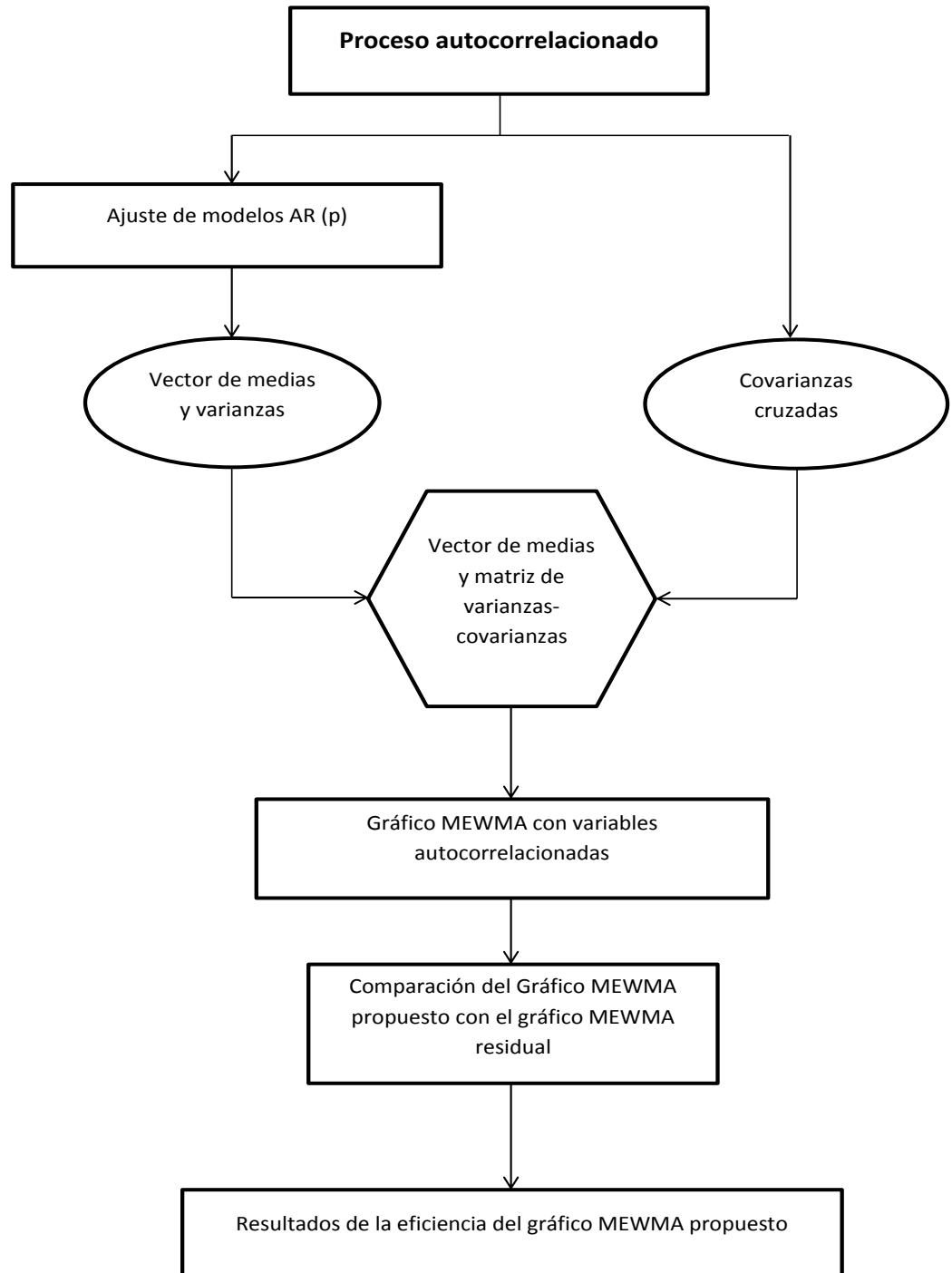


Figura 5.1: Construcción del modelo de monitorización

Considere un modelo autoregresivo de orden d (AR(d)) para cada variable p_{it} definido como:

$$\Phi_{d,i}(B)p_{it} = c + a_{it} \quad (5.1)$$

donde $\Phi_d(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_d B^d$, B es el operador de retardo y a_{it} es un ruido blanco ($a_{it} \sim N(0, \sigma)$ a_{it_1} independiente de a_{it_2} , $\forall t_1 \neq t_2$).

El modelo AR(d) es estacionario si y solamente si las raíces del polinomio $\Phi_d(B)$ están fuera del círculo unitario. En los procesos donde no se verifica esta condición, deben aplicarse estrategias para poder lograr la estacionariedad.

En muchos casos reales en la industria los procesos suelen ajustarse a un modelo autoregresivo de orden uno. Por ello, pasaremos a considerar el caso particular donde $d=1$. El modelo AR(1) estacionario puede ser representado como:

$$p_{it} - \mu = \phi_{1i}(p_{it-1} - \mu) + a_t \quad (5.2)$$

La media y la autocovarianza del modelo son calculadas respectivamente a través de las siguientes expresiones:

$$\mu_i = \frac{c}{1 - \phi_{1i}} \quad (5.3)$$

$$\gamma_i(k) = \frac{\sigma^2 \phi_{1i}^k}{1 - \phi_{1i}^2} \quad (5.4)$$

donde $k = 0, 1, 2, \dots$; $\sigma_{p_i}^2 = \frac{\sigma^2}{1 - \phi_{1i}^2}$ y σ^2 es la varianza del ruido blanco a_t .

La formula 5.2 representa el valor de cada variable i (o el promedio) en cada instante t de muestreo, tomando como ejemplo una de las variables, tenemos:

$$\begin{aligned} t = 1 & \quad p_1 = \mu + a_1 \\ t = 2 & \quad p_2 = \mu + \phi(p_1 - \mu) + a_2 \\ t = 3 & \quad p_3 = \mu + \phi(p_2 - \mu) + a_3 \\ t = 4 & \quad p_4 = \mu + \phi(p_3 - \mu) + a_4 \\ & \quad \dots \\ t = n & \quad p_n = \mu + \phi(p_{n-1} - \mu) + a_n \end{aligned}$$

Los promedios y las respectivas varianzas pueden ser calculados como:

$$\begin{aligned}
 t = 2 \quad \bar{p} &= \frac{p_1 + p_2}{2} & \sigma_{\bar{p}}^2 &= \frac{4 + 4\phi}{2} \sigma_p^2 \\
 t = 3 \quad \bar{p} &= \frac{p_1 + p_2 + p_3}{3} & \sigma_{\bar{p}}^2 &= \frac{9 + 12\phi + 6\phi^2}{3} \sigma_p^2 \\
 t = 4 \quad \bar{p} &= \frac{p_1 + p_2 + p_3 + p_4}{4} & \sigma_{\bar{p}}^2 &= \frac{16 + 24\phi + 16\phi^2 + 4\phi^3}{4} \sigma_p^2 \\
 & \dots & & \\
 t = n \quad \bar{p} &= \frac{p_1 + p_2 + \dots + p_n}{n} & \sigma_{\bar{p}}^2 &= \frac{1}{n} \left[\sigma_p^2 + 2 \sum_{k=1}^{n-1} \left(1 - \frac{k}{n}\right) \gamma_i(k) \right]
 \end{aligned}$$

El parámetro $\sigma_{\bar{p}}^2$ no depende de k y recoge la varianza y la autocovarianza de la variable p . Dicha fórmula fue usada por Vasilopoulos y Stamboulis (Vasilopoulos y Stamboulis, 1978) y por Costa et al. (Costa et al., 2005) en sus estudios de control estadístico de procesos univariados. De modo a uniformizar la nomenclatura que usamos, pasaremos a representar $\sigma_{\bar{p}}^2$ por γ_i donde i representa cada variable del proceso, y $\gamma_i(0) = \sigma_p^2$. Así, la fórmula anterior pasa a representarse:

$$\gamma_i = \frac{1}{n} \left[\gamma_i(0) + 2 \sum_{k=1}^{n-1} \left(1 - \frac{k}{n}\right) \gamma_i(k) \right] \quad (5.5)$$

(3) Cálculo de covarianzas cruzadas

Para cada par de variables p_{it} y p_{jt} ($\forall i \neq j$), las covarianzas cruzadas fueron calculadas para las m variables del proceso. El número de funciones de correlación cruzada es $C_2^m = \frac{m!}{2!(m-2)!}$.

La función de covarianza cruzada de orden k entre p_{it} y p_{jt} está definida por:

$$\gamma_{i,j}(k) = E[(p_{it} - \mu_i)(p_{jt-k} - \mu_j)] \quad (5.6)$$

para $k = 0, \pm 1, \pm 2, \dots$

Para $k = 0$ la función de covarianza cruzada se convierte en una covarianza instantánea entre p_{it} y p_{jt} , es decir, mide la relación del mismo instante t de dichas dos variables.

Para el cálculo de las covarianzas cruzadas entre las variables p_{it} y p_{jt} , que no depende de k , se usó el mismo raciocinio que el caso de la varianza.

Nuestro objetivo es que los parámetros recojan la máxima variabilidad dinámica en los datos, por tanto, consideramos todas las variaciones posibles entre cada par de variables del proceso.

Consideremos los cuatro primeros valores ($t=1,2,3,4$) de las variables p_{it} y p_{jt} , con media μ_i y μ_j respectivamente. Definimos todas las discrepancias posibles respecto a la media (llamamos discrepancias al producto entre las diferencias de cada variable con su respectiva media) en cada instante t , como se muestra a continuación:

$$\begin{aligned}
 t = 1 \quad p_{1i} \quad p_{1j} &: (p_{1i} - \mu_i)(p_{1j} - \mu_j) \\
 t = 2 \quad p_{2i} \quad p_{2j} &: (p_{2i} - \mu_i)(p_{2j} - \mu_j), (p_{2i} - \mu_i)(p_{1j} - \mu_j) \\
 t = 3 \quad p_{3i} \quad p_{3j} &: (p_{3i} - \mu_i)(p_{3j} - \mu_j), (p_{3i} - \mu_i)(p_{2j} - \mu_j), \\
 & \quad (p_{3i} - \mu_i)(p_{1j} - \mu_j) \\
 t = 4 \quad p_{4i} \quad p_{4j} &: (p_{4i} - \mu_i)(p_{4j} - \mu_j), (p_{4i} - \mu_i)(p_{3j} - \mu_j), \\
 & \quad (p_{4i} - \mu_i)(p_{2j} - \mu_j), (p_{4i} - \mu_i)(p_{1j} - \mu_j)
 \end{aligned}$$

.....

A partir del instante $t > 1$ es posible calcular el producto de las discrepancias entre los valores del mismo instante y de instantes anteriores entre cada par de variables, lo cual representan las relaciones dinámicas entre las variables.

El promedio de las discrepancias para los cuatro instantes presentados es:

$$\begin{aligned}
 & \frac{(p_{1i} - \mu_i)(p_{1j} - \mu_j) + (p_{2i} - \mu_i)(p_{2j} - \mu_j) + (p_{3i} - \mu_i)(p_{3j} - \mu_j) + (p_{4i} - \mu_i)(p_{4j} - \mu_j)}{4} + \\
 & \frac{(p_{4i} - \mu_i)(p_{3j} - \mu_j) + (p_{4i} - \mu_i)(p_{2j} - \mu_j) + (p_{4i} - \mu_i)(p_{1j} - \mu_j)}{4} + \\
 & \frac{(p_{3i} - \mu_i)(p_{2j} - \mu_j) + (p_{3i} - \mu_i)(p_{1j} - \mu_j) + (p_{2i} - \mu_i)(p_{1j} - \mu_j)}{4}
 \end{aligned}$$

que representa a la vez la variabilidad estática y dinámica entre las variables i y j en los cuatro primeros instantes.

De modo general, para $t=n$ periodos se define $\gamma_{i,j}$ como:

$$\gamma_{i,j} = \underbrace{\sum_{t=1}^n (p_{it} - \mu_i)(p_{jt} - \mu_j)}_n \underbrace{+ \sum_{k=1}^{n-1} (p_{it} - \mu_i)(p_{jt-k} - \mu_j)}_n \underbrace{+ \sum_{k=1}^{n-2} (p_{it-k} - \mu_i)(p_{jt-k-1} - \mu_j)}_n$$

$$\underbrace{E[(p_{it} - \mu_i)(p_{jt} - \mu_j)]}_{E[(p_{it} - \mu_i)(p_{jt} - \mu_j)]} \underbrace{E[(p_{it} - \mu_i)(p_{jt-k} - \mu_j)]}_{E[(p_{it} - \mu_i)(p_{jt-k} - \mu_j)]} \underbrace{E[(p_{it-k} - \mu_i)(p_{jt-k-1} - \mu_j)]}_{E[(p_{it-k} - \mu_i)(p_{jt-k-1} - \mu_j)]}$$

$$\gamma_{i,j} = \gamma_{i,j}(0) + \underbrace{E[(p_{it} - \mu_i)(p_{jt-k} - \mu_j)] + E[(p_{it-k} - \mu_i)(p_{jt-k-1} - \mu_j)]}_{\varphi} \quad (5.7)$$

donde $\gamma_{i,j}(0)$ es la covarianza instantánea entre p_i y p_j ;

$E[(p_{it} - \mu_i)(p_{jt-k} - \mu_j)]$ es la covarianza entre los valores de p_i en el instante t y los valores anteriores de p_j ;

$E[(p_{it} - \mu_i)(p_{jt-k} - \mu_j)]$ es la covarianza entre cada valor anterior de p_i con todos los valores anteriores de p_j .

φ es la parte que recoge la estructura dinámica entre p_i y p_j . Si $k=0$, la relación entre las variables i y j es estática, y por lo tanto, $\gamma_{i,j} = \gamma_{i,j}(0)$.

Los parámetros γ_i y $\gamma_{i,j}$ recogen la variabilidad máxima de las variables del proceso, siendo que γ_i recoge la varianza y la autocovarianza de cada variable p_{it} , y $\gamma_{i,j}$ recoge la covarianza y las covarianzas cruzadas entre p_{it} y p_{jt} .

(4) Vector de medias y la matriz de varianzas-covarianzas cruzadas

En el estudio realizado en (2) se calcularon las medias y las varianzas de cada serie. El conjunto de las medias calculadas constituye un vector de dimensión m representado por:

$$\vec{p}_t = (p_{1t}, p_{2t}, \dots, p_{mt}) \quad (5.8)$$

donde m es el número de variables del proceso y t el instante de muestreo.

Las varianzas (formula (5.5)) y las covarianzas cruzadas (formula (5.7)) calculadas forman la siguiente matriz:

$$\Sigma_{\gamma} = \begin{pmatrix} \gamma_1 & \dots & \gamma_{1,m} \\ \gamma_{2,1} & \gamma_2 & \dots \gamma_{2,m} \\ \vdots & \gamma_{i,j} & \gamma_i \dots \\ \gamma_{m,1} & \dots & \dots \gamma_m \end{pmatrix} \quad (5.9)$$

Con los cuales se ajusta el gráfico MEWMA que se presenta a continuación:

(5) Gráfico de control MEWMA con variables autocorrelacionadas

Los parámetros calculados a través de las fórmulas (5.5) y (5.7) recogen el efecto de la autocovarianza del proceso (o autocorrelación, basta que se divida por las respectivas desviaciones típicas) que forman la matriz Σ_γ usada para ajustar el gráfico. Además, se incluye en el estadístico MEWMA la diferencia entre los valores del vector de medias actual y el anterior ($\vec{p}_t - \vec{p}_{t-1}$) para que el gráfico sea muy rápido en detectar el cambio más reciente del proceso, como hicieron Patel y Devicha (Patel y Divecha, 2013). En los casos donde no hay cambio en el vector de medias, $\vec{p}_t - \vec{p}_{t-1}$ es nulo, y por lo tanto, volvemos al estadístico MEWMA de Lowry et al. (Lowry et al., 1992) habitual. El estadístico del gráfico MEWMA propuesto se define como:

$$\vec{z}_t = \Lambda \vec{p}_t + (1 - \Lambda) \vec{z}_{t-1} + (\vec{p}_t - \vec{p}_{t-1}), t \geq 1 \quad (5.10)$$

La matriz diagonal Λ y el estadístico T_{3t}^2 son calculados de modo similar que en el gráfico MEWMA y MMOEWMA presentados en el capítulo 3, considerando la hipótesis de igualdad de los pesos λ para las m variables ($\lambda_1 = \dots = \lambda_m = \lambda$) y cambiando Σ por Σ_γ en el cálculo de $\Sigma_{\vec{z}_t}$, o sea, $\Sigma_{\vec{z}_t} = (\frac{\lambda}{2-\lambda} + \frac{2\lambda(1-\lambda)}{2-\lambda})\Sigma_\gamma$. Si considera que el proceso está fuera de control si:

$$T_{3t}^2 = \vec{z}_t' \Sigma_{\vec{z}_t}^{-1} \vec{z}_t > h_3$$

donde $h_3 > 0$ es calculado de modo que se obtenga un determinado ARL bajo control.

En este trabajo el límite superior h_3 fue calculado aplicando el método de la cadena de Markov usado por Patel y Devicha (Patel y Divecha, 2013) y por Lucas y Saccucci (Lucas y Saccucci, 1990) para el cálculo de ARL.

En la línea de estos autores, la región de control definido por la norma $\|\vec{z}_t\|$ fue dividida en $h+1$ estados, cada uno con anchura $g = \frac{2UCL}{2h+1}$, donde UCL representa el último límite superior de los $h+1$ intervalos considerados, o sea, la región de control corresponde al intervalo de cero hasta UCL (0, UCL).

Considere X_t una variable aleatoria que define el estado de una observación en la región de control en el tiempo t . La probabilidad de que una observación transite de un estado $X_{t-1} = i$ a otro $X_t = j$ es representada por $p(i, j) = P[X_{t-1} = i | X_t = j]$. Patel y Devicha (Patel y Divecha, 2013),

demonstraron que dicha probabilidad puede ser calculado como:

$$p(i, j) = P \left[\frac{(j - 0,5)^2 g^2}{\lambda^2} < \chi^2(m, c) < \frac{(j + 0,5)^2 g^2}{\lambda^2} \right]$$

donde $\chi^2(m, c)$ es una variable aleatoria con distribución chi-cuadrado no centrada, con m grados de libertad y $c = [(1 - \lambda)ig/\lambda + d]^2$; d representa el valor del cambio de la media del proceso. Por lo tanto, si no hay cambio en el proceso, el valor de d y j son nulos, y la probabilidad de transición se define como:

$$p(i, 0) = P[\chi^2(m, c) < ((0,5)^2 g^2)/\lambda^2]$$

que corresponde a la probabilidad de que una observación se mantenga bajo control.

En este trabajo, hemos considerado un $ARL(d = 0) = 370$, que corresponde a una probabilidad de falsas alarmas $\alpha = 1 - p(i, 0) = 0,0027$. Con esta probabilidad, el valor $((0,5)^2 g^2)/\lambda^2$ fue igualado al valor crítico correspondiente a $\chi^2(m, c)$, donde el valor de g fue calculado. El valor de UCL fue determinado para distintos valores de h y finalmente el límite superior h_3 fue calculado como:

$$h_3 = \frac{UCL^2}{\left(\frac{\lambda}{2-\lambda} + \frac{2\lambda(1-\lambda)}{2-\lambda} \right)}$$

En la línea de la nomenclatura usada para los gráficos anteriores, se puede identificar este gráfico propuesto como: Multivariate Autocorrelated Adapted EWMA chart (MAAEWMA).

5.2.1. Análisis del desempeño del gráfico MAAEWMA

El desempeño del gráfico MAAEWMA fue analizado mediante una simulación de procesos bivariantes y autocorrelacionados, donde las correspondientes variables son series temporales autoregresivas de orden uno y correlacionadas entre sí. La simulación fue implementada en R, siguiendo los pasos:

1. Fueron simuladas series p_{it} y p_{jt} de tamaño 10000, con las cuales los parámetros del gráfico MAAEWMA y MEWMA residual fueron calculados, con $\lambda = 0,1$ (mejor para detectar pequeños cambios (Aparisi y Haro, 2003)). Los límites h y h_3 fueron calculados para un $ARL_0 = 370$;

2. Las series anteriores fueron subdivididas en 10 partes, formando 10 subseries de tamaño 1000. Para cada par (p_{it}, p_{jt}) de las subseries obtenidas fueron simulados cambios de magnitud $\delta = 1, 2, 3$ en la media del proceso a partir de la observación 50;
3. Las series autoregresivas fueron simuladas considerando cuatro coeficientes autoregresivos: 0.3, 0.5, 0.7, 0.9 para representar procesos con baja, media y alta autocorrelación.
4. Para analizar el comportamiento de los dos gráficos se empleó un diseño factorial completo con tres factores: dos de ellos corresponden a los coeficientes del modelo AR(1) de cada una de las variables consideradas, y cada uno con cuatro niveles (0.3, 0.5, 0.7, 0.9). El tercer factor corresponde al cambio δ que presenta tres niveles $\delta = 1, 2, 3$;

El conjunto de procesos bivariantes fue formado considerando una combinación de los cuatro coeficientes de ambas variables, formando 16 pares de series autoregresivas y correlacionadas. Para cada uno de los 16 pares, los tres cambios $\delta = 1, 2, 3$ fueron considerados. Para cada cambio δ el proceso fue simulado 10 veces, y en cada caso el RL (Run of Length) fue calculado. El ARL fue determinado como el promedio de los RLs obtenidos en los grupos de 10 pares simulados. La combinación de los tres factores formó un conjunto de datos con $4*4*3*10=480$ observaciones.

El experimento fue analizado mediante un análisis de la varianza (ANOVA) para estudiar la influencia de los tres factores sobre el RL obtenido en el gráfico MAAEWMA (figuras 5.2-5.5).

Para los tres factores el efecto resultó muy significativo ($p < 0,0001$). Su efecto se ilustra en las figuras mencionadas.

Los resultados muestran una diferencia significativa en los valores de RL (y por tanto ARL) para procesos con diferentes coeficientes autoregresivos que sufrieron el mismo cambio δ en la media.

La estructura autoregresiva de ambas series y la correlación entre ellas está seguramente detrás de la presencia de múltiples interacciones significativas (figura 5.5).

El rendimiento del gráfico MAAEWMA fue mejor para procesos con media y alta autocorrelación. En los 480 casos simulados, el peor resultado se obtuvo

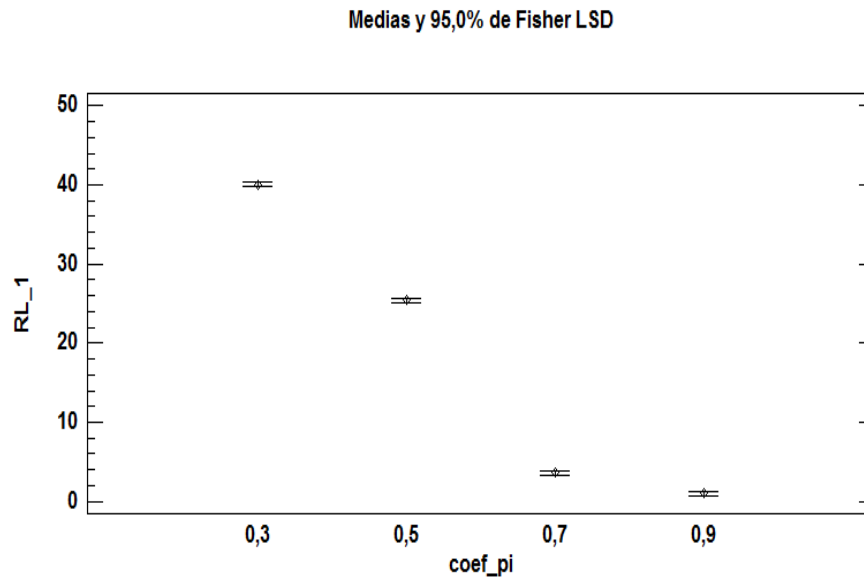


Figura 5.2: Efecto del coeficiente autoregresivo de la variable p_i en el RL (Run of Length) para los tres cambios considerados

en los procesos donde ambas variables tienen $\phi = 0,3$. Los mejores resultados del gráfico MAAEWMA respecto al MEWMA residual se obtuvieron en los procesos con $\phi > 0,3$, siendo que la mayor diferencia fue obtenida en los procesos con $\phi = 0,7$ y $\phi = 0,9$. Con lo cual se puede afirmar que el gráfico MAAEWMA presenta mejor desempeño para procesos con media y alta autocorrelación (ver figuras 5.6-5.9).

Consideremos la ganancia en términos de ARL definido de la siguiente manera:

$$\Delta \text{ARL}_1 \% = \frac{\text{ARL}_{1A} - \text{ARL}_{1B}}{\text{ARL}_{1A}} * 100 \quad (5.11)$$

donde ARL_{1A} y ARL_{1B} son los respectivos ARL del proceso fuera de control del gráfico MEWMA construido sobre los residuos y del gráfico MAAEWMA. Dicha ganancia es un indicador de la mejora relativa que se obtendría al aplicar el gráfico propuesto respecto a otra metodología.

El promedio de la ganancia obtenida para todos los casos $\Delta \overline{\text{ARL}}_1 \%$ es mayor que zero ($\text{sig} = 0,0033$), indicando que el ARL_1 del gráfico MAAEWMA es inferior que los de MEWMA residual.

Una prueba de hipótesis t de student con medidas apareadas fue considerada para testar la diferencia entre ARL_{1A} y ARL_{1B} correspondiente a las medias de cada 10 pares de los correspondientes RL_{1A} y RL_{1B} . La tabla 5.1 muestra que hubo una diferencia significativa (p – valor $< 0,0001$) entre dichos

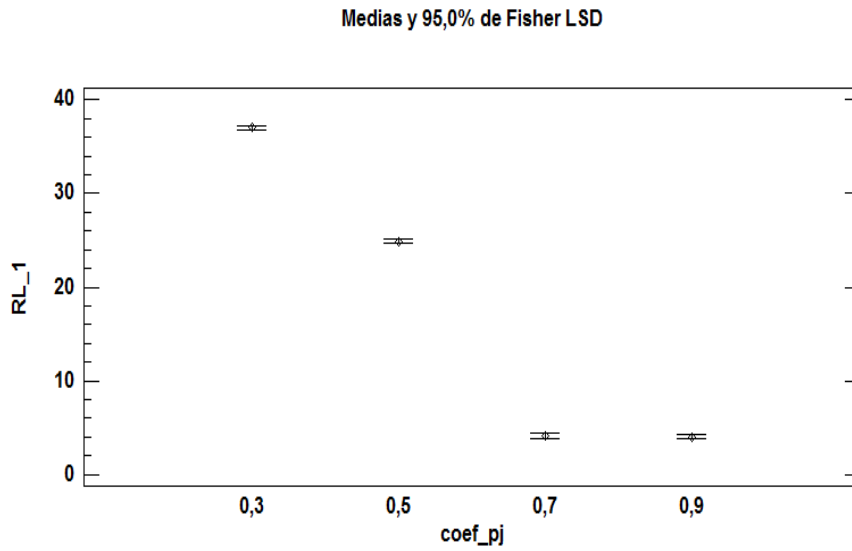


Figura 5.3: Efecto del coeficiente autoregresivo de la variable p_j en el RL (Run of Length) para los tres cambios considerados

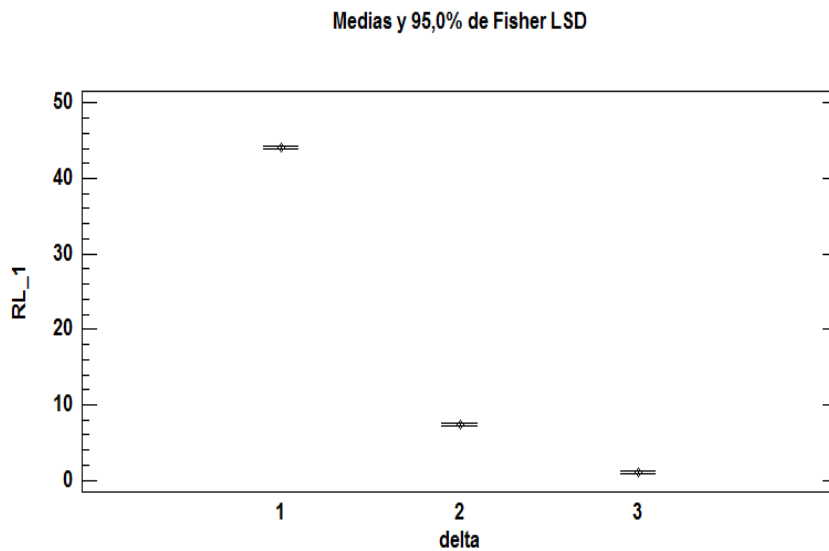


Figura 5.4: Comparación del RL (Run of Length) para los tres cambios considerados

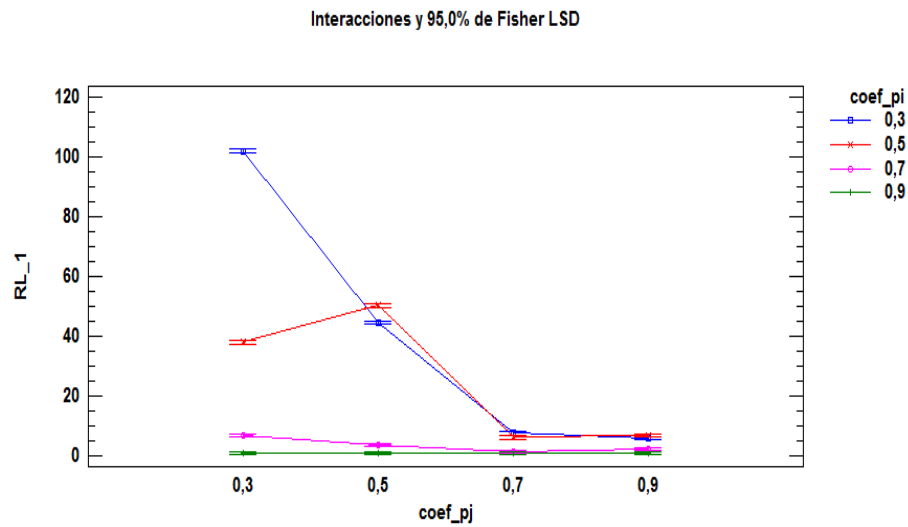


Figura 5.5: Interacciones de procesos con diferentes coeficientes autoregresivos

ARLs. Siendo que en los procesos con autocorrelacion media y alta resultó que $ARL_{1A} > ARL_{1B}$.

El valor negativo de $\Delta ARL_1 \%$ obtenido para procesos con $\phi = 0,3$ (Tabla 5.1) muestra que el desempeño del gráfico MAAEWMA es menor respecto al MEWMA residual para procesos con baja autocorrelación. Sin embargo, su desempeño se torna mayor para cambios de mayor magnitud.

El gráfico MAAEWMA fue desarrollado para monitorizar procesos con autocorrelación significativa. En los procesos con baja autocorrelación no hay una fuerte dependencia entre las observaciones, y en este caso los gráficos MEWMA propuestos por Lowry et al. (Lowry *et al.*, 1992) presentan ligera ventaja respecto al gráfico MAAEWMA.

Las figuras 5.6-5.9 ilustran los resultados de algunos casos simulados con diferentes combinaciones de los coeficientes:

Figura 5.6: $\phi_1(p_i) = \phi_1(p_j) = 0,9$,

Figura 5.7: $\phi_1(p_i) = 0,7$ y $\phi_1(p_j) = 0,5$,

Figura 5.8: $\phi_1(p_i) = 0,5$ y $\phi_1(p_j) = 0,3$ y,

Figura 5.9: $\phi_1(p_i) = 0,3$ y $\phi_1(p_j) = 0,3$,

Con un cambio $\delta = 1$ en la media de p_i a partir de la observación 50.

En los tres primeros casos (Figuras 5.6-5.8) el gráfico MAAEWMA (chart(a))

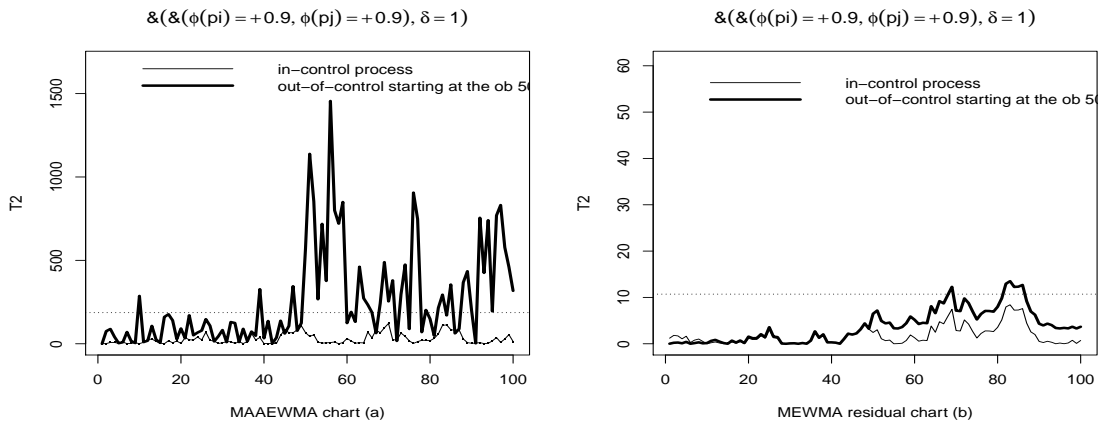


Figura 5.6: Monitorización de un proceso autocorrelacionado bivariate con coeficientes $p_i(\phi = 0,9)$ y $p_j(\phi = 0,9)$

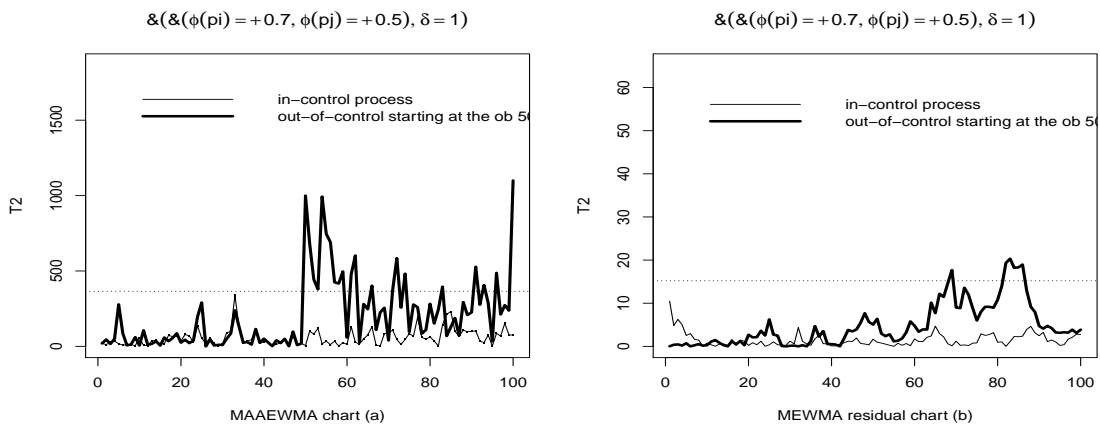


Figura 5.7: Monitorización de un proceso autocorrelacionado bivariate con coeficientes $p_i(\phi = 0,7)$ y $p_j(\phi = 0,5)$

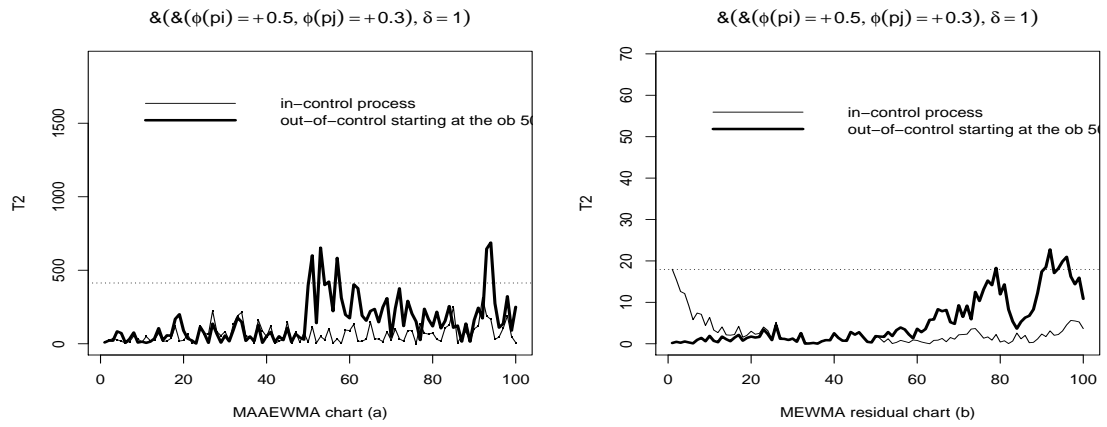


Figura 5.8: Monitorización de un proceso autocorrelacionado bivariate con coeficientes $p_i(\phi = 0,5)$ y $p_j(\phi = 0,3)$

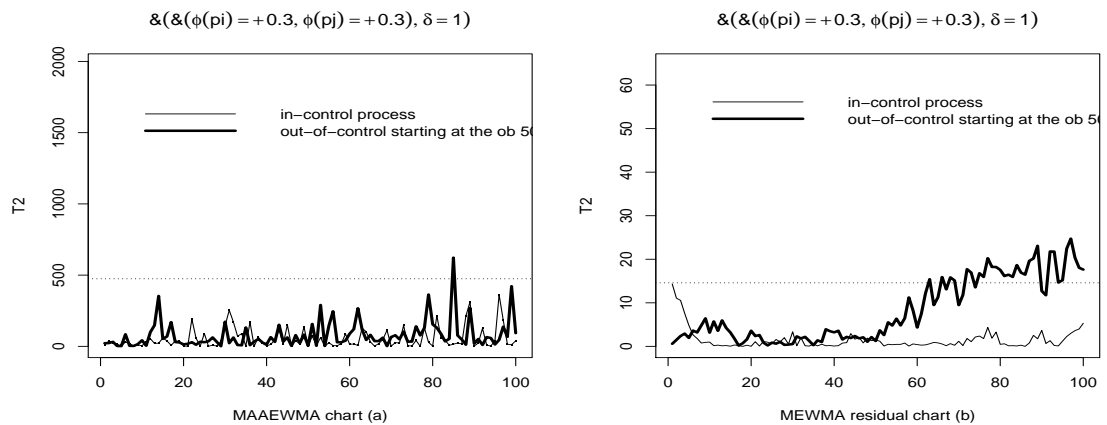


Figura 5.9: Monitorización de un proceso autocorrelacionado bivariate con coeficientes $p_i(\phi = 0,3)$ y $p_j(\phi = 0,3)$

detecta los cambios inmediatamente en el instante en que éste ocurre (RL=1). En los mismos casos, el gráfico MEWMA residual detecta los cambios después de 20 (RL=20), 13 (RL=13) y 17 (RL=17) observaciones, respectivamente.

En los 10 casos simulados para el proceso $\phi(p_i) = \phi(p_j) = 0,9$ se obtuvo $ARL(MAAEWMA) = 1$ y $ARL(MEWMA_{residual}) = 240,3$ (ver Tabla 5.1), es decir, en los 10 casos el gráfico MAAEWMA detectó el cambio en la primera observación. En cambio, el mejor desempeño del gráfico MEWMA residual fue de RL=20, además de presentar una mayor variabilidad en los resultados obtenidos.

Respecto al cuarto caso (figura 5.9), el gráfico MAAEWMA no presenta mejor desempeño que el gráfico MEWMA residual, como comentamos anteriormente, el gráfico MAAEWMA no presenta mejor desempeño para procesos con autocorrelación débil.

Bajos coeficientes autoregresivos significa una autocorrelación débil, y por lo tanto, el desempeño del gráfico MAAEWMA es reducido. No obstante, cuando la estructura de autocorrelación es más consistente, y los coeficientes autoregresivos presentan valores medios y altos, el gráfico MAAEWMA presenta ventajas significativas por considerar la presencia de autocorrelación de la manera más eficiente.

5.2.2. Conclusión

El gráfico MAAEWMA fue desarrollado para: (1) modelar la estructura multivariante y autocorrelacionada del proceso; (2) lograr eficiencia en detectar cambios en la media del proceso. Los resultados muestran que ambos aspectos mencionados fueron alcanzados.

El gráfico fue ajustado bajo modelos AR(1) y a través de RL y sus respectivos ARL su eficiencia fue testada.

Un diseño de experimentos factorial completo fue realizado para analizar el impacto de los coeficientes autoregresivos en el ARL del gráfico para el proceso fuera de control en diferentes magnitudes de cambio de la media del proceso.

El desempeño del gráfico MAAEWMA depende del valor de los coeficientes autoregresivos. Mejores resultados se obtuvieron para procesos con medios y altos coeficientes autoregresivos, comparado con el gráfico MEWMA residual. Este resultado es extremadamente importante ya que el valor del coeficiente autoregresivo expresa la magnitud de la estructura de autocorrelación. Cuando los coeficientes presentan valores altos el gráfico MAAEWMA conlleva de la

Tabla 5.1: Comparación de los ARL de gráfico MAAEWMA (ARL_B) respecto al gráfico MEWMA (ARL_A)

$\phi_1(p_i)$	$\phi_1(p_j)$	observaciones de la serie	nr. de series temporales	δ	ARL_A	ARL_B	ΔARL_1	p-valor ($RL_B - RL_A$)
0,3	0,3	10000	1	0	372,1	371,231	...	
		1000	10	1	19,8	264,7	-1236,86	0,000***
		1000	10	2	5,8	39,2	-575,86	0,000***
		1000	10	3	3,7	1,8	51,35	0,000***
0,3	0,5	10000	1	0	372,1	370,2	...	
		1000	10	1	22,1	75,8	-242,98	0,000***
		1000	10	2	5,6	17,14	-206,07	0,000***
		1000	10	3	2,8	1,1	60,71	0,000***
0,3	0,7	10000	1	0	372,1	371,231	...	
		1000	10	1	50,3	19,6	61,03	0,000***
		1000	10	2	6	2,4	60,00	0,000***
		1000	10	3	3,8	1	73,68	0,000***
0,3	0,9	10000	1	0	372,1	371,23	...	
		1000	10	1	54,5	14,4	73,57	0,000***
		1000	10	2	8,8	2,3	73,86	0,000***
		1000	10	3	5,2	1	80,769	0,000***
0,5	0,3	10000	1	0	372,1	371,23	...	
		1000	10	1	238,6	111,5	53,26	0,000***
		1000	10	2	10	1,5	85,00	0,000***
		1000	10	3	5,2	1,1	78,84	0,000***
0,5	0,5	10000	1	0	372,1	371,23	...	
		1000	10	1	299,8	144,6	51,76	0,000***
		1000	10	2	15,4	5,2	66,23	0,000***
		1000	10	3	7,1	1,1	84,50	0,000***
0,5	0,7	10000	1	0	372,1	371,23	...	
		1000	10	1	23,4	16,7	28,63	0,000***
		1000	10	2	5,4	1,2	77,77	0,000***
		1000	10	3	3,1	1	67,741	0,000***
0,5	0,9	10000	1	0	372,1	371,23	...	
		1000	10	1	147,25	18,375	87,52	0,000***
		1000	10	2	14,6	1,2	91,78	0,000***
		1000	10	3	7	1	85,71	0,000***
0,7	0,3	10000	1	0	372,1	371,23	...	
		1000	10	1	26,4	18,9	28,40	0,000***
		1000	10	2	8,3	1,1	86,74	0,000***
		1000	10	3	4,4	1	77,27	0,000***
0,7	0,5	10000	1	0	372,1	371,23	...	
		1000	10	1	60,3	9,1	84,90	0,000***
		1000	10	2	8,9	1,1	87,64	0,000***
		1000	10	3	4,3	1	76,74	0,000***
0,7	0,7	10000	1	0	372,1	371,23	...	
		1000	10	1	9,9	2,3	76,76	0,000***
		1000	10	2	2,2	1	54,54	0,000***
		1000	10	3	1,3	1	23,07	0,000***
0,7	0,9	10000	1	0	372,1	371,23	...	
		1000	10	1	269,8	5,1	98,10	0,000***
		1000	10	2	83,5	1	98,80	0,000***
		1000	10	3	13	1	92,30	0,000***
0,9	0,3	10000	1	0	372,1	371,23	...	
		1000	10	1	181,7	1,2	99,33	0,000***
		1000	10	2	12,4	1	91,93	0,000***
		1000	10	3	6,4	1	84,37	0,000***
0,9	0,5	10000	1	0	372,1	371,23	...	
		1000	10	1	269,3	1,2	99,55	0,000***
		1000	10	2	16,8	1	94,04	0,000***
		1000	10	3	9,3	1	89,24	0,000***
0,9	0,7	10000	1	0	372,1	371,23	...	
		1000	10	1	269,3	1,1	99,59	0,000***
		1000	10	2	118,9	1	99,15	0,000***
		1000	10	3	12,9	1	92,24	0,000***
0,9	0,9	10000	1	0	371,231	371,23	...	
		1000	10	1	240,3	1	99,58	0,000***
		1000	10	2	119,1	1	99,16	0,000***
		1000	10	3	14,3	1	93,01	0,000***

*** = p - valor $< 0,0001$

mejor forma la autocorrelación produciendo mejores resultados.

5.3. Modelo dinámico sobre estructuras latentes

El método estadístico propuesto en este trabajo corresponde al modelo de monitorización y predicción de procesos multivariantes y autocorrelacionados. El gráfico MAAEWMA presentado en el apartado anterior fue desarrollado para lograr el primer objetivo. En este apartado, presentamos el modelo de predicción que se basa en el ajuste de un modelo dinámico sobre estructuras latentes (MDEL).

El gráfico MAAEWMA fue desarrollado bajo las hipótesis de la existencia de correlación y autocorrelación en los datos. Además de dichas suposiciones, el modelo de predicción propuesto se aplica a los procesos con datos multivariantes de entrada y de salida (multi-input, multi-output (MIMO)).

Los modelos predictivos orientados a los procesos MIMO en el entorno dinámico fueron estudiados por varios autores, los cuales presentan diferentes enfoques. En este trabajo analizamos los modelos partial least square (PLS) y su versión dinámica (DPLS), cuyos detalles presentamos en el capítulo 3.

El DPLS ajusta un modelo de regresión en el entorno de variables latentes, buscando la dirección de máxima covarianza entre los scores. A diferencia del PLS convencional, las variables latentes son calculadas de modo que la estructura dinámica entre las variables en los dos bloques se refleja en los scores.

Los estudios sobre DPLS presentados en el capítulo 3 se basan fundamentalmente en el análisis de la estructura dinámica de los datos, que culmina con la extensión de la matriz de entrada (input data) mediante la inclusión de un número k de variables decaladas, y a continuación se aplica el PLS a la nueva estructura de datos. Esta estrategia garantiza que la estructura dinámica en los datos se refleja en el modelo de regresión ajustado a los scores. Es decir, la relación interna del modelo (expresión 3.37) es representada por un modelo de regresión dinámico, donde cada variable latente respuesta es explicada por la correspondiente variable latente explicativa y sus respectivos retardos.

Del análisis realizado en los diferentes estudios desarrollados sobre el modelo DPLS se puede destacar lo siguiente:

- (1) La dimensión de la matriz de datos de entrada extendida depende de la estructura dinámica de los datos.

Mediante un estudio inicial se identifica un número k de variables decaladas a incluir en el modelo, de modo que la calidad del modelo DPLS depende de ello.

Diferentes enfoques fueron presentados en el capítulo 3 para la determinación de número óptimo de lags a incluir en el modelo, y destacamos los métodos de Ku et al. (Ku *et al.*, 1995), Rato y Reis (Rato y Reis, 2013) y Vanhatalo et al. (Vanhatalo *et al.*, 2017), los cuales no son muy sencillos de aplicar, de modo que tornan el modelo DPLS susceptible a errores de especificación de variables explicativas.

El error de especificación de las variables afecta al modelo de dos modos: No incluir variables con información significativa al modelo afecta negativamente a la variabilidad explicada de las componentes determinadas. Es decir, la variabilidad explicada de las variables latentes determinadas será menor respecto a la que se obtendría considerando dichas variables. Por otro lado, si se incluyen variables que no aportan información significativa al modelo, se aumenta la proporción del error al modelo, afectando la calidad predictiva del modelo.

- (2) El modelo DPLS se ajusta en la dirección de máxima covarianza entre los scores. A pesar de considerarse variables decaladas en la matriz de entrada, el modelo no logra maximizar la estructura de correlación dinámica entre las variables, ya que el único objetivo del algoritmo NIPALS aplicado al modelo (ver capítulo 3) es maximizar la covarianza entre los scores de las variables explicativas y respuestas.

Nuestra propuesta metodológica se basa en ajustar un modelo predictivo a las variables latentes determinadas en las matrices de datos decaladas, donde la estructura dinámica entre las variables es identificada mediante gráficos de la función de correlación cruzada entre los residuos pre-blanqueados de cada par de variables. Las variables latentes son determinadas con vista a explotar la máxima varianza en los datos de entrada y de salida.

El modelo se aplica a los casos donde además de la estructura de correlación entre los datos de entrada (variables del proceso) y de salida (variables del producto) hay una fuerte dependencia dinámica del producto a las variables del proceso, por lo que, las siguientes hipótesis son consideradas:

- (1) Hay una estructura de correlación dinámica significativa entre las variables de cada bloque (datos de entrada y de salida);

- (2) La calidad del producto depende de las características del proceso.

Los métodos de proyección sobre estructuras latentes logran reducir la dimensión de los datos en pocas variables latentes que explican la mayor variabilidad de los datos. Nuestra propuesta se orienta a estos métodos con vista a reducir la gran cantidad de datos que caracterizan un sistema de producción en indicadores globales de la calidad del producto y del proceso. Por tanto, es fundamental que los scores recojan la mayor variabilidad existente en los datos para que puedan representarlos. En vista a ello, el modelo propuesto consiste en aplicar un análisis de componentes principales dinámico (DPCA) en cada bloque de variables de modo a determinar los escores que maximizan la variabilidad dinámica en los datos, y asumiendo la hipótesis (2) se ajusta un modelo de regresión dinámico entre dichos scores resultantes.

El método empieza por analizar la estructura dinámica en los datos, con lo cual se identifica el número de lags de las variables a incluir en el modelo. Para ello, se aplica un filtro a cada par de variables y se identifica la estructura dinámica en el gráfico de correlaciones cruzadas. Este procedimiento puede parecer complejo, pero, como afirma Vanhatalo et al. (Vanhatalo *et al.*, 2017), normalmente la estructura dinámica de las variables es la misma en cada bloque, de modo que suele ser suficiente analizar la estructura de un par de ellas para comprender el conjunto.

En los escores obtenidos de las matrices de datos extendidas (las que recogen las variables originales y las decaladas) se ajusta una función de transferencia (FT), donde se analiza el impacto de las variables latentes explicativas y sus retardos sobre las variables latentes respuestas. A diferencia del modelo DPLS donde la regresión entre los scores es construido a partir del cálculo de las variables latentes mediante el algoritmo NIPALS, en este caso, los scores son calculados de manera independiente en los dos bloques de variables, priorizando la maximización de la varianza en cada bloque, y a continuación, se aplica la función de transferencia mediante la metodología de Box-Jenkins (Box y Jenkins, 1976) para analizar la relación dinámica entre los scores obtenidos.

La metodología de Box-Jenkins permite identificar la estructura dinámica de las variables latentes a través del gráfico de funciones de correlaciones cruzadas ajustado a los residuos pre-blanqueados. A diferencia del modelo DPLS donde se incluye las componentes autoregresivas de la variable respuesta recurriendo al modelo ARX (Dong y Qin, 2015), en este caso, el impulso de respuesta identificado en la función de correlación cruzada identifica el orden autoregresivo de la variable respuesta a incluir en el modelo.

La figura 5.10 presenta un esquema que ilustra los procedimientos usados

en la construcción del modelo dinámico sobre estructuras latentes (MDEL).

5.3.1. Construcción del modelo propuesto

Sea $X = [x_1, x_2 \dots x_m]$ una matriz de datos de entrada de dimensión $n * m$ e $Y_t = [y_1, y_2 \dots y_p]$ la correspondiente matriz de datos de salida de dimensión $n * p$ de un proceso multivariante y autocorrelacionado. Donde x_j ($j=1,2,\dots,m$) e y_k ($k=1,2,\dots,p$) son las respectivas variables de X y de Y , y x_{ij} e y_{ik} las observaciones de las variables x_j e y_k en el instante i ($i=1,2,\dots,n$).

Como ejemplo se podría pensar en que el proceso consiste en la manufactura de un determinado producto caracterizado por muchas variables y que sus propiedades cambian a lo largo del tiempo.

El estudio empieza por un análisis de la estructura dinámica de las variables de las matrices X e Y . Sea f y h el número de lags significativos en la matriz X e Y , respectivamente. Se definen las matrices $X_{i-1}, X_{i-2}, \dots, X_{i-f}$ a partir de X , e $Y_{i-1}, Y_{i-2}, \dots, Y_{i-h}$ a partir de Y , donde $\tilde{X} = [X_i \ X_{i-1} \ \dots \ X_{i-f}]$ e $\tilde{Y} = [Y_i \ Y_{i-1} \ \dots \ Y_{i-h}]$ son las respectivas matrices extendidas de X e Y .

Sea $T = [t_1 \ t_2 \ \dots \ t_a]$ y $W = [w_1 \ w_2 \ \dots \ w_a]$ la matriz de scores y loadings que recogen la variabilidad explicada óptima de la matriz \tilde{X} , y $U = [u_1 \ u_2 \ \dots \ u_b]$ y $Q = [q_1 \ q_2 \ \dots \ q_b]$ la matriz de scores y loadings que recogen la variabilidad explicada óptima de la matriz \tilde{Y} . Los scores de T y U son calculados como se muestra a continuación:

$$t_{ia} = w'_a \tilde{X} = w_{aj}x_{ij} + w_{a(j+f)}x_{(i-f)j} \quad (5.12)$$

y

$$u_{ib} = q'_b \tilde{Y} = q_{bk}y_{ik} + q_{b(k+h)}y_{(i-h)k} \quad (5.13)$$

donde se considera el siguiente problema de maximización:

max: $w' \tilde{X}$ s.a $\|w\| = 1$, y max: $q' \tilde{Y}$ s.a $\|q\| = 1$.

En un caso particular donde $m=p=2$, $f=h=1$ y $a=b=1$, tendríamos:

$$t_{i1} = w_{11}x_{i1} + w_{12}x_{i2} + w_{13}x_{(i-1)1} + w_{14}x_{(i-1)2}$$

y

$$u_{i1} = q_{11}y_{i1} + q_{12}y_{i2} + q_{13}y_{(i-1)1} + q_{14}y_{(i-1)2}$$

Los scores t y u son calculados en la dirección de máxima varianza de los datos en los respectivos bloques extendidos con las variables decaladas. Para un número de componentes principales a de \tilde{X} y b de \tilde{Y} se obtiene la siguiente descomposición:

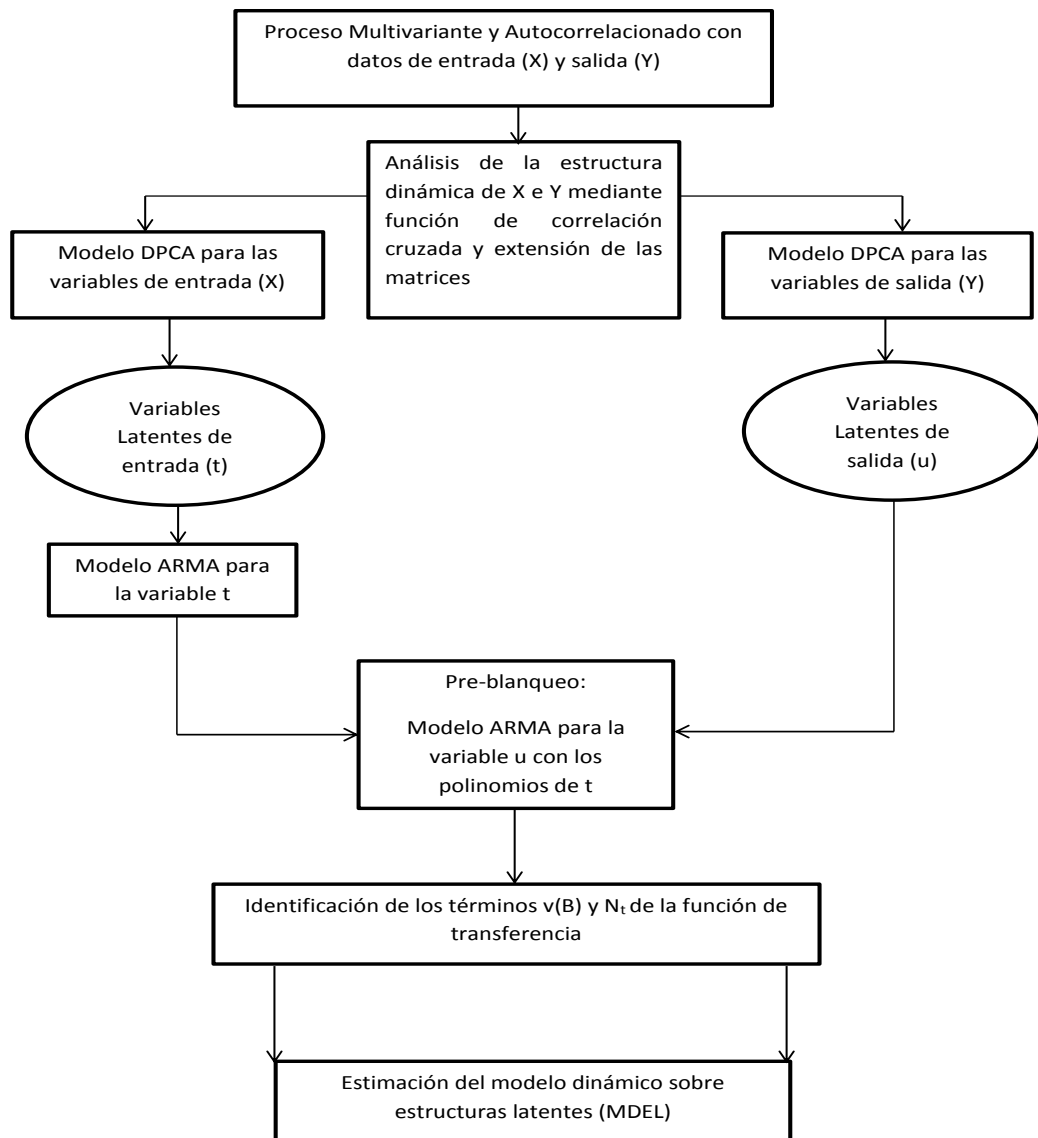


Figura 5.10: Esquema de construcción del modelo dinámico propuesto

$$\tilde{X} = TW' + E \quad (5.14)$$

y

$$\tilde{Y} = UQ' + F \quad (5.15)$$

donde E es la parte de la variabilidad de \tilde{X} no explicada por T, y F la parte de la variabilidad de \tilde{Y} no explicada por U.

Si $\tilde{X}_a = TW'$ e $\tilde{Y}_b = UQ'$, entonces la calidad de ajuste del modelo DPCA en cada bloque puede ser evaluado por:

$$\text{RESS}_a = \| \tilde{X} - \tilde{X}_a \|$$

y

$$\text{RESS}_b = \| \tilde{Y} - \tilde{Y}_b \|$$

La calidad predictiva del modelo DPCA en los dos bloques \tilde{X} e \tilde{Y} puede ser analizado mediante el método de validación cruzada (descrito en el apartado 3.2.1.5), con lo cual se obtienen las matrices de los valores predichos $\hat{\tilde{X}}$ e $\hat{\tilde{Y}}$, donde el *predict residual sum of square* (PRESS) puede ser calculado como:

$$\text{PRESS}_a = \| \tilde{X} - \hat{\tilde{X}} \|$$

y

$$\text{PRESS}_b = \| \tilde{Y} - \hat{\tilde{Y}} \|$$

y finalmente se calcula el estadístico de bondad de ajuste del modelo definido por:

$$Q_a^2 = 1 - \frac{\text{PRESS}_a}{\text{RESS}_{a-1}}$$

y

$$Q_b^2 = 1 - \frac{\text{PRESS}_b}{\text{RESS}_{b-1}}$$

5.3.1.1. La Función de transferencia

Un sistema dinámico entre una serie respuesta y_t y una serie explicativa x_t consiste en una relación funcional, en donde y_t se define como función de x_t (Yaffee y McGee, 1999). El impacto de los valores presentes y pasados de x_t sobre y_t puede ser medido a través de una función de transferencia.

La función de transferencia es un modelo estadístico que describe la relación dinámica entre una variable respuesta y_t y una o varias variables explicativas x_t (Tsay, 2009).

En este trabajo, la función de transferencia no se aplica a las variables originales y_t y x_t , sino a sus versiones latentes, las cuales designamos por u_i y t_i .

La serie u_i correspondiente a la variable latente de la matriz de salida extendida \tilde{Y} y t_i corresponde a la variable latente de la matriz de entrada extendida \tilde{X} . La función de transferencia que describe la relación lineal entre estas dos series puede ser definida por la siguiente expresión:

$$\begin{aligned} u_i &= c + v_0 t_i + v_1 B t_{i-1} + v_2 B^2 t_{i-2} + \dots + N_i \\ &= c + v(B) t_i + N_i \end{aligned} \quad (5.16)$$

Donde los escalares v_0, v_1, v_2, \dots representan el impacto de t_{i-f} sobre u_i , B es el operador de retardo, es decir, $B t_i = t_{i-1}$; y N_i es el residuo del modelo, que puede estar muy correlacionado con t_i , y $v(B)$ es un polinomio en B de grado infinito. Para que se cumpla la estacionariedad es necesario que $\sum_{f=0}^{\infty} |v_f| < \infty$.

La función de transferencia tiene dos componentes: la componente $v(B)$ que representa el impacto de t_i sobre u_i y la N_i que es la componente residual del modelo. De manera análoga a los modelos ARIMA univariados, estas componentes pueden ser escritas como fracción de dos polinomios que dependen de B de la siguiente manera:

$$v(B) = \frac{w_s(B)}{\delta_r(B)} B^b \quad (5.17)$$

donde b es un número entero no negativo, $w_s(B) = w_0 - w_1 B - w_2 B^2 - \dots - w_s B^s$ y $\delta_r(B) = 1 - \delta_1 B - \delta_2 B^2 - \dots - \delta_r B^r$.

Por otro lado, la estructura autocorrelacionada de N_i puede ser modelada como un ARMA(p, q), como representamos a continuación:

$$\phi_p(B) N_i = \theta_q(B) a_i \quad (5.18)$$

donde $\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ y $\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$ son polinomios de B de grados p y q , respectivamente, y $a_i \sim N(0, \sigma_a)$.

Por tanto, si las condiciones de estacionariedad y invertibilidad se cumplen en el modelo 5.18, se puede escribir,

$$N_i = \frac{\theta_q(B)}{\phi_p(B)} a_i$$

entonces, la expresión 5.16 puede ser escrita de la siguiente manera:

$$u_i = c + \frac{w_s(B)}{\delta_r(B)} B^b t_i + \frac{\theta_q(B)}{\phi_p(B)} a_i \quad (5.19)$$

5.3.1.2. La metodología de Box-Jenkins

La metodología de Box y Jenkins (Box y Jenkins, 1976) para modelación de series temporales con función de transferencia consiste fundamentalmente en:

(1) ajustar un modelo ARMA a la serie explicativa; (2) aplicar el mismo filtro a un modelo ARMA para la serie respuesta, y ajustar una función de correlación cruzada a los residuos de los modelos resultantes, donde se identifica los ordenes de la componente $v(B)$; (3) Identificar los ordenes de la componente N_i a partir de los ordenes p y q del modelo ARMA ajustado a la variable respuesta con sus propios polinomios, y finalmente estimar el modelo. A continuación presentamos los detalles del método:

(1) Ajustar un modelo ARMA a la serie explicativa del modelo

Las series deben ser estacionarias. Por tanto, a un proceso ARIMA(p,d,q) se le aplican las d diferencias regulares necesarias para lograr la estacionariedad. Es decir, si la serie W_t sigue un proceso ARIMA(p,d,q), la transformación $\nabla^d(W_t) = w_t$ corresponde a un proceso ARMA(p,q). Dicha transformación se aplica a las series U_i y T_i para lograr las correspondientes series estacionarias u_i y t_i .

(2) Identificación de las componentes de la función de transferencia

(2.1) Pre-blanqueo;

El pre-blanqueo consiste en aplicar un filtro a la serie respuesta a través de un modelo ARMA ajustado a la serie explicativa. Con ello se estiman los coeficientes de la función de transferencia: v_0, v_1, \dots ,

Sean $\phi_{p,t}(B)$ y $\theta_{q,t}(B)$ los respectivos polinomios autoregresivos de orden p , y de medias móviles de orden q de un modelo ARMA(p,q) ajustado a la serie t_i . Se aplican dichos polinomios para ajustar un modelo ARMA a la variable respuesta u_i con objeto de eliminar el efecto de la autocorrelación en la correlación cruzada entre las series en causa.

El pre-blanqueo garantiza que las correlaciones cruzadas se reflejan con precisión en los pesos de la respuesta al impulso. La idea es eliminar la alta dependencia entre los valores de t_i . Así, de un modelo ARMA ajustado a la serie t_i , se puede traducir que

$$\eta_i = \frac{\phi_{p,t}(B)}{\theta_{q,t}(B)} t_i$$

y multiplicando el coeficiente de t_i en la expresión 5.19 se obtiene:

$$\frac{\phi_{p,t}(B)}{\theta_{q,t}(B)} u_i = \frac{\phi_{p,t}(B)}{\theta_{q,t}(B)} c + v(B) \frac{\phi_{p,t}(B)}{\theta_{q,t}(B)} t_i + \frac{\phi_{p,t}(B)}{\theta_{q,t}(B)} N_i \quad (5.20)$$

donde teniendo en cuenta que $u_i^* = \frac{\phi_{p,t}(B)}{\theta_{q,t}(B)}u_i$ (u_i^* es el residuo del modelo ARMA ajustado a la serie respuesta u_i con los polinomios $\phi_{p,t}$ y $\theta_{q,t}$), $c^* = \frac{\phi_{p,t}(1)}{\theta_{q,t}(1)}c$, la ecuación pasa a ser:

$$u_i^* = c^* + v(B)\eta_i + a_i \quad (5.21)$$

en este caso η_i y a_i son independientes, y ambas son series de ruidos blancos.

(2.2) Análisis de la función de correlación cruzada;

Sea η_{t-j} con $j \geq 0$, la covarianza cruzada entre u_t^* y η_{t-j} es definida por:

$$\text{Cov}(u_t^*, \eta_{t-j}) = v_j \sigma_{\eta_{t-j}}^2$$

y la función de correlación cruzada pasa a ser

$$\rho_{(u_t^*, \eta_{t-j})} = \rho_{u_t^*, \eta_t}(j) = \frac{\text{Cov}(u_t^*, \eta_{t-j})}{\sigma_{u_t^*} \sigma_{\eta_{t-j}}}, \text{ lo que resulta:}$$

$$v_j = \frac{\sigma_{u_t^*}}{\sigma_{\eta_t}} \rho_{u_t^*, \eta_t}(j) \quad (5.22)$$

Es decir, el coeficiente de la función de transferencia es proporcional a la función de correlación cruzada entre las series respuesta y explicativa pre-blanqueadas en el retardo j .

Los ordenes r , s , b de la función de transferencia (ver expresión 5.17) suelen identificarse a través del gráfico de la función de correlación cruzada. Analizando la estructura de las correlaciones significativas, dicho elementos son interpretados de la siguiente manera:

- b indica el primer lag significativo, y representa el primer decalado de la variable regresora en entrar en el modelo;
- s representa el número de lags crecientes después de b ;
- r indica el número de bloques con crecimiento exponencial en los lags crecientes.

(3) Estimación del modelo;

La estimación del modelo consiste en calcular los valores correspondientes a cada coeficiente de la componente $v(B)$ y N_i obedeciendo los respectivos ordenes b , s , r y p y q .

Con auxilio a un software dichos coeficientes son calculados y el modelo estimado de función de transferencia es representado por:

$$u_i = c + \frac{\hat{w}_s(B)}{\hat{\delta}_r(B)} B^b t_i + \frac{\hat{\theta}_q(B)}{\hat{\phi}_p(B)} a_i$$

La metodología de Box y Jenkins presentada, corresponde al caso del modelo de función de transferencia con solo un regresor. Sin embargo, se puede ajustar el modelo con varios regresores realizando de manera sucesiva el pre-blanqueo de la variable respuesta con cada regresor.

De modo sintético, las relaciones externas en cada bloque de matrices extendidas son expresas por:

$$\tilde{X} = tw' + E, \quad E = \tilde{X} - tw'$$

y

$$\tilde{Y} = uq' + F, \quad F = \tilde{Y} - uq'$$

y la relación interna, que modela las variables latentes entre los scores de los dos bloques es datos por:

$$u = \frac{w_s(B)}{\delta_r(B)} B^b t + \frac{\theta_q(B)}{\phi_p(B)} a$$

de las dos relaciones resulta que:

$$F = \tilde{Y} - \left(\frac{w_s(B)}{\delta_r(B)} B^b t + \frac{\theta_q(B)}{\phi_p(B)} a \right) q'$$

que representa la parte de la variabilidad de los datos no explicada ni por el modelo DPCA, ni por la función de transferencia. Por tanto, un buen ajuste del modelo dinámico sobre estructuras latentes (MDEL) implica un valor de F pequeño. Para ello se puede evaluar la calidad del modelo en dos fases: La calidad de ajuste del modelo DPCA y la calidad de ajuste de la función de transferencia ajustada a las variables latentes.

5.3.2. Simulación de procesos multivariantes y autocorrelacionados

Para testar el desempeño del modelo dinámico sobre estructuras latentes propuesto, se desarrolló una simulación de procesos multivariantes y autocorrelacionados con datos multivariantes de entrada y salida (multi-inpt, multi-output:MIMO) referidos en este trabajo.

La simulación fue implementada en el software R, y consistió en determinar un conjunto suficientemente grande de datos con las características mencionadas anteriormente (ver código R en el Apéndice A.2). El modelo propuesto y el

DPLS fueron ajustados a los datos simulados, y sus resultados fueron comparados. La estructura de autocorrelación en el proceso fue considerada tomando las variables del proceso y del producto como series temporales autoregresivas.

En cada bloque de variables, cada serie simulada fue relacionada con la serie simulada en el paso anterior, para garantizar la correlación entre las variables. Respecto a los bloques, las variables de entrada X fueron simuladas primero y, a continuación, las variables de salida Y fueron simuladas considerando que en cada instante i sufrieron un impacto de las variables del proceso en el instante anterior $i - 1$. Es decir, el proceso fue simulado de manera que X_{i-1} influya significativamente en Y_i .

De modo análogo a la monitorización, en la simulación de las series temporales fueron considerados los siguientes coeficientes autoregresivos 0.3, 0.5, 0.7, 0.9, para representar procesos con baja, media y alta autocorrelación. Se demostró en el apartado anterior mediante un diseño de experimentos que los procesos con estos coeficientes presentan diferentes resultados respecto al desempeño del modelo ajustado. A diferencia del caso anterior (en la monitorización), donde fueron simulados 16 pares de series temporales a través de la combinación de dichos coeficientes, en este caso, se obtuvieron 16 pares de bloques X e Y. Es decir, en cada caso simulado, se obtuvieron dos bloques de variables, cada uno formado por 5 series autoregresivas correlacionadas entre si.

Los pasos usados en la simulación de los datos son:

1. Variables de entrada (Bloque X)

- a) Fue simulada una serie autoregresiva de tamaño 500, que designamos por x ;
- b) De la variable x anterior, fueron considerados los primeros 100 elementos, formando la variable x_1
- c) Una segunda variable x_2 fue determinada como la diferencia entre los valores de los segundos 100 elementos de x y los valores de x_1 ; de manera análoga, las variables x_3 , x_4 , x_5 fueron determinadas;
- d) La matriz X fue formada por las variables x_1, x_2, x_3, x_4, x_5 .

2. Variables de salida (Bloque Y)

- a) Fue simulada una serie autoregresiva y , de tal manera que en cada instante i sufre un impacto del valor de x en instante $i - 1$;
- b) Las variables y_1, y_2, y_3, y_4, y_5 fueron determinadas de manera análoga que en el bloque X;
- c) La matriz Y fue formada por las variables y_1, y_2, y_3, y_4, y_5 .

Para cada matriz X simulada con un conjunto de variables con un determinado coeficiente autoregresivo (uno de los cuatro mencionados anteriormente), cuatro matrices Y fueron determinadas, formando un conjunto de 16 pares de bloques $X * Y$, como consecuencia de la combinación de los coeficientes autoregresivos seleccionados.

A estos conjuntos de datos se ajustaron los modelos MDEL y DPLS con el fin de comparar los resultados.

La inclusión de las variables decaladas en los modelos fue realizada como explicamos anteriormente. La estructura autoregresiva de los bloques X e Y indica que $f=h=1$ (solo el primer lag es significativo en las dos matrices), entonces fueron determinadas las matrices \tilde{Y} y \tilde{X} que recogen las respectivas variables de X e Y y sus decalados de primer orden, donde se ajustaron los modelos MDEL y DPLS.

5.3.3. Resultados

5.3.3.1. Ajuste del modelo dinámico sobre estructuras latentes (MDEL)

En este apartado se presentan los procedimientos usados en el ajuste del modelo dinámico sobre estructuras latentes, y se ilustran los resultados del modelo ajustado al proceso simulado con coeficientes autoregresivos igual a 0.9.

Las figuras 5.11 y 5.12 presentan los gráficos de correlaciones simples y parciales de las variables t_1 , t_2 y u_1 , u_2 , respectivamente. En la figura 5.11 se observa un decaimiento progresivo y amortiguado de las correlaciones simples, y las dos primeras correlaciones parciales significativas, indicando que t_1 y t_2 se ajustan a un proceso AR(2). En la figura 5.12 hay un decaimiento progresivo y amortiguado de las correlaciones simples, y solo una correlación parcial significativa en los dos gráficos, indicando que las series u_1 y u_2 se ajustan a un proceso AR(1).

De acuerdo con la metodología de Box y Jenkins (Box y Jenkins, 1976) aplicada en este trabajo, las series u_1 y u_2 fueron pre-blanqueadas usando los polinomios de las respectivas series explicativas t_1 y t_2 . Sean a_1 y a_2 los respectivos ruidos blancos de t_1 y t_2 , y a_1^* y a_2^* los ruidos obtenidos de los modelos pre-blanqueados de u_1 y u_2 , respectivamente.

La figura 5.13 muestra la estructura de correlaciones cruzadas entre los residuos a_1 y a_1^* y entre a_2 y a_2^* . Los gráficos indican que el impacto de t_2 sobre u_1 se refleja un instante después de su ocurrencia (lag=1). Por otro lado, el impacto de t_1 sobre u_1 se observa en el mismo instante (lag=0), y en los dos

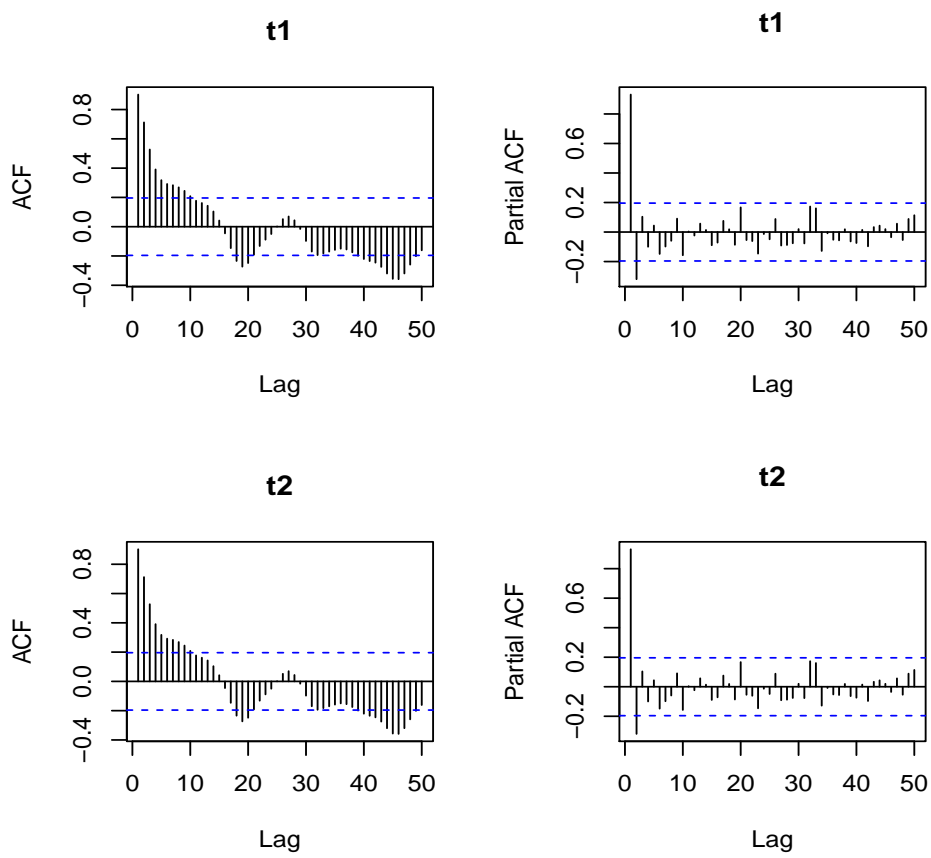


Figura 5.11: Representación de las variables latentes t_1 y t_2

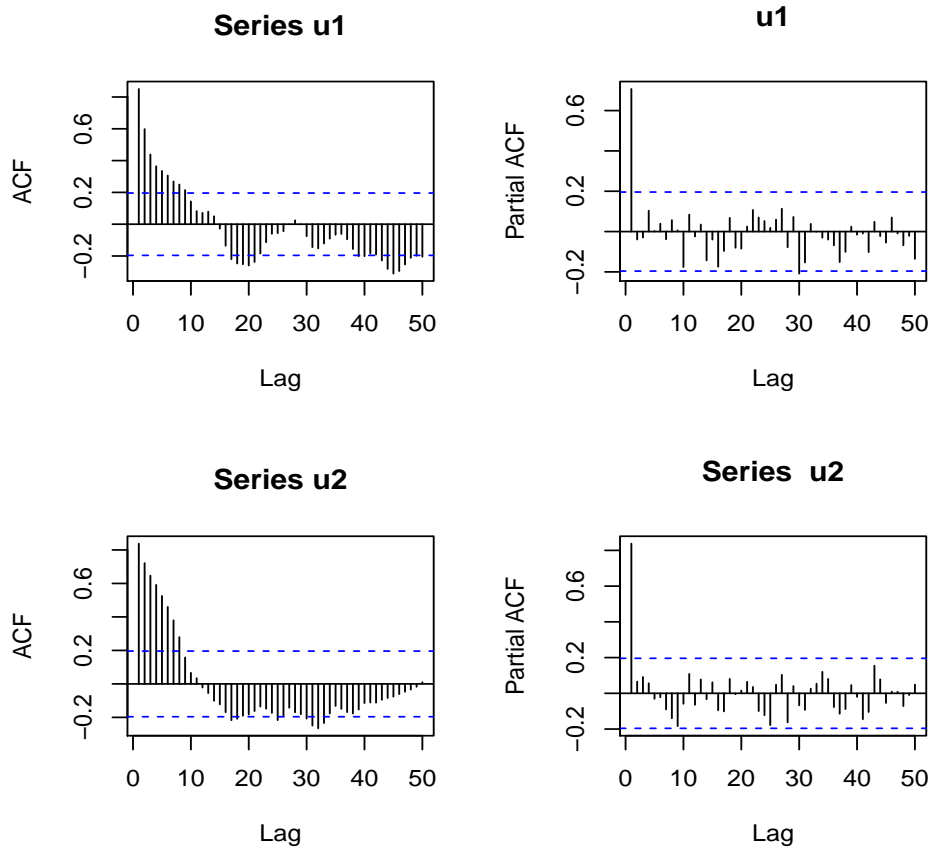


Figura 5.12: Representación de las variables latentes u_1 y u_1

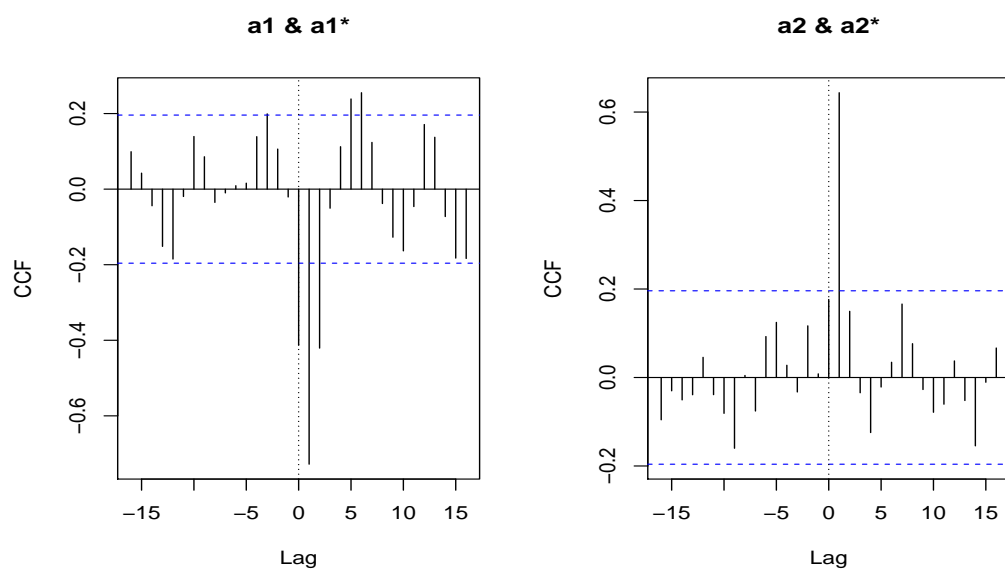


Figura 5.13: Correlaciones cruzadas entre los residuos a_1 y a_1^* y a_2 y a_2^*

siguientes (lag=1,2), siendo más evidente en lag=1.

Con esta estructura de correlación cruzada, los ordenes b, s y r de V(B) de los dos modelos de función de transferencia fueron identificados: Para la función $u_2 = f(t_2)$, el primer lag significativo es 1, por tanto b=1, y por ser el único lag significativo, s=1. No se observan descensos exponenciales en las correlaciones cruzadas, por lo que, r=0.

Para la función $u_1 = f(t_1)$, a pesar de observarse una correlación cruzada significativa en lag=0, el mayor impacto de t_1 sobre u_1 se observa en lag=1, por lo que se consideró b=1. Después de b, hay un lag significativo, entonces s=2, y por no haber caídas exponenciales en la estructura de correlaciones cruzadas, r=0.

Tabla 5.2: Estimación de los parámetros del modelo de función de transferencia entre u_1 y t_1

z test of coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
ar1	-0.60728	0.11271	-5.3880	7.125e-08	***
T1-MA0	-1.37713	0.35799	-3.8468	0.0001197	***
T1-MA1	-1.94728	0.37035	-5.2579	1.457e-07	***
T1-MA2	0.60934	0.35948	1.6951	0.0900643	.

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Tabla 5.3: Estimación de los parámetros del modelo de función de transferencia entre u_2 y t_2

z test of coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
ar1	-0.034092	0.137054	-0.2487	0.80356	
intercept	0.630371	0.587268	1.0734	0.28309	
T1-MA0	2.022107	0.317215	6.3746	1.835e-10	***
T1-MA1	0.814919	0.319084	2.5539	0.01065	*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Los ordenes p y q del modelo ARMA para la componente residual N_i fueron

identificados mediante los modelos ARMA ajustados a las series respuestas u_1 y u_2 con sus propios filtros. Como comentamos anteriormente, las series u_1 y u_2 se ajustan a un proceso AR(1), por lo que las funciones de transferencia de u_1 y u_2 quedan definidas por las siguientes expresiones analíticas:

$$u_{1i} = (w_0 + w_1B + w_2B^2)t_{i-1} + \frac{1}{1 - \phi_1B}a_i$$

$$u_{2i} = (w_0 + w_1B)t_{i-1} + \frac{1}{1 - \phi_1B}a_i$$

Los coeficientes estimados y su significación en los dos modelos son presentados en las tablas 5.2 y 5.3.

Los resultados muestran que el coeficiente w_2 de la función de transferencia $u_1 = f(t_1)$ no es significativo (a un nivel de significación de 5%). Por tanto, la expresión analítica se reduce a:

$$u_{1i} = (-1,37713 - 1,94728B)t_{i-1} + \frac{1}{1 + 0,60728B}a_i$$

la cual indica que cada observación de u_1 en cada instante i sufre una influencia negativa de los valores de t_1 en los instantes $i=1$ y $i=2$, además de una influencia negativa de su propio retardo de orden uno (es decir, sufre también influencia negativa de u_{1i-1}).

Respecto a la función de transferencia $u_2 = f(t_2)$, los únicos coeficientes significativos son w_0 y w_1 (tabla 5.3), entonces, la expresión analítica se reduce a:

$$u_{2i} = (2,022107 + 0,814919B)t_{i-1} + a_i$$

y muestra que las observaciones de u_2 sufren una influencia positiva de los valores de t_1 en los primeros dos instantes anteriores $i=1,2$.

5.3.3.2. Ajuste del modelo DPLS

De modo análogo al modelo MDEL, en este apartado presentamos los resultados del modelo DPLS ajustado a los mismos datos X e Y simulados con un coeficiente autoregresivo igual a 0.9. Las figuras 5.14 y 5.15 presentan los gráficos de correlaciones simples y parciales de las variables t y u respectivamente.

Los gráficos en la figura 5.14 muestran un descenso progresivo de las correlaciones simples, y las dos primeras correlaciones parciales significativas, lo que indica que las dos variables t_1 y t_2 se ajustan a un proceso AR(2).

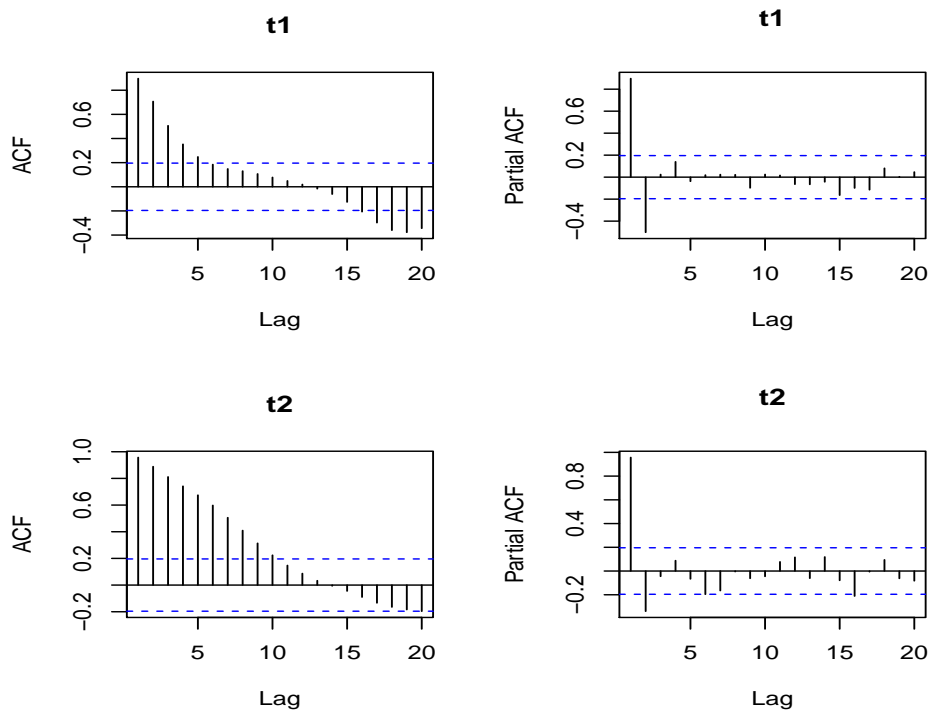


Figura 5.14: Representación de las variables latentes t_1 y t_2

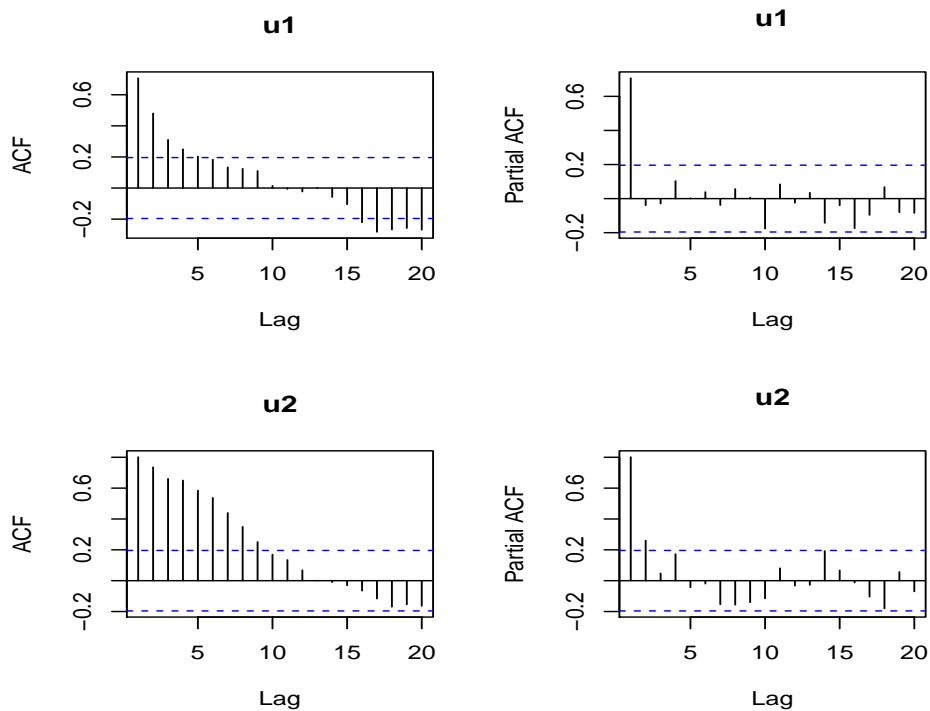


Figura 5.15: Representación de las variables latentes u_1 y u_2

Tabla 5.4: Estimación de los coeficientes del modelo DPLS

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	0.001886	0.150700	0.013	0.99004	
u1_1	-0.123930	0.097207	-1.275	0.20542	
t1	0.648668	0.121800	5.326	6.62e-07	***
t1_1	0.513859	0.158279	3.247	0.00161	**

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	0.007009	0.070696	0.099	0.92123	
u2_1	-0.040984	0.016355	-2.506	0.01390	*
t2	0.427777	0.144257	2.965	0.00381	**
t2_1	0.379158	0.132765	2.856	0.00526	**

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Los gráficos en la figura 5.15 muestran un descenso progresivo de la correlaciones simples, y solo las primeras correlaciones parciales son significativas, indicando que las variables latentes u_1 y u_2 se ajustan a un proceso AR(1).

La tabla 5.4 recoge los coeficientes estimados de los modelos ajustados a u_1 y u_2 , y sus respectivas significaciones. La variable latente u_1 fue ajustada con los regresores:

$$u1_1, t1 \text{ y } t1_1$$

que representan respectivamente los valores del instante anterior de la variable u_1 , los valores de la variable latente t_1 y los valores del instante anterior de la variable t_1 .

De manera análoga,

$$u2_1, t2 \text{ y } t2_1$$

representan los regresores de u_2 .

A partir de los resultados presentados en la tabla 5.4 las expresiones analíticas de los modelos de u_1 y u_2 son representadas por:

$$u_{1i} = 0,648t_{1i} + 0,513t_{1i-1} + \epsilon_{1i}$$

y

$$u_{2i} = -0,041u_{2i-1} + 0,427t_{2i} + 0,379t_{2i-1} + \epsilon_{2i}$$

Los valores de la variable latente u_1 en un instante i sufre un impacto positivo de los valores de la variable t_1 del mismo instante y del instante anterior $i-1$. Por otro lado, los valores de la variable latente u_2 sufre una influencia negativa de sus valores anteriores ($i-1$), una influencia positiva de los valores de t_2 del mismo instante y del instante anterior.

5.3.3.3. Comparación del ajuste y predicción del MDEL y DPLS

Los modelos MDEL y DPLS fueron ajustados al mismo conjunto de datos simulados y los resultados fueron comparados usando los siguientes dos principales criterios:

- (1) Calidad de ajuste de las variables latentes

Las variables latentes fueron determinadas en cada método usando criterios diferentes. En el modelo DPLS las variables latentes fueron determinadas en la dirección de máxima covarianza entre scores de los bloques, en cambio, en el MDEL dichas variables fueron determinadas explorando la máxima varianza en cada bloque, donde fue ajustada la función de transferencia.

Para comparar la calidad de ajuste de las variables latentes fue usado el estadístico R^2 que evalúa la variabilidad explicada de los scores en los 16 modelos ajustados. Los valores obtenidos se muestran en la tabla 5.5 y son comentados más adelante.

- (2) Calidad predictiva del modelo ajustado a las variables latentes

En el modelo DPLS las variables latentes son ajustadas mediante un modelo de regresión lineal, donde cada variable latente respuesta es explicada por una variable latente de los datos de entrada y sus respectivos retardos. En cambio, el modelo MDEL ajusta las variables latentes a una función de transferencia, donde se analiza el impacto de la serie explicativa y sus retardos en la serie respuesta.

La calidad predictiva de las dos metodologías fue testada mediante el método de validación cruzada que consistió en: los datos fueron divididos en dos grupos ($k=2$), donde uno de ellos fue usado para ajustar el modelo (grupo de entrenamiento) y el otro para testar la predicción

(grupo de pruebas). Cada observación del grupo de prueba fue usada en el modelo ajustado y fue calculado su correspondiente residuo. Para el conjunto de los datos de prueba fue calculado el Mean Absolute Error (MAE) con lo cual se compara la capacidad predictiva de las dos metodologías. Estos valores se recogen en la tabla 5.5 y son comentados más adelante.

En todos los casos, fueron extraídos dos componentes (variables latentes) en cada bloque. Con ello, no se busca un número óptimo de componentes que maximiza la información en los datos originales, sino una comparación de los resultados obtenidos en las dos metodologías.

Las variables latentes del modelo dinámico sobre estructuras latentes (MDEL) presentan mayor variabilidad explicada comparado con el modelo DPLS. Los resultados de la prueba *t* de student realizada indican que el \bar{R}^2 del MDEL es mayor, tanto en el bloque X ($t=9.48$, $p - valor < 0,001$), como el bloque Y ($t=5.94$, $p - valor < 0,001$). Los intervalos LSD presentados en las figuras 5.16 y 5.17 ilustran este resultado. La tabla 5.4 muestra que en los 16 modelos ajustados el R^2 de MDEL ha sido mayor que el de DPLS en los dos bloques de variables.

Las variables latentes determinadas en el modelo dinámico sobre estructuras latentes explican en media 82.2 % en el bloque X y 82.3 % en el bloque Y. Mientras que al usar el modelo DPLS los correspondientes valores son 62.5 % y 67.6 % (ver tabla 5.5 para información detallada). Se verifica que la variabilidad explicada por las variables latentes en el modelo propuesto es mayor que en el modelo DPLS. Además, la tabla 5.5 muestra que en los dos bloques X e Y los valores de la varianza explicada es más homogénea en el modelo MDEL que en el modelo DPLS, indicando una mayor robustez y estabilidad en su desempeño (las desviaciones típicas para las medias de las varianzas explicadas son 0.025 y 0.014 para el modelo MDEL y 0.104 y 0.101 para DPLS).

Respecto a la calidad predictiva de los modelos, el MAE fue calculado a través de los residuos estandarizados obtenidos al usar los modelos ajustados a los datos de prueba en la validación cruzada. En los 32 modelos ajustados se obtuvo un MAE medio de 0.129, mucho menor que el MAE medio de 0.893 obtenido en el modelo DPLS ($t=-11.49$, $p - valor < 0,001$), la figura 5.18 ilustra los intervalos LSD que destacan dichas diferencias. De modo análogo al \bar{R}^2 , los valores de MAE muestran que el MDEL es más estable en la predicción comparado al modelo DPLS (El MAE medio del modelo MDEL fue calculado con una desviación típica de 0.002, y el MAE del DPLS tiene una desviación típica de 0.375).

Un ANOVA fue aplicado para analizar el efecto de la autocorrelación en la

Tabla 5.5: Comparación del ajuste y predicción de los modelos MDEL y DPLS

coef.X	coef.Y	Modelo	MDEL			DPLS		
			R ² X	R ² Y	MAE	R ² X	R ² Y	MAE
0.3	0.3	u ₁ =f(t ₁)	0.8044	0.8337	0.1277881	0.5449	0.6295	1.26401
		u ₂ =f(t ₂)			0.1283099			0.52041
0.3	0.5	u ₁ =f(t ₁)	0.8044	0.8096	0.1301638	0.5232	0.6607	1.26462
		u ₂ =f(t ₂)			0.1310276			0.52193
0.3	0.7	u ₁ =f(t ₁)	0.8044	0.8345	0.1291469	0.5341	0.6134	1.26081
		u ₂ =f(t ₂)			0.1277414			0.52232
0.3	0.9	u ₁ =f(t ₁)	0.8044	0.8347	0.127612	0.5106	0.5162	1.26271
		u ₂ =f(t ₂)			0.1281777			0.52118
0.5	0.3	u ₁ =f(t ₁)	0.8140	0.8242	0.1259834	0.5508	0.6685	1.26033
		u ₂ =f(t ₂)			0.1311101			0.52900
0.5	0.5	u ₁ =f(t ₁)	0.8140	0.8316	0.1318307	0.5327	0.7205	1.26022
		u ₂ =f(t ₂)			0.1282074			0.52001
0.5	0.7	u ₁ =f(t ₁)	0.8140	0.8356	0.132531	0.5713	0.6417	1.26021
		u ₂ =f(t ₂)			0.1319707			0.52911
0.5	0.9	u ₁ =f(t ₁)	0.8140	0.8134	0.1325209	0.5470	0.5003	1.26732
		u ₂ =f(t ₂)			0.127483			0.52781
0.7	0.3	u ₁ =f(t ₁)	0.8049	0.8051	0.1315716	0.6318	0.7110	1.26012
		u ₂ =f(t ₂)			0.1293678			0.52203
0.7	0.5	u ₁ =f(t ₁)	0.8049	0.8095	0.130561	0.6461	0.7288	1.26241
		u ₂ =f(t ₂)			0.1290117			0.52512
0.7	0.7	u ₁ =f(t ₁)	0.8049	0.8038	0.1295587	0.6528	0.7162	1.26350
		u ₂ =f(t ₂)			0.1315897			0.52002
0.7	0.9	u ₁ =f(t ₁)	0.8049	0.8079	0.134164	0.622	0.5402	1.26462
		u ₂ =f(t ₂)			0.1318124			0.52031
0.9	0.3	u ₁ =f(t ₁)	0.8651	0.8417	0.1301093	0.7848	0.8221	1.26801
		u ₂ =f(t ₂)			0.1317815			0.52022
0.9	0.5	u ₁ =f(t ₁)	0.8651	0.8386	0.1298668	0.792	0.824	1.26231
		u ₂ =f(t ₂)			0.1291468			0.52992
0.9	0.7	u ₁ =f(t ₁)	0.8651	0.8410	0.1260220	0.788	0.813	1.26241
		u ₂ =f(t ₂)			0.1300121			0.52455
0.9	0.9	u ₁ =f(t ₁)	0.8651	0.8054	0.1254203	0.775	0.718	1.26311
		u ₂ =f(t ₂)			0.1291012			0.52381

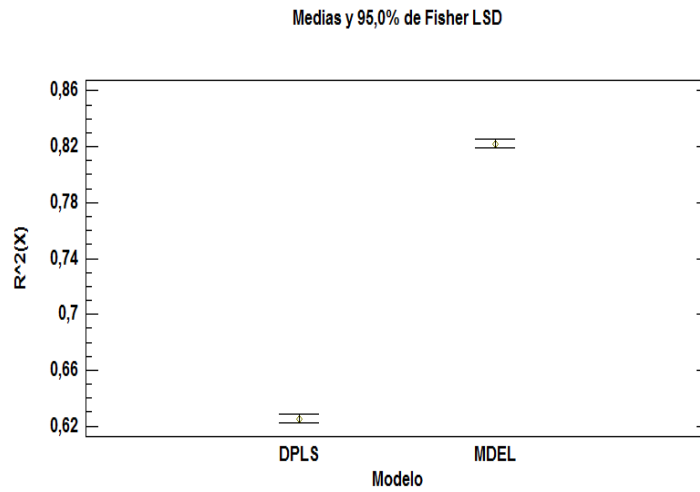


Figura 5.16: Comparación del \bar{R}^2 de las variables latentes del bloque X obtenidas en el MDEL y DPLS

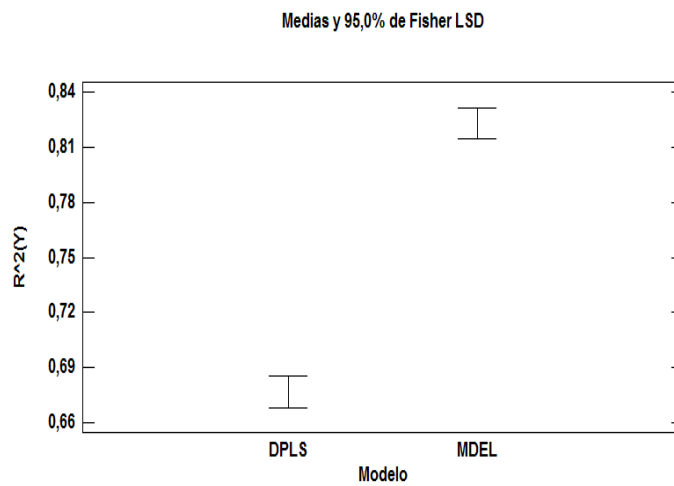


Figura 5.17: Comparación del \bar{R}^2 de las variables latentes del bloque Y obtenidas en el MDEL y DPLS

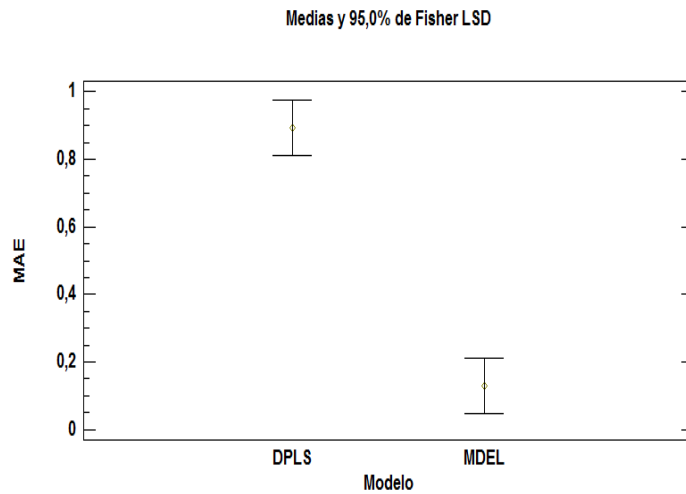


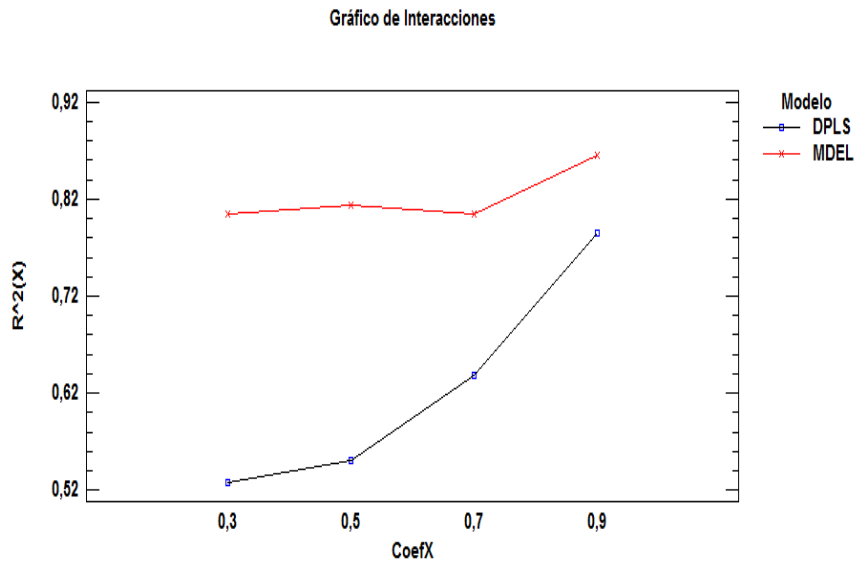
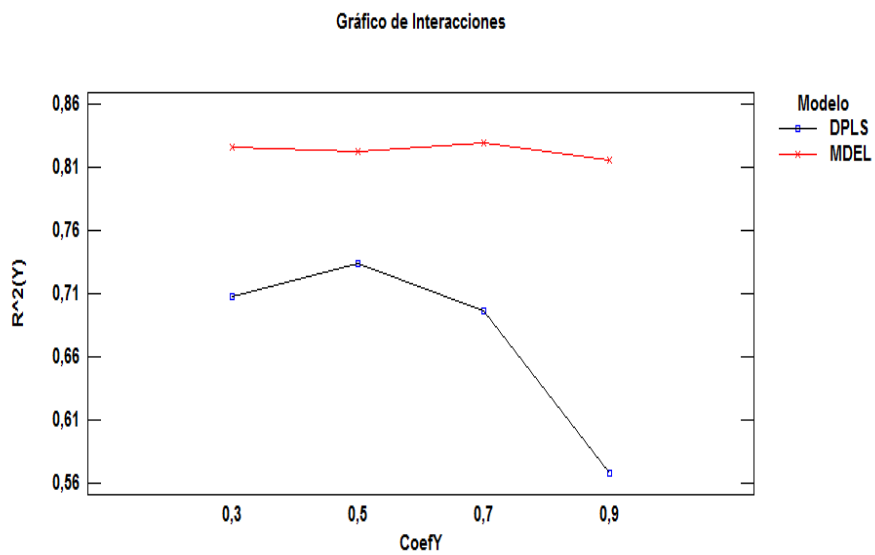
Figura 5.18: Comparación del MAE de MDEL y DPLS

variabilidad explicada por las variables latentes, y en la capacidad predictiva de los dos métodos. Los resultados muestran que el coeficiente autoregresivo no afecta a la calidad de ajuste del MDEL. Sin embargo, el R^2 de las variables latentes en el modelo DPLS depende de la estructura de autocorrelación de las variables. Las figuras 5.19 y 5.20 muestran una gran variabilidad de los valores de R^2 del modelo DPLS para los diferentes coeficiente autoregresivos de las variables de entrada (X) y de salida (Y). Dicha dependencia del R^2 del modelo DPLS a los coeficientes autoregresivos del proceso se confirma en la prueba F de Fisher donde la interacción entre los factores modelo y coeficiente es muy significativa (Factores: Modelo*Coeficiente X: $F=287.53$, $p - \text{valor} < 0,001$; Factores: Modelo*Coeficiente Y: $F=30.28$, $p - \text{valor} < 0,001$).

Los coeficientes autoregresivos no afectan a la calidad predictiva en los dos métodos. No obstante, en todos los procesos, con baja, media y alta autocorrelación, el modelo MDEL presentó mejor calidad predictiva comparado con el modelo DPLS ($F=87.46$, $p - \text{valor} < 0,001$).

De modo ilustrativo, las figuras 5.21-5.23 presentan los gráficos de los dos modelos ajustados a los procesos con coeficientes autoregresivos igual a 0.3, 0.5 y 0.9 que representan procesos con baja, media y alta correlación, respectivamente.

En todas las figuras, el gráfico MDEL corresponde a la primera variable latente del modelo dinámico sobre estructuras latentes y su respectiva predicción; asimismo, el gráfico DPLS corresponde a la primera variable latente obtenida en el modelo DPLS y su respectiva predicción. Las líneas continuas y interrumpidas indican los valores observados y predichos, respectivamente.

Figura 5.19: Efecto de la autocorrelación al $\bar{R}^2(X)$ Figura 5.20: Efecto de la autocorrelación al $\bar{R}^2(Y)$

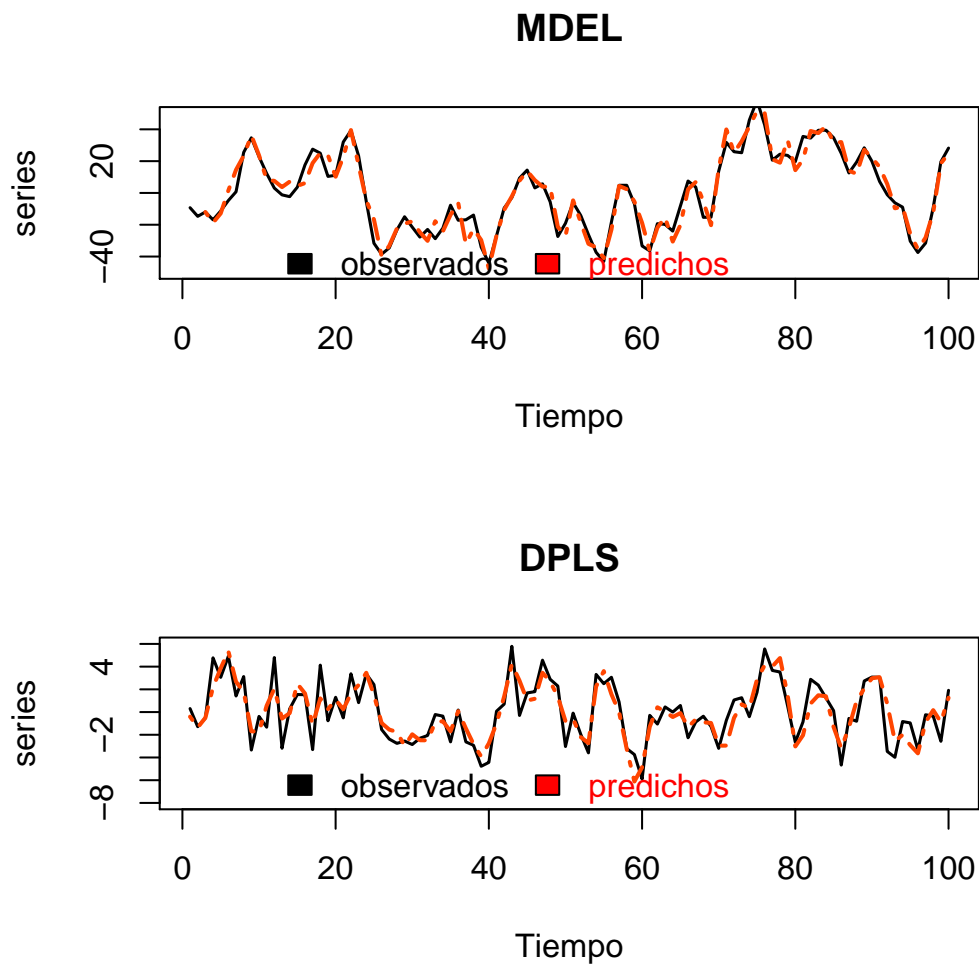


Figura 5.21: Ajuste del modelo dinámico y del DPLS para procesos con coeficientes autoregresivos igual a 0.3

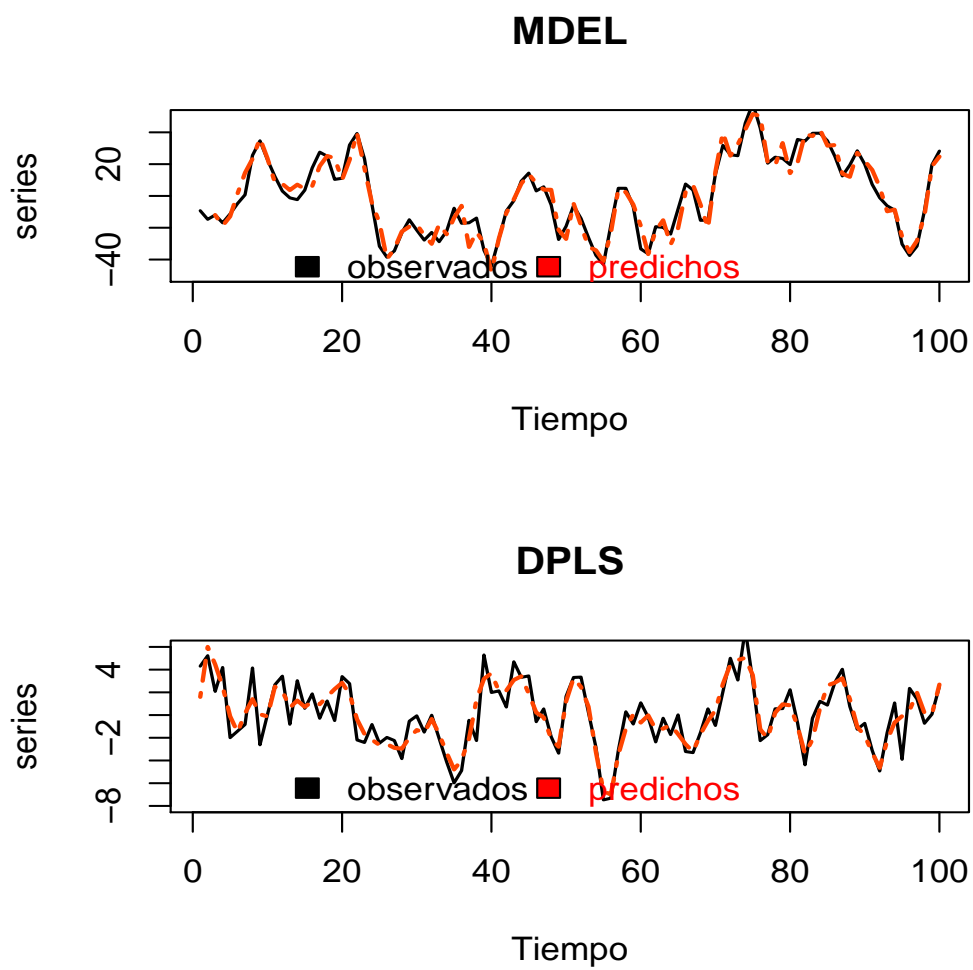


Figura 5.22: Ajuste del modelo dinámico y del DPLS para procesos con coeficientes autoregresivos igual a 0.5

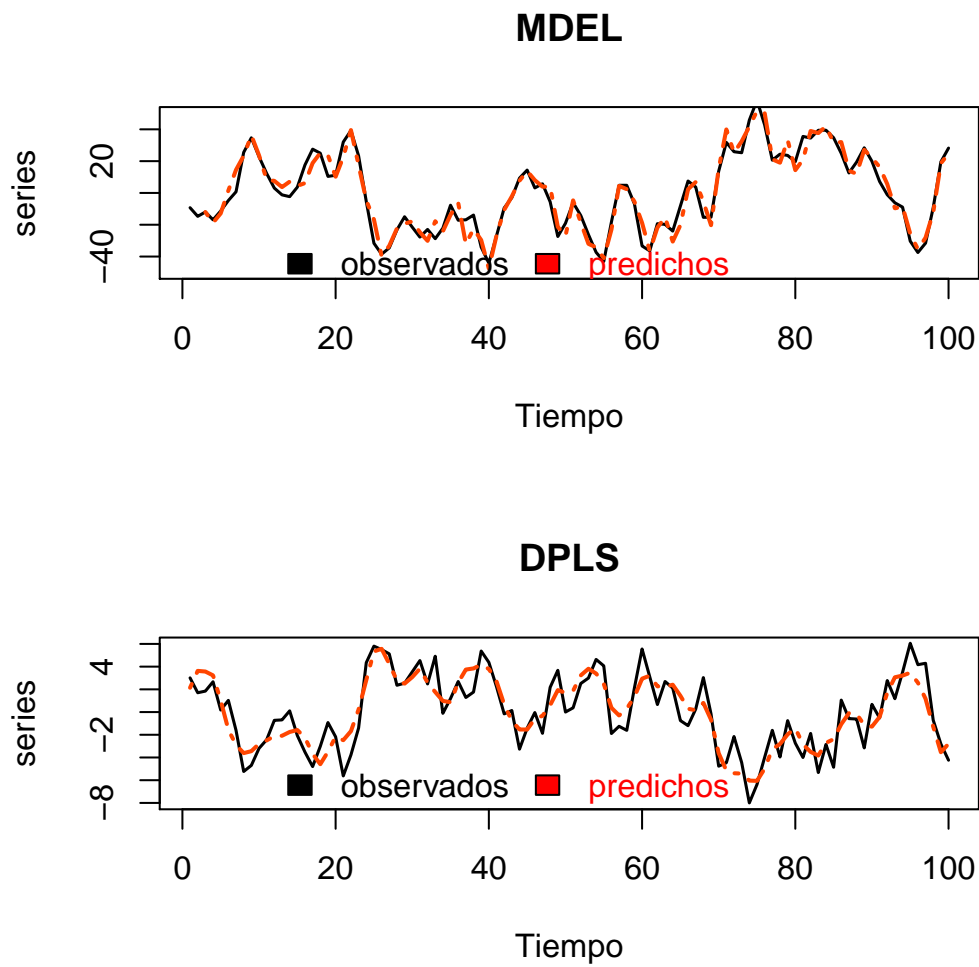


Figura 5.23: Ajuste del modelo dinámico y del DPLS para procesos con coeficientes autoregresivos igual a 0.9

Las figuras muestran que en todos los casos presentados el modelo dinámico sobre estructuras latentes se ajusta mejor a los datos comparado con el modelo DPLS. La discrepancia entre los valores observados y predichos en el modelo MDEL es muy pequeña, lo que corrobora con los valores bajos de MAE observados.

5.3.4. Conclusión

El modelo dinámico sobre estructuras latentes fue desarrollado para lograr dos principales objetivos: (1) modelar de manera adecuada los procesos multivariantes y autocorrelacionados con datos de entrada y salida, y (2) lograr mejores resultados respecto a un método aplicado a estos procesos.

El primer objetivo fue logrado mediante la aplicación de un modelo DP-CA a los datos de entrada y salida, y sobre los scores resultantes fue ajustada una función de transferencia. Con el modelo DPCA se explota la variabilidad máxima en cada bloque, y la función de transferencia modela la estructura dinámica de los scores obtenidos mediante el método de Box-Jenkins.

Respecto al segundo objetivo, los resultados del modelo dinámico fueron comparados con el modelo DPLS, donde la calidad de ajuste de las variables latentes fue analizado a través del R^2 , y la capacidad predictiva de los modelos fue testada mediante el método de validación cruzada, donde el MAE de cada modelo fue determinado.

El modelo dinámico sobre estructuras latentes presentó mejor ajuste a los datos simulados, con un R^2 de las variables latentes superior que el obtenido en el modelo DPLS. Además, presentó un MAE medio inferior, con una desviación típica muy baja, indicando que el modelo propuesto es más robusto y estable en la predicción de nuevas observaciones comparando al modelo DPLS.

C apitulo 6

APLICACI ON DE LOS M ETODOS ESTAD ISTICOS PROPUESTOS AL PROCESO DE PRODUCCI ON EN ACUICULTURA

6.1. Introducci on

El m etodo estad stico propuesto en este trabajo fue elaborado con vista a monitorizar y predecir procesos multivariantes y autocorrelacionados en la industria. El desempe o de los modelos desarrollados fue testado mediante su ajuste a los datos simulados, y a continuaci on fueron comparados con otros m etodos conocidos y aplicados a estos procesos. En este capitulo se ilustran los resultados de aplicaci on de dichos modelos a un caso real, donde el gr afico MAAEWMA y el modelo din amico sobre estructuras latentes (MDEL) fueron usados para monitorizar y predecir el proceso de cr a de especies acu ticas en acuicultura.

La estructura de datos del proceso de cr a de especies acu ticas en acuicultura corresponde a la clase de procesos multivariantes y autocorrelacionados, con datos de entrada y salida, referidos en este trabajo. Las caracter sticas del proceso en acuicultura forman dos grupos: (1) Los par metros de control del crecimiento y calidad de la especie cultivada (par metros biom tricos), y (2) los par metros que caracterizan el ambiente acu tico (par metros f sico-qu mico y biol gicos, tambi n conocidos por par metros ambientales).

Las caracter sticas del proceso en acuicultura indican que hay una dependencia entre los par metros de cada grupo (correlaci on), y sus propiedades cambian con el tiempo (autocorrelaci on). Adem s, la calidad de la especie depende del ambiente acu tico a que est n expuestos (influencia del ambiente acu tico a la calidad de la especie), de modo que los datos del proceso cumplen con las hip tesis referentes a los procesos MIMO tratados en este trabajo.

Los datos usados en este capitulo fueron registrados en el proceso de cultivo de tilapia en la empresa Aquapesca.Lda en Mozambique. La aplicaci on del m etodo estad sticos busca mejorar la monitorizaci on de los par metros ambientales y la predicci on del  ndice de crecimiento de la tilapias a trav s del

ambiente acuático.

La monitorización del ambiente acuático en acuicultura se suele basar en el control univariado de la conformidad de los parámetros ambientales. El proceso consiste en medir cada parámetro a lo largo del tiempo y verificar si se cumplen los requisitos técnicos establecidos para cada parámetro. Por tanto, la conformidad del proceso en acuicultura se refiere al cumplimiento univariado de los requisitos técnicos establecidos a los parámetros ambientales.

No obstante, la hipótesis de correlación entre los parámetros ambientales sugiere la aplicación de un método de monitorización multivariante con vista a explotar a la vez la información de cada parámetro y la originada por la interacción entre ellos. Un proceso puede presentar de manera independiente todas sus variables bajo control, mientras que la estructura de correlación entre ellas está fuera de control, pudiendo comprometer a la calidad del producto.

La monitorización mediante el gráfico MAAEWMA desarrollado en el capítulo 5 se basa en el control de la estabilidad del proceso mediante el estadístico T_3^2 , donde se asume que el proceso está bajo control si el valor del estadístico está en la banda $[0, h_3]$. De modo general, un proceso se considera estable si sus parámetros (normalmente la media y/o la varianza) no cambian a lo largo del tiempo.

El control de la estabilidad y de la conformidad son ambos relevantes en la monitorización del ambiente acuático en acuicultura. A pesar de que en muchos procesos de cría de especies acuáticas solo se consideran el control de la conformidad (como es el caso del cría de la tilapia en Mozambique), el control de la estabilidad del proceso también es relevante, ya que cambios repentinos de los parámetros ambientales (por ejemplo cambios de 5^0c en la temperatura del agua) causa estrés en los peces, y en casos más críticos, la muerte (Martinez, 2006).

Para lograr ambos objetivos, el gráfico MAAEWMA fue aplicado para monitorizar la estabilidad y la conformidad del ambiente acuático, donde además del límite de estabilidad h_3 , fueron aplicados los límites de conformidad (que explicaremos con detalles en su momento).

6.2. La Acuicultura

La actividad de criar peces y otros organismos acuáticos es muy antigua. Desde los tiempos remotos el hombre mostró la necesidad de optimizar los criterios de producción de pescado para su subsistencia. Los registros indican que los chinos ya practicaban acuicultura varios siglos antes de nuestra era, y

los egipcios ya creaban la tilapia del nilo (*Sarotherodon niloticus*) hace 4000 años.

Según la FAO (Organización de las Naciones Unidas para la Agricultura y la Alimentación) la acuicultura se define como la producción de organismos acuáticos, la cría de pescado, moluscos, crustáceos, anfibios, reptiles y el cultivo de plantas acuáticas para el consumo humano.

Para Acuña (Acuña, n/a) la acuicultura está mucho más relacionada con la agricultura y la ganadería que con la pesca, ya que implica la cría y el manejo de los recursos acuáticos vivientes en un medio ambiente restringido. Mientras que en la pesca y en la caza se recogen animales a partir de acceso libre, en la acuicultura implica la existencia de derechos de tenencia y de propiedad de dichos recursos. La posesión de los medios de producción y los derechos de propiedad sobre la producción son tan importantes para el éxito de la acuicultura, como la tenencia de la tierra lo es para la agricultura.

La producción en acuicultura implica una intervención de la capacidad humana en el proceso de cría para obtener mejores resultados. Para ello, la acuicultura industrializada moderna es una actividad con un fuerte componente científico-técnico y se encuentra asociada a diversas disciplinas tales como la biología, la ingeniería y la economía. Las ramas de la biología más directamente implicadas en la producción acuícola son la fisiología, la etología, la genética, la ecología, la patología y la biotecnología (Acuña, n/a).

Actualmente, la acuicultura es considerada responsable de la producción de la mitad de los pescados consumidos por la población mundial. La producción de pescados a través de la acuicultura se triplicó entre 1995 y 2007 (University, 2009). Y esto puede estar asociado al hecho de que la acuicultura sea la actividad zootécnica de más rápido crecimiento a escala global, con tasas de expansión que se han sostenido en 6,9% en promedio, entre 1970 y 2006. El crecimiento exponencial de esta actividad en los últimos 50 años permitió pasar de menos de un millón de toneladas en 1950 a 51.7 millones de toneladas en 2006 y contribuye actualmente con el 50% de los productos acuáticos que se consumen en el mundo.

La acuicultura tiene orígenes asiáticos, siendo que los registros indican que la China es uno de los pioneros. Actualmente esta actividad se ha expandido a todos los continentes, y según datos de la FAO más de 440 especies de ambientes dulce-acuícolas, salobres y marinos han sido cultivadas en el mundo entre 1950 y 2006, con un valor comercial en este último año, de aproximadamente 91.200 millones de dólares. La América Latina, con una expansión superior a 22% anual, presenta la tasa de crecimiento más acelerado de la actividad en los últimos años, siendo que Chile, Brasil, Ecuador y México contribuyen con

el 85 % de la producción.

Según Labarta (Labarta, 2007), los niveles de la acuicultura en la unión europea se aproxima al millón y medio de toneladas, representando más del 15 % de la producción pesquera. Las especies principales son los moluscos, con especial importancia del mejillón, ostra y almejas. El salmón es el pescado que más se destaca en acuicultura marina, seguido de la dorada, lubina y rodaballo. La producción española de acuicultura alcanza el 3 % en la producción mundial y 25 % de la europea.

Respecto a la producción de los peces marinos, España ha alcanzado un 72 % en el período 2000 a 2005, aunque participen en solo 10 % en el total de la acuicultura marina.

6.2.1. La Acuicultura en Mozambique

Mozambique, oficialmente República de Mozambique, es un país localizado al sudeste de África, bañado por el Oceano Índico y hace frontera con Tanzania al norte; Malawi y Zambia al nordeste; Zimbabwe al oeste y Suazilandia y Sudáfrica al sudoeste (en rojo en la fig.6.1).



Figura 6.1: Localización de Mzambique en el globo

Constituido por 11 provincias y con una área total de 801590km^2 tiene una población, estimada en 2017, de aproximadamente 28.5 millones de habitantes. La agricultura y la pesca son las actividades históricamente de mayor adherencia por los habitantes, siendo que la agricultura suele ser practicada en el interior y la pesca en el litoral. Estas actividades, a pesar de ser practicadas de manera rudimentaria, constituyen la base de subsistencia de la mayor parte de la población del país.

Las políticas actuales del gobierno muestran un mayor interés en la mejora y crecimiento de las actividades del sector pesquero, siendo que para ello, fue elaborado un plan estratégico (*Plano director das Pescas 2010-2019*), que ofrece metas a largo plazo de la contribución del sector de la pesca para los objetivos nacionales. La seguridad alimentaria y la reducción de la pobreza son las prioridades. La acuicultura y el mejoramiento de las capacidades humanas e institucionales, son considerados como prioridades para lograr tales objetivos (MinistériodasPescas, 2012).

El Ministerio de la Pesca de Mozambique es el sector del gobierno responsable del desarrollo de la acuicultura. Este elaboró el plan referido en el párrafo anterior, que está alineado a la estrategia de la reducción de la pobreza PARPA II. 2011, cuyo objetivo general es la producción y el aumento de la productividad de la agricultura y de la pesca. El PARPA menciona especialmente los desafíos de la agricultura y de las pescas a nivel familiar y los considera como fundamentales para la seguridad alimentaria y nutrición de la población. El documento indica las siguientes prioridades para la promoción de la producción y productividad agrícola y de la pesca:

1. Mejorar y ampliar el acceso a los factores de la producción
2. Facilitar el acceso al mercado
3. Mejorar la gestión sustentable de los recursos naturales (tierra, agua, pesca y bosques)

El gobierno de Mozambique ha reconocido la importancia socioeconómica y el alto potencial del sector de acuicultura, siendo por eso aprobado la *Estrategia para el desarrollo de la acuicultura en Mozambique 2008-2017* (MinistériodasPescas, 2012), con el objeto de garantizar que los recursos naturales con potencialidades para acuicultura sean efectivamente explorados para apoyar el crecimiento y el desarrollo del país y contribuir para la reducción de niveles de pobreza. Los tres objetivos estratégicos son:

1. Incrementar el desarrollo sostenible de la acuicultura, que tiene como resultado *augmentar el rendimento de los acuicultores como contribución del sector de pesca para el plan de acción, para reducción de la pobreza absoluta (2006-2009)*
2. Aumentar los niveles actuales de producción de camarón marino y otras especies acuáticas (animales y vegetales) destinadas a la exportación y garantizar la seguridad alimentaria de las comunidades. Para ello, se esperan los siguientes resultados:
 - a) El aumento del volumen de producción de acuicultura;
 - b) La mejora del consumo de proteínas en la alimentación de las comunidades, garantizando así su seguridad alimentaria;

- c) El crecimiento de los ingresos de exportación como contribución del sector de la pesca para el PIB.
3. Establecer un cuadro legal, normativo e institucional de gestión de la acuicultura apropiado y eficaz, teniendo como resultados la gestión, administración y exploración sustentable de los recursos de acuicultura mejorada.

Además de la elaboración de un plan estratégico, hubo financiación de proyectos de gran alcance (como es el caso del proyecto de cría de tilapia en la empresa Aquaquel, de Nicoadala, entre otros grupos asociados en las provincias de Zambezia y Sofala), de soporte técnico y suministro de pienso a los cultivadores de pequeña escala.

La acción del Gobierno motiva cada vez más a los jóvenes en crear proyectos relevantes en la cría de especies acuáticas, particularmente en la cría de camarón y de la tilapia, con lo cual se puede afirmar que la expansión de la actividad de acuicultura, presentada como uno de los objetivos en el plan estratégico mencionado, se está cumpliendo. No obstante, lo mismo no se puede decir respecto a la capacidad científico-tecnológica para monitorizar y controlar el proceso de cría de las especies. Existe la necesidad de usar métodos eficientes para monitorizar y predecir el ambiente acuático y la calidad de la especie.

Actualmente la acuicultura en Mozambique es fuente de rendimiento económico familiar y empresarial, y contribuye significativamente al crecimiento de la economía del país a través de exportaciones del pescado y de creación de puestos de trabajo. El camarón y la tilapia son las especies más cultivadas.

La cría de tilapia (a pesar de ser una especie más resistente a situaciones adversas que el camarón) es una actividad delicada y que necesita de cuidados y controles regulares muy profundos. El éxito de esta actividad depende del control de muchos parámetros físico-químicos (oxígeno, pH, temperatura, salinidad, alcalinidad, amonio, nitrato, nitrito, fósforo, transparencia, silicio, hierro, entre otros). Una pequeña variación de uno de los parámetros físico-químicos (como es el caso de la temperatura y oxígeno, por ejemplo) afecta al ambiente acuático, causando crisis que pueden llevar a muerte (Alves y Mello, 2007).

La tilapia es actualmente la especie más cultivada en la provincia de Zambezia-Mozambique¹. Además de Aquapesca, estanques de cría de tilapia se pueden encontrar en Inhagome-Quelimane en la empresa Aquaquel y en otras localidades de la provincia. Sin embargo, la productividad de esta actividad puede ser

¹http://www.inaqua.gov.mz/Documentos/Estrategia/PlanoAquaComercial_CC12.pdf, consultado a 30.02.2015

mejorada significativamente empleando mejores métodos de control del proceso.

Como consecuencia de que la industrialización de la actividad es relativamente reciente en Mozambique, los métodos de monitorización aún no son consistentes para un nivel de producción a gran escala. Así, al aplicar los métodos de monitorización y previsión propuestos en este trabajo se pretende contribuir al desarrollo de la acuicultura en Mozambique, explotando aspectos relevantes de la influencia de la variabilidad del ambiente acuático en la calidad de la tilapia cultivada.

6.3. Métodos estadísticos en acuicultura

La relevante contribución de la acuicultura en la alimentación mundial, despertó interés en la investigación orientada a la aplicación de técnicas estadísticas, con vista a explotar la información importante registradas en estos procesos. Este aspecto podría estar asociado a la necesidad de mejorar los métodos aplicados en el cultivo, principalmente respecto al control de calidad, con vista a lograr mejores resultados en el proceso de cría.

Diversos estudios orientados al ambiente acuático, infecciones, nutrición, etc., han sido realizados. El estudio de Miranda (Miranda *et al.*, 2010) consiste en analizar los parámetros ambientales de manera cualitativa y aplica criterios técnicos de acuicultura para su evaluación. Jimenez y Barba (Jiménez y Barba, 2000) analiza una serie de parámetros físico-químicos mediante métodos normalizados, con objeto de conocer si el valor de estos parámetros se encuentra dentro del intervalo que marca la legislación vigente.

Las técnicas estadísticas descriptivas e inferencias univariantes fueron aplicadas en muchos estudios para analizar los niveles de los parámetros, comparar grupos, calcular correlaciones (Sotolu y Faturoti, 2009).

El estudio del crecimiento de la especie en acuicultura es llevado a cabo a través de los parámetros biometricos. Las curvas de crecimiento en peso y en talla de Von Bertalanffy, las funciones de relación peso-talla ha sido aplicada por Garcia et al. (Garcia *et al.*, s/a) para analizar el crecimiento de la tilapia Nilótica (*Oreochromis niloticus*). Sus resultados indican un crecimiento isométrico de la especie, considerando que la relación entre peso y longitud sigue un modelo exponencial de crecimiento paralelo.

A su vez Guerra y Manriquez (Guerra y Manriquez, 1979) afirma que las expresiones utilizadas por Garcia et al. (Garcia *et al.*, s/a) sólo se aplican con rigor cuando se analiza el crecimiento de uno o varios individuos que se pe-

san y se miden a lo largo de toda su vida. No obstante, este procedimiento rara vez es posible y, por lo general, se procede a determinar el valor de los parámetros de la regresión a base de comparar los pesos y las tallas de una gran cantidad de ejemplares de diferente tamaño obtenido por uno o diversos métodos de muestreo. Y en su trabajo, ajusta tres modelos de regresión simple para analizar la relación entre el peso y la talla de *Octopus vulgaris* de forma separada a los machos, las hembras y ambos los sexos. Sus resultados indican que el modelo de regresión simple explica mejor la relación entre el peso y la talla del animal estudiado.

Un estudio que aplica análisis factorial y de componentes principales (PCA: su sigla en inglés) para analizar la estructura de correlación entre los parámetros ambientales en las aguas costeras de la región de Murcia ha sido llevado a cabo por Martínez et al. (Martínez *et al.*, n/a). Los resultados muestran que con un PCA se ha podido reducir el ambiente acuático en tres componentes que explican una variabilidad de 70 % de los datos originales. Las variables clorofila, ortofosfatos y nitrógeno total presentan una correlación positiva, y a su vez están relacionadas negativamente con la salinidad en la primera componente. En la segunda componente el fitoplancton total y los silicatos presentan una mayor relación. En tercera, la temperatura y la tasa de sólidos en suspensión se presentan de forma muy homogénea, y este aspecto se justifica por se tratar de aguas someras.

Yan y Shi (Yan y Shi, 2014) afirman que los métodos tradicionales de cultivo no pueden satisfacer la creciente escala de la acuicultura, y presentan en su trabajo una combinación de la tecnología para una creación inteligente de las especies acuáticas. Este método analiza los factores ambientales que afectan al crecimiento de las especies, buscando el mejor entorno para su cultivo.

Nuevos métodos para el manejo y creación de especies marinas basadas en las tecnologías informáticas fueron desarrollados. En 2014 Xu (Xu y Zhang, 2014) presentó un método de monitorización de especies acuáticas basado en la plataforma androide usando un sistema con muchos nodos de sensores, donde los valores de los parámetros ambientales (pH, temperatura, nivel de agua, el oxígeno disuelto y otros parámetros ambientales) son recogidos para ser transformados en datos digitales. Un estudio parecido al de Xu fue realizado en 2015 por Jinfeng (Jinfeng y Shun, 2015) que propuso un sistema de monitorización de la calidad del agua basado en una metodología digital denominada ZigBee. El sistema recopila, transmite las pantallas, registra los parámetros de calidad del agua (la temperatura, concentración de oxígeno disuelto, pH de valor, y del nivel de agua) y controla el aumento del oxígeno de la máquina. Los resultados indican que el sistema funciona de forma estable y se puede regular automáticamente la calidad del agua.

La alta calidad del agua juega un papel importante para la cosecha de los organismos acuáticos. Por ello, Liu (Liu *et al.*, 2016) propuso en 2016 un sistema de control que contiene un subsistema de monitorización de agua para el usuario, así como un servidor para almacenar y analizar los datos monitorizados.

Los métodos actuales de monitorización en acuicultura tienen una tendencia tecnológica y informática muy desarrollada. Zhou (Zhou y Xing, 2013) aplica un sistema inteligente de monitorización en acuicultura basado en el diseño de un hardware y software para el control de los procesos de la industria de acuicultura. Se aplica un servidor web, servidor de base de datos y el navegador para establecer la plataforma de gestión para el control del medio ambiente y el proceso de producción. A través de un control en tiempo real de oxígeno disuelto, temperatura y pH en el estanque, este sistema estabiliza estos parámetros en cada valor óptimos propios, y mejora considerablemente la productividad global. Los resultados de las pruebas muestran que el sistema es de manejo fácil y proporciona una medida directa y práctica para la acuicultura, y ahorra energía.

Por otro lado, técnicas estadísticas más elaboradas fueron aplicados. En 2014, Zhou et.al (Zhou *et al.*, 2014) realizaron un estudio para describir la composición y abundancia de microinvertebrados, y su relación con los factores ambientales aplicando técnicas estadísticas multivariantes. Consideraron varios factores ambientales: la profundidad del agua, temperatura del agua, pH, salinidad, oxígeno disuelto, nitrógeno inorgánico disuelto, fósforo reactivo soluble, y silicato. El análisis de correspondencia aplicado evidenció que el nitrógeno inorgánico disuelto y oxígeno disuelto, fueron los factores ambientales que más influyen en conjuntos de macroinvertebrados en Bohai Bay-norte de China.

Respecto a los modelos de predicción, los modelos ARIMA y de redes neuronales fueron aplicados en muchos estudios para predecir la calidad del agua en acuicultura. Como es el caso de Palani, Liong y Tkalich (Palani *et al.*, 2008) que aplican las redes neuronales artificiales para ajustar un modelo predictivo de la calidad del agua caracterizada por los parámetros: temperatura, oxígeno disuelto, salinidad y clorofila-a, en la región costera de Singapur. El estudio fue orientado para evidenciar las relaciones lineales y no lineales de las variables y analizar las causas de dicha relaciones a partir de la base de datos. Con un R^2 entre 0,8 a 0,9 obtenido en la simulación del grupo testado y ajustado en el modelo, se consideró un buen resultado para predecir la calidad del agua.

Han (Han *et al.*, 2011) propuso una metodología que consiste en una función de base radical de red neuronal de estructura flexible para predecir la calidad del agua. Han afirma que la metodología puede variar su estructura de forma

dinámica con el fin de mantener la exactitud de la predicción. Los resultados del algoritmo fue comparado con otras técnicas referentes a sistemas dinámicos no lineales y los resultados preliminares muestran que esta metodología puede ser empleada para construir una estructura de función de base radial con un numero menor de neuronas ocultas. A su vez, West (West y Dellana, 2011) realizó un análisis empírico sobre redes neuronales sobre estructuras con memoria con objeto de predecir la calidad del agua, y sus resultados muestran que no ha podido superar las predicciones de los modelos JENN y GMNN de redes neuronales. Sin embargo, considera un buen ajuste del modelo a los datos.

Con todo, Faruk (Faruk, 2009) realizó un estudio donde aplica a la vez un modelo ARIMA y las redes neuronales artificiales para predicción de la calidad del agua. Faruk (Faruk, 2009) considera que para estudiar el patrón de la variabilidad de estos procesos y obtener buenas predicciones son necesarias las dos metodologías, ya que con un ARIMA no se puede explotar la relación no lineal entre las variables y una red artificial no es capaz de explotar en simultaneo la relación lineal y no lineal entre ellas. Los resultados muestran que la mezcla de dichas metodologías presentan mejores resultados comparativamente a cada una en separado.

Modelos Estadísticos de proyecciones sobre estructuras latentes fueron aplicados por Liu (Liu *et al.*, 2014) en acuicultura con el propósito de superar las bajas predicciones y mala robustez de los métodos de previsión tradicionales en la calidad del agua. El estudio muestra que la dimensión de la acuicultura ecológica representados por los datos ambientales se redujo con un análisis de componentes principales y el modelo de mínimos cuadrados de vectores soporte máquina (LSSVM) se aplicó para predecir los parámetros estudiados. Los resultados, indican que los modelos presentan buena capacidad predictiva y que tienen alta precisión en los valores pronósticos.

6.4. Parámetros que caracterizan el proceso en acuicultura

El proceso de cría de especies acuáticas se puede considerar una actividad que comporta riesgos, ya que se basa en mantener con vida las especies en el cautiverio y garantizar las condiciones para su desarrollo a través del manejo de su ambiente acuático, asumiendo cambios climáticos, de nivel de oxígeno, entre otros parámetros que condicionan la calidad del agua. La calidad de la alimentación suministrada es un factor relevante en este proceso. Cultivar una especie en ambientes diferentes o aplicar diferentes métodos de alimentación, puede resultar en rendimientos diferentes.

De acuerdo con Martínez (Martínez, 2006), la calidad del agua es determinada por sus propiedades físico-químicas. La temperatura, oxígeno, pH y la transparencia son los parámetros más importantes, ya que influyen en la productividad y reproductividad del pez, por lo que estos deben mantenerse bajo control para garantizar buen cultivo de la tilapia (Martínez, 2006).

En vista a su relevancia, en este trabajo fueron considerados los cuatro parámetros físico-químicos mencionados por Martínez para evaluar la calidad del ambiente acuático en el proceso de cultivo de la tilapia en la empresa Aquapesca de Mozambique. Para evaluar el crecimiento de la tilapia fueron considerados dos parámetros biométricos: peso y longitud standard. Por lo tanto, según la nomenclatura usada en el método estadístico desarrollado, los parámetros ambientales corresponden a las variables del proceso (X), y los biométricos a las variables del producto (Y).

En vista a los objetivos de la empresa Aquapesca, asociados a la investigación de alternativas alimenticias, fue considerado en el experimento un factor *tratamiento* que recoge los datos del proceso de cría de tilapia usando cuatro diferentes alternativas alimenticias.

Los tratamientos C1 (donde fue suministrado el pienso), C6 (fertilizante orgánico (compost)) y C7 (fertilizante inorgánico(urea)), fueron aplicados en estanques de $600m^2$ de superficie. El tratamiento C2, fue usado como referencia (control) en un estanque con dimensión de $250m^2$. En todos casos se consideró una densidad de $100g/m^2$, lo que corresponde a un *pez/m^2*. De acuerdo con las dimensiones de los estanques y su biomasa total, el sistema de cultivo se puede considerar intensivo (Nicovita, 2002).

El cultivo de la tilapia de Mozambique se realizó en agua salobre, y Harada y King demostraron que el cultivo en este ambiente puede ser más rentable que en agua dulce (Martínez, 2003):

”Los primeros estudios sobre la posibilidad de cultivar tilapia en agua salada se realizaron en Hawai en 1950, se mantuvieron cultivos intensivos de *O. mossambicus* en estanques con agua salobre, con una salinidad de entre 10–15%. Los resultados de sobrevivencia y de crecimiento que se obtuvieron en esas condiciones con esta especie fueron superiores a los logrados en agua dulce, con lo cual quedó demostrada la facilidad de su cultivo en agua salobre”.

6.4.1. Parámetros físico-químicos

El estudio de los parámetros físico-químicos resulta muy relevante, ya que el crecimiento de los peces depende en gran parte de la calidad del agua. Se asume

que para lograr una buena producción es necesario mantener las condiciones físico-químicas del agua dentro de los límites de tolerancia para la especie a cultivar (Martinez, 2006). En este trabajo fueron considerados cuatro parámetros físico-químicos (oxígeno, temperatura, pH, transparencia), seleccionados según su relevancia en la contribución en la calidad del agua como se comentó anteriormente.

1. Oxígeno disuelto

Según Alves y Mello (Alves y Mello, 2007) oxígeno es el más importante entre los parámetros físico-químicos en el cultivo de especies acuáticas. La variación del oxígeno en acuicultura proporciona cambios importantes en el proceso. Niveles más altos de oxígeno son los más deseados. Sin embargo, se consideran cuatro intervalos en los que estando este parámetro en uno de ellos se espera un determinado resultado en la cría de peces (tabla 6.1). El grado de saturación del oxígeno disuelto es inversamente proporcional a la altitud y directamente proporcional a la temperatura y pH (Nicovita, 2002).

Tabla 6.1: efectos de cada intervalo de fluctuación de oxígeno

oxígeno(mg/l)	efectos
0 - 0.3	los peces pequeños sobreviven en cortos períodos
0.3 - 2.0	letal a exposiciones prolongadas
3.0 - 4.0	los peces sobreviven pero, crecen lentamente
≥ 4.5	deseable para el crecimiento del peces

La exposición de la especie a bajos niveles de oxígeno causa problemas en el cultivo, como la disminución de la tasa de crecimiento, relación alimento/aumento del peso, inapetencia y letargia, enfermedades a nivel de branquia, susceptibilidad a enfermedades y disminución de la capacidad reproductiva (Nicovita, 2002).

2. Temperatura

De acuerdo con Vinatea (Vinatea, 2007), la temperatura no es considerada un parámetro químico de calidad del agua, si no un factor físico. Es uno de los parámetros que limitan la gran variedad de los procesos biológicos, desde la velocidad de una simple reacción química hasta la distribución ecológica de la especie animal.

Peces y camarones son animales pecilotermos² y altamente termófilos

²animales pecilotermos son aquellos que su temperatura corporal dependen de la temperatura del ambiente

(dependientes y sensibles a los cambios de la temperatura)(Nicovita, 2002), al contrario de los mamíferos y aves, la temperatura de su sangre no está internamente regulada. Como consecuencia, la temperatura ambiental tiene un profundo efecto sobre el crecimiento, tasa de alimentación y el metabolismo de dichos animales.

Para Hardy (Hardy, 1981), los animales pecilotermos se encuentran subordinados a su ambiente, ya que su actividad y supervivencia están permanentemente dependientes a la temperatura de su hábitat. El rango óptimo de la temperatura para el cultivo de tilapias es de 28 a 32°C, admitiendo una tolerancia de hasta 5°C por debajo de este rango óptimo. Morales (Morles, 1996), afirma que cuanto mayor sea la temperatura, mayor será la velocidad del crecimiento de los animales cultivados, siempre que las demás variables se conserven óptimas.

3. Nivel de pH

El pH tiene un profundo efecto sobre el metabolismo y los procesos fisiológicos de todos los organismos acuáticos, además de influenciar a muchos procesos químicos, por ejemplo, en la disponibilidad de los nutrientes que están directamente relacionados con la productividad primaria.

El término pH se refiere a la concentración de iones de hidrógeno (H⁺) en el agua, indicando su grado de acidez o alcalinidad. Alves y Mello (Alves y Mello, 2007) consideran que el nivel ideal de pH para el cultivo de tilapias es de 7 a 9. Valores de pH fuera de este rango causan cambios en el comportamiento de los peces, como letargia, inapetencia, retraso del crecimiento y en la reproducción.

Según Nicovita (Nicovita, 2002) valores de pH cercanos a 5 producen mortalidad en un período de 3 a 5 horas por fallas respiratorias; además, causan pérdidas de pigmentación e incremento en la secreción de la piel. A niveles de pH ácidos, el ion Fe⁺⁺ se vuelve soluble afectando las células de los arcos branquiales, disminuyendo los procesos de respiración, causando la muerte por asfixia por falta de oxígeno.

4. Transparencia

Los viveros con aguas muy transparentes tienen poco fitoplancton³, lo que indica una limitada cantidad de alimento natural para la tilapia. Así que, este parámetro constituye un vector importante para el control de la alimentación de las tilapias en el cautiverio. Para una buena producción,

³Producción de materia orgánica a partir de sales inorgánicas por microalgas a través del proceso de fotosíntesis.

considerase un nivel ideal de transparencia del agua de 30 a 50 cm. La medición de este parámetro se realiza a través del disco de secchi, que consiste en meter el aparato en el agua hasta el último nivel del agua en que el disco se puede observar, y se registra el nivel registrado.

6.4.2. Parámetros Biométricos

En acuicultura, biometría se refiere al proceso de toma de medidas de los organismos acuáticos en cultivo para observar sus características físicas y su estado de salud. Por lo tanto, las variables usadas en este proceso se denominan parámetros biométricos. El peso, la longitud standard y longitud total, son los parámetros biométricos más usados en el control de crecimiento de la tilapia. Por motivos comerciales y asociados a los objetivos de la empresa Aquapesca, en este trabajo fueron considerados dos parámetros biométricos: el peso y la longitud standard.

1. Peso

El peso ha sido registrado a través de balanza electrónica, con una precisión de 0.5g. La medición consistió en pesar una muestra aleatoria extraída en cada estanque de cría y estimar el peso medio (PM) de la muestra. Por lo que, un dato del peso en cada biométrica, corresponde al promedio del peso de la muestra de tilapias extraída en dicha biometría.

2. Longitud standard

La longitud standard de la tilapia se refiere a la distancia medida desde la punta de la boca hasta el pedúnculo (inicio de la aleta caudal). Tras pesar la muestra seleccionada en cada estanque, la longitud de cada individuo fue registrada mediante una regla graduada en milímetros. Los datos de la longitud fueron registrado para cada individuo.

6.4.3. Alternativas en la alimentación

Uno de los factores restrictivos en acuicultura en Mozambique es la falta de pienso suficiente para garantizar una buena alimentación a las especies en cultivo. Asociado a esto está el factor económico y el tiempo que tarda para obtener dicha alimentación, ya que hay que recurrir a otros países como Sudáfrica para comprar este producto. Una alternativa para superar este inconveniente sería aplicar el método de fertilización de los estanques, como forma de activar el alimento natural (plantón) suficiente para alimentar la tilapia. En palabras

de Lagos et al. (Lagos *et al.*, 2000):

”La tilapia es un pez que aprovecha eficientemente una variedad de alimentos naturales. El pez produce proteína al consumir el fitoplancton presente en los estanques fertilizados con abonos orgánicos de muy bajo costo. El alto costo de los alimentos concentrados es un limitante para lograr una producción rentable en los cultivos acuícola intensivos”.

El informe de la FAO de 2016 (FAO, 2016), muestra que la tilapia del Nilo puede utilizar de manera eficiente los alimentos naturales y se pueden lograr rendimientos de 3000 kg por hectárea en estanques bien fertilizados sin ningún alimento suplementario. Esta estrategia alimenticia depende de la aplicación de fertilizantes inorgánicos y/u orgánicos para estimular la producción de organismos alimenticios vivos y plantas en el sistema de cultivo.

Tacon (FAO, 2016) mostró que tanto los fertilizantes inorgánicos como los orgánicos actúan estimulando directamente la producción de fitoplancton dentro del estanque. Los fertilizantes inorgánicos actúan a través de la productividad primaria del estanque, mientras que el abono orgánico puede, además, estimular directamente los niveles tróficos al proveer materia orgánica y detritos. Para Tacon los abonos orgánicos son especialmente adecuados para la cría de la tilapia, pues, más allá de su valor como fertilizantes, representan una fuente inmediata de alimento, ya que esta especie puede alimentarse de desechos y subproductos de plantas.

Por lo tanto, con vista a analizar el efecto de fertilizantes (orgánico e inorgánico) en el cultivo de *Oreochromis mossambicus*, se llevó a cabo un proceso de cría de tilapia en la empresa Aquapesca de Mozambique, donde se consideraron cuatro métodos diferentes de alimentación.

En cuatro estanques diferentes, en cada uno se aplicó un método de alimentación. Los estanques designados por la empresa como C1, C2, C6 y C7 corresponden a los viveros donde se aplicaron los métodos de alimentación con pienso, control, fertilizante inorgánico y fertilizante orgánico, respectivamente. Se denominó por tipo de tratamiento al factor que recoge los datos de estos cuatro métodos alimenticios usados.

6.5. Estructura y toma de datos

La toma de datos en acuicultura se basa fundamentalmente en la medición de los parámetros ambientales y biométricos a lo largo del tiempo. En este

trabajo ha sido necesario definir unas pautas de experimentación para garantizar que dichas mediciones se realicen en condiciones idénticas en todos los tratamientos.

La experimentación es una práctica muy importante para la investigación y desarrollo en muchos campos de la ciencia. Una buena planificación y ejecución, proporciona información de calidad que permita comprender mejor el sistema y tomar decisiones de cómo optimizarlo (Ferré y Rius, 2002).

En todo el proceso de cultivo, el trabajo se orienta en analizar aspectos asociados a la variabilidad del ambiente acuático en los cuatro tratamientos (C1, C2, C6 y C7) y su influencia en los machos de la tilapia de Mozambique (*Oreochromis mossambicus*), por lo tanto, la **tilapia** y su respectivo **ambiente acuático** corresponden las unidades experimentales. La calidad del agua y el crecimiento de la tilapia fueron evaluados a través de los respectivos parámetros físico-químicos y biométricos.

Los parámetros biométricos fueron registrados semanalmente, orientados por un programa de control del crecimiento de las tilapias elaborado por la empresa. El valor del peso corresponde al promedio estimado para cada muestra aleatoria extraída en cada biometría; y la longitud standard de la tilapia fue medida en cada individuo de dicha muestra.

Por otro lado, los parámetros físico-químicos: oxígeno, temperatura y pH fueron medidos dos veces al día (6h y 15h) y la transparencia a las 12h.

Los intervalos de tiempo considerados en la medición para cada grupo de parámetros se fundamentan a los criterios técnicos de acuicultura, que busca captar la variabilidad diaria del ambiente acuático en los estanques, y estimar la tasa de crecimiento semanal de la especie.

La correspondencia entre los datos (cada dato del peso corresponde a un dato de la longitud y de cada parámetros ambiental) fue lograda considerado lo siguiente:

- (1) Para cada muestra extraída, el peso medio fue estimado. Para corresponder a este dato, un promedio de las longitudes fue calculado a partir de las longitudes medidas a cada individuo de la muestra, formando en cada biometría un par (PM, LM) para cada tratamiento.
- (2) A partir de los datos diarios de los parámetros físico-químicos fueron calculados promedios semanales, lo cual corresponde a un indicador de la variabilidad del ambiente acuático en dicha semana. Por lo tanto, la

base de datos es constituida por observaciones semanales (Y,X) donde Y representa los parámetros biométricos (variables del producto) y X los parámetros ambientales (variables del proceso).

Por lo tanto, la base de datos es constituida por siete (07) variables (dos parámetros biométricos, cuatro parámetros ambientales y el factor tratamiento con cuatro niveles) y veintiuna (21) observaciones para cada tratamiento (fueron realizadas veintiuna biometrías), formando una matriz con 84 líneas y 7 columnas.

6.5.1. Muestreo

6.5.1.1. Métodos de asignación de las tilapias en los estanques

Según Gomés (Gómez, 2005), la cría de la tilapia para su manejo se clasifica en *pre-engorde* y *engorde*. El peso de los peces de la etapa de pre-engorde (juveniles) varían de 10 a 100g de peso. La etapa de engorde se refiere a los peces con 100g de peso hasta la cosecha.

El presente estudio empezó con asignación de tilapias de 100g en los estanques de cultivo, que de acuerdo con Gomés, se refiere a la etapa de engorde.

La asignación de las tilapias en los estanques se realizó mediante muestras de 20 individuos ($n = 20$), con una biomasa (peso total) de aproximadamente 2000g. Considerando los tamaños de los estanques, 30 muestras (cada una con 20 individuos) fueron asignadas en los estanques C1, C6 y C7 de dimensión $600m^2$. 13 muestras fueron asignadas en el estanque C2 (12 muestras con $n = 20$ individuos y 1 muestra con $n = 10$) de dimensión $250m^2$. Los individuos fueron asignados con un peso medio de 100g con una desviación típica de 10.71,6.93,7.75 y 7.15g en los estanques C1, C2, C6 y C7 respectivamente.

6.5.1.2. Métodos de selección de la muestra

En cada biometria se analizó la tasa de mortalidad y una actualización de la biomasa del estanque. Las tilapias extraídas y observadas en cada estanque y en cada biometria correspondían al menos un 10% de la biomasa actualizada. Es decir, en la primera biometría se midió el peso y la longitud de al menos 60 tilapias en C1, C6 y C7 y de 30 tilapias en C2. El tamaño de las muestras siguientes dependió de la biomasa actualizada, siendo esta calculada con base en los datos de la tasa de mortalidad registrada.

El procedimiento de extracción de las muestras se puede considerar aleatorio, ya que en el proceso de arrastro (el que se aplica para extraer las tilapias en los estanques) los individuos tienen la misma probabilidad de ser seleccionadas.

De hecho, considerando no significativa la tasa de mortalidad en cada estanque y el número de individuos en cada muestra seleccionada, se puede inferir que en cada biometría, en los tratamientos C1, C6 y C7 se podrían determinar aproximadamente $\binom{600}{60} \approx 2,8 * 10^{83}$ muestras diferentes, y $\binom{250}{30} \approx 5,3 * 10^{38}$ muestras diferentes en C2. Sea m_a una muestra de uno de los estanques C1, C6 o C7 y (m_b) una muestra de C2. La probabilidad de que m_a sea seleccionada en una determinada biometría es:

$$P(m_a) = \frac{1}{2,8 * 10^{83}}$$

y la probabilidad de que (m_b) sea seleccionada en una determinada biometría es:

$$P(m_b) = \frac{1}{5,3 * 10^{38}}$$

Resulta imposible garantizar que los mismos elementos (peces) sean evaluados en sucesivas extracciones.

6.6. Análisis e Interpretación de los datos

En este apartado se presentan los resultados obtenidos en la monitorización del ambiente acuático caracterizado por los parámetros físico-químicos, y la predicción del índice de crecimiento de la tilapia medidos a través de los parámetros biométricos. Como se comentó anteriormente, los datos fueron registrados en un proceso de cría de tilapia en la empresa Aquapesca, llevado a cabo en cuatro diferentes tratamientos (tipo de alimentación). Desde luego, resulta necesario un análisis sobre el efecto del factor tratamiento en el crecimiento de la tilapia. El contenido de este apartado fue organizado en:

1. Análisis del efecto del tipo de tratamiento en el crecimiento de la tilapia;
2. Monitorización del ambiente acuático mediante el gráfico MAAEWMA;
3. Predicción del índice de crecimiento de la tilapia mediante el modelo dinámico sobre estructuras latentes (MDEL).

La figura 6.2 ilustra la evolución de los parámetros biometricos a lo largo de las 21 semanas de la cría de la tilapia. Se observa en los cuatro tratamientos una evolución creciente del peso y de la longitud. Y como es natural, esto indica un crecimiento de las tilapias a lo largo del tiempo. Sin embargo, dicho crecimiento no es lineal, siendo que en algunos instantes los parámetros presentan valores inferiores respecto a los registrados en los instantes anteriores. Este aspecto puede estar asociado a la aleatoriedad usada en la selección de los individuos en el proceso de muestreo. Como se explicó anteriormente, los datos fueron extraídos de una muestra aleatoria de tamaño no inferior a 10 % de la

biomasa total. Por lo que puede ocurrir que en una biometría sean seleccionados individuos con peso relativamente menor que a los individuos seleccionados en la biometría anterior (a pesar de empezar el estudio con individuos del mismo peso medio, no hay garantías de que haya un crecimiento uniforme, ya que el desarrollo de cada individuo depende de su adaptación al ambiente acuático a que están expuestos y la facilidad de acceso a la alimentación).

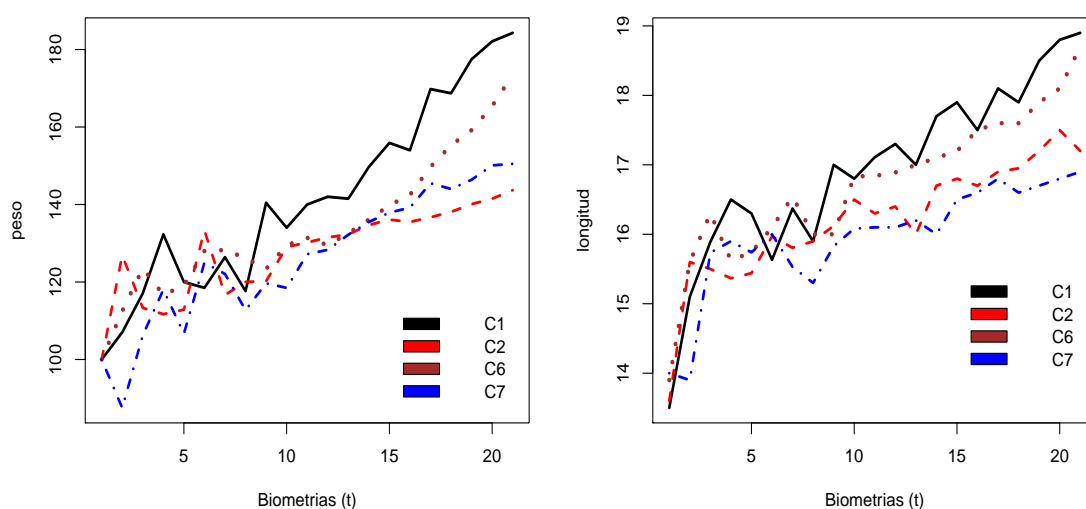


Figura 6.2: parámetros biometricos

6.6.1. Análisis del efecto del tipo de tratamiento en el crecimiento de la tilapia

El análisis del efecto del tratamiento al crecimiento de la tilapia fue realizado mediante un análisis de la varianza (ANOVA) a un nivel de significación de 5 %, donde se evidenciaron diferencias significativas del peso ($F\text{-Fisher}=3.071$, $p\text{-valor}=0.0324$) y de la longitud ($F\text{-Fisher}=3.958$, $p\text{-valor}=0.011$). Las mayores diferencias en los dos parámetros fueron observadas entre los tratamientos C1 (alimentación con pienso) y C7 (fertilizante orgánico).

Los intervalos HSD presentados en la figura 6.3 muestran que las diferencias del peso y de la longitud media entre C7 y C1 son inferiores a cero, indicando que en C1 hubo mayor crecimiento de las tilapias relativamente a C7. En todos los casos, el C1 se presenta mejor que los demás tratamientos, siendo que las menores diferencias se obtiene al compararlo con C6.

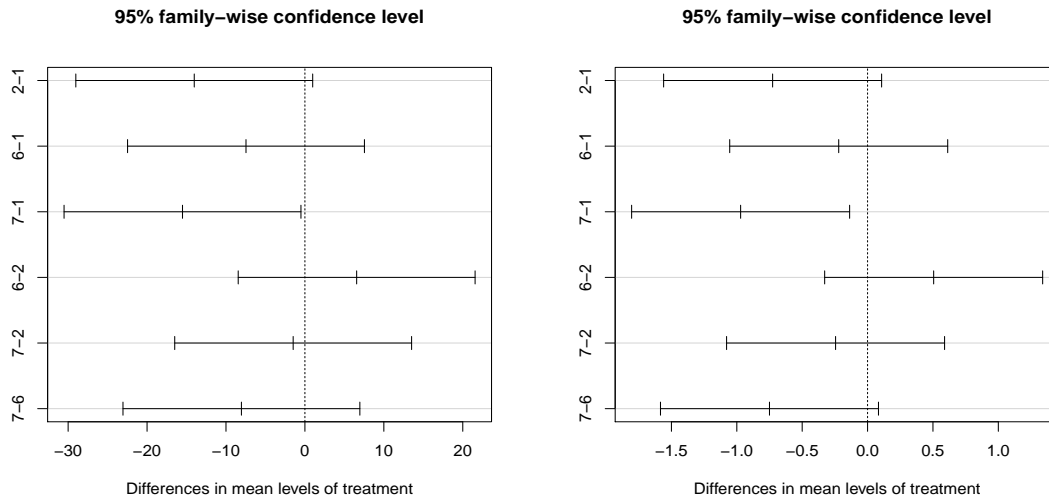


Figura 6.3: Comparación del peso (a la izquierda) y de la longitud (a la derecha) en los cuatro tratamientos (i-j, marca la comparación entre tratamientos C_i y C_j)

Estos resultados indican que la alimentación con el pienso ha resultado más eficaz que otros tratamientos utilizados. No obstante, la mejora de dicho tratamiento respecto al tratamiento C6 es muy pequeña, indicando que el fertilizante inorgánico puede ser una alternativa alimenticia en el cultivo de *Oreochromis mossambicus* considerado, por criterios económicos.

Las diferencias significativas del crecimiento de las tilapias se asocian a los diferentes métodos de tratamiento alimenticio usado en cada estanque. Sin embargo, el ambiente acuático puede haber influenciado en dichas diferencias. Para comprobar que las diferencias del crecimiento de la tilapia fue causada por los tipos de tratamientos usados, sin influencia del ambiente acuático, un análisis de similitud del ambiente acuático fue llevado a cabo mediante un análisis discriminante lineal de Fisher. Para ello, los cuatro parámetros ambientales (temperatura, oxígeno, pH, y transparencia) fueron usados como variables independientes y el factor tratamiento con sus cuatro niveles, la variable dependiente (la categoría a ser discriminada).

Tres funciones discriminantes fueron determinadas (nr de funciones discriminantes = $\min(\text{nr de variables independiente}, \text{nr de categorías} - 1)$) con los estadísticos λ de Wilks y los respectivos p-valores no significativos (Tabla 6.2).

La figura 6.4 presenta los individuos clasificados en las seis combinaciones de pares de las funciones determinadas. En los seis casos, no hay una estructura clara en la que se puede agrupar los individuos de cada tratamiento, por lo que no hay evidencias para rechazar la hipótesis de que los cuatro tratamientos fue-

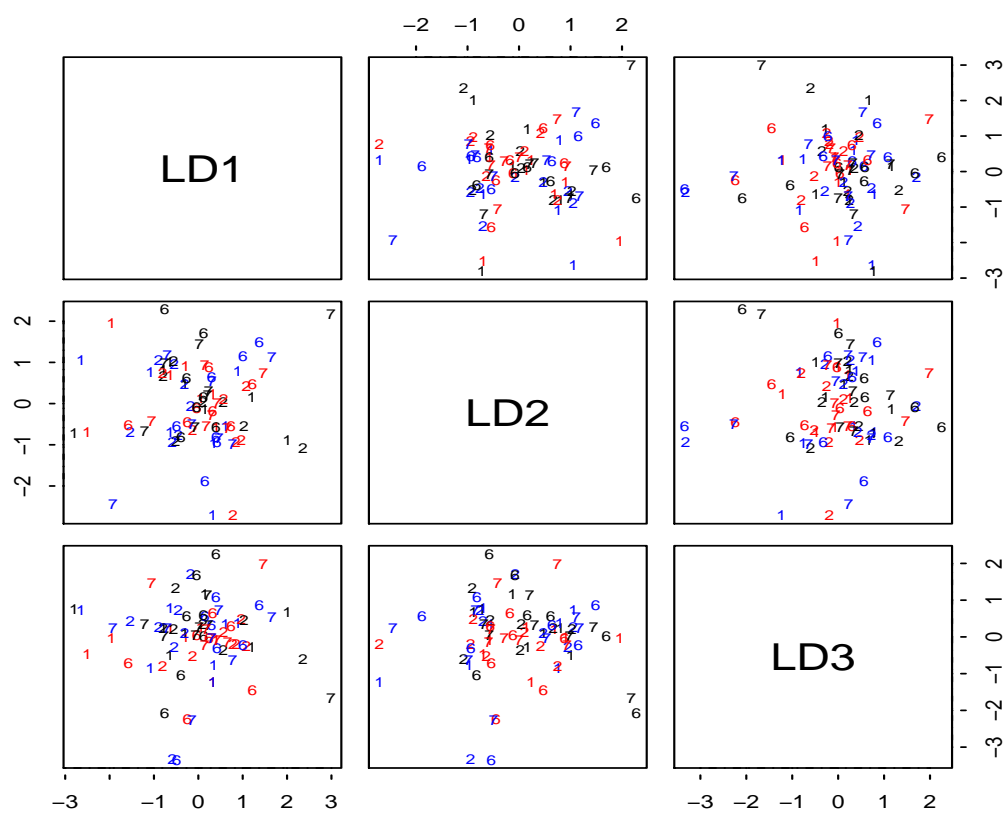


Figura 6.4: Clasificación de los individuos de los cuatro tratamientos

LD	λ de Wilks	χ^2	g.l	sig
1	0,815	5,815	9	0,758
2	0,918	2,445	4	0,654
3	0,979	0,604	1	0,437

Tabla 6.2: significación de la funciones discriminante

ron implementados en cultivos con ambientes similares. Por tanto, la diferencia del rendimiento entre ellos se atribuye solamente al efecto de cada método alimenticio usado, confirmando la no significación de las funciones discriminantes.

6.6.2. Monitorización del ambiente acuático en Acuicultura

6.6.2.1. Enfoque Univariante

Como comentamos en el apartado 6.1, la monitorización del ambiente acuático se lleva a cabo mediante gráficos de control de la estabilidad y conformidad. Para ilustrar las diferencias en los resultados entre el método univariante y el multivariante, empezamos por presentar los resultados de monitorización de la estabilidad y conformidad mediante gráficos de control univariantes, y en el apartado siguiente se presentan los resultados del gráfico MAAEWMA aplicado a los mismos datos.

Los límites de control de estabilidad (LCE) fueron calculados usando el método 3-sigma del gráfico Xbar de Shewart (Shewhart, 1931), para lograr un ARL bajo control de 370 (o sea, $\alpha = 0,0027$), como se muestra a continuación:

$$\begin{aligned} \text{LCE}_l &= \bar{p} - 3\sigma \\ \text{LCE}_u &= \bar{p} + 3\sigma \end{aligned} \tag{6.1}$$

donde \bar{p} es la media general (promedio de las medias semanales) del parámetro ambiental en estudio, con su respectiva desviación típica σ .

Los límites de control de conformidad (LCC) corresponden a los límites calculados a través de los límites técnicos de tolerancia mencionados en el apartado 6.4.1 en los cuales los parámetros ambientales deben mantenerse para garantizar un buen cultivo de la tilapia. Usando la formula de Mohammadian y Amiri (Mohammadian y Amiri, 2012), para un $\alpha = 0,0027$, los LCC fueron calculados como se muestra en la expresión 6.2:

6.6. Monitorización del ambiente acuático mediante el gráfico MAAEWMA11

Tabla 6.3: Límites de control para los parámetros ambientales

Parámetro	LTT		LCE(C1,C2,C6,C7)							
	LTT _l	LTT _u	LCE _l				LCE _u			
Temperatura	23	32	18.3	20.0	17.8	17.5	36.1	36.0	36.2	36.6
Oxígeno	5	9	1.0	2.0	3.0	3.0	8.9	8.7	7.6	7.4
pH	7	9	6.8	5.5	6.0	6.0	8.2	9.8	9.7	9.7
Transparencia	30	50	20.3	17.2	11.4	20.7	38.6	42.5	47.5	38.1

$$\begin{aligned}
 LCC_l &= LTT_l - \left(z_p - \frac{3}{n}\right)\sigma \\
 LCC_u &= LTT_u + \left(z_p - \frac{3}{n}\right)\sigma
 \end{aligned}
 \tag{6.2}$$

Donde:

LCC_l y LCC_s son los límites de conformidad inferior y superior, respectivamente;

LTT_l y LTT_s son los límites de tolerancia técnicos usados en acuicultura (ver en la tabla 6.3);

z_p el valor de la normal tipificada correspondiente al parámetro ambiental p en análisis. Dependiendo del tamaño de la muestra podría sustituirse por una $t_{n-1,p}$;

α nivel de significación;

n : número de observaciones tomadas en cada instante (semana);

σ desviación típica de las observaciones respecto al promedio semanal calculado.

La tabla 6.3 recoge los valores de los límites de tolerancia usados en el cálculo de los límites de control de conformidad, y los límites de estabilidad calculados. Los límites de estabilidad inferior y superior están presentados en orden de los tratamientos C1,C2,C6 y C7.

Las figuras 6.5-6.8 muestran la variabilidad semanal de los cuatro parámetros ambientales registrados en los cuatro tratamientos considerados. Como

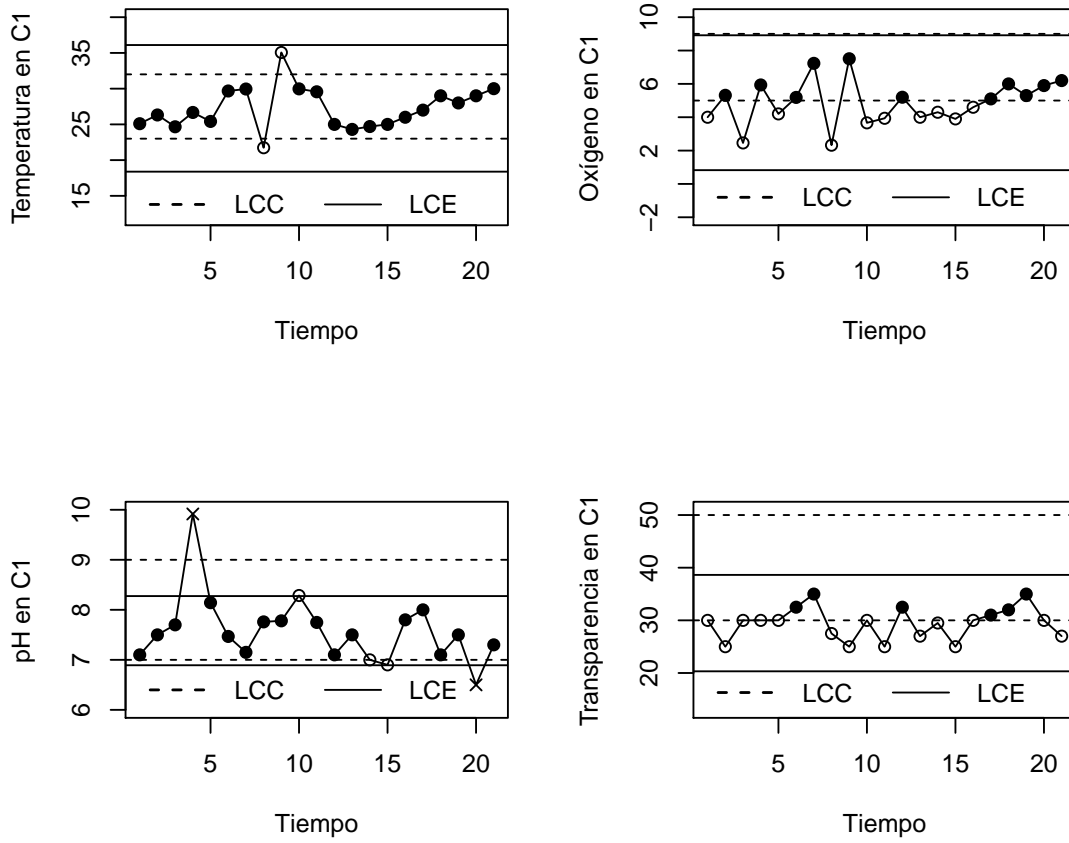


Figura 6.5: Variabilidad de los parámetros ambientales en el tratamiento C1

6.6. Monitorización del ambiente acuático mediante el gráfico MAAEWMA13

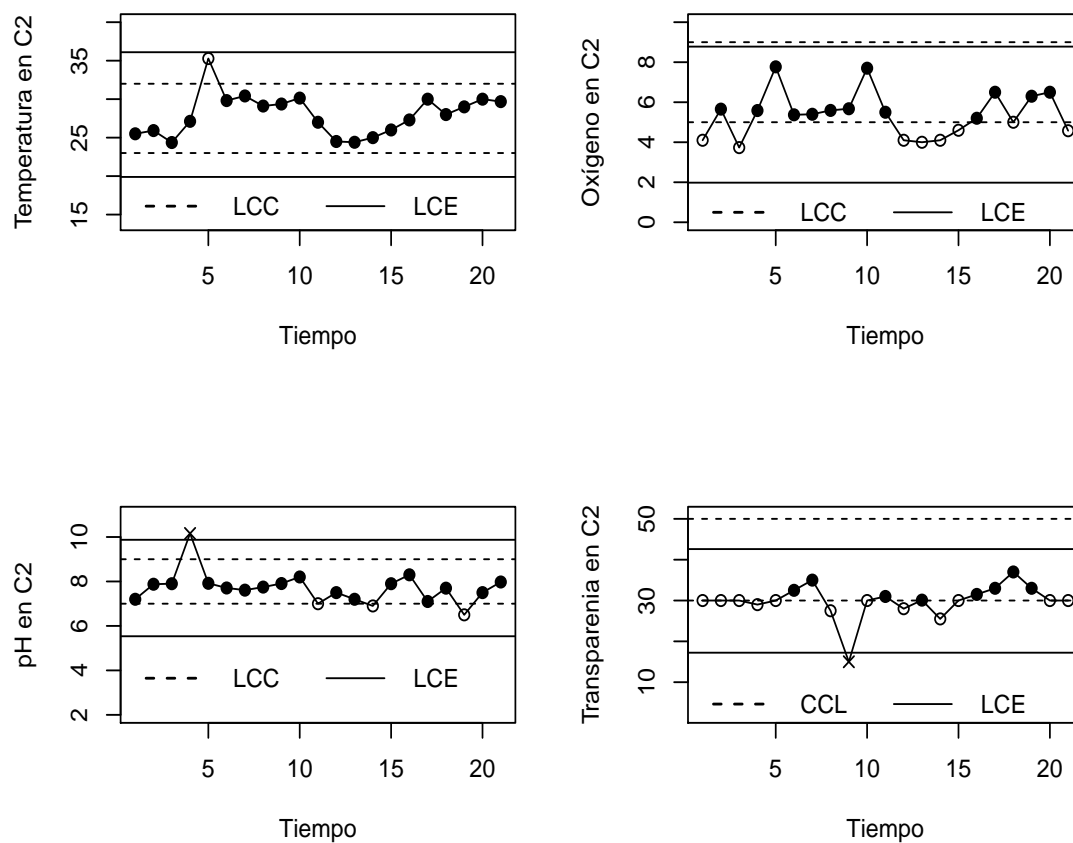


Figura 6.6: Variabilidad de los parámetros ambientales en el tratamiento C2

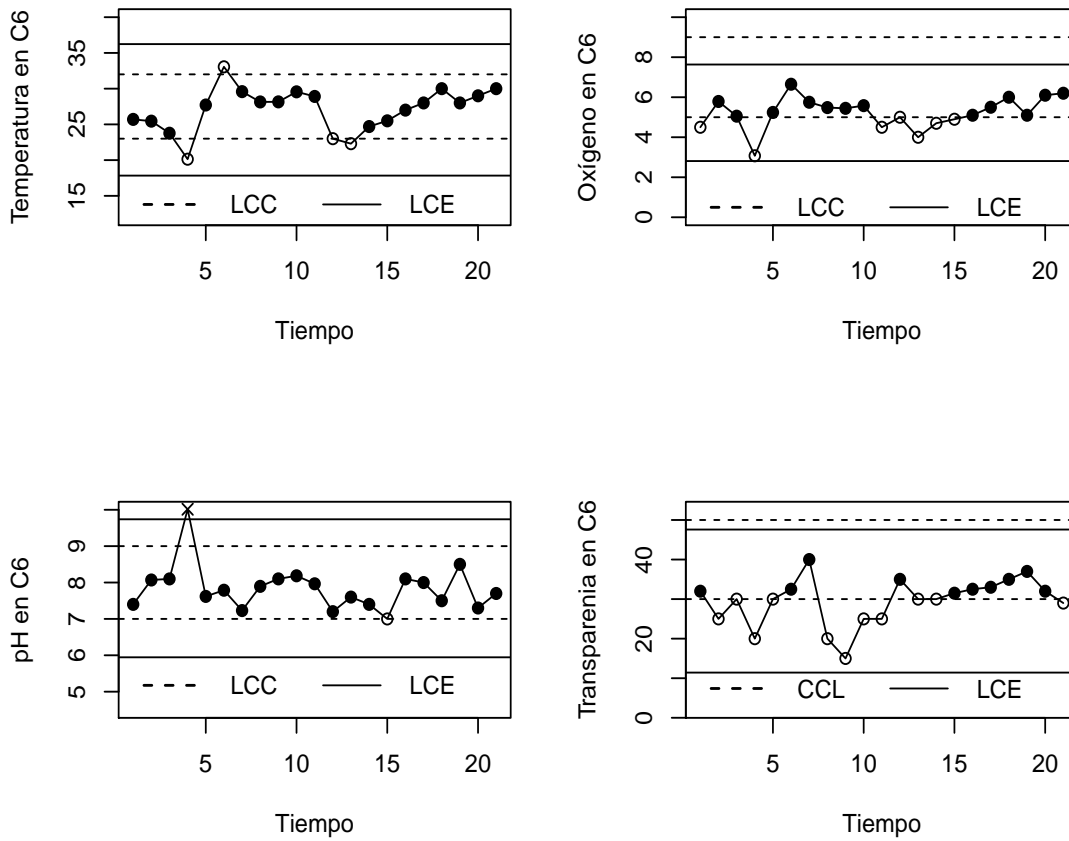


Figura 6.7: Variabilidad de los parámetros ambientales en el tratamiento C6

6.6. Monitorización del ambiente acuático mediante el gráfico MAAEWMA15

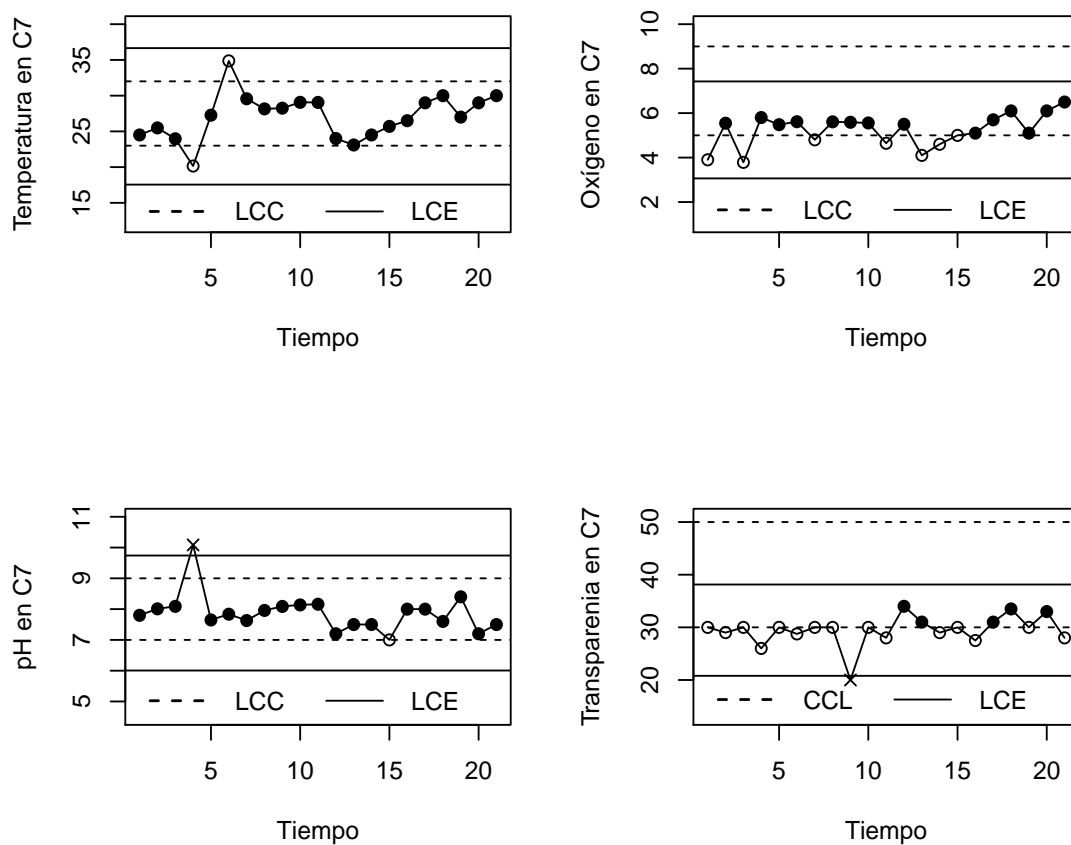


Figura 6.8: Variabilidad de los parámetros ambientales en el tratamiento C7

explicamos en el apartado 6.5, los parámetros: oxígeno, temperatura y pH fueron registrados dos veces al día (a las 6 y 15 horas) y la transparencia, una vez al día (a las 12 horas), por tanto, los puntos en los gráficos corresponden a los promedios semanales de cada uno de dichos parámetros. Es decir: los gráficos de oxígeno, temperatura y pH representan los promedios de 14 observaciones, y los gráficos de la transparencia representan los promedios de 7 observaciones. En cada uno de los promedios semanales, fueron calculadas las correspondientes desviaciones típicas con las cuales los límites de control de conformidad (LCC) fueron calculados usando la expresión 6.2.

En dichas figuras, las líneas horizontales continuas representan los límites de estabilidad, y las líneas interrumpidas corresponden a los límites de conformidad. Los símbolos (●), (○) y (×) están asociados a las observaciones que cumplen ambos requisitos de conformidad y estabilidad, las que fallan uno de los requisitos, y las que fallan ambos los requisitos, respectivamente.

(1) Cuanto a la estabilidad de los parámetros:

Los gráficos indican que la temperatura estuvo dentro de los límites de control en todo el proceso de cultivo, en todos los tratamientos. Sin embargo, en la cuarta semana hubo un cambio significativo del oxígeno (en el tratamiento C6) y del pH (en todos los tratamientos); asimismo, el nivel del pH ha vuelto a registrar un cambio significativo en la semana 10 en el tratamiento C1. La transparencia del agua registró bajos niveles en la octava semana en los tratamientos C2 y C7.

(2) Respecto a la conformidad:

El oxígeno y la transparencia fueron los parámetros menos conformes con los requisitos técnicos de acuicultura. A pesar de su aparente estabilidad, estos parámetros estuvieron en la mayor parte del proceso abajo de sus respectivos niveles recomendados, con lo cual los gráficos indican que el proceso de cría de la tilapia fue realizado en un ambiente con déficit de oxígeno y con aguas turbias.

Las semanas 4 y 9 son consideradas las más críticas para los niveles del pH y de la transparencia. En todos los tratamientos el pH varía significativamente en la cuarta semana y presenta un nivel muy alto respecto a lo recomendado para la cría de la tilapia. El bajo nivel de la transparencia registrado en la semana 9 tuvo mayor evidencia en los tratamientos C2 y C7.

De modo sintético, el estado de cada parámetro, en los cuatro tratamientos, se presenta en la tabla 6.4, donde se observa que en la mayor parte del proceso los parámetros no estaban conformes con los requisitos establecidos

6.6. Monitorización del ambiente acuático mediante el gráfico MAAEWMA

en la cría de la tilapia. Además, el pH y la transparencia registraron cambios significativos en las semanas 4, 9 y 10.

Si se considera que el proceso está bajo control si y solamente si todos sus parámetros ambientales están conformes y estables, el proceso estaría bajo control en las observaciones presentadas en la celda OBC*Total de la tabla 6.4, para cada tratamiento. Por tanto, siguiendo esta metodología, en los tratamientos C1 y C6 el proceso estuvo bajo control solamente en seis semanas, y en C2 y C7 en cinco semanas.

6.6.2.2. Enfoque multivariante: Aplicación del gráfico MAAEMA

El análisis univariante anterior se torna muy complejo cuando hay que considerar muchas variables en el estudio, ya que implicaría analizar la conformidad y estabilidad de cada variables y a partir de ello decidir el estado global del proceso. La probabilidad de falsas alarmas del proceso es tan grande cuanto mayor es el número de gráficos univariantes construidos. Es decir: sea α la probabilidad de falsas alarmas del gráfico de control de cada variable p_i de un proceso con m variables ($i = 1, \dots, m$). La probabilidad de falsas alarmas de los m gráficos de control independientes construidos para cada variable es: $1 - (1 - \alpha)^m$.

Este trabajo se orienta a los métodos multivariantes, donde además del efecto de cada variable, buscase modelar la estructura de correlación entre ellas. Defendemos la idea de que a un proceso multivariante se le debe aplicar un método multivarainte correspondiente, de cara a explotar la relevante información presentada en la estructura compleja de los datos correspondientes. Para ello, el gráfico MAAEWMA desarrollado en el capítulo 5 fue aplicado para monitorizar en simultaneo los cuatro parámetros ambientales considerados en cada tratamiento.

El ajuste del gráfico MAAEWMA fue realizado siguiendo los siguientes pasos:

1. Tipificar cada parámetro ambiental

Los parámetros ambientales fueron tipificados para eliminar el efecto de escala de medición de cada variable en el gráfico, ya que los parámetros ambientales tienen medidas diferentes. Este procedimiento no afecta a la estructura de autocorrelación en las variables, y a las variables tipificadas fueron ajustados modelos AR(1) (identificados a través de sus funciones de correlación simple y parciales (figuras A.3-A.6));

Tabla 6.4: Resumen del estado del proceso

C1				
	Parámetro	Conformidad	Estabilidad	Total
	Temperatura	8,9	—	8,9
OFC	Oxígeno	1,3,5,8,10,11,13-16	—	1,3,5,8,10,11,13-16
	pH	4,14,15	4,10	4,10,14,15
	Transparencia	1-5,8-11,13-16,20-21	—	1-5,8-11,13-16,20-21
	Total	1-5,8-11,13-16,20-21	4,10	1-5,8-11,13-16,20-21
OBC		6,7,12,17,18,19	1-3,5-9,11-21	6,7,12,17,18,19
C2				
	Parámetro	Conformidad	Estabilidad	Total
	Temperatura	5	—	5
OFC	Oxígeno	1,3,12-15,18,21	—	1,3,12-15,18,21
	pH	4,11,14,19	4	4,11,14,19
	Transparencia	1-5,8-10,12,14,15,20,21	—	1-5,8-10,12,14,15,20,21
	Total	1-5,8-12,14,15,18-21	4	1-5,8-12,14,15,18,20,21
OBC		6,7,13,16,17	Todas-4	6,7,13,16,17
C6				
	Parámetro	Conformidad	Estabilidad	Total
	Temperatura	4,6,12,13	—	4,6,12,13
OFC	Oxígeno	1,4,11-15	—	1,4,11-15
	pH	4,15	4	4,15
	Transparencia	2-5,8-11,13,14,21	—	2-5,8-11,13,14,21
	Total	1-6,8-15,21	4	1-6,8-15,21
OBC		7,16-20	Todas-4	7,16-20
C7				
	Parámetro	Conformidad	Estabilidad	Total
	Temperatura	4,6	—	4,6
OFC	Oxígeno	1,4,11-15	—	1,4,11-15
	pH	4,15	4	4,15
	Transparencia	1-11,14-16,19,21	9	1-11,14-16,19,21
	Total	1-11,14-16,19,21	9	1-11,14-16,19,21
OBC		12,13,17,18,20	Todas-(4,9)	12,13,17,18,20

OFC=Observación fuera de control

OBC=Observación bajo de control

6.6. Monitorización del ambiente acuático mediante el gráfico MAAEWMA

2. Calcular los parámetros del modelo (vector de medias y la matriz de varianza-covarianza cruzada)

Las medias de cada parámetro fueron calculadas a través de la formula 5.3 (como las variables fueron tipificadas, sus medias son nulas), y las formulas 5.5 y 5.7 fueron usadas para calcular las varianzas y covarianzas cruzadas (incluyendo en el modelo las correlaciones cruzadas significativas (figura A.7 a figura A.12));

3. Ajustar el gráfico MAAEWMA

Los estadísticos z_i y T^2 fueron calculados, y ambos limites de estabilidad ($LCE=h_3$ calculado a través del método de la cadena de Markov, aplicado como presentamos en el capitulo 5) y de conformidad (LCC calculado a través de la formula 6.3) fueron considerados.

De modo ilustrativo, se presentan a continuación los parámetros calculados y los principales resultados obtenidos:

- Tratamiento C1:

$$\vec{m}_{c_1} = (0, 0, 0, 0)$$

$$\Sigma_{\hat{\gamma}_{c_1}} = \begin{pmatrix} 0,816 & 0,125 & -0,055 & 0,010 \\ 0,125 & 1,084 & -0,094 & 0,073 \\ -0,055 & -0,094 & 0,736 & -0,179 \\ 0,010 & 0,073 & -0,179 & 0,626 \end{pmatrix}$$

- Tratamiento C2:

$$\vec{m}_{c_1} = (0, 0, 0, 0)$$

$$\Sigma_{\hat{\gamma}_{c_2}} = \begin{pmatrix} 0,828 & 0,335 & 0,150 & 0,093 \\ 0,335 & 1,132 & 0,004 & -0,031 \\ 0,150 & 0,004 & 0,714 & -0,070 \\ 0,093 & -0,031 & -0,070 & 0,909 \end{pmatrix}$$

- Tratamiento C6:

$$\vec{m}_{c_1} = (0, 0, 0, 0)$$

$$\Sigma_{\hat{\gamma}_{c_6}} = \begin{pmatrix} 0,766 & 0,347 & 0,160 & 0,183 \\ 0,347 & 1,421 & -0,006 & -0,103 \\ 0,160 & -0,006 & 0,677 & -0,204 \\ 0,183 & -0,103 & -0,204 & 0,872 \end{pmatrix}$$

- Tratamiento C_7 :

$$\vec{m}_{c_1} = (0, 0, 0, 0)$$

$$\Sigma_{\hat{\gamma}_{c_7}} = \begin{pmatrix} 0,818 & 0,241 & 0,051 & 0,021 \\ 0,241 & 0,769 & 0,030 & 0,045 \\ 0,051 & 0,030 & 0,666 & -0,186 \\ 0,021 & 0,045 & -0,186 & 0,524 \end{pmatrix}$$

El limite de estabilidad $h_3 = 247,339$ fue calculado para un $\lambda = 0,1$ y $ARL = 370$ para los cuatro tratamientos.

Los limites de control de conformidad LCC fueron calculado como se muestra a continuación:

$$\begin{aligned} LCC_l &= z'_l \Sigma_{\gamma} z_l \\ LCC_u &= z'_u \Sigma_{\gamma} z_u \end{aligned} \quad (6.3)$$

donde:

l es el vector que recoge los limites de conformidad inferiores de los parámetros ambientales, y u corresponde al vector que recoge los limites de conformidad superiores;

z_l y z_u son los respectivos vectores MAAEWMA de l y u ;

Σ_{γ} es la matriz de varianza-covarianza cruzada usada en el gráfico MAAEWMA.

Estos limites de conformidad calculados mediante la expresión 6.3 corresponden a la tolerancia mínima y máxima del ambiente acuático adaptada al entorno multivariante. En este caso LCC_l y LCC_u corresponden a los valores de T^2 de los limites de conformidad inferior y superior, respectivamente. Por tanto, la conformidad del proceso se verifica si $LCC_l < T^2 < LCC_u$.

6.6. Monitorización del ambiente acuático mediante el gráfico MAAEWMA

Tabla 6.5: Resultados Numéricos del proceso de monitorización mediante gráfico MEWMA en C1

Biometria	MAAEWMA vector (Tratamiento C1)				MAAEWMA statistics
	z_1	z_2	z_3	z_4	T^2
1	-0.798	-0.709	-0.768	0.188	14.579 [×]
2	0.034	1.248	0.395	-3.089	121.469
3	-1.119	-3.079	0.921	0.349	73.617
4	0.261	2.208	7.255	0.331	593.521*
5	-0.633	-0.511	1.738	0.315	36.787
6	2.252	0.957	-0.305	1.939	84.486
7	2.295	3.918	-1.211	3.483	237.638
8	-3.406	-3.579	0.608	-1.599	176.362
9	5.773	4.279	0.634	-3.142	480.985*
10	2.014	-1.655	2.063	0.301	106.707
11	1.628	-1.174	0.401	-2.988	129.275
12	-1.546	0.746	-1.489	2.079	80.142
13	-1.941	-1.083	-0.253	-1.634	69.489
14	-1.576	-0.594	-1.686	0.086	53.495
15	-1.301	-1.171	-1.891	-2.870	185.347
16	-0.570	-0.087	0.802	0.546	14.485 [×]
17	0.121	0.642	1.331	1.164	45.323
18	1.455	1.930	-1.344	1.753	69.144
19	0.692	0.783	-0.122	3.627	155.797
20	1.325	1.626	-3.014	0.168	105.731
21	1.929	1.985	-0.556	-1.797	96.923

*=fuera del limite de estabilidad,

[×]=fuera de los limites de conformidad

Tabla 6.6: Resultados Numéricos del proceso de monitorización mediante gráfico MEWMA en C2

Biometria	MAAEWMA vector (Tratamiento C2)				MAAEWMA statistics
	z_1	z_2	z_3	z_4	T^2
1	-1.016	-1.241	-0.768	0.023	17.105 [×]
2	-0.697	1.514	1.110	0.023	45.003
3	-1.853	-2.004	1.087	0.023	57.803
4	0.236	1.320	7.252	-0.449	538.726*
5	6.222	5.060	0.669	0.0464	364.169*
6	1.824	0.533	0.051	1.225	35.590
7	2.153	0.529	-0.224	2.346	73.437
8	1.059	0.816	0.167	-1.313	25.437
9	1.185	0.900	0.596	-7.149	413.337*
10	1.690	4.413	1.384	0.305	136.852
11	-0.724	0.291	-2.014	0.749	43.577
12	-2.539	-2.197	-0.527	-0.718	64.497
13	-2.489	-2.267	-1.333	0.301	71.610
14	-1.929	-1.985	-2.099	-1.899	104.173
15	-1.106	-1.017	0.765	0.314	23.002 [×]
16	-0.107	0.074	1.822	0.994	46.312
17	1.876	2.345	-1.597	1.641	92.701
18	0.282	-0.437	0.138	3.441	90.645
19	0.994	1.866	-3.195	1.373	136.812
20	1.672	2.114	-0.276	-0.108	39.978
21	1.344	-1.392	1.040	-0.095	45.632

*=fuera del limite de estabilidad,
[×]=fuera de los limites de conformidad

6.6. Monitorización del ambiente acuático mediante el gráfico MAAEWMA

Tabla 6.7: Resultados Numéricos del proceso de monitorización mediante gráfico MEWMA en C6

Biometria	MAAEWMA vector (Tratamiento C6)				MAAEWMA statistics
	z_1	z_2	z_3	z_4	T^2
1	-0.477	-0.988	-0.767	0.456	10.950 [×]
2	-0.634	2.226	1.358	-1.870	71.034
3	-1.722	0.245	1.353	-0.099	78.977
4	-4.028	-4.723	7.315	-3.399	866.017*
5	1.095	0.871	-0.651	0.101	21.137
6	4.494	4.291	-0.082	0.929	214.499
7	1.974	1.801	-1.850	3.374	135.747
8	0.919	1.023	0.342	-3.425	127.448
9	0.872	0.860	0.966	-4.899	230.710
10	1.737	1.135	1.178	-1.331	52.791
11	1.225	-1.619	0.419	-1.273	72.956
12	-2.692	-0.303	-2.024	2.097	141.670
13	-3.011	-2.788	-0.658	0.320	97.186
14	-1.299	-0.920	-1.263	0.296	26.627 [×]
15	-0.723	-0.395	-2.471	0.772	62.075
16	0.277	0.101	1.121	1.060	32.103
17	0.900	1.070	0.733	1.170	29.283 [×]
18	2.146	2.242	-0.895	1.774	75.068
19	0.724	-0.124	2.302	2.352	138.686
20	1.335	2.360	-1.618	0.582	66.651
21	1.918	2.482	-0.277	-0.429	57.601

*=fuera del limite de estabilidad,

[×]=fuera de los limites de conformidad

Tabla 6.8: Resultados Numéricos del proceso de monitorización mediante gráfico MEWMA en C7

Biometria	MAAEWMA vector (Tratamiento C7)				MAAEWMA statistics
	z_1	z_2	z_3	z_4	T^2
1	-0.898	-2.033	-0.129	0.203	38.548
2	-0.269	2.512	0.546	-0.489	71.867
3	-1.249	-2.536	0.770	0.234	72.870
4	-3.620	3.0594	7.137	-2.538	819.970*
5	0.995	1.945	-1.050	0.363	48.343
6	5.673	2.152	-0.377	-0.518	286.420*
7	2.020	-0.251	-1.007	0.373	53.554
8	1.006	1.935	0.108	0.354	35.325 [×]
9	1.001	1.739	0.529	-6.582	655.357*
10	1.450	1.525	0.668	0.668	39.919
11	1.360	-1.092	0.714	-0.764	45.578
12	-1.890	1.285	-2.393	3.414	223.228
13	-2.363	-2.659	-1.298	1.153	110.978
14	-1.373	-1.175	-1.228	-0.292	38.492
15	-0.564	-0.046	-2.77	0.412	81.424
16	-0.049	0.200	0.577	-1.340	24.279 [×]
17	1.506	1.810	0.540	1.147	57.908
18	2.043	2.792	-0.778	2.816	171.853
19	0.046	-0.119	1.825	0.252	43.086
20	1.294	2.623	-2.127	2.321	148.352
21	1.852	3.579	-1.059	-1.248	181.642

*=fuera del limite de estabilidad,
[×]=fuera de los limites de conformidad

6.6. Monitorización del ambiente acuático mediante el gráfico MAAEWMA

Los límites de conformidad inferior (LCC_l) y superior (LCC_u) determinados para cada tratamiento se presentan en la tabla 6.9.

Las tablas 6.5-6.8 presentan los resultados numéricos del proceso de monitorización realizado. Los valores marcados por (*) corresponden a los valores de T^2 superiores a h_3 , es decir, las observaciones de las semanas donde el ambiente acuático no estuvo estable. Los valores marcados por (x) no cumplen la condición $LCC_l < T^2 < LCC_u$, y corresponden a las observaciones no conformes con los requisitos establecidos en la cría de tilapia.

Las figuras 6.9-6.12 ilustran los resultados de la monitorización del ambiente acuático mediante el gráfico MAAEWMA en los cuatro tratamientos, donde cada observación corresponde al valor del estadístico T^2 que representa el índice de la variabilidad global del ambiente acuático.

Además del efecto principal e interactivo entre los parámetros ambientales, el gráfico MAAEWMA modela la autocorrelación en las observaciones. Por tanto, cada observación del gráfico contiene información correspondiente a dichos tres aspectos, y se destaca lo siguiente:

1. El proceso puede estar fuera de control debido al cambio de una de las variables del proceso.

En este caso, el cambio señalado por el gráfico MAAEWMA también será identificado por el gráfico univariante correspondiente a dicha variable: Tal es el caso de las observaciones 4 y 9, donde hubo un cambio significativo del pH y de la transparencia y fueron identificados en los dos métodos;

2. El proceso puede estar fuera de control debido a un cambio causado por la correlación entre las variables.

En este caso, los gráficos univariantes no detectan el cambio: Tal es el caso de la observación 17, la cual todos los gráficos univariantes no señalaron cambio del proceso, no obstante, el gráfico MAAEWMA muestra que el ambiente no estuvo conforme en esta semana.

3. Falsas alarmas

Claro et. al (Claro *et al.*, 2007) demostraron que cuando no se modela la autocorrelación de una variables en un gráfico de control, dicha característica propia del proceso es señalada como un fallo del proceso, y por lo tanto, los gráficos diseñados bajo la hipótesis de independencia entre

Tabla 6.9: Límites de control y observaciones fuera de control

	Conformidad			Estabilidad	
	LCC _l	LCC _s	OFC	h ₃	OFC
C1	29.7	1379.5	1,16	247.339	4,9
C2	34.2	765.7	1,8,15	247.339	4,5,9
C6	30.0	912.1	1,5,14,17	247.339	4
C7	37.5	1718.9	8,16	247.339	4,6,9

OFC=Observación fuera de control

observaciones son susceptibles a falsas alarmas para procesos autocorrelacionados.

Corresponde a este caso, las observaciones señaladas fuera de control en los gráfico univariantes y que están dentro de los límites del gráfico MAAEWMA. Siempre asumiendo que el gráfico MAAEWMA dispone de más información del proceso respecto a la suma de los gráficos univariantes.

En la tabla 6.9 se presentan las observaciones fuera de control tanto por la conformidad como por la estabilidad, y se puede comentar lo siguiente:

(1) Cuanto a la estabilidad:

En todos los tratamientos, el ambiente acuático estuvo fuera de control en la cuarta semana. El cambio del proceso podría ser causado por el cambio significativo del nivel del pH en esta semana.

En la semana 9 hubo un cambio significativo del ambiente acuático en los tratamientos C1, C2 y C7. Comparando a los gráficos univariantes, ninguno de ellos señala esta observación en C1.

El cambio del ambiente acuático en C2 en la semana 5, puede estar relacionado con el efecto de la interacción entre la temperatura y oxígeno, y el cambio del ambiente en C7 en la semana 6 podría estar asociado a la interacción de la temperatura con otros parámetros.

De manera general, el gráfico MAAEWMA presentó mayor capacidad en detectar cambios de estabilidad del ambiente acuático (figuras 6.9 a 6.12) comparado a los gráficos univariantes (presentados en las figuras 6.5 a 6.8). Este resultado se atribuye a la modelación de la estructura de

6.6. Monitorización del ambiente acuático mediante el gráfico MAAEWMA

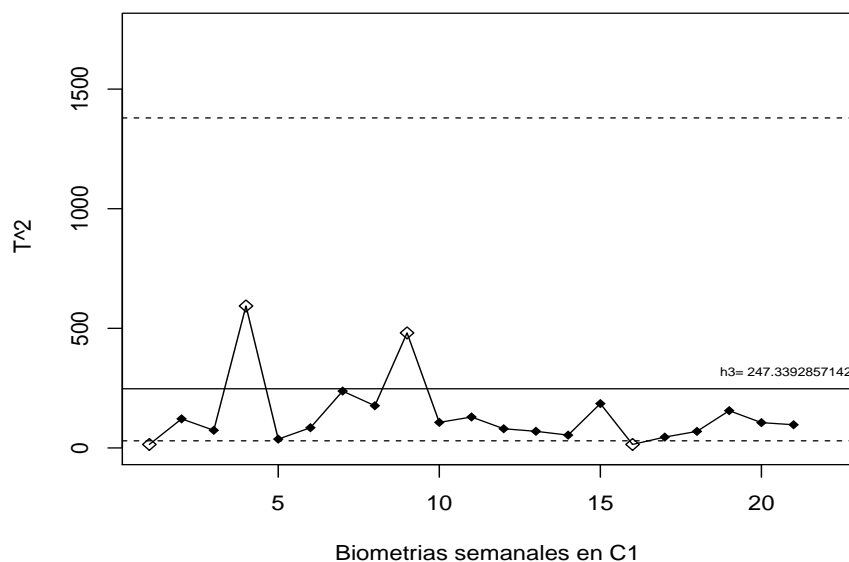


Figura 6.9: Monitorización del ambiente acuático en el tratamiento C1 mediante el gráfico MAAEWMA

correlación entre los parámetros en el gráfico MAAEWMA.

(2) Cuanto a la conformidad

El gráfico MAAEWMA presenta más observaciones bajo control comparado a los gráficos univariantes. La autocorrelación en los datos no modelada por los gráficos univariantes puede justificar este resultado. Como nos referimos anteriormente, los gráficos univariantes desarrollados bajo la hipótesis de independencia son susceptibles a falsas alarmas al aplicarlos a los procesos autocorrelacionados.

Las observaciones no conformes señaladas por el gráfico MAAEWMA en todos los tratamientos (presentadas en la tabla 6.9) están relacionadas con bajos niveles del oxígeno y de la transparencia. La observación 17 fuera de control en el tratamiento C6 no fue señalada por ningún gráfico univariate.

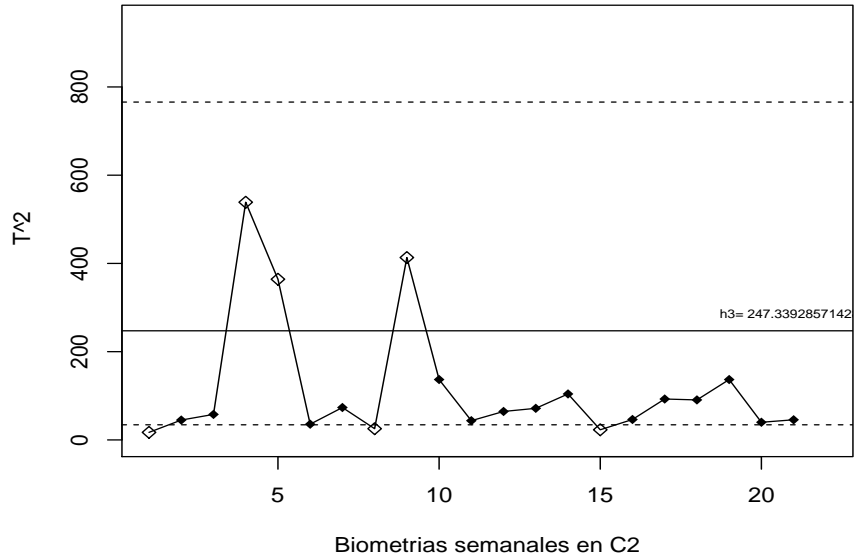


Figura 6.10: Monitorización del ambiente acuático en el tratamiento C2 mediante el gráfico MAAEWMA

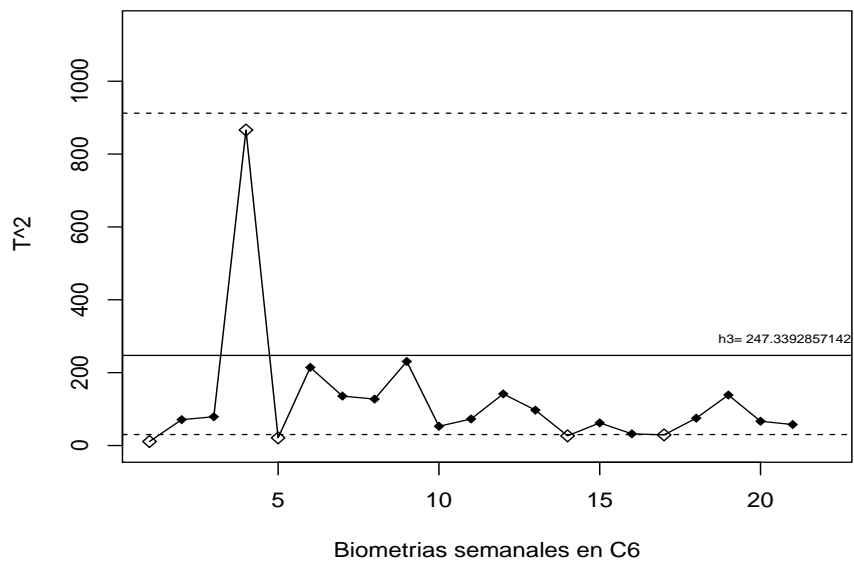


Figura 6.11: Monitorización del ambiente acuático en el tratamiento C6 mediante el gráfico MAAEWMA

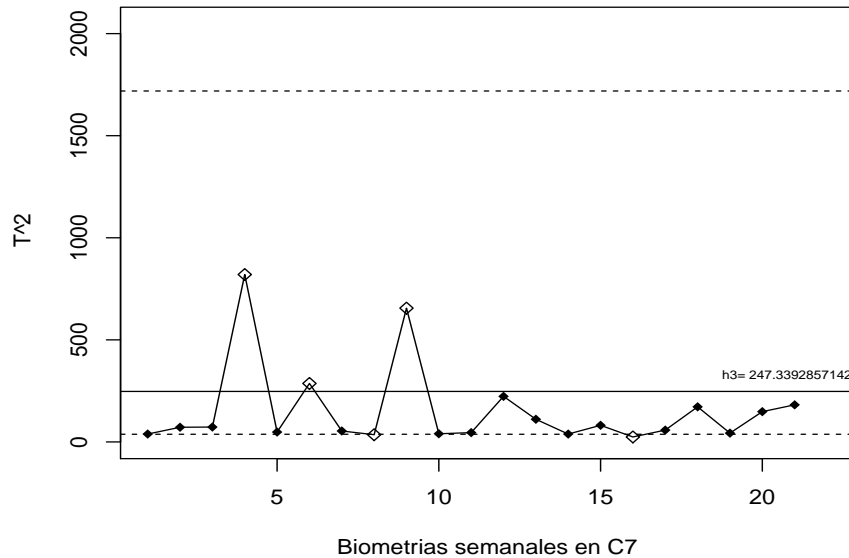


Figura 6.12: Monitorización del ambiente acuático en el tratamiento C7 mediante el gráfico MAAEWMA

6.6.3. Predicción del índice de crecimiento de la tilapia mediante el modelo dinámico sobre estructuras latentes

El gráfico MAAEWMA usado en el apartado anterior mostró cambios significativos del ambiente acuático a lo largo del proceso de cría de la tilapia en los cuatro tratamientos considerados. Dichos cambios están asociados o bien por la falta de estabilidad en los parámetros ambientales, o bien por sus niveles no adecuados para la cría de especies en acuicultura. Como nos referimos anteriormente, los peces (especialmente la tilapia) son muy dependientes del ambiente acuático a que están sujetos, siendo que los niveles bajos o altos de los parámetros ambientales pueden afectar a la productividad.

En este apartado se analiza el impacto de los niveles registrados del ambiente acuático en el crecimiento de la tilapia. Para ello, se aplica el modelo dinámico sobre estructuras latentes desarrollado en el capítulo 5, donde los parámetros biométricos (peso y longitud estandar) corresponden a las variables del producto (bloque Y), y los cuatro parámetros ambientales (temperatura, oxígeno, pH y transparencia) forman el bloque de las variables del proceso (bloque X). Los resultados presentados en el apartado 6.6.1 muestran un crecimiento diferente de las tilapias en los cuatro tratamientos, por ello, en

cada uno de ellos fue ajustado un modelo dinámico.

6.6.3.1. Ajuste del modelo

De manera análoga al modelo MAAEWMA, los datos de ambos bloques fueron tipificados para eliminar el efecto de escala de medición de las variables (parámetros biométricos y ambientales). El ajuste del modelo dinámico fue realizado siguiendo los pasos presentados en el capítulo 5, que consisten en:

1. Aplicar un análisis de componentes principales dinámico en cada bloque de variables;
2. Sobre las variables latentes obtenidas en 1, aplicar la metodología de Box-Jenkins para ajustar la función de transferencia:
 - 2.1 Ajustar un modelo ARMA a cada variable latente explicativa t_i (componentes principales de X) y aplicar el mismo filtro a su correspondiente variable latente respuesta u_i (componente principal de Y);
 - 2.2 Analizar la función de correlación cruzada entre cada par de variables t_i y u_i e identificar los ordenes b, r, s de la correspondiente componente $v(B)$;
 - 2.3 Identificar los ordenes p y q de la componente N_i a partir de los ordenes del modelo ARMA ajustado a u_i ;
 - 2.4 Estimar el modelo de función de transferencia con los ordenes identificados;
 - 2.5 Validar el modelo.

La estructura dinámica en cada bloque X e Y fue analizada, como proponemos en el modelo dinámico (capítulo 5), según las correlaciones cruzadas entre cada par de residuos pre-blanqueados de las variables. La figura 6.13 (y las figuras A.7 a las figuras A.12) ilustra dichas correlaciones cruzadas, pudiéndose constatar que:

1. Variables del producto: Y

El peso y longitud presentan correlaciones significativas en las observaciones del mismo instante en los tratamientos C1 y C2. En C6, además de las correlaciones entre observaciones del mismo instante, los valores de la longitud están correlacionadas con los valores del peso de dos instantes anteriores;

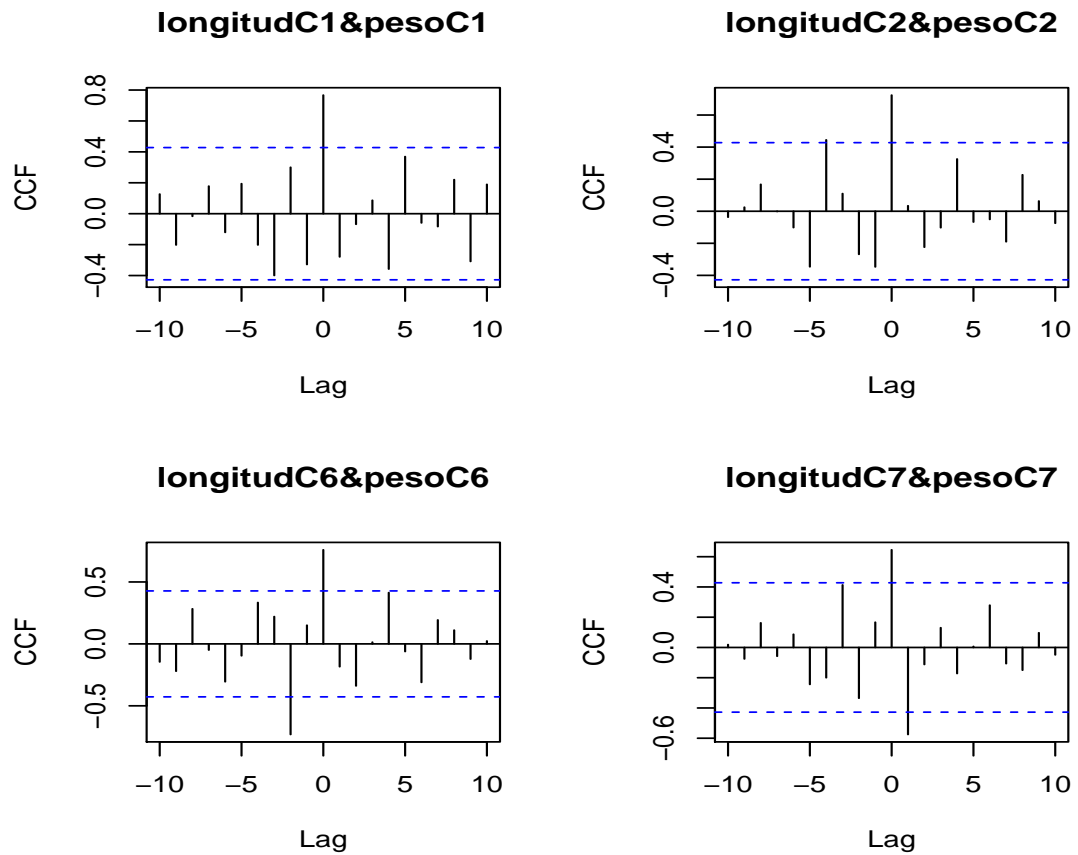


Figura 6.13: Función de correlación cruzada entre peso y longitud en los cuatro tratamientos

En C7 hay correlaciones significativas entre observaciones del peso y longitud en el mismo instante, además, los valores del peso están correlacionados con los valores de la longitud del instante anterior.

Por tanto, en los tratamientos C1 y C2 se verifica $\tilde{Y} = Y$, ya que no hay lags significativos para incluir en la matriz Y . En el tratamiento C6 la matriz \tilde{Y} es formada por los datos de Y y los lags de segundo orden de las variables; en C7, la matriz \tilde{Y} es formada por los datos de Y y los lags de primer orden de las correspondientes variables.

2. Variables del proceso X

De un modo general, los parámetros ambientales presentan correlaciones significativas cruzadas entre observaciones del mismo instante (figuras A.7 a figuras A.12), de modo que en todos los tratamientos se consideró

$\tilde{X} = X$, es decir, no hubo necesidad de incluir lags en la matriz X .

6.6.3.2. Resultados del modelo ajustado

Se aplicó un análisis de componentes principales dinámico (DPCA) a los dos bloques de variables, y en los cuatro tratamientos fue determinada una variable latente u_i en el bloque Y y dos variables latentes (t_{1i} y t_{2i}) en el bloque X .

La variable u_i fue determinada con una variabilidad explicada de 97.79 % en C1, 96.10 % en C2, 97.61 % en C6 y 95.43 % en C7. En virtud a la alta correlación entre el peso y la longitud, dicha variable recoge la mayor varianza explicada por estos parámetros. Por tanto, u_i es un indicador del tamaño de la tilapia. Los valores positivos de u_i indican un mayor tamaño de la tilapia, y a los valores negativos indican lo contrario, con respecto al promedio general. Las variaciones de u_i a lo largo del tiempo (i) indican el crecimiento de los ejemplares.

Las variables latentes t_{1i} y t_{2i} fueron determinadas con una variabilidad explicada de 70.022 % en C1, 74.47 % en C2, 84.14 % en C6 y 74.25 % en C7. En todos los tratamientos, t_{1i} es caracterizada por una correlación fuerte y positiva entre el oxígeno y la temperatura, y t_{2i} es caracterizada por una correlación fuerte y negativa entre pH y la transparencia (figura 6.14). Por tanto, los valores positivos de la componente t_{1i} indican niveles altos de la temperatura y del oxígeno, y a los valores negativos indican lo contrario. Los valores positivos de la componente t_{2i} indican un ambiente acuático muy transparente (o con un bajo nivel de turbidez) con niveles bajos del pH, y a sus valores negativos indican lo contrario.

Las figuras 6.15 a 6.18 presentan las variables latentes determinadas en los cuatro tratamientos. Se observa que u_i tiene una tendencia creciente en todos los tratamientos. Para lograr la estacionariedad en la media, se aplicó una diferencia regular ($u_i - u_{i-1} = (1 - B)u_i$) y la serie resultante fue designada por u_{1i} la cual representa el crecimiento de la tilapia de un instante al otro.

A las series t_{1i} y t_{2i} fueron ajustados modelos AR(1) en todos los tratamientos (figura 6.19 y figuras A.13 a figuras A.15). Los mismos filtros fueron usados para pre-blanquear la serie u_i en cada tratamiento, con los cuales fueron calculadas las funciones de correlación cruzadas entre los residuos de u_i y los residuos de t_{1i} y t_{2i} (figuras A.16 y figura A.17), donde fueron identificados los ordenes r,s y b de la componente $v_1(B)$ y $v_2(B)$ asociado a las respectivas variables explicativas t_{1i} y t_{2i} en la función de transferencia.

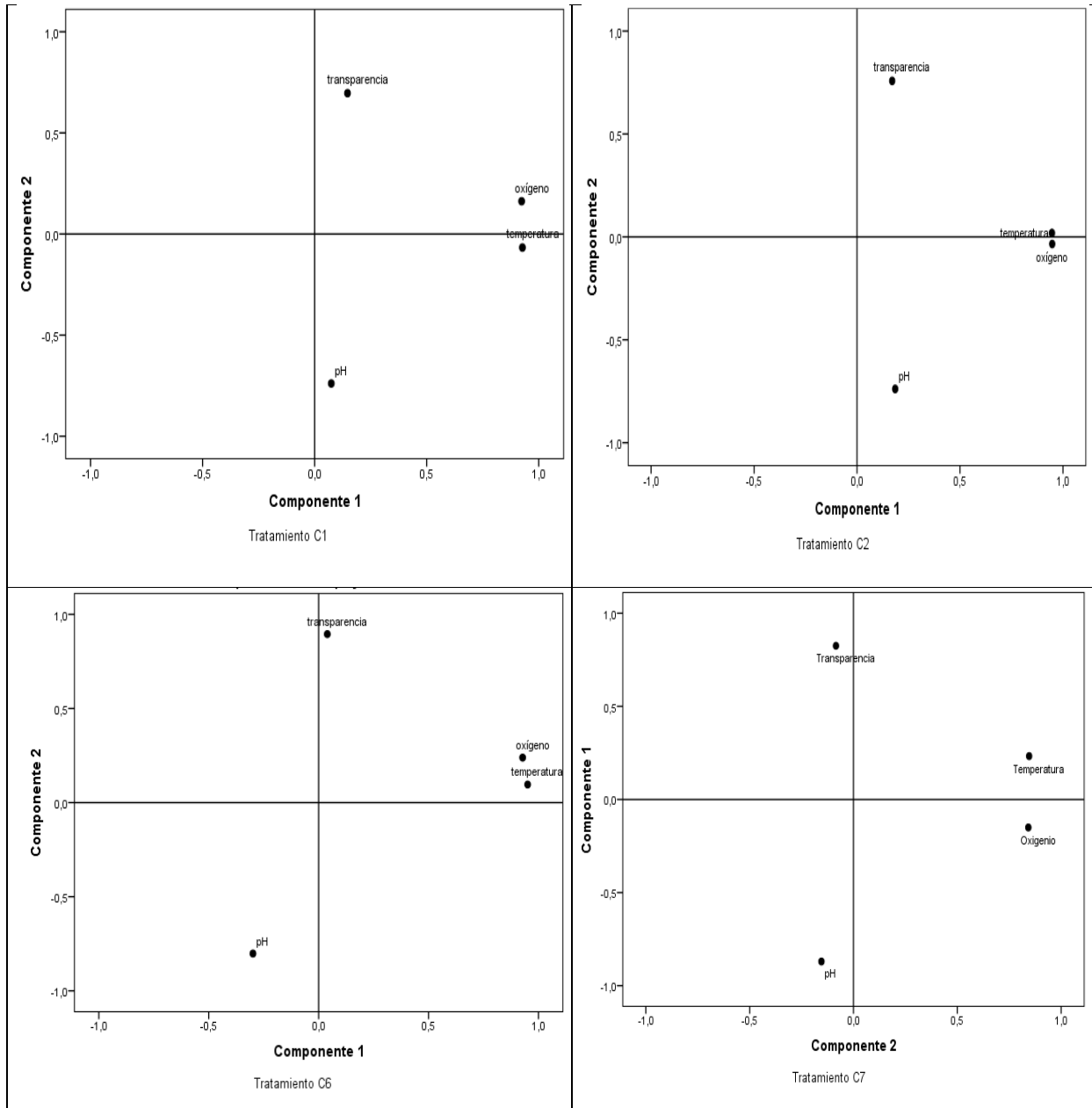


Figura 6.14: Representación de las componentes principales

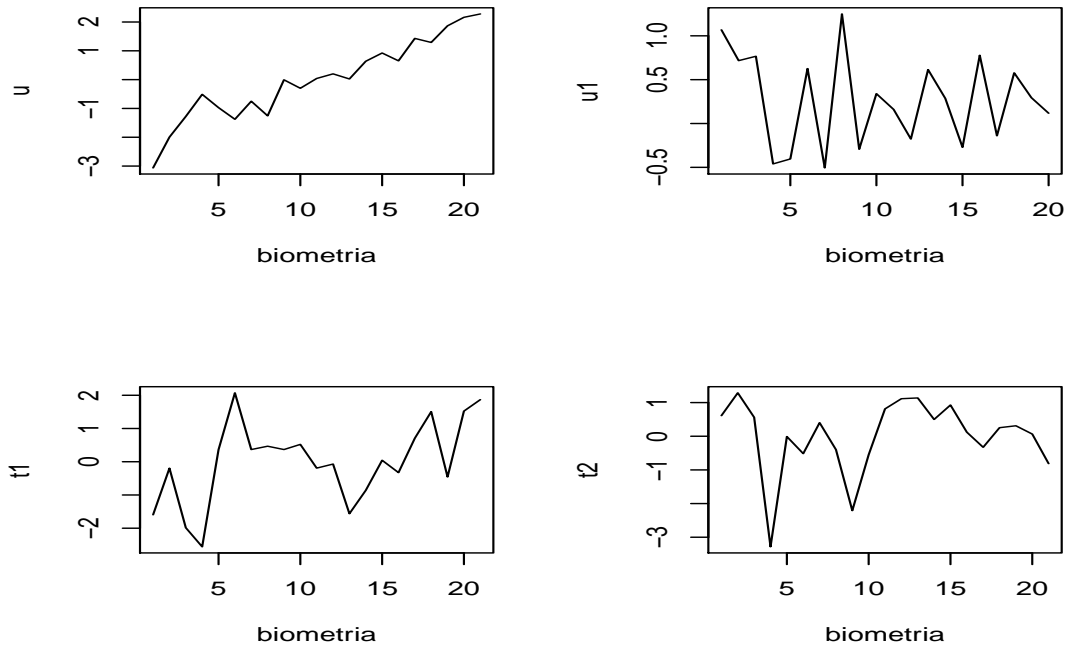


Figura 6.15: Representación de las variables latentes en el tratamiento C1

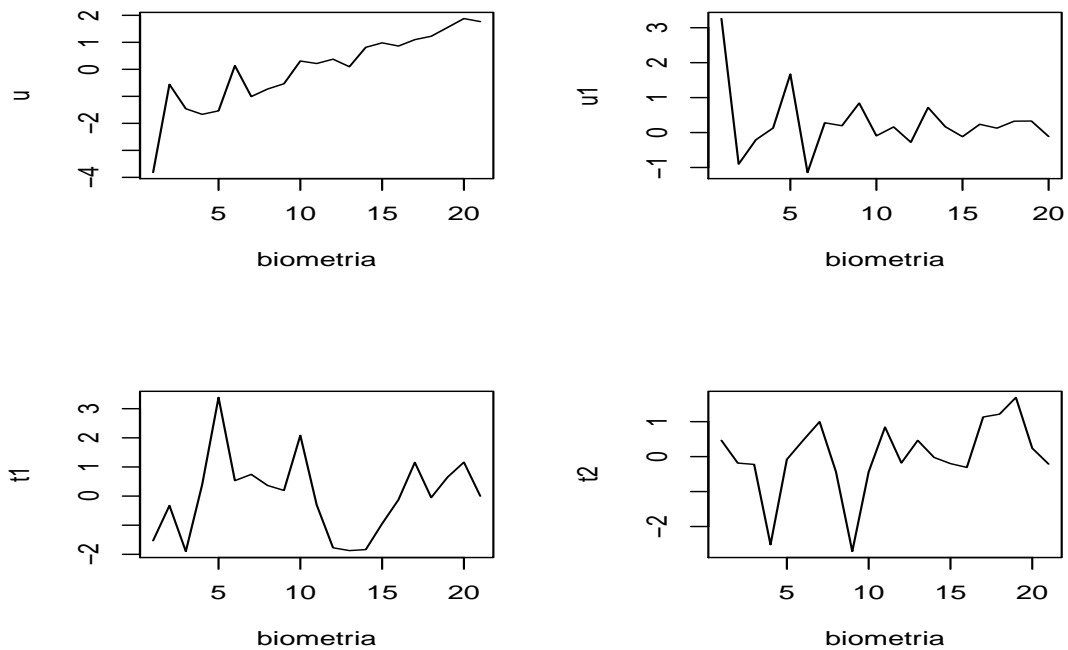


Figura 6.16: Representación de las variables latentes en el tratamiento C2

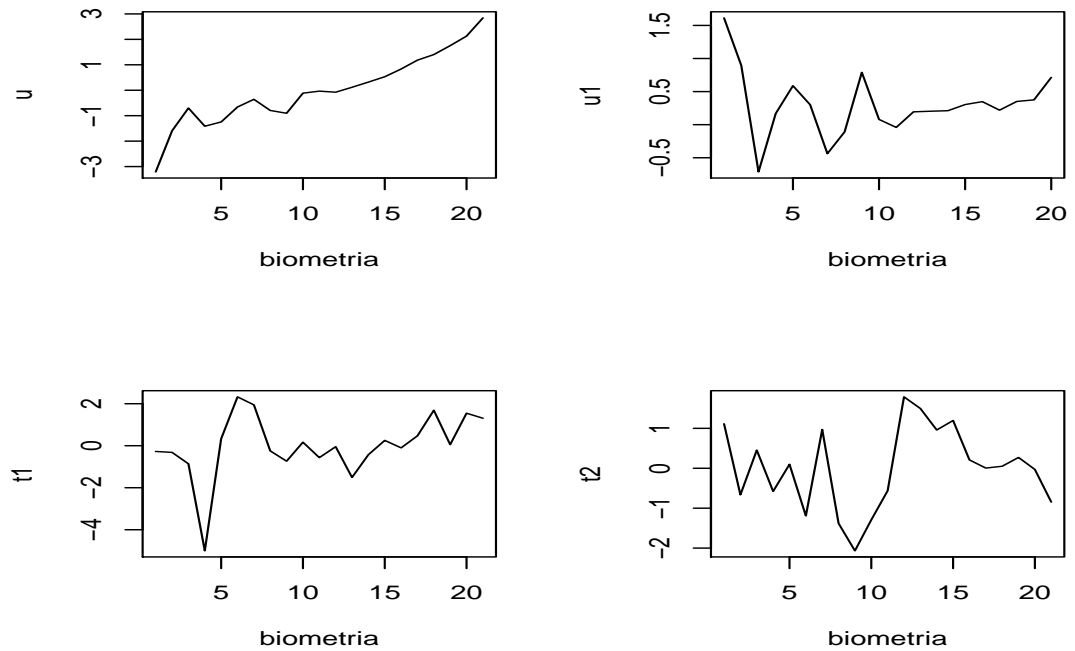


Figura 6.17: Representación de las variables latentes en el tratamiento C6

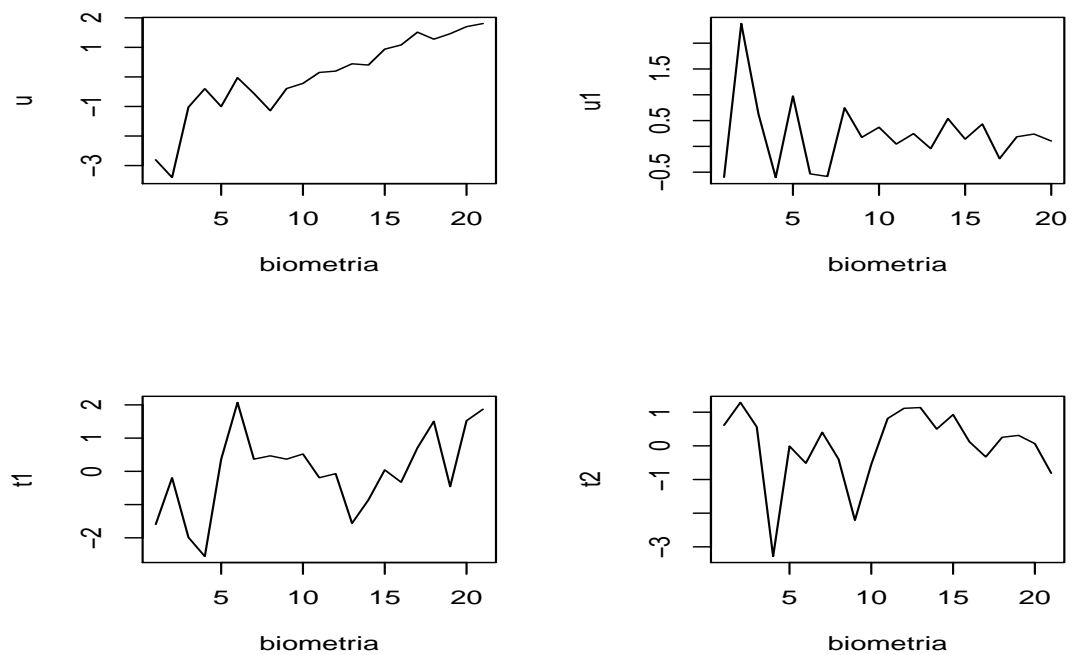


Figura 6.18: Representación de las variables latentes en el tratamiento C7

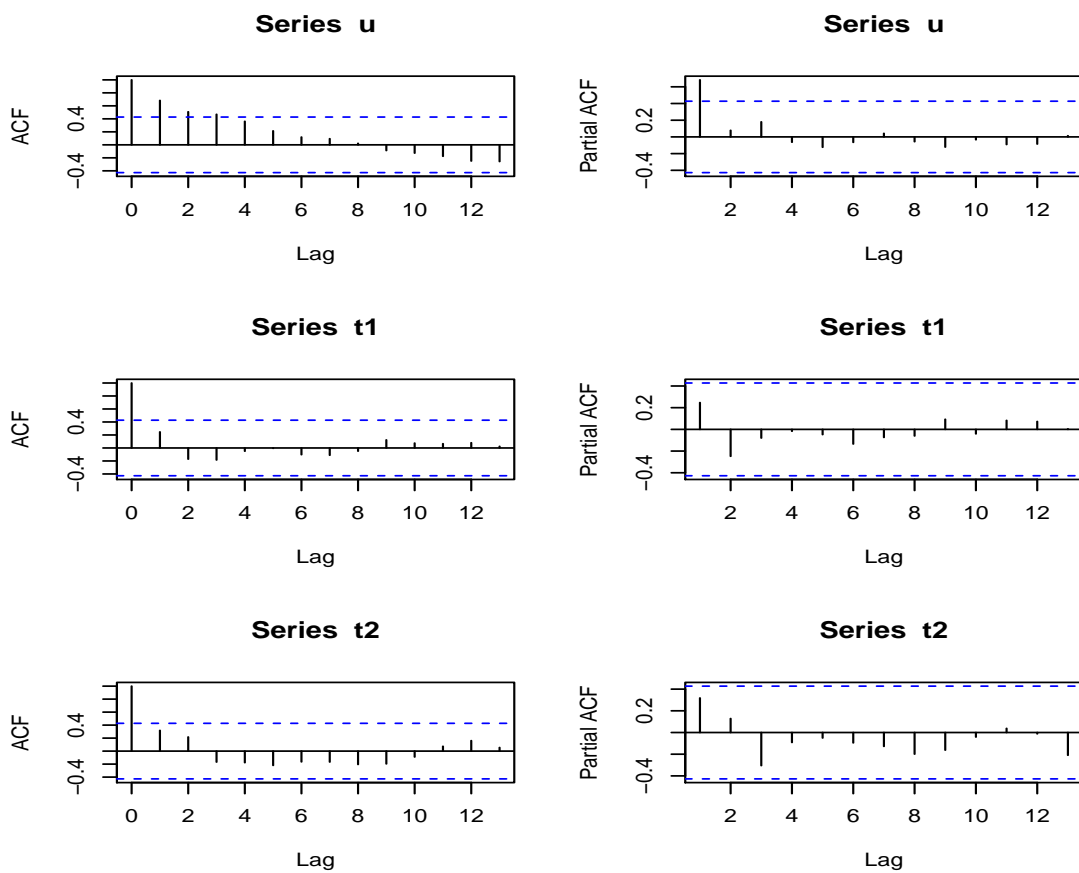


Figura 6.19: Correlaciones simple y parciales de las variables latentes en el tratamiento C7

1. En el tratamiento C1, no hay correlaciones significativas entre u_i y t_{1i} (la correlación en lag=0 parece importante, no obstante, el coeficiente asociado a ella resultó no significativo); y entre u_i y t_{2i} , la primera correlación significativa corresponde al lag=3, por lo que se obtiene $r=0$, $s=1$ y $b=3$ para la componente $v_2(B)$ asociada a t_{2i} .
2. De modo análogo al tratamiento C1, no hay correlación significativa entre u_i y t_{1i} en C2 (el lag=4 parece importante, pero el coeficiente asociado a ella no es significativo). La función de correlación cruzada entre u_i y t_{2i} indica una estructura $r=0$, $s=1$ y $b=1$ para $v_2(B)$.
3. En C6, la correlación más importante entre u_i y t_{1i} está en lag=0, entonces, la componente $v_1(B)$ presenta los ordenes $r=0$, $s=1$ y $b=0$; y no hay correlación significativa entre u_i y t_{2i} ;
4. La estructura de correlación entre u_i y t_{1i} en C7 sugiere una estructura $r=1$, $s=0$ y $b=0$; y entre u_i y t_{2i} $r=0$, $s=1$ y $b=2$.

Respecto a los ordenes p y q de la componente N_i de la función de transferencia, fueron identificados a través del modelo correspondiente a u_i con sus propios polinomios, es decir, el modelo correspondiente a u_i sin aplicar el preblanqueo.

Las funciones de correlaciones simples y parciales presentadas en las figura 6.19 (y figuras A13 a figuras A.15) muestran que u_i corresponde a un proceso AR(1). Considerando la diferencia regular aplicada a la serie, se obtiene un modelo $u \sim \text{ARIMA}(1, 1, 0)$, donde $p=1$ y $q=0$ son tomados como ordenes de N_i en todos los tratamientos.

Las tablas 6.10 y 6.11 presentan los parámetros estimados del modelo de función de transferencia. Considerando solamente las estimaciones significativas, las expresiones analíticas de los modelos ajustados en cada tratamiento son representados en la tabla 6.12. A partir de ellas se puede sintetizar lo siguiente:

- Los niveles observados del pH y de la transparencia (t_{2i}) en los tratamientos C1 y en C2 tuvieron influencia significativa en el índice de crecimiento de la tilapia .

Los valores del pH y de la transparencia en cada instante en C1 tuvieron un impacto positivo tres semanas posteriores en el crecimiento de la tilapia. Este resultado muestra que los niveles altos de la transparencia

Tabla 6.10: Resultados Numéricos de las estimación del Modelo dinámico sobre estructuras latentes en los tratamientos C1 y C2

Tratamiento C1						
Componentes de la FT			Estimativa	SE	t	sig
$v_1(B)$	Numerador	w_0	0,059	0,041	1,450	0,178
		w_1	0,023	0,043	0,548	0,596
		b	0			
$v_2(B)$	Numerador	w_0	0,065	0,022	2,931	0,015**
		w_1	0,016	0,019	0,814	0,434
		b	3			
η_t	Denominador	ϕ_1	0,893	0,119	-7,500	0,000***
		d	1			
Bondad de ajuste						
R^2	0,880					
RMSE	0,156					
Observación atípica						
observación 4	tipo: aditivo					
Estadístico de Ljung-Box						
Estadístico	gl	sig				
14,617	17	0,623				
Tratamiento C2						
Componentes de la FT			Estimativa	SE	t	sig
$v_1(B)$	Numerador	w_0	0,050	0,080	0,632	0,539
	Denominador	δ_1	0,325	1,256	0,259	0,800
		b	3			
$v_2(B)$	Numerador	w_0	0,253	0,097	2,605	0,023**
		w_1	0,150	0,079	1,892	0,083*
		b	1			
η_t	Denominador	ϕ_1	0,755	0,180	-4,206	0,001***
		d	1			
Bondad de ajuste						
R^2	0,831					
RMSE	0,406					
Observación atípica						
observación 9	tipo: cambio de nivel					
Estadístico de Ljung-Box						
Estadístico	gl	sig				
18,437	17	0,362				

Tabla 6.11: Resultados Numéricos de las estimación del Modelo dinámico sobre estructuras latentes en los tratamientos C6 y C7

Tratamiento C6						
Componentes de la FT			Estimativa	SE	t	sig
$v_1(B)$	Numerador	w_0	0,205	0,089	2,300	0,039**
		w_1	0,002	0,121	0,019	0,985
		b	0			
$v_2(B)$	Numerador	w_0	0,116	0,083	1,388	0,188
		w_1	-0,091	0,784	-0,116	0,909
		b	0			
η_t	Denominador	ϕ_1	0,648	0,218	2,974	0,011**
		d	1			
Bondad de ajuste						
R^2	0,875					
RMSE	0,372					
Observación atípica						
observación 5	tipo: cambio de nivel					
Estadístico de Ljung-Box						
Estadístico	gl	sig				
15,802	17	0,538				
Tratamiento C7						
Componentes de la FT			Estimativa	SE	t	sig
$v_1(B)$	Numerador	w_0	-0,074	0,035	-2,099	0,060*
	Denominador	δ_1	0,974	0,018	53,152	0,000***
		b	0			
$v_2(B)$	Numerador	w_0	0,047	0,042	1,126	0,284
		w_1	-0,145	0,051	-2,851	0,016**
		b	2			
η_t	Denominador	ϕ_1	0,808	0,183	-4,410	0,001***
		d	1			
Bondad de ajuste						
R^2	0,911					
RMSE	0,251					
Observación atípica						
observación 4	tipo: aditivo					
Estadístico de Ljung-Box						
Estadístico	gl	sig				
17,048	17	0,451				

* <0.1 , ** <0.05 , *** <0.001

y bajos del pH (siempre dentro de sus límites de control) favoreció al crecimiento de la tilapia en C1.

En C2, los niveles del pH y de la transparencia en un instante tuvieron impacto positivo al crecimiento de la tilapia en la semana siguiente, y un impacto negativo dos semanas posteriores, con un saldo de crecimiento que resulta positivo (a partir de los coeficientes del modelo).

- Los niveles del oxígeno y de la temperatura registrados en cada instante en C6 influenciaron de manera positiva en el índice del crecimiento registrado en el mismo instante. Es decir, los valores altos de dichos parámetros tuvieron un impacto positivo en el crecimiento de la tilapia en la misma semana.
- Las componentes (t_{1i} y t_{2i}) son significativas en C7, lo que indica que los niveles de los cuatro parámetros ambientales influenciaron en el crecimiento de la tilapia en C7.

Los niveles registrados del oxígeno y de la temperatura en cada semana tuvieron un impacto negativo en el índice de crecimiento de la tilapia registrado en la semana siguiente.

Los niveles del pH y de la transparencia de cada semana influenciaron de manera positiva en los valores del índice de crecimiento de la tilapia tres semanas después.

6.6.3.3. Validación del modelo

Los modelos ajustados en los cuatro tratamientos fueron validados mediante un análisis de los residuos ϵ_i para averiguar si se cumple la condición: $\epsilon_i = iiN(0, \sigma^2)$. Para ello, fue considerado el contraste de la aleatoriedad de Ljung-Box para contrastar la hipótesis de que ϵ_i es independiente y normalmente distribuido. En la última línea de las tablas 6.10 y 6.11 se presentan los resultados del contraste, donde se puede observar que los p-valores son claramente superiores a 0.05, con lo cual a un nivel de confianza de 95 % no hay evidencias para rechazar la hipótesis de independencia y normalidad de los residuos.

Por otro lado, las correlaciones simples y parciales presentadas en la figura A.18 muestran que no hay correlaciones significativas en los residuos ϵ_i (en

Tabla 6.12: Expresiones analíticas de los modelos dinámicos en todos los tratamientos

1. Modelo en C1:

$$\begin{aligned}(1 - \phi_1 B)u_i &= w_0 B^3 t_{i2} + a_i \\(1 - 0,893B)u_i &= 0,065t_{2(i-3)} + a_i \\u_i &= 0,893u_{(i-1)} + 0,065t_{2(i-3)} + a_i\end{aligned}$$

2. Modelo en C2:

$$\begin{aligned}(1 - \phi_1 B)u_i &= (w_0 - w_1 B)B^1 t_{i2} + a_i \\(1 - 0,755B)u_i &= (0,253 - 0,150B)t_{2(i-1)} + a_i \\u_i &= 0,755u_{(i-1)} + 0,253t_{2(i-1)} - 0,150t_{2(i-2)} + a_i\end{aligned}$$

3. Modelo en C6:

$$\begin{aligned}(1 - \phi_1 B)u_i &= w_0 B t_{i1} + a_i \\(1 - 0,648B)u_i &= 0,205t_{1i} + a_i \\u_i &= 0,648u_{(i-1)} + 0,205t_{1i} + a_i\end{aligned}$$

4. Modelo en C7:

$$\begin{aligned}(1 - \phi_1 B)u_i &= \frac{w_0}{1 - \delta_1 B} B t_{i1} + (w_1 B) B^2 t_{i2} + a_i \\(1 - 0,808B)u_i &= \frac{-0,074}{1 + 0,974B} t_{1(i-1)} + (0,145B)t_{2(i-2)} + a_i \\u_i &= 0,808u_{(i-1)} + \frac{-0,074}{(1 + 0,974B)} t_{1(i-1)} + 0,145t_{2(i-3)} + a_i\end{aligned}$$

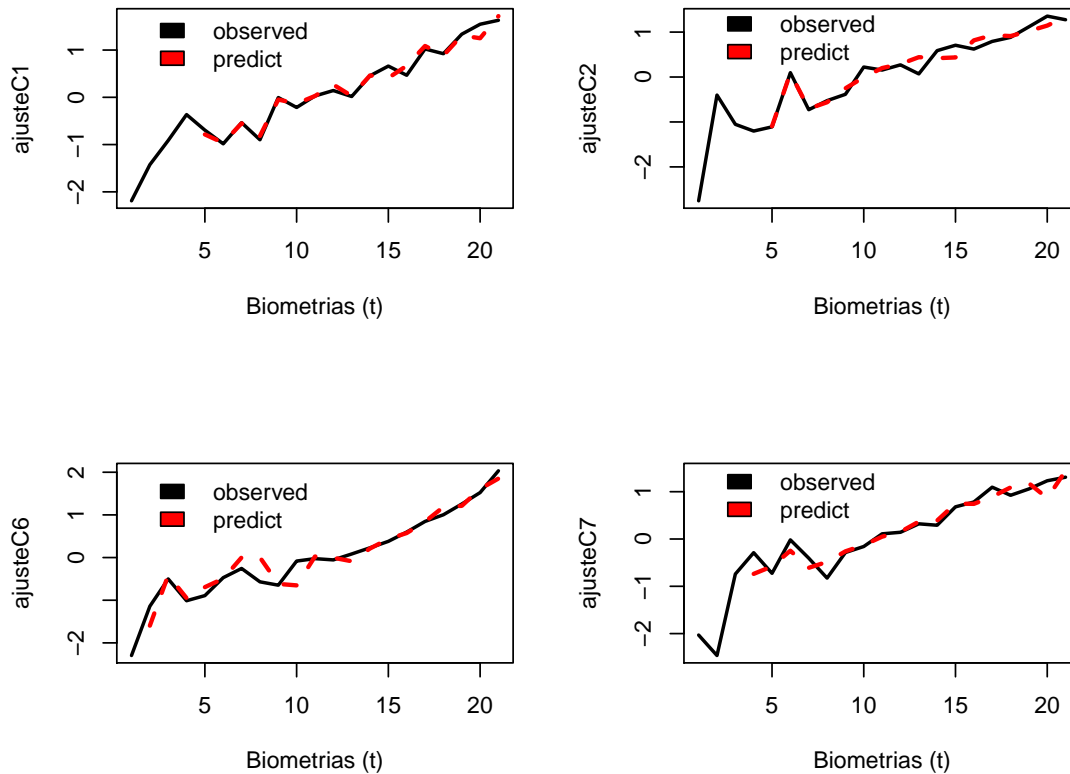


Figura 6.20: Predicción del índice del crecimiento de la tilapia en los cuatro tratamientos

algunos casos las correlaciones han sido influenciadas por observaciones atípicas), por lo que se asume la independencia del residuo.

La figura 6.20 ilustra los gráficos de los valores observados de u_t y su respectiva predicción usando los modelos ajustados en cada tratamiento. Se observa que no hay una gran discrepancia entre los valores observados y los predichos, indicando que el ruido ϵ_t (que corresponde la parte de la variabilidad de u_t no explicada por el modelo) es pequeña, y que el modelo dinámico ajustado en cada tratamiento puede ser usado para predecir el índice de crecimiento de la tilapia a través de la variabilidad del ambiente acuático sin cometer gran error.

6.7. Conclusión del capítulo

En este capítulo se buscó aplicar el método estadístico desarrollado en el capítulo 5 en la monitorización y predicción de procesos de cultivo en acuicul-

tura. Los modelos desarrollados fueron aplicados para responder a la necesidad de monitorización del ambiente acuático y la predicción del crecimiento de la tilapia en el proceso de cría en los estanques de acuicultura en Mozambique.

El gráfico MAAEWMA fue aplicado para monitorizar la conformidad y la estabilidad del ambiente acuático en los cuatro tratamientos considerados. Los resultados muestran que en algunas semanas el ambiente acuático no se ha mantenido estable, ni conforme con los requisitos exigidos para la cría de especies acuáticas. Dicho cambio del ambiente acuático afectó al crecimiento de la tilapia.

El impacto del ambiente acuático sobre el crecimiento de la tilapia fue analizado mediante el modelo dinámico sobre estructuras latentes.

El modelo dinámico muestra que los niveles registrados del oxígeno y de la temperatura tuvieron un impacto positivo en el crecimiento de la tilapia en C6, y un impacto negativo en C7. Los niveles registrados del pH y de la transparencia tuvieron un impacto positivo al crecimiento de la tilapia en C1 y C7. El impacto del pH y de la transparencia en C2 es tal que en una semana se registran mayores crecimientos, y en otra semana una reducción de dicho crecimiento.

La interpretación de tales resultados desde el punto de vista de la acuicultura excede el alcance de esta tesis.

C apitulo 7

CONCLUSIONES GENERALES Y FUTURAS LINEAS DE INVESTIGACI ON

7.1. Conclusiones

En este capitulo presentamos las principales conclusiones alcanzadas a lo largo del trabajo y que se alinean a los objetivos planteados en el capitulo 2. El m etodo estad istico propuesto est a constituido por dos modelos: el modelo de monitorizaci on y el modelo de predicci on de procesos multivariantes y autocorrelacionados. Los modelos fueron testados con un conjunto de datos simulados, y sus resultados fueron comparados con otros m etodos usados en este campo.

Como parte final del trabajo, el m etodo estad istico fue aplicado a un estudio real, donde el gr afico MAAEWMA y el modelo MDEL fueron usados para monitorizar y predecir los procesos de cultivo de especies acu aticas en acuicultura.

Los principales resultados se sintetizan en los siguientes puntos:

1. Respecto a la bibliogr afica consultada (objetivo espec ifico 1 y 2)

Los antecedentes presentados en el capitulo 3 se basan en el an alisis y presentaci on de m etodos estad isticos usados en la monitorizaci on y predicci on de procesos. Los m etodos fueron presentados seg un un orden l ogico basado en la creciente complejidad de los mismos y en su creciente capacidad para modelizar procesos complejos.

La presente investigaci on consisti o en desarrollar un gr afico de control multivariante que incluye el efecto de la autocorrelaci on en los par ametros del gr afico a trav es de los modelos autoregresivos.

El modelo de predicci on que se pretende desarrollar, explota la relaci on entre las variables de entrada y salida de un proceso. Para ello, los m etodos estad isticos sobre estructuras latentes fueron analizados, y se

presentaron el análisis de componentes principales (PCA) y el partial least square (PLS), y sus versiones dinámicas (DPCA y DPLS).

2. Respecto a los métodos desarrollados (objetivos específicos 3, 4 y 5)

El modelo de monitorización propuesto corresponde al gráfico MEWMA con los parámetros (vector de medias y la matriz de varianzas-covarianzas) calculados mediante modelos autoregresivos.

Los modelos autoregresivos fueron usados para modelar la autocorrelación en las variables del proceso. El estadístico fue adaptado para detectar rápidamente los cambios en la media. De acuerdo con la nomenclatura usada, el gráfico resultante fue designado como multivariate, autocorrelated and adapted EWMA chart (MAAEWMA). El desempeño del gráfico fue analizado mediante una simulación de series temporales autoregresivas bi-variadas de orden uno, y su resultado fue comparado con el modelo MEWMA ajustado a los residuos de las series simuladas.

La simulación se estructuró de acuerdo con un diseño factorial completo, teniendo como factores los coeficientes autorregresivos y los niveles de cambio en la media. Los procesos fueron simulados con los siguientes coeficientes autoregresivos: 0.3, 0.5, 0.7 y 0.9 que representan procesos con baja, media y alta autocorrelación. Los gráficos MAAEWMA y MEWMA residual fueron ajustados al proceso bajo control, y a continuación se simuló un cambio en la media del proceso, con lo cual los gráficos fueron testados.

El desempeño de los gráficos en detectar el cambio fue analizado a través del ARL usando el método ANOVA, y los resultados de los gráficos fueron comparados. El gráfico MAAEWMA resultó ser más rápido en señalar cambios en el proceso respecto al gráfico MEWMA ajustado a los ruidos, y dicho mejor rendimiento es significativo para procesos con media y alta autocorrelación.

Respecto a la predicción, el modelo dinámico sobre estructuras latentes fue desarrollado combinando el método dynamic PCA (DPCA) y la función de transferencia (FT). El modelo DPCA fue aplicado en cada bloque de variables para obtener variables latentes que recogen la máxima variabilidad dinámica en los datos. Sobre las variables latentes obtenidas, fue ajustada la función de transferencia que modela la relación dinámica entre los scores, y el modelo resultante fue designado como modelo

dinámico sobre estructuras latentes (MDEL).

El funcionamiento del MDEL fue testado mediante un ajuste a los datos simulados. Se desarrollo una simulación de procesos multivariantes, con coeficientes 0.3, 0.5, 0.7 y 0.9 que representan procesos con baja, media y alta autocorrelación. A los datos simulados fueron ajustados los modelos MDEL y DPLS, y la calidad de ajuste de las variables latentes fue evaluada mediante el R^2 , y la calidad predictiva fue testada mediante el método de validación cruzada, donde el mean absolute error MAE fue calculado.

En los 16 casos simulados el MDEL presentó mayor variabilidad explicada en las variables latentes, con un promedio de R^2 mayor comparado al DPLS. En los 32 modelos ajustados a las variables latentes, el MAE de MDEL fue menor respecto al MAE de DPLS, por lo que el modelo dinámico sobre estructuras latentes presentó mejor ajuste y predicción comparado al DPLS.

3. Respecto a la aplicación de los métodos propuestos al proceso de acuicultura (objetivo específico 6)

El método estadístico desarrollado se adecua a la estructura de los datos en acuicultura. Los modelos MAAEWMA y MDEL fueron usados para monitorizar y predecir el proceso de producción de tilapia en la empresa Aquapesca,Lda en Mozambique.

El gráfico MAAEWMA fue usado para monitorizar la conformidad y la estabilidad del ambiente acuático caracterizado por cuatro parámetros físico-químicos. El gráfico mostró que en algunas semanas el ambiente acuático no estuvo estable, ni conforme con los requisitos técnicos establecidos en acuicultura.

El MDEL fue aplicado para evaluar el impacto de los niveles observados de los parámetros ambientales en el crecimiento de la tilapia. El análisis de componentes principales dinámico usado mostró que el ambiente acuático presenta dos características: correlaciones fuertes y positivas entre el oxígeno y la temperatura; y correlaciones fuertes y negativas entre el pH y la transparencia. El ambiente acuático tubo un impacto significativo en el crecimiento de la tilapia.

7.2. Futuras líneas de investigaciones

El gráfico MAAEWMA propuesto fue desarrollado para monitorizar procesos multivariantes y autocorrelacionados con características que se ajustan a modelos autoregresivos de orden uno. Una primera línea de investigación consiste obviamente en considerar que el gráfico puede ser generalizado para monitorizar procesos multivariantes que se ajustan a un proceso ARIMA (p,d,q).

De manera análoga, el funcionamiento del modelo dinámico sobre estructuras latentes fue testado mediante procesos autoregresivos de primer orden. Sin embargo, el desarrollo de este modelo no se limita a estos procesos, por lo que puede ser extendido a los procesos que se ajustan a modelos ARIMA(p,d,q).

El límite de control del gráfico MAAEWMA fue calculado para un ARL bajo control igual a 370, el gráfico fue usado para detectar un cambio simulado en la media del proceso. Otra línea de investigación consistiría en la mejora de los parámetros del gráfico mediante algoritmos de optimización con vista a minimizar el ARL del proceso fuera de control, haciéndolo más rápido en detectar cambios del proceso.

Bibliografía

- ABDI, H. (2007). RV coefficient y congruence coefficient. In: Salkind NJ. *Encyclopedia of Measurement y Statistics*. Thousand Oaks: Sage Publications , 849–853.
- ABDI, H. Y WILLIAMS, L. (2010). Principal Component Analysis. *John Wiley & Sons* .
- ACUÑA, C. (n/a). Biotecnología aplicada a la acuicultura. *Consejo Argentino para la información y el desarrollo de la biotecnología* .
- ALTAF, B. A. (2013). *Application of Dynamic Partial Least Squares to Complex Processes*. School of Chemical Engineering and Advanced Materials. Newcastle University.
- ALVES, S. Y MELLO, L. (2007). *Manual para Monitoramento Hidrobiológico em Fazendas de cultivo de camarão*. Recife.
- APARISI, F. Y HARO, S. (2003). A comparison of t2 control charts with variable sampling schemes as opposed to MEWMA chart. *International Journal of Production Research* .
- BAILLO, A. Y GRANÉ, A. (2007). *100 Problemas resueltos de Estadística Multivariante implementados en Matlab*. Delta Publicaciones.
- BARCELÓ, S., VIDAL-PUIG, S. Y A., F. (2011). Comparison of Multivariate Statistical Methods for Dynamic Systems Modeling. *Quality and Reliability Engineering International* .
- BOX, G. Y JENKINS, G. (1976). Time Series Analysis: Forecasting and Control. *Holden-Day, San Francisco, USA* .
- BROOK, D. Y EVANS, D. A. (1972). An approach to the probability distribution of cusum run length. *Biometrika* .
- CLARO, F., COSTA, A. F. B. Y MACHADO, M. (2007). Ewma and control charts for the monitoring of autocorrelated processes. *Produção, v. 17, n. 3, p. 536-546, Set./Dez* .
- COSTA, A., EPPRECHT, E. Y CARPINETTI, L. (2005). Controle estatístico de qualidade. 2. ed. *São Paulo: Editora Atlas* .

- CROSIER, R. B. (1988). Multivariate generalizations of cumulative sum quality-control schemes. *Technometrics* , 291–303.
- CROWDER, S. V. (1972). A simple method for studying run length distributions of exponentially weighted moving average charts. *Technometrics* , 401–407.
- DAYAL, B. Y MACGREGOR, J. (1996). Introduction to Multi and Megavariate Data Analysis using Projection Methods (PCA and PLS). Identification of finite impulse models: Methods and robustness issues. *Industrial and Engineering Chemistry Research* .
- DONG, Y. Y QIN, S. J. (2015). Dynamic Inner Partial Least Squares for Dynamic Data Modeling. *International Federation of Automatic Control* , 117–122.
- EFRON, B. (1982). The Jackknife, the Bootstrap and other Resampling Plans. *CMBF-NSF Regional Conference Series in Applied Mathematics: New York SIAM vol.83*, 353–360.
- ERIKSSON, L., JOHANSSON, E., KETTANEH-WOLD, N. Y WOLD, S. (1984). Introduction to Multi and Megavariate Data Analysis using Projection Methods (PCA & PLS). *Chemometrics: Mathematics and Statistics in Chemistry, Kowalski BR (ed.). D. Reidel Publishing Company: Dordrecht, Holland* .
- ESBENSEN, K. Y GELADI, P. (1987). Principal Component Analysis. *Elsevier Science Publishers* , 37–52.
- FALTIN, F., MASTRENGELO, C., RUNGER, G. Y RYAN, T. (1997). Statistical process control procedures for correlated observations. *Considerations in the monitoring of autocorrelated and independent data. Journal of Quality Technology, v. 29, p. 131-1337* .
- FAO (2016). Fertilizantes y fertilización. (<http://www.fao.org/fishery/culturedspecies/Oreochromis-niloticus>) .
- FARUK, D. (2009). A hybrid neural network and arima model for water quality time series prediction. *Water Resources Reseach Center, Adan Menderes University, Aydn, Turkey. Elsevier Ltd* .
- FERRÉ, J. Y RIUS, F. (2002). Introducción al diseño estadístico de experimentos .
- FERRER, A. (2003). Control Estadístico MegaVariante para los procesos del siglo XXI. *27 Congreso Nacional de Estadística e Investigación Operativa* .

- FERRER, A. (2007). Multivariate Statistical Process Control Based on Principal Component Analysis (MSPC PCA): Some Reflections and a Case Study in Autobody Assembly Process. *Quality Engineering* , 311–325.
- GARCIA, A., TUME, J. Y JUÁREZ, V. (s/a). Determination of growth parameters of nile tilapia (*oreochromis niloticus*) in a Magazine with geomembrane pond stocked with water and subsoil .
- GARCÍA-DÍAZ, J. Y APARISI, F. (2003). Optimización de los gráficos de control estadístico de procesos EWMA y MEWMA mediante algoritmos genéticos. *27 Congreso Nacional de Estadística e Investigación Operativa* .
- GÓMEZ, B. (2005). Cultivo de tilapia. Manual para la Construcción de jaulas y corrales. *Disponible en http://www.sra.gob.mx/internet/informacion-general/programas/fondo-tierras/m_anuales/Cultivo-tilapia-jaulas-flotantes.pdf* .
- GUERRA, A. Y MANRIQUEZ, M. (1979). Parámetros biométricos de octopus vulgaris .
- HAN, H.-G., CHEN, Q.-L. Y QIAO, J.-F. (2011). An efficient self-organizing RBF neural network for water quality prediction. *Elsevier Ltd.* .
- HARDY (1981). *Manual para Monitoramento Hidrobiológico em Fazendas de cultivo de camarão*. Recife.
- HARRIS, T. Y ROSS, W. (1991). Statistical Process Control procedures for correlated observations. *The Canadian Journal of Chemical Engineering*, v. 69, p. 48-57 .
- HOLGATE, P. (1964). Estimation for the bivariate Poisson distribution. *Biometrika* , 241–245.
- HOSKULDSSON, A. (1988). PLS regression methods. *Chemometrics* **vol.2**, 211–228.
- HOSKULDSSON, A. (1996). Prediction Methods in Science and Technology. *Thor Publishing, Copenhagen* **vol. 1**(9).
- HOTELLING, H. (1947). Multivariate Quality Control. In C. Eisenhart, M. W. Hastay, and W.A. Wallis, eds. *Techniques of Statistical Analysis*. *New York: McGraw-Hill* .
- JACKSON, J. (1991). A Users Guide to Principal Components. *New York: John Wiley & Sons* .
- JIMÉNEZ, A. Y BARBA, A. (2000). Determinación de los parámetros fisicoquímicos de calidad del agua. *Gestión Ambiental*, vol. 2(23) pag. 12-19 .

- JINFENG, L. Y SHUN, C. (2015). A zigbee-based aquiculture water quality monitoring system. *International Journal of u- and e- Service, Science and Technology* .
- JOLLIFFE, I. (2002). Principal Component Analysis. *New York. Springer* .
- KASPAR, M. Y RAY, W. (1993). Dynamic pls modelling for process control. *Chemical Engineering Science* , 3447–3461.
- KOWALSK, B., GERLACH, R. Y WOLD, H. (1982). Chemical systems under indirect observation, in k. joreskog and h. wold (eds.), systems under indirect observation. *North-Holland, Amsterdam* , 191–209.
- KRESTA, J., MACGREGOR, J. Y MARLIN, T. (1995). Multivariate statistical monitoring of process oper-ating performance. *Can. J. Chem. Eng* , 35–47.
- KU, W., STORER, R. H. Y GEORGAKIS, C. (1995). Disturbance detection and isolation by dynamic principal component analysis. *Chemometrics and Intelligent Laboratory Systems* , 179–196.
- LABARTA, U. (2007). El desarrollo de la acuicultura en españa. *ARBOR Ciencia, Pensamiento y Cultura, CLXXXIII 727, septiembre-octubre* , 669–685.
- LAGOS, M., HÉCTOR, M., MEYER, D. Y CASTILLO, R. (2000). Comparación de la sobrevivencia y crecimiento de dos líneas de tilapia cultivadas bajo dos sistemas de manejo .
- LAKSHMINARAYANAN, S., SHAH, S. Y NANDAKUMAR, K. (1997). Modeling and control of multivariable processes: dynamic pls approach. *AIChE Journal* , 2307–2322.
- LIU, S., XU, L., LI, Z. Y LI, D. (2014). Forecasting model for ph value of aquaculture water quality based on PCA-MCAFA-LSSVM. *Nongye Jixie Xuebao. Transactions of the Chinese Society for Agricultural Machinery Vol.45*, 239–246.
- LIU, X., CHEN, H. Y ZHU, C. (2016). Design of self-learning cruising type water quality monitoring system based on gps. *Chinese Society of Agricultural Engineering* .
- LOWRY, C. A., WOODALL, W. H., CHAMP, C. W. Y RIGDON, S. E. (1992). A multivariate exponentially weighted moving average control chart. *Technometrics* (34), 46–53.
- LUCAS, J. Y SACCUCCI, M. (1990). Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements. *Technometrics* .

- MACGREGOR, J. (1996). Using On-Line Process Data to Improve Quality. Is there a Role for Statisticians?. Are They Up for the Challenge? *ASQC Statistics Division Newsletter* , 6–13.
- MACGREGOR, J. Y T.KOURTI (1995). Statistical process control of multivariate processes. *Pergamon* **vol.3**, 403–414.
- MARTÍNEZ, L., GIMÉNEZ, A., OLIVER, N. Y GILABERT, J. (n/a). Análisis de parámetros físicos, químicos y biológicos en las aguas costeras de la región de murcia. *III Jornadas de Introducción a la Investigación de la UPCT* .
- MARTINEZ, M. (2006). *Manejo del Cultivo de Tilapia*. University of Hawai'i.
- MARTINEZ, T. (2003). *Adaptación y crecimiento de las tilapias Oreochromis niloticus, Oreochromis aureus, Oreochromis mossambicus x Oreochromis niloticus en agua salada*. Universidad de Colima. Disponible en <http://digeset.ucol.mx.tesis-posgrado.Pdf>.
- MINISTÉRIODASPESCAS (2012). Plano Diretor das Pescas 2010–2019. *Ministério das Pescas, Direcção Nacional de Economia Pesqueira. Mozambique* .
- MIRANDA, M. R., FOSSE-FILHO, J. C. Y AMARAL, A. A. (2010). Parámetros físico-químicos da água e qualidade microbiológica do pescado na criação integrada entre piscicultura e suinocultura. *V Jornada de Iniciação Científica, III Jornada de Iniciação em Desenvolvimento Tecnológico e Inovação* .
- MOHAMMADIAN, F. Y AMIRI, A. (2012). Economic-Statistical Design of Acceptance Control Chart. *Quality and Reliability Engineering International* , 53–61.
- MONTGOMERY, D. (2001). *Introduction to statistical quality control*. 4end edn, John Wiley, New York.
- MONTGOMERY, D. Y MASTRANGELO, C. (1991). Some statistical process control methods for autocorrelated data. *Journal of Quality Technology*, v. 23, p. 179–193 .
- MORETTIN, P. Y TOLOI, C. (2004). *Análise de séries temporais*. 1.ed. São Paulo: Editora Edgard Blucher Ltda., 535p .
- MORLES (1996). *Manual para Monitoramento Hidrobiológico em Fazendas de cultivo de camarão*. Recife.
- NICOVITA (2002). *Manual de Crianza de Tilapia*. Disponible en <http://www.nicovita.com.pe>.

- ONG, Z. (2009). Structural damage detection using autoregressive-model-incorporating multivariate exponentially weighted moving control chart. *Elsevier Ltd* .
- OTOOLE, A., ABDI, H., DEFFENBACHER, K. Y VALENTIN, D. (1993). A low dimensional representation of faces in the higher dimensions of the space. *J Opt Soc Am [Ser A]* , 405–411.
- PAGE, E. (1954). Continuous Inspection Schemes. *Biometrika* **vol.41**(1).
- PALANI, S., LIONG, S.-Y. Y TKALICH, P. (2008). An ann application for water quality forecasting. *Marine Pollution Bulletin* .
- PATEL, A. Y DIVECHA, J. (2013). Modified MEWMA Control Scheme for an Analytical Process Data. *Volume 13 Issue 3 Version 1.0, Global Journals Inc. (USA)*. .
- PERES-NETO, P., JACKON, D. Y SOMERS, K. (2005). How many principal components ? stopping rules for determinin the number of non-trivial axes revisited. *Comput Stat DataAnal* , 974–997.
- QIN, S. Y MCAVOY, T. (1996). Nonlinear fir modeling via a neural net pls approach. *Computers & chemical engineering*, 20(2) , 147–159.
- QUENOUILLE, M. (1956). Notes on bias and estimation. *Biometrika* , 353–360.
- RATO, T. Y REIS, M. (2013). Defining the structure of dpca model and its impact on process monitoring and prediction activities. *Chemometrics and Intelligent Laboratory Systems* .
- REYNOLDS-JR., M. Y LU, C. (1997). Control charts for monitoring processes with autocorrelated data. *Nonlinear Analysis, Theory, Methods & Applications*. v. 30, p. 4059-4067 .
- ROBERTS, S. (1959). Control chart tests based on geometric moving average. *Technometrics* , 239–250.
- SHEWHART, W. A. (1931). Economic Control of Quality of Manufactured product. *New York: D. Van Nostrand Co., Inc.* .
- SHLENS, J. (2005). A Tutorial on Principal Component Analysis. *Version 2* .
- SOTOLU, A. Y FATUROTI, E. (2009). Growth performance and haematology of Clarias gariepinus (Burchell, 1822) fed varying inclusions of Leucaena leucocephala seed meal based-diets. *ARBOR Ciencia, Pensamiento y Cultura, CLXXXIII 727 septiembre-octubre* .
- SUPERVILLE, C. Y ADAMS, B. (1994). An evaluation of forecast-based quality control schemes. *Communications in Statistics-Simulation and computation* v. 23, p. 645-661 .

- TENENHAUS, M. (1998). La Regression PLS: Theorie et Pratique. *Technip, Paris* .
- TESTIK, M. (2004). Model Inadequacy and Residuals Control Charts for Autocorrelated Processes. *Wiley InterScience* .
- TSAY, R. (2009). Transfer Function Model. *Spring Quarter* .
- UNIVERSITY, S. (2009). Half Of Fish Consumed Globally Is Now Raised On Farms, Study Finds. *ScienceDaily. Disponible en www.sciencedaily.com.releases* .
- VAN-BRACKLE-III, L. Y REYMOLDS-JR., M. (1997). Ewma and cusum control charts in the presence of correlation. *Communication in Statistics-Simulation and computation* , v. 26, p. 979-1008 .
- VANHATALO, E. Y KULAHCI, M. (2015). Impact of correlation on principal components and their used in statistical process control. *Qual. Reliab. Eng.Int.32 (4)* .
- VANHATALO, E., KULAHCI, M. Y B.BERGQUIST (2017). On the structure of dynamic principal component analysis used in statistical process monitoring. *Chemometrics and Intelligent Laboratory Systems* , 1–11.
- VASILOPOULOS, A. Y STAMBOULIS, P. (1978). Modification of Control Chart Limits in the Presence of Data Correlation. *Journal of Quality Technology* .
- VINATEA (2007). *Manual para Monitoramento Hidrobiológico em Fazendas de cultivo de camarão*. Recife.
- WARDELL, D., MOSCOWITZ, H. Y R.D, P. (1994). Run-Length distributions of special-cause control charts for correlated processes. *Technometrics*, v. 36, p. 3-17 .
- WEST, D. Y DELLANA, S. (2011). An empirical analysis of neural network memory structures for basin water quality forecasting. *International Journal of Forecasting. Elsevier Ltd* .
- WIEL, S. V. (1996). Monitoring processes that wander using integrated moving average models. *Technometrics*, v. 38, p. 139-15 .
- WOLD, H. (1982). Soft modelling, The basic design and some extensions, in: K.G. Joreskog, H. Wold Eds., Systems Under In direct Observation. *North-Holland, Amsterdam vols. I and II*, 211–228.
- WOLD, S. (1978). Cross-validatory estimation of the number of components in factor and principal component models. *Technometrics*, 20(4) , 397–405.

- WOLD, S., ESBENSEN, K. Y GELADI, P. (1987). Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 2(1-3) , 37–52.
- WOLD, S., JOHANSSON, E. Y COCCHI, M. (1993). PLS-partial least squares Z. projections to latent structures, in: H. Kubinyi Ed., 3D QSAR in Drug Design, Theory, Methods, and Applications. *ESCOM Science Publishers, Leiden* , 523–550.
- WOLD, S., MARTENS, H. Y RUSSWURM, H. (1983). Food research and data analysis. *Applied Science Publishers, London* .
- WOLD, S., RUHE, A., WOLD, H. Y DUNN-III, W. (1984). The collinearity problem in linear regression, The partial least squares approach to generalized inverses. *SIAM J. Sci. Stat. Comput.* 5 Z , 735–743.
- XU, H. Y ZHANG, Y. (2014). Aquiculture remote monitoring system based on android platform. *International Conference on Green Materials and Environmental Engineering, GMEE 2014; Hong Kong; China; Code 110619* , 956–961.
- YAFFEE, R. Y MCGEE, M. (1999). *Introduction to Time Series Analysis and Forecasting with Applications of SAS and SPSS*. ACADEMIC PRESS, INC. San Diego London Boston New York Sydney Tokyo Toronto.
- YAN, B. Y SHI, P. (2014). Intelligent monitoring system for aquiculture based on internet of things. *Nongye Jixie Xuebao. Transactions of the Chinese Society for Agricultural Machinery* **Vol.45**, 259–265.
- ZHOU, L. Y XING, Y. (2013). Intelligent aquaculture monitoring system based on fieldbus. *International Conference on Materials Science, Machinery and Energy Engineering, MSMEE* .
- ZHOU, R., QIN, X., PENG, S., SHI, H. Y DENG, S. (2014). Macroinvertebrate investigation and their relation to environmental factors in Bohai Bay. *Shengtai Xuebao. Acta Ecologica Sinica* **Vol. 34**, 50–58.

Cápítulo **A**

APÉNDICES: Figuras

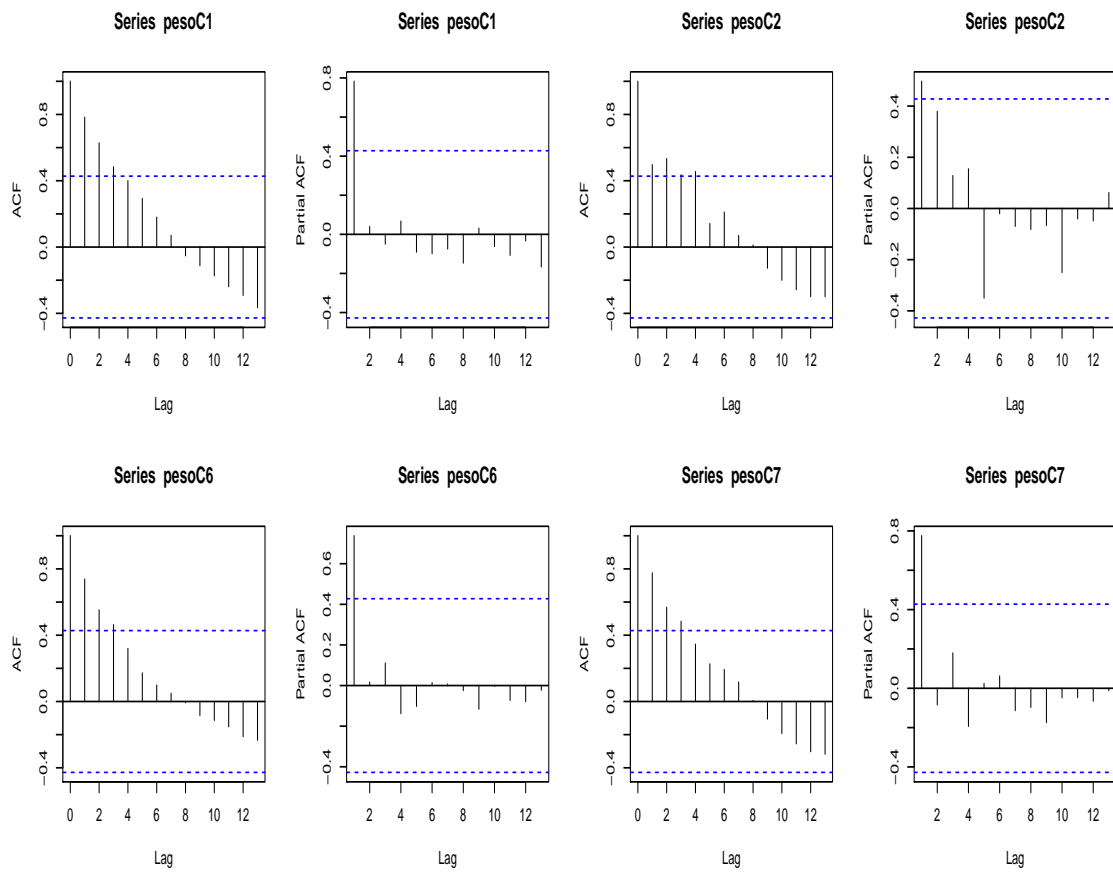


Figura A.1: Funciones de correlaciones simple y parciales del peso en los cuatro tratamientos

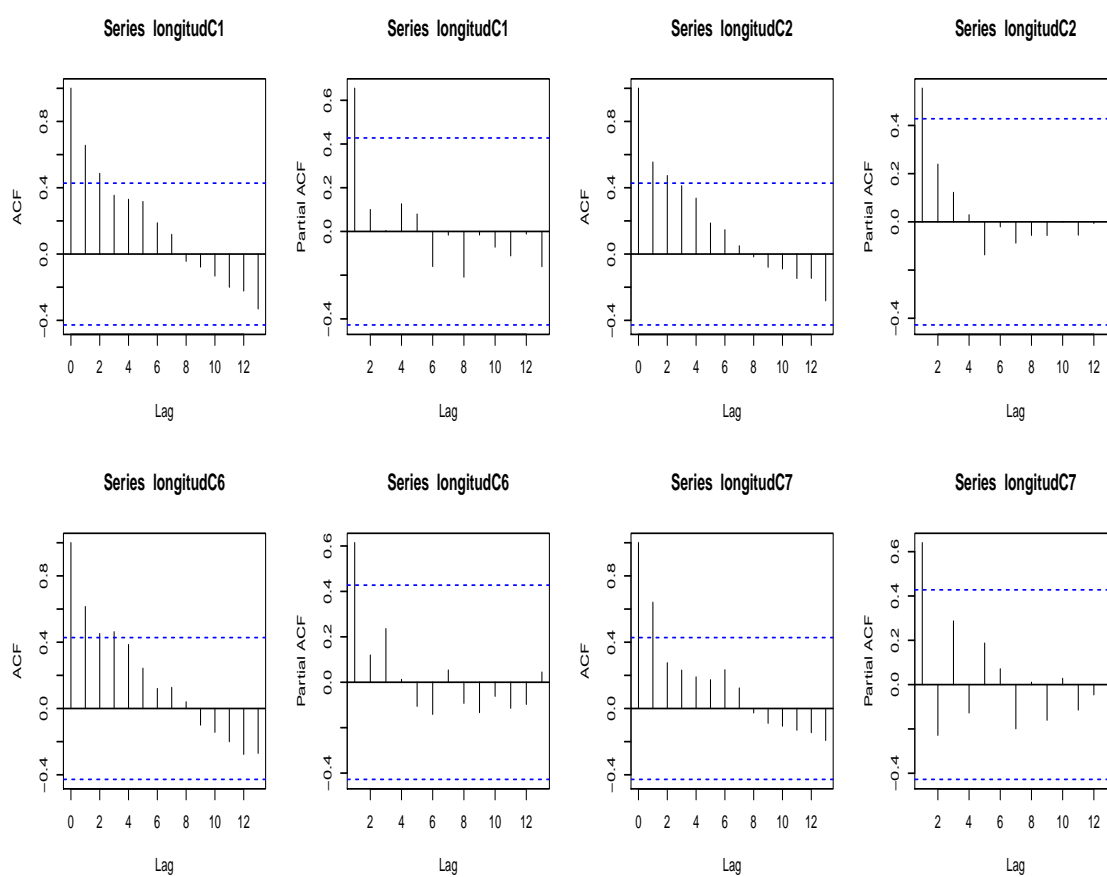


Figura A.2: Funciones de correlaciones simple y parciales de la longitud en los cuatro tratamientos

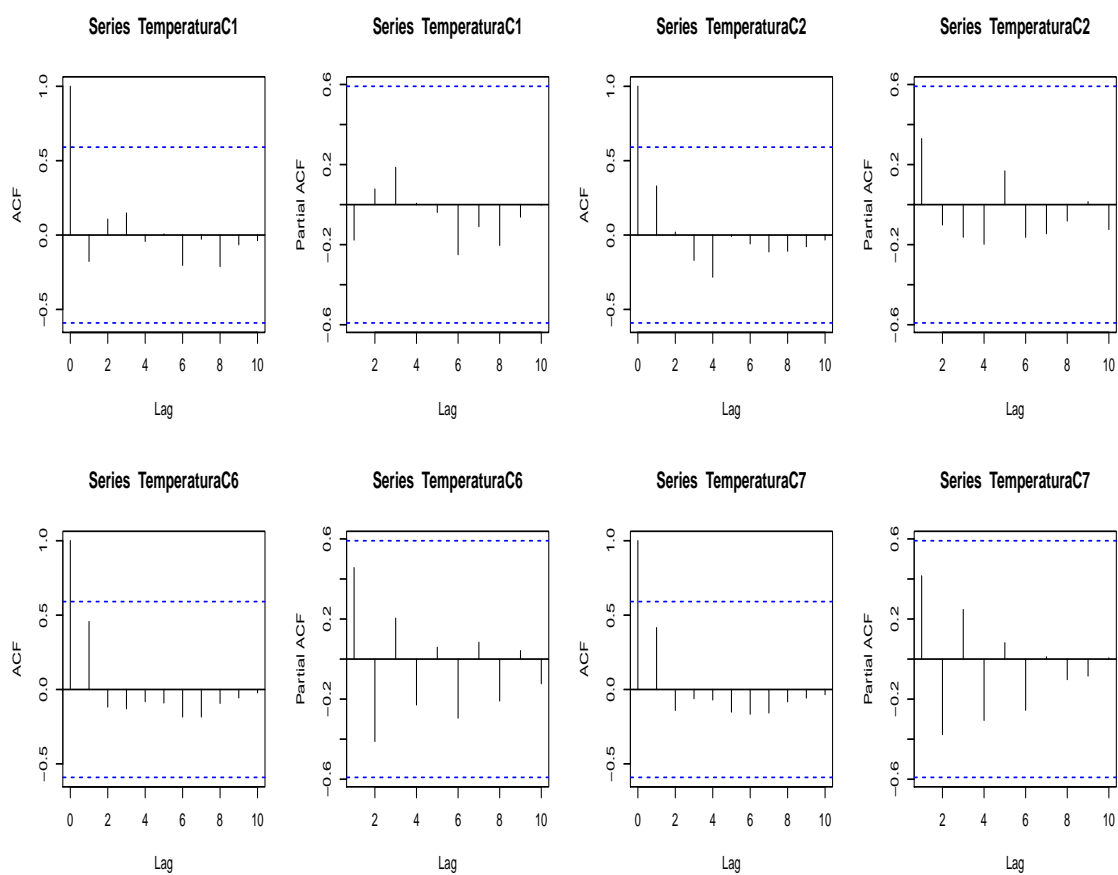


Figura A.3: Funciones de correlaciones simple y parciales de la temperatura en los cuatro tratamientos

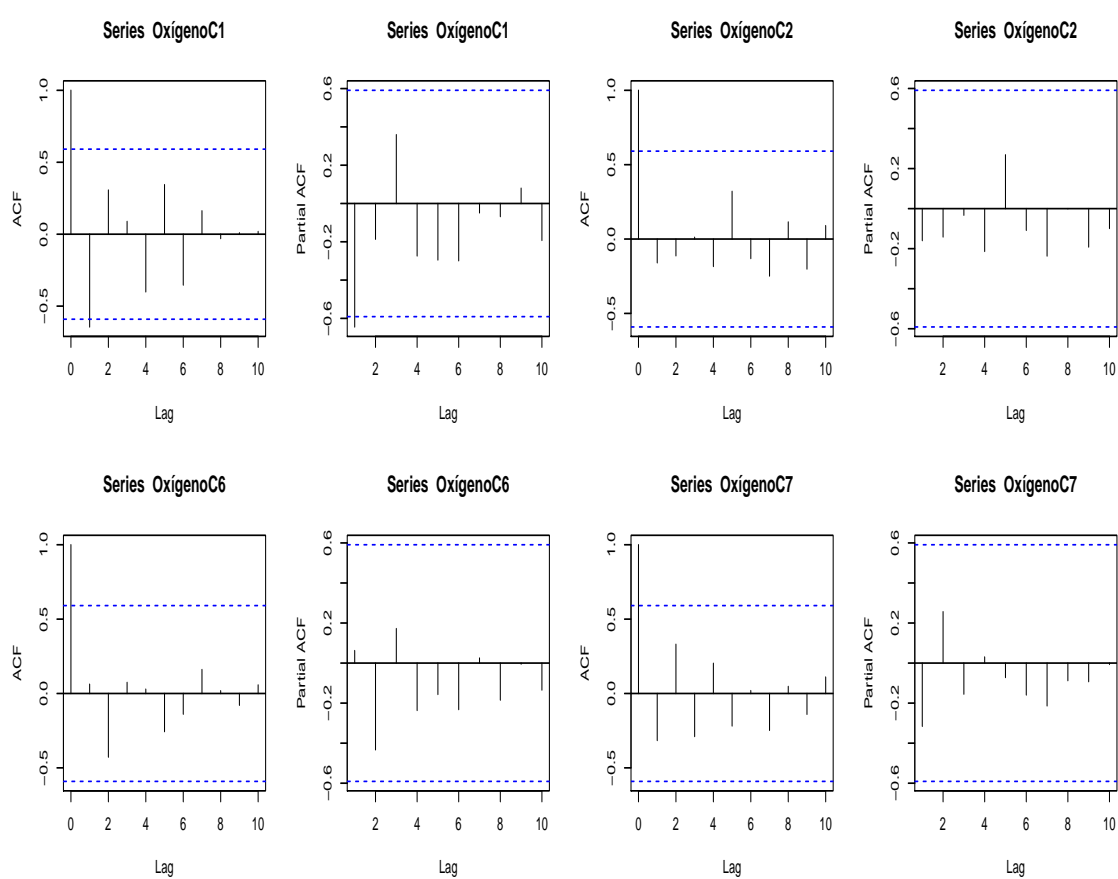


Figura A.4: Funciones de correlaciones simple y parciales de la oxígeno en los cuatro tratamientos

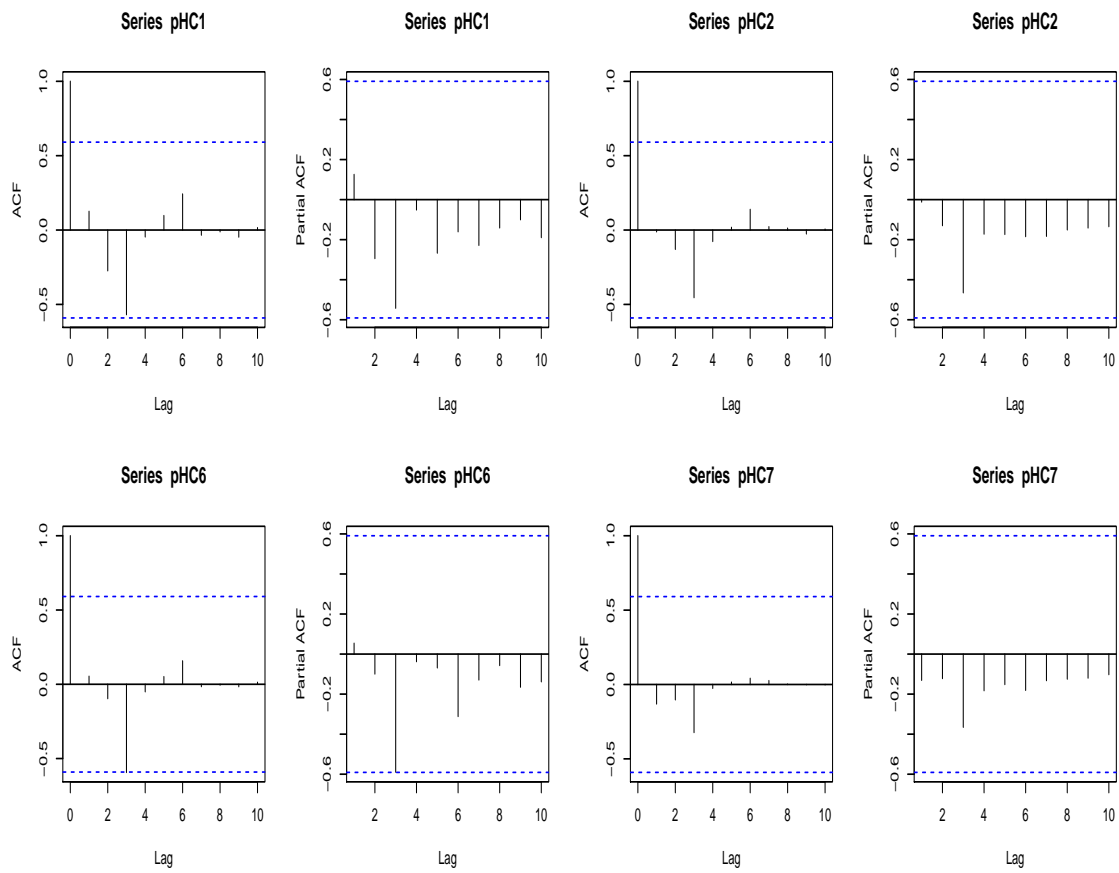


Figura A.5: Funciones de correlaciones simple y parciales de la pH en los cuatro tratamientos

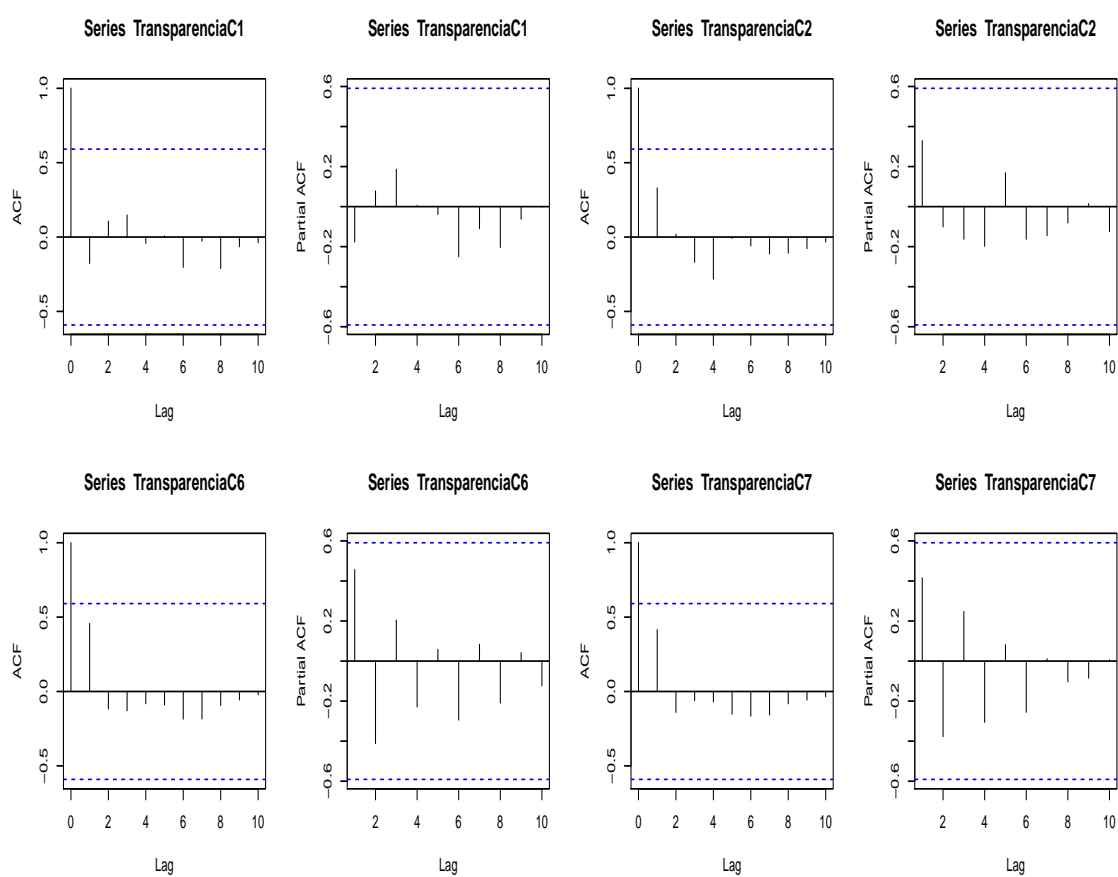


Figura A.6: Funciones de correlaciones simple y parciales de la transparencia en los cuatro tratamientos

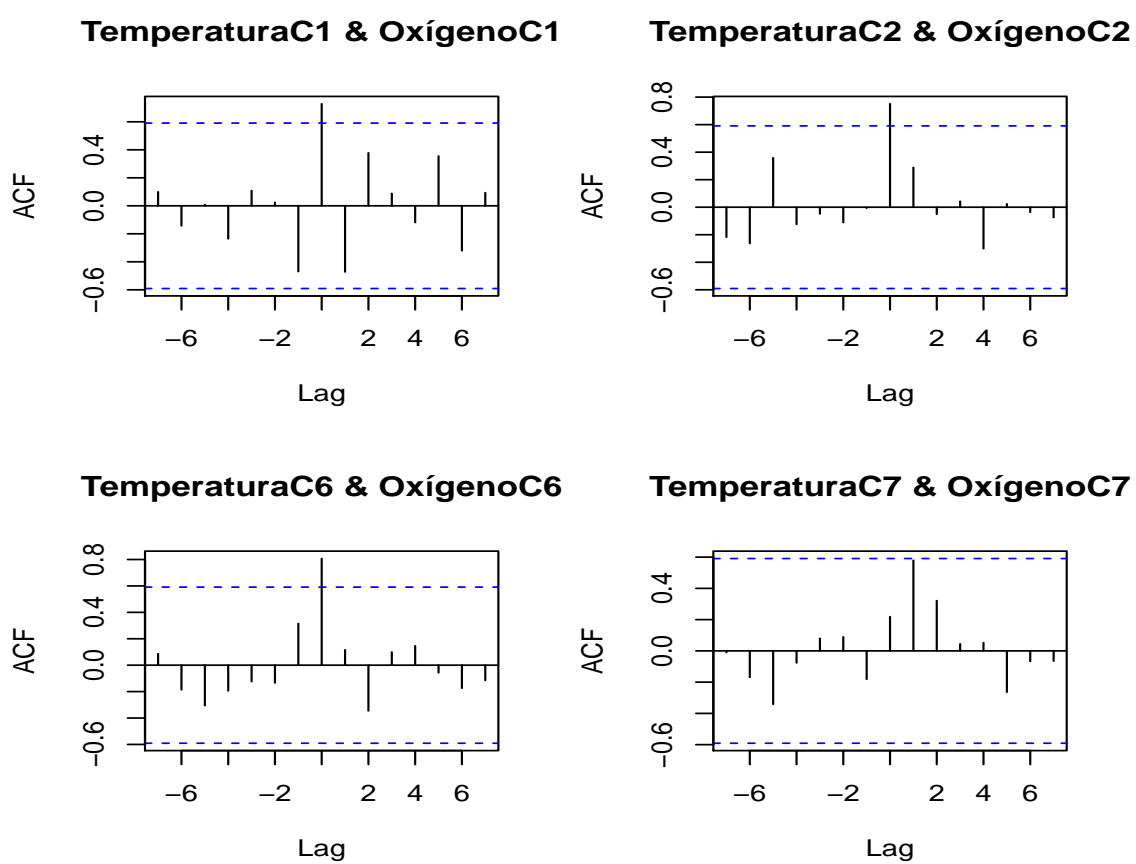


Figura A.7: Función de correlación cruzada entre temperatura y oxígeno en los cuatro tratamientos

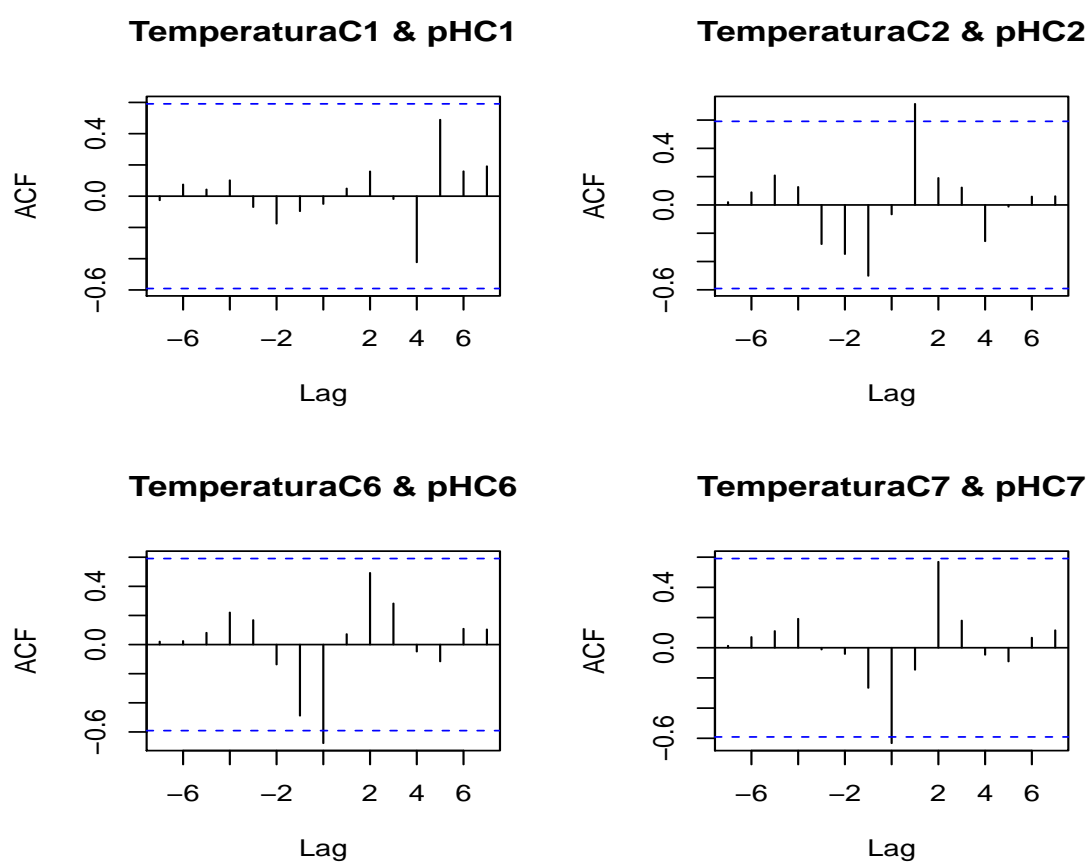


Figura A.8: Función de correlación cruzada entre temperatura y pH en los cuatro tratamientos

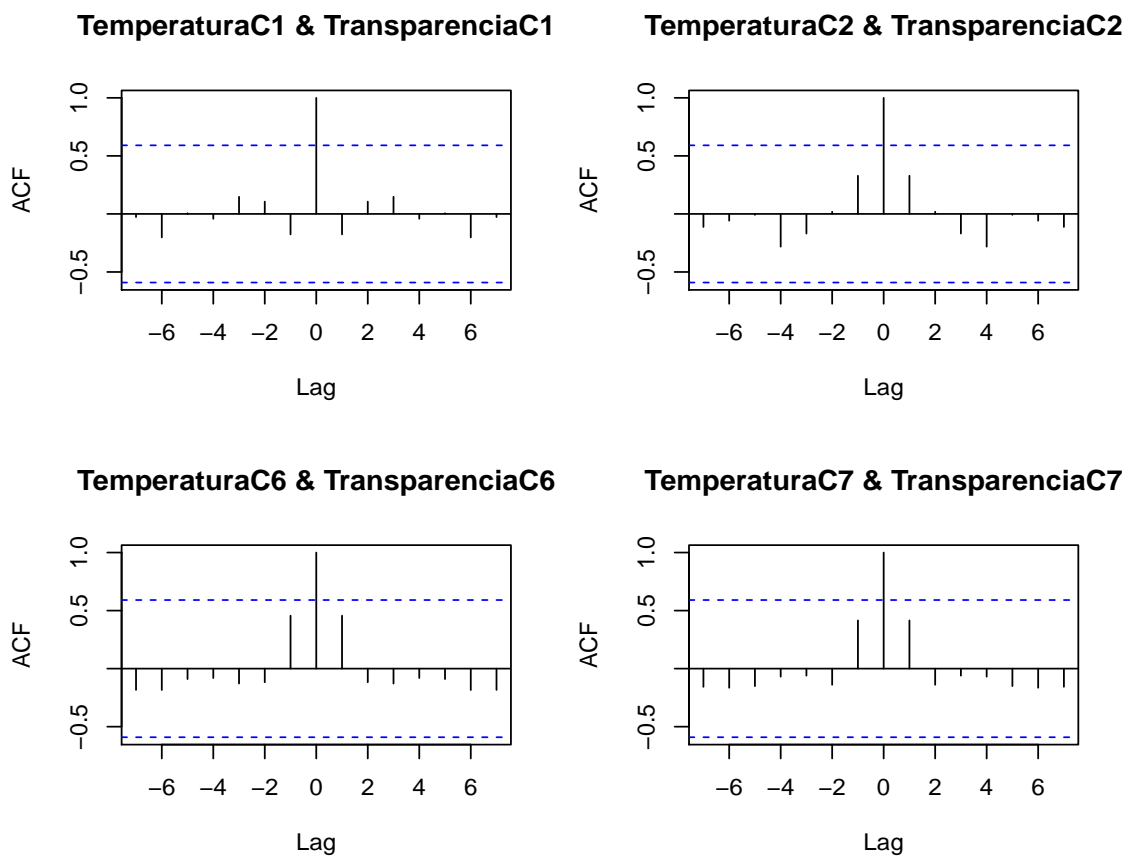


Figura A.9: Función de correlación cruzada entre temperatura y transparencia en los cuatro tratamientos

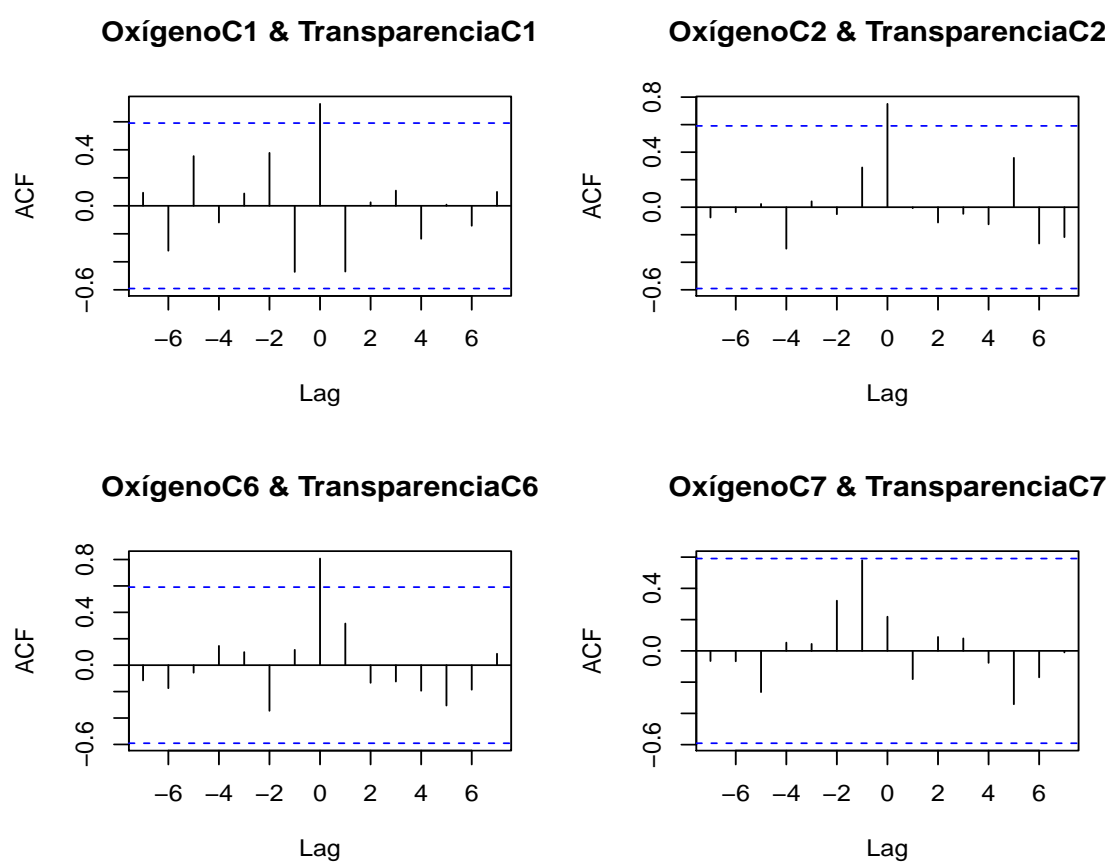


Figura A.10: Función de correlación cruzada entre oxígeno y pH en los cuatro tratamientos

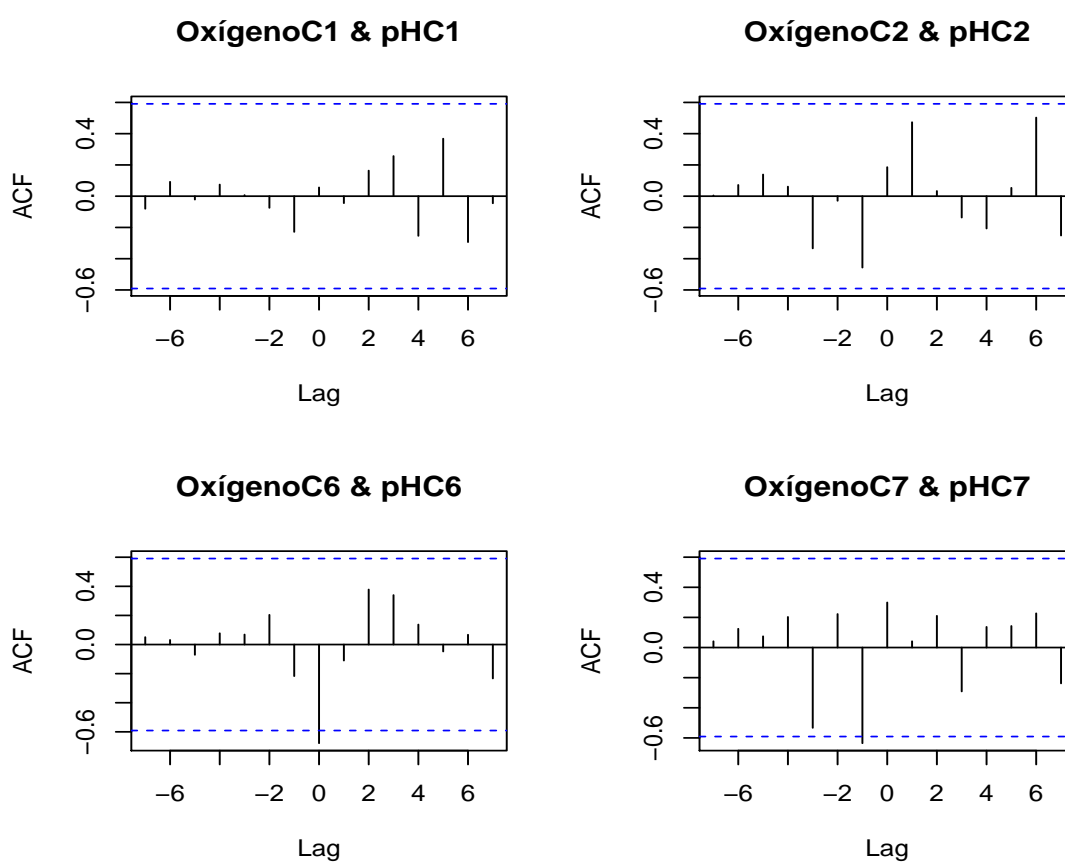


Figura A.11: Función de correlación cruzada entre oxígeno y transparencia en los cuatro tratamientos

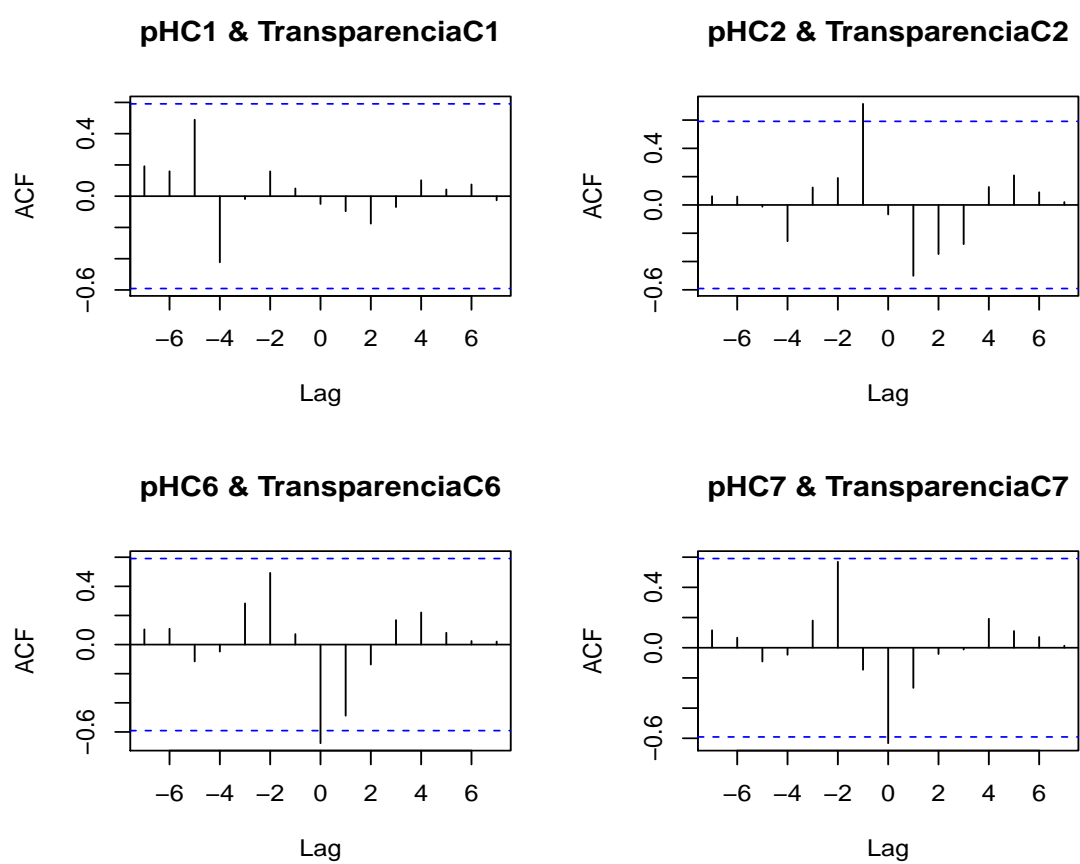


Figura A.12: Función de correlación cruzada entre pH y transparencia en los cuatro tratamientos

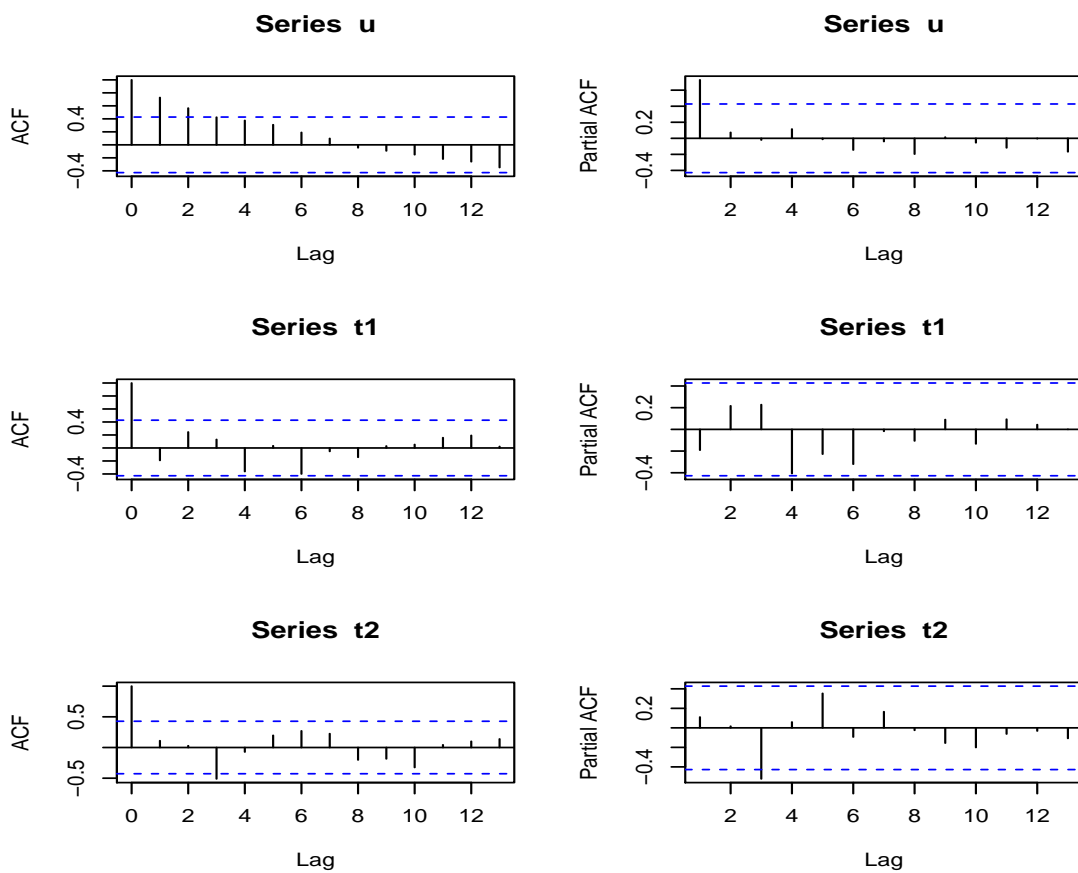


Figura A.13: Correlaciones simple y parciales de las variables latentes en el tratamiento C1

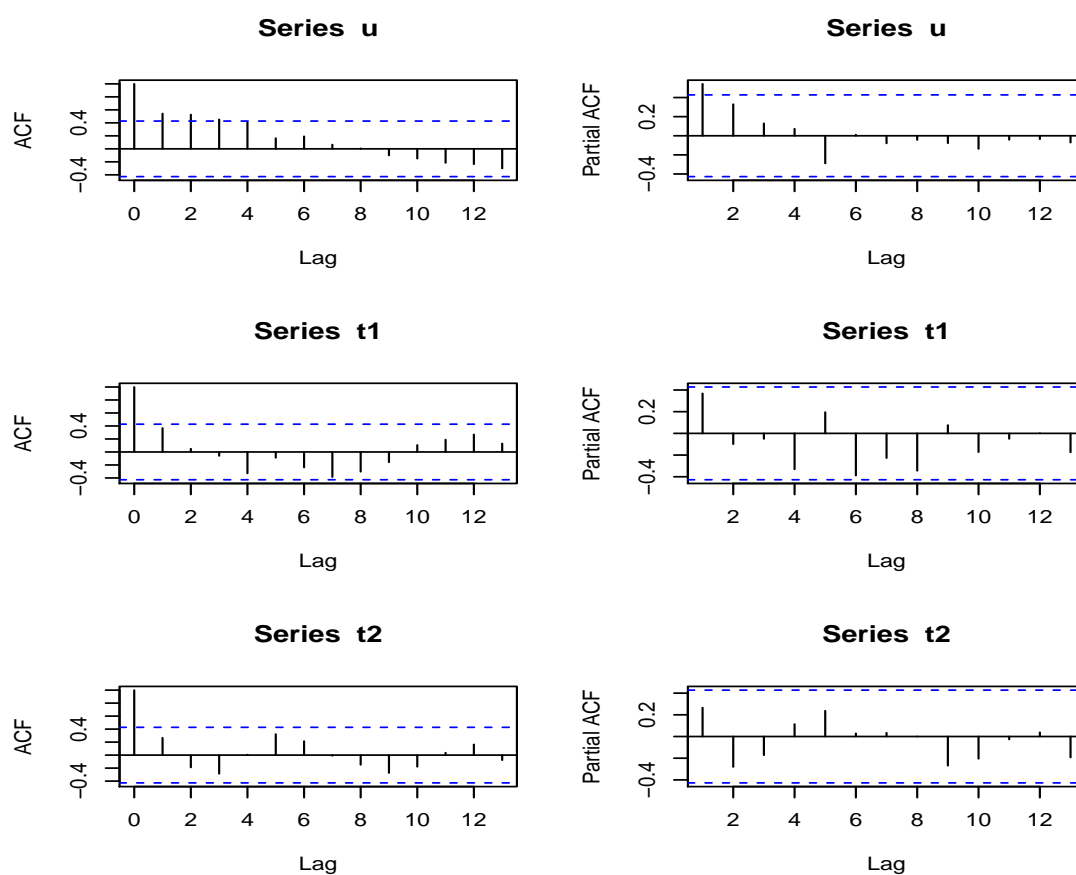


Figura A.14: Correlaciones simple y parciales de las variables latentes en el tratamiento C2

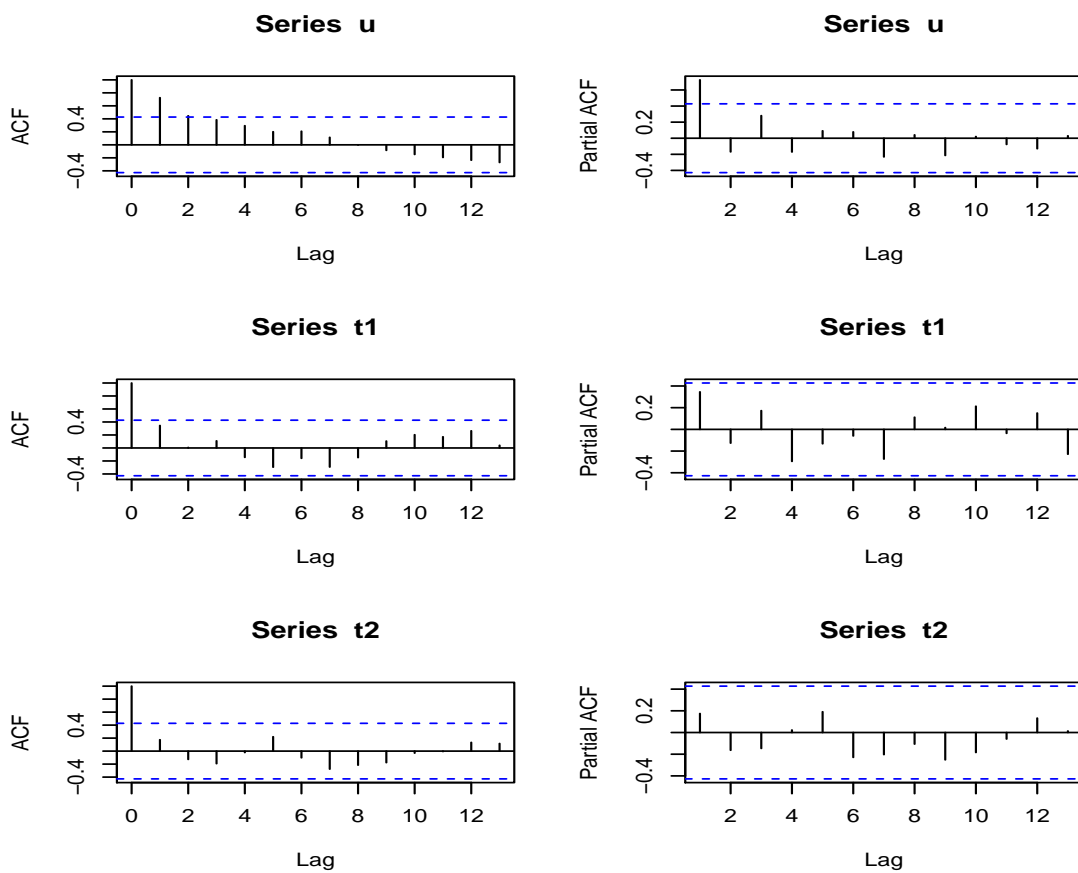


Figura A.15: Correlaciones simple y parciales de las variables latentes en el tratamiento C6

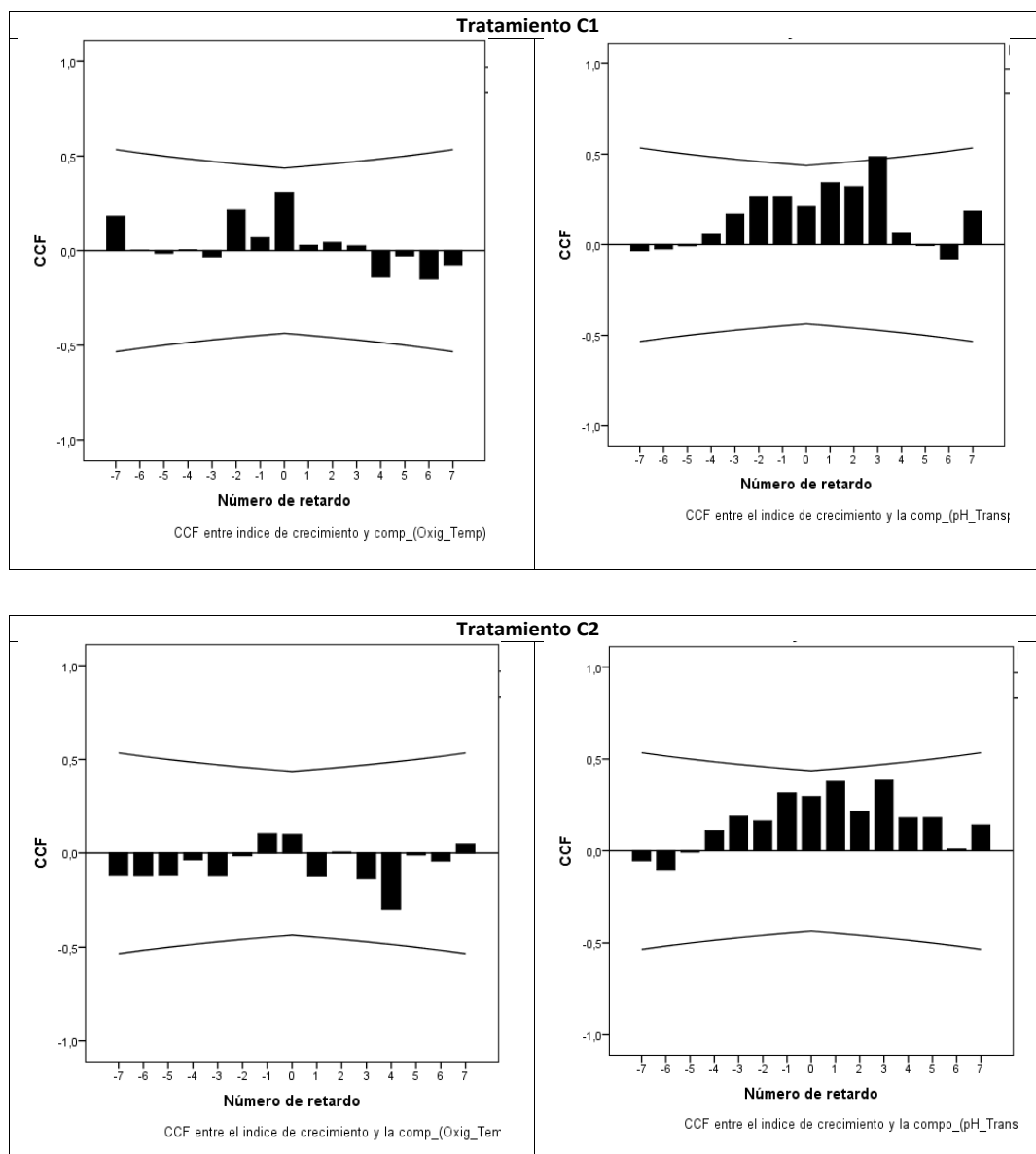


Figura A.16: Función de correlación cruzada entre el índice del crecimiento de la tilapia (u) y las componentes t_1 y t_2) en los tratamientos C1 y C2.

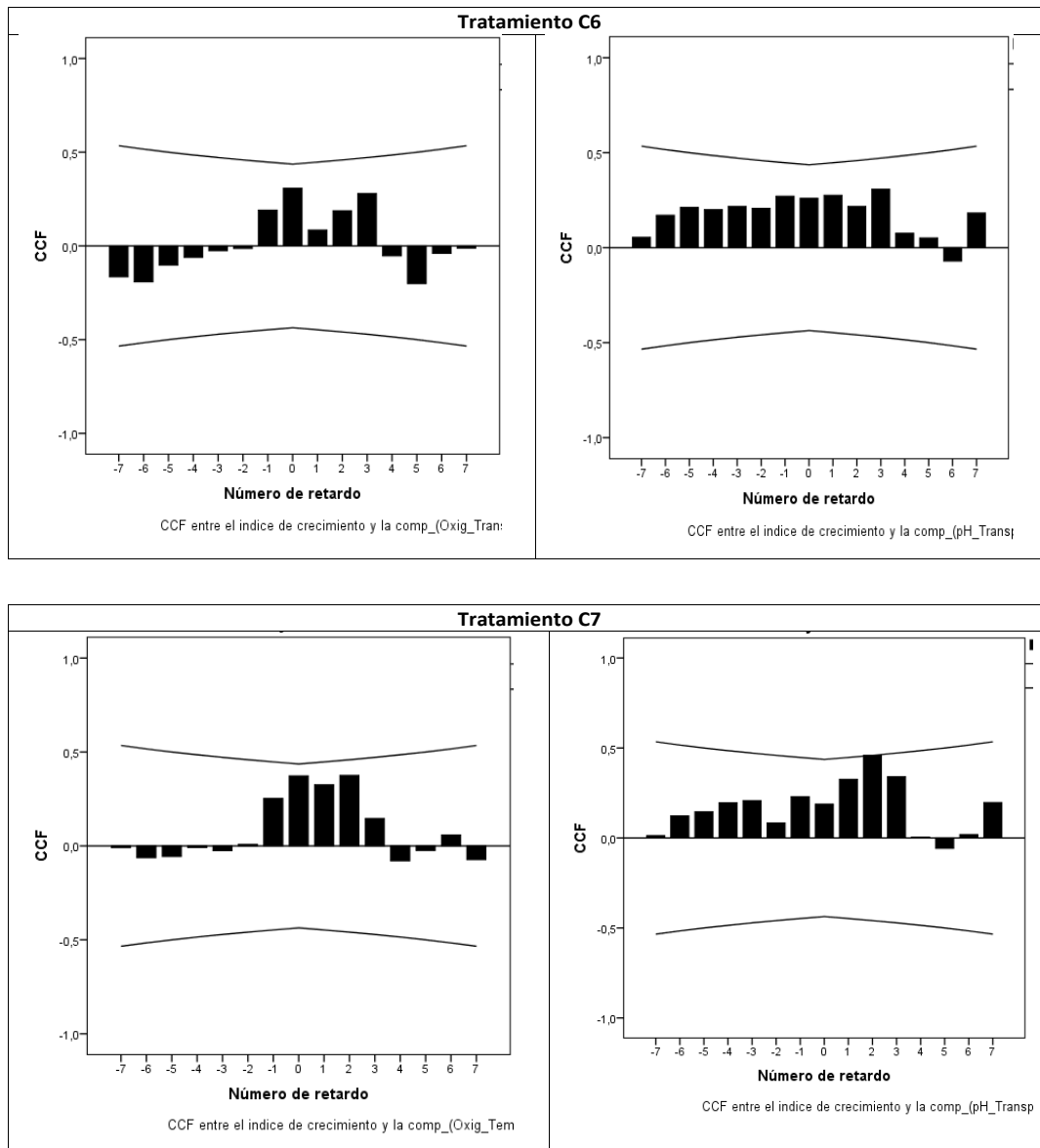


Figura A.17: Función de correlación cruzada entre el índice del crecimiento de la tilapia (u) y las componentes t_1 y t_2 en los tratamientos C6 y C7.

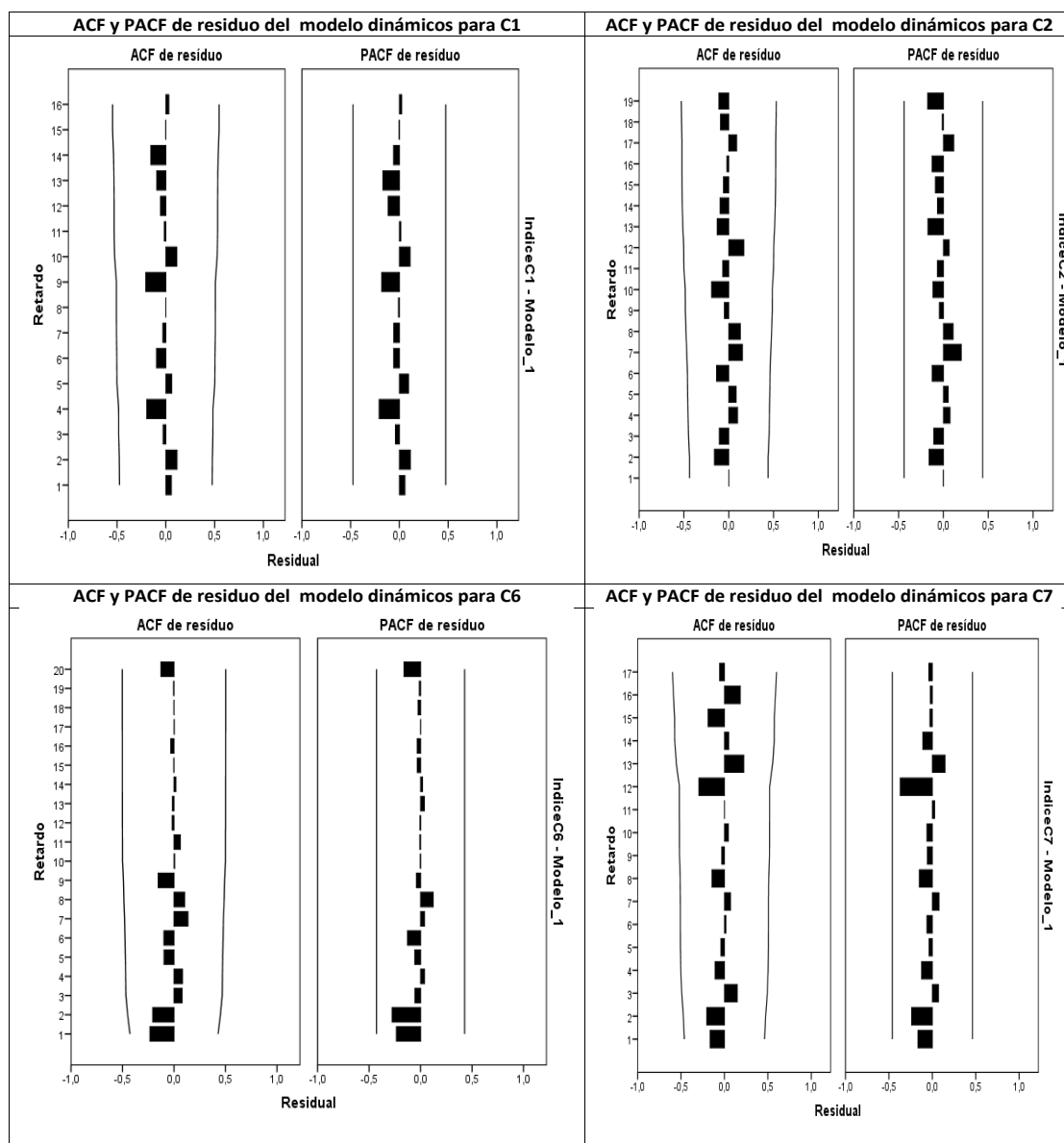


Figura A.18: Función de autocorrelación simple y parcial para los residuos del modelo dinámico sobre estructuras latentes ajustados en los cuatro tratamientos

C apitulo B

C ODIGO R DESARROLLADO

B.1. C odigo R para el ajuste del gr afico MA-AEWMA

```

set.seed(19)
x<- arima.sim(list(order = c(1,0,0), ar=0.9),mean=0,n = 10000)
library(tseries)
arma(x,order = c(1,0,0))
set.seed(19)
y=x+arima.sim(list(order = c(1,0,0),ar=0.9),mean=0,n = 10000)
arma(y,order = c(1,0,0))
rx=residuals(arma(x,order = c(1,0,0)))
ry=residuals(arma(y,order = c(1,0,0)))
rx1=rx[2:10000]
ry1=ry[2:10000]
mean(x)
mean(y)
varx=var(rx1)/(1-0.9^2)
vary=var(ry1)/(1-0.9^2)
cx1=acf(x,lag.max=2,type='covariance',plot=T)
cy1=acf(y,lag.max=2,type='covariance',plot=T)
cx=cx1[[1]][-1]
cy=cy1[[1]][-1]
c1=numeric(length(cx))
for(i in 1:length(c1)){
  c1[i]=(1-(i/length(c1)))*cx[i]
}
Cx=1/(length(c1))*(varx+2*sum(c1))
c2=numeric(length(cy))
for(i in 1:length(c2)){
  c2[i]=(1-(i/length(c2)))*cy[i]
}
Cy=1/(length(c2))*(vary+2*sum(c2))
c1=ccf(x,y,lag.max=2,type='covariance',plot=T)

```

```

cc=cc1[[1]][-3]
cf=numeric(length(cc))
for(i in 1:length(cf)){
  cf[i]=(1-(i/length(cf)))*cc[i]
}
CCF=(1/length(cf))*(cov(x,y)+2*sum(cf))
tiempo=1:10000
m0=c(mean(x),mean(y))
a=c(Cx,CCF)
b=c(CCF,Cy)
X=(0.1/(2-0.1)+2*0.1*(1-0.1)/(2-0.1))*rbind(a,b)
inversa=solve(X)
Y=cbind(x,y)
z=numeric(length(Y))
dim(z)=c(10000,2)
t=numeric(length(Y))
for(i in 1:length(tiempo)){
  if(i==1) z[i,]=0.1*Y[1,]+(1-0.1)*m0+Y[i,]
  else if (i>1) z[i,]=0.1*Y[i,]+(1-0.1)*z[i-1,]+(Y[i,]-Y[i-1,])
}
t=numeric(10000)
for(i in 1:length(tiempo)){
  if(i==1) t[i]=t(z[1,])%%inversa%%z[1,]
  else if (i>1) t[i]=t(z[i,])%%inversa%%z[i,]
}
t
#Calculo del h3 usado en el limite de control
#g=2*UCL/(2*m+1) el ancho de los m+1 estados transitorios considerados.
#p(i,0)=P[X^2(p,c)<((0.5^2)*g^2/lambda^2)]=0.99973 (alpha=1-p(i,0)=0.00027)
la probabilidad de que una observación esté bajo control
#((0.5^2)*g^2/lambda^2)=16,2
ARL=370
chi=16.2
lambda=0.1
m=6
g=(sqrt(chi)*lambda)/0.5
UCL=g*(2*m+1)/2
h3=(UCL^2)/((0.1/1.9)+2*0.1*(0.9/1.9))
#win.graph()
plot(tiempo[1:100],t[1:100],type='l',ylim=c(0,h3+1))
abline(h=h3,lty=3)
#text((max(tiempo)+1),UCL,paste('UCL=',UCL),pos=3,font=2,cex=0.8)

###Simulación del cambio del proceso a partir de la observación 50

```

```

set.seed(19)
x=numeric(1000)
for(i in 1:1000){
  if (i<50)x[i]<- arima.sim(list(order = c(1,0,0), ar=0.9),mean=0,n =1)
  else x[i]<- arima.sim(list(order = c(1,0,0), ar=0.9),mean=1,n =1)
}
set.seed(19)
y<-x+arima.sim(list(order = c(1,0,0), ar=0.9),mean=0,n = 1000)
m0=c(mean(x),mean(y))
a=c(Cx,CCF)
b=c(CCF,Cy)
X=(0.1/(2-0.1)+2*0.1*(1-0.1)/(2-0.1))*rbind(a,b)
inversa=solve(X)
x1=array(x,dim=c(100,10))
y1=array(y,dim=c(100,10))
Y=cbind(x1[,1],y1[,1])
z=numeric(length(Y))
dim(z)=c(100,2)
for(i in 1:100){
  if(i==1) z[i,]=0.1*Y[1,]+(1-0.1)*m0+Y[1,]
  else if (i>1) z[i,]=0.1*Y[i,]+(1-0.1)*z[i-1,]+(2-0.1)*(Y[i,]-Y[i-1,])
}
t1=numeric(10000)
for(i in 1:length(tiempo)){
  if(i==1) t1[i]=t(z[1,])%*%inversa%*%z[1,]
  else if (i>1) t1[i]=t(z[i,])%*%inversa%*%z[i,]
}
t1
lambda=0.1
m=6
g=(sqrt(chi)*lambda)/0.5
UCL=g*(2*m+1)/2
h3=(UCL^2)/((0.1/1.9)+2*0.1*(0.9/1.9))
win.graph()
par(mfrow=c(1,2))
tiempo=1:100
#plot(tiempo,H,type='l',ylim=c(0,UCL+500))
H=data.frame(t[1:100],t1[1:100])
matplot(tiempo,H,type=c("o","l"),
  main=(expression(phi('pi')==+.9&phi(pj)==+.9&delta==1)),
  lty=1,xlab="MAAEWMA chart (a)",ylab="T2",col=c("black","black"),
  ylim=c(0,UCL+1500),lwd=c(1,3), pch=".")
legend(0,1750,bty="n",
  c("in-control process",

```

```

    "out-of-control starting at the ob 50"),
    border="black",lty=c(7,7),lwd=c(1,3))
abline(h=h3,lty=3)
text((min(tiempo)+1),h3,paste('h3=',h3),pos=3,font=2,cex=0.8)
RL=which(t1[50:100]>h3)
length(RL)
c=numeric(length(RL))
for(i in 1:length(RL)){
  if(i==1) c[i]=RL[1]
  else if(i>1) c[i]=RL[i]-RL[i-1]
}
ARL=sum(c)/length(c)
length(RL)
print(ARL)

```

B.2. Código R para el ajuste del modelo dinámico sobre estructuras latentes (MDEL)

```

###MODELO DE PREVISION
library(MASS)
library(tseries)
library(forecast)
library(MTS)
library(timeSeries)
library(devtools)
library(TSA)
library(PCA4TS)
library(C50)#Para validación cruzada
library(rpart)#Para validación cruzada
library(caret)
library(mlbench)
###BLOQUE DE X
#Determinación de la matriz de datos de entrada
set.seed(6)
x<-arima.sim(list(order = c(1,0,0), ar = 0.9), n = 500)
x1=matrix(x,100,5)
X=numeric(500)
dim(X)=c(100,5)
for (i in 1:100){
  for(j in 1:5){
    if(j==1) X[,1]=x1[,1]
    else if (j>1) X[i,j]=X[i,j-1]-x1[i,j]
  }
}

```


B.2. Código R para el ajuste del modelo dinámico sobre estructuras latentes

```
}
# La matriz que recoge las variables de X decaladas en orden 1

Z=numeric(500)
dim(Z)=c(100,5)
for(i in 1:100){
  if(i==1) Z[1,]=0
  else if (i>1) Z[i,]=X[i-1,]
}

Xs=data.frame(X,Z)
PCA1=prcomp(Xs,rank. = 2)
summary(PCA1)
scores=PCA1$x
t1=scores[,1]
t2=scores[,2]

win.graph()
par(mfrow=c(2,2))
#plot(c(1:100),c1,type="l",xlab = "Time")
#plot(c(1:100),c2,type="l",xlab = "Time")
acf(c1,50,main="t1")
pacf(c2,50,main="t1")
acf(c1,50,main="t2")
pacf(c2,50,main="t2")

##BLOQUE Y
#Determinación de la matriz de datos de salida
set.seed(6)
y=numeric(500)
for(i in 1:500){
  if(i==1) y[i]<-arima.sim(list(order = c(1,0,0), ar = 0.9),n=1)
  else if (i>1) y[i]<-arima.sim(list(order = c(1,0,0),
                                     ar=0.9),n = 1)+3*x[i-1]
}

y1=matrix(y,100,5)
Y=numeric(500)
dim(Y)=c(100,5)
for (i in 1:100){
  for(j in 1:5){
    if(j==1) Y[,1]=y1[,1]
    else if (j>1) Y[i,j]=Y[i,j-1]-y1[i,j]
```

```

    }
  }
  # La matriz que recoge las variables de Y decaladas (lag=1)
  Z1=numeric(500)
  dim(Z1)=c(100,5)
  for(i in 1:100){
    if(i==1) Z1[1,]=0
    else if (i>1) Z1[i,]=Y[i-1,]
  }
  Ys=data.frame(Y,Z1)

  PCA=prcomp(Ys,rank. = 2,retx = TRUE)
  summary(PCA)
  scores=PCA$x
  u1=scores[,1]
  u2=scores[,2]

  win.graph()
  par(mfrow=c(2,2))
  #plot(c(1:100),c1,type="l",xlab = "Time")
  #plot(c(1:100),c2,type="l",xlab = "Time")
  acf(u1,50)
  pacf(d1,50, main="u1")
  acf(u2,50)
  pacf(u2,50)

  ##Modelo
  library(tseries)
  #Para t1
  mt1=arima(t1,order = c(1,0,0))
  res_t1=residuals(mt1)

  #Para t2
  mt2=Arima(t2,order = c(1,0,0))
  res_t2=residuals(mt2)
  #Para u1
  mu1=Arima(u1,model=mt1)
  res_u1=residuals(mu1)

  #Para u2
  mu2=Arima(u2,model=mt2)
  res_u2=residuals(mu2)

  ##Modelo de la funcion de transferencia

```

```

#Identificación de la función de transferencia a través de la CCF
win.graph()
par(mfrow=c(1,2))
ccf(res_u1,res_t1,main="a1 & a1*", ylab="CCF")
abline(v=0,lty=3)
ccf(res_u2,res_t2,main="a2 & a2*",ylab="CCF")
abline(v=0,lty=3)

#Modelo para u1=f(t1)
###AJUSTE DE MODELO DE TRANSFERENCIA
#MODELO PARA u1=f(t1)
datos=data.frame(u1,t1)
#Entrenamiento y prueba del modelo
set.seed(1)
Folds<- 2
datos$kfold<- sample(1:Folds, nrow(datos), replace = TRUE)
for (i in 1:Folds)
{
  Test          <- subset(datos, kfold == i)
  Entrenamiento <- subset(datos, !kfold == i)
  Modelo        <- arimax(Entrenamiento$u1,order=c(1,1,0),
                          xtransf=Entrenamiento$c1,
                          transfer=list(c(0,2)))
  fit.train     <- Arima(Entrenamiento$u1,order=c(1,0,0),
                        xreg=Entrenamiento$t1)
  fit.test      <- Arima(Test$u1,xreg=Test$t1, model = fit.train,
                        n.ahead = 1,se.fit = TRUE)
}
#SIGNIFICACIÓN DEL MODELO
require(lmtest)
coeftest(Modelo)

#CALCULO DE RMSE
resid=residuals(fit.test)
resid1=as.numeric(scale(resid))
rmse=numeric(length(resid1))
mae=numeric(length(resid1))
for(i in 1:length(rmse)){
  rmse[i]=(resid1[i])^2
  suma=sum(rmse)
  RMSE=suma/length(rmse)
  mae[i]=abs(resid1[i])
  MAE=sqrt((sum(mae)))/length(rmse)
}

```

MAE

```

#MODELO PARA u2=f(t2)
#Entrenamiento y prueba del modelo
datos1=data.frame(u2,t2)
set.seed(1)
Folds<- 2
datos1$kfolds<- sample(1:Folds, nrow(datos1), replace = TRUE)
for (i in 1:Folds)
{
  Test1          <- subset(datos1, kfolds == i)
  Entrenamiento1 <- subset(datos1, !kfolds == i)
  Modelo1        <- arimax(Entrenamiento1$u2,order=c(1,1,0),
                           xtransf=Entrenamiento1$t2,
                           transfer=list(c(0,1)))

  fit.train1     <- Arima(Entrenamiento1$u2,order=c(1,0,0),
                           xreg=Entrenamiento1$t2)
  fit.test1      <-Arima(Test1$u2,xreg=Test1$t2,
                           model = fit.train1,
                           n.ahead = 1,se.fit = TRUE)
}
coeftest(Modelo1)
#CALCULO DE RMSE
resid2=residuals(fit.test1)
resid21=as.numeric(scale(resid2))
rmse2=numeric(length(resid21))
mae1=numeric(length(resid21))
for(i in 1:length(rmse2)){
  rmse2[i]=(resid21[i])^2
  suma1=sum(rmse2)
  RMSE1=suma1/length(rmse2)
  mae1[i]=abs(resid21[i])
  MAE1=sqrt((sum(mae1)))/length(rmse2)
}
MAE1
#MÉTODO COMPARATIVO (MODELO DPLS)
Xs
Ys
Xs1=data.frame(Xs,Z1)

library(plsdepot)
pls1 = plsreg2(Xs1,Y, comps = 2,crosval = TRUE)
summary(pls1)

```

```

pls1$reg.coefs
pls1$Q2cum
pls1$Q2
pls1$expvar
pls1$y.scores
d1=pls1$y.scores[,1]
d1_1=(c(s1[-1],0))
t1=pls1$x.scores[,1]
t1_1=(c(t1[-1],0))
d2=pls1$y.scores[,2]
d2_1=(c(u2[-1],0))
t2=pls1$x.scores[,2]
t2_1=(c(t2[-1],0))

win.graph()
par(mfrow=c(2,2))
acf(t1,main="t1")
pacf(t1,main="t1")
acf(t2,main="t2")
pacf(t2,main="t2")
win.graph()
par(mfrow=c(2,2))
acf(d1,main="u1")
pacf(d1,main="u1")
acf(d2,main="u2")
pacf(d2,main="u2")

#Modelo u1=f(t1)
datos2=data.frame(d1,d1_1,t1,t1_1)
reg=train(d1~d1_1+t1+t1_1, datos2, method="lm",tuneLength = 2,
          trControl = trainControl(method = "cv"))
summary(reg1)
#Modelo u2=f(t2)
datos21=data.frame(d2,d2_1,t2,t2_1)
reg1=train(d2~d2_1+t2+t2_1, datos21, method="lm",tuneLength = 2,
          trControl = trainControl(method = "cv"))

##GRAFICOS PARA EL MDEL
Modelo u1=f(t1)
modelo_ft1      <- arimax(u1,order=c(3,1,0),
                        xtransf=c1,
                        transfer=list(c(0,2)))

```

```

ajustados_ft1 <- fitted(modelo_ft1)

win.graph()
par(mfrow=c(2,1))
H1=data.frame(u1,ajustados_ft1)
matplot(c(1:100),H1,type="l",lty = c(1,6),
        lwd = c(1.5,2), col = c("black","orangered"),
        xlab ="Tiempo",ylab="series",
        ylim=c(-50,50),main="MDEL")
legend(10,-20,c("observados","predichos"),c("black","red"),
      bty="n",text.col=c("black","red"),horiz=1)
help("legend")
Modelo u2=f(t2)
modelo_ft2      <- arimax(u2,order=c(1,0,0),
                        xtransf=c2,
                        transfer=list(c(1,0)))

residuos2 <- resid(modelo_ft2)
ajustados_ft2 <- fitted(modelo_ft2)
H2=data.frame(u2,ajustados_ft2)
matplot(c(1:100),H2,type="l",lty = c(1,2),
        lwd = c(1,2), col = c("black","orangered"),
        xlab ="Time",ylab = "series")

##GRAFICOS PARA MODELO DPLS
Modelo u1=f(t1) Para DPLS
Modelo1      <- lm(d1~d1_1+t1+t1_1, datos2)
predichos1=predict(Modelo1)
G1=data.frame(d1,predichos1)
win.graph()
par(mfrow=c(2,1))
matplot(c(1:100),G1,type="l",lty = c(1,6),
        lwd = c(1.5,2), main="DPLS",
        col = c("black","orangered"),
        xlab = "Tiempo",ylab = "series",ylim = c(-8,6))
legend(10,-3,c("observados","predichos"),c("black","red"),
      bty="n",text.col=c("black","red"),horiz=1)

#Modelo u2=f(t2)
Modelo2      <- lm(d2~d2_1+t2+t2_1, datos21)
predichos2=predict(Modelo2)
G2=data.frame(d2,predichos2)
matplot(c(1:100),G2,type="l",lty = c(1,2),lwd = c(1,2),
        col = c("black","orangered"), xlab = "Time",ylab = "series")

```

B.3. Código R para el gráfico MAAEWMA: Caso acuicultura

```

##Grafico MAAEWMA para C1
(de manera análoga al código usado en C1, en los tratamientos C2,
C6 y C7 fueron usados este código apenas
cambiando los respectivos datos)
##calculo de la autovarianza de la temperatura en C1
library(tseries)
TempC1=as.numeric(scale(TemperaturaC1))
t=arma(TempC1,order = c(1,0))
rt=residuals(arma(TempC1,order = c(1,0)))
rt1=rt[2:11]
vart=var(rt1)/(1-0.1^2)
ct1=acf(TempC1,lag.max=3,type='covariance',plot=FALSE)
ct=ct1[[1]][-1]
c1=numeric(length(cx))
for(i in 1:length(c1)){
  c1[i]=(1-(i/length(c1)))*ct[i]
}
Ct=1/(length(c1))*(vart+2*sum(c1))
##calculo de la autovarianza del oxigeno en C1
OxíC1=as.numeric(scale(OxígenoC1))
o=arma(OxíC1,order = c(1,0))
ro=residuals(arma(OxíC1,order = c(1,0)))
ro1=ro[2:11]
varo=var(ro1)/(1-0.1^2)
co1=acf(OxíC1,lag.max=1,type='covariance',plot=FALSE)
co=co1[[1]][-1]
c1=numeric(length(co))
for(i in 1:length(c1)){
  c1[i]=(1-(i/length(c1)))*co[i]
}
Co=1/(length(c1))*(varo+2*sum(c1))
##calculo de la autovarianza del pH en C1
PHC1=as.numeric(scale(pHC1))
p=arma(PHC1,order = c(1,0))
rp=residuals(arma(PHC1,order = c(1,0)))
rp1=rp[2:11]
varp=var(rp1)/(1-0.1^2)
cp1=acf(PHC1,lag.max=2,type='covariance',plot=FALSE)
cp=cp1[[1]][-1]
c1=numeric(length(cp))
for(i in 1:length(c1)){

```

```

    c1[i]=(1-(i/length(c1)))*cp[i]
  }
Cp=1/(length(c1))*(varp+2*sum(c1))
##calculo de la autovarianza del transparencia en C1
TranspC1=as.numeric(scale(TransparenciaC1))
tr=arma(TranspC1,order = c(1,0))
rtr=residuals(arma(TranspC1,order = c(1,0)))
rtr1=rtr[2:11]
vartr=var(rtr1)/(1-0.1^2)
ctr1=acf(TranspC1,lag.max=2,type='covariance',plot=FALSE)
ctr=ctr1[[1]][-1]
c1=numeric(length(ctr))
for(i in 1:length(c1)){
  c1[i]=(1-(i/length(c1)))*ctr[i]
}
Ctr=1/(length(c1))*(vartr+2*sum(c1))
## calculo de las covarianzas cruzadas
# covarianza cruzada entre temperatura y oxigeno
cc1=ccf(TempC1,OxíC1,lag.max=2,type='covariance',plot=FALSE)
cc=cc1[[1]][-3]
cf=numeric(length(cc))
for(i in 1:length(cf)){
  cf[i]=(1-(i/length(cf)))*cc[i]
}
CCF=(1/length(cf))*(cov(TempC1,OxíC1)+2*sum(cf))
# covarianza cruzada entre temperatura y pH
cc2=ccf(TempC1,PHC1,lag.max=2,type='covariance',plot=FALSE)
cc1=cc1[[1]][-3]
cf1=numeric(length(cc1))
for(i in 1:length(cf1)){
  cf1[i]=(1-(i/length(cf1)))*cc1[i]
}
CCF1=(1/length(cf1))*(cov(TempC1,PHC1)+2*sum(cf1))
# covarianza cruzada entre temperatura y Transparencia
cc3=ccf(TempC1,TranspC1,lag.max=2,type='covariance',plot=FALSE)
cc2=cc1[[1]][-3]
cf2=numeric(length(cc2))
for(i in 1:length(cf2)){
  cf2[i]=(1-(i/length(cf2)))*cc2[i]
}
CCF2=(1/length(cf2))*(cov(TempC1,TranspC1)+2*sum(cf2))
# covarianza cruzada entre Oxigeno y pH
cc4=ccf(OxíC1,PHC1,lag.max=2,type='covariance',plot=FALSE)
cc3=cc4[[1]][-3]

```



```

cf3=numeric(length(cc3))
for(i in 1:length(cf3)){
  cf3[i]=(1-(i/length(cf3)))*cc3[i]
}
CCF3=(1/length(cf3))*(cov(OxígenC1,PHC1)+2*sum(cf3))
# covarianza cruzada entre Oxígeno y Transparencia
cc5=ccf(OxígenC1,TranspC1,lag.max=3,type='covariance',plot=FALSE)
help("ccf")
cc4=cc5[[1]][-3]
cf4=numeric(length(cc4))
for(i in 1:length(cf4)){
  cf4[i]=(1-(i/length(cf4)))*cc4[i]
}
CCF4=(1/length(cf4))*(cov(OxígenC1,TranspC1)+2*sum(cf4))
# covarianza cruzada entre Transparencia y pH
cc6=ccf(TranspC1,PHC1,lag.max=2,type='covariance',plot=FALSE)
cc5=cc6[[1]][-3]
cf5=numeric(length(cc5))
for(i in 1:length(cf5)){
  cf5[i]=(1-(i/length(cf5)))*cc5[i]
}
CCF5=(1/length(cf5))*(cov(PHC1,TranspC1)+2*sum(cf5))
##Gráfico de control
m=c(0,0,0,0)
A=matrix(c(Ct,CCF,CCF1,CCF2,CCF,
Co,CCF3,CCF4,CCF1,CCF3,Cp,CCF5,CCF2,CCF4,CCF5,Ctr),4,4)
X=(0.1/(2-0.1)+2*0.1*(1-0.1)/(2-0.1))*A
inversa=solve(X)
Y=cbind(TempC1,OxígenC1,PHC1,TranspC1)
z=numeric(length(Y))
dim(z)=c(21,4)
for(i in 1:length(biometria)){
  if(i==1) z[i,]=0.1*Y[1,]+(1-0.1)*m+Y[i,]

  else if (i>1) z[i,]=0.1*Y[i,]+(1-0.1)*z[i-1,]+(2-0.1)*(Y[i,]-Y[i-1,])
}
z[,1]
data.frame(z[,1],z[,2],z[,3],z[,4])
d=data.frame(z[,1],z[,2],z[,3],z[,4])
t=numeric(21)
for(i in 1:length(biometria)){
  if(i==1) t[i]=t(z[1,])%%inversa%%z[1,]
  else if (i>1) t[i]=t(z[i,])%%inversa%%z[i,]
}

```

```

t
##Calculo del h3 usado en el limite de control
#g=2*UCL/(2*m+1) el ancho de los m+1 estados transitorios considerados.
#p(i,0)=P[X^2(p,c)<((0.5^2)*g^2/lambda^2)]=0.99973
(alpha=1-p(i,0)=0.00027) la probabilidad de que una
observación esté bajo control
#((0.5^2)*g^2/lambda^2)=16,2
ARL=370
chi=16.2
lambda=0.1
m=7
g=(sqrt(chi)*lambda)/0.5
UCL=g*(2*m+1)/2
h3=(UCL^2)/((0.1/1.9)+2*0.1*(0.9/1.9))
win.graph()
par(mfrow=c(2,2))
plot(biometria,t,type='l',ylim=c(0,h3+700),xlim = c(1,22),
xlab = "Biometrias semanales en C1", ylab = " T^2")
abline(h=h3,lty=2)
text((max(biometria)),h3,paste('h3=',h3),pos=3,font=1,cex=0.6)
for(i in 1:length(t)){
  if(t[i]>h3)temp<-23
  else if(t[i]<h3)temp<-18
  points(biometria[i],t[i],pch=temp)}

```

B.4. Código R para MDEL: Caso acuicultura

```

#Modelo para tratamiento C1
(de manera análoga en los tratamientos C2,
C6 y C7 fue usado este código apenas
cambiando los respectivos datos)
d1=data.frame(pesoC1,longitudC1)
datosYC1=scale(d1)
PCA1=prcomp(datosYC1,rank. = 1)
summary(PCA1)
uC1=as.numeric(PCA1$x)
d2=data.frame(TemperaturaC1,OxígenoC1,pHC1,TransparenciaC1)
datosXC1=scale(d2)
win.graph()
PCA2=prcomp(datosXC1,rank. = 2,retx = TRUE)
summary(PCA2)
scores=PCA2$x

```

B.4. Código R para el ajuste del modelo dinámico sobre estructuras latentes

```
t1=as.numeric(scores[,1])
t2=as.numeric(scores[,2])
PCA2$rotation[,1]
P1=PCA2$rotation[,1]
P2=PCA2$rotation[,2]
plot(P1,P2,xlim=c(-1,1),ylim=c(-1,1))
abline(h=0)
abline(v=0)

win.graph()
par(mfrow=c(3,2))
acf(uC1)
pacf(uC1)
acf(t1)
pacf(t1)
acf(t2)
pacf(t2)

library(tseries)
u1=numeric(20)
for(i in 1:20){
  u1[i]=uC1[i+1]-uC1[i]
}
win.graph()
par(mfrow=c(2,2))
plot(biometria,uC1,type="l")
plot(c(1:20),u1,type="l",xlab="biometria")
plot(biometria,t1,type="l")
plot(biometria,t2,type="l")
a=arma(u1,order=c(1,0))
summary(a)
b=arma(t1,order=c(1,0))
summary(b)
c=arma(t2,order=c(1,0))
summary(c)
res=residuals(a)[-c(1:2)]
res1=residuals(b)[-c(1:2)]
res2=residuals(c)[-c(1:2)]
win.graph()
par(mfrow=c(1,2))
CCF1=ccf(res,res1)
CCF1=ccf(res,res2)
#función de transferencia
library(MTS)
```

```
library(timeSeries)
library(forecast)
tf=tfm2(u1,t1,t2,orderN=c(1,1,0),order1=c(1,0,0),order2=c(0,1,2))
win.graph()
par(mfrow=c(2,2))
ccf(u1,t1)
res=residuals(tf)
hist(res)
acf(res,xlim=c(1,10))
pacf(res,xlim=c(1,10))
shapiro.test(res)
```