UNIVERSIDAD DE CHILE

FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS

DEPARTAMENTO DE INGENIERÍA CIVIL

IMPLEMENTACIÓN DE UN ALGORITMO DE MONITOREO DE SALUD ESTRUCTURAL BASADO EN OBJETOS SIMBÓLICOS Y CLASIFICACIÓN POR AGRUPAMIENTO

TESIS PARA OPTAR AL GRADO DE MAGÍSTER EN INGENIERÍA SÍSMICA

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL

GUSTAVO PATRICIO LAGOS FLORES

PROFESOR GUÍA:

RUBEN BOROSCHEK KRAUSKOPF

MIEMBROS DE LA COMISIÓN:

JOAO PEDRO SANTOS

MARCOS ORCHARD CONCHA

SANTIAGO DE CHILE

2017

RESUMEN DE LA TESIS PARA OPTAR AL TITULO DE: Ingeniero Civil y grado de Magister en Ingeniería Sísmica
Por: Gustavo Patricio Lagos Flores
Fecha: 23/05/2017
Profesor Guía: Rubén Boroschek K.

IMPLEMENTACIÓN DE UN ALGORITMO DE MONITOREO DE SALUD ESTRUCTURAL BASADO EN OBJETOS SIMBÓLICOS Y CLASIFICACIÓN POR AGRUPAMIENTO

El presente trabajo de Tesis muestra la implementación y análisis de variados métodos de aprendizaje de máquinas y minería de datos, desde la fase de extracción de características sensibles usando objetos simbólicos y clasificación mediante algoritmos de agrupamiento, para el estudio y monitoreo de la condición estructural de obras civiles, con énfasis en la detección temprana de la ocurrencia de daños estructurales.

El monitoreo de salud estructural mediante algoritmos de minería de datos, reconocimiento de patrones y aprendizaje de máquinas es un campo moderno y activo en la Ingeniería Civil. El flujo general es que a partir de mediciones de aceleración en sitio y utilizando metodologías de identificación de sistemas, se extrae la información modal que representa algún modelo clásico de la dinámica estructural. En este trabajo se busca extender el tipo de información que se puede extraer desde las series de aceleración, estudiando el uso de las series de aceleración en bruto y a su vez mediante la extracción de características sensibles usando ajustes de modelos autoregresivos.

La metodología global que se utiliza está basada en el agrupamiento de objetos simbólicos que representan el comportamiento estadístico de características sensibles, y se implementa tanto en series de aceleración obtenidas en laboratorio como en un edificio en operación instrumentado. En el segundo capítulo se estudian las series de aceleración en bruto como entrada para la transformación a objetos simbólicos con uso de histogramas e intervalos intercuartiles, concluyéndose que la energía de entrada es altamente determinante en los grupos obtenidos por los algoritmos de clasificación. Aún así, se puede extraer información del estado estructural si se analizan series de aceleración obtenidas desde un misma condición operacional y ambiental. En el tercer capítulo se estudia la extracción de otro tipo de característica basada en modelos autoregresivos, con las que se generan series de coeficientes de modelos AR(p) ajustados a las series de aceleración originales, encontrándose que los parámetros AR son mucho más sensibles a los cambios estructurales que la aceleración y que dicha sensibilidad puede aumentarse sin pérdida de robustez si se consideran líneas base de referencia. En el cuarto capítulo se analiza el uso de las series de coeficientes AR como entrada para la condensación a objetos simbólicos con los que realizar el agrupamiento, consiguiendo una mejora considerable en la separación de con respecto al uso de las series de aceleración en bruto.

Agradecimientos

Quiero en primer lugar agradecer a mi familia por todo el apoyo brindado y por la paciencia mostrada durante mi largo proceso de estudios, que viene a concluir pasados casi 10 años desde que termine mi etapa escolar. Nadie desconoce que el tiempo que se utiliza en el aprendizaje equivale a una merma en el tiempo que se puede dedicar a la familia. Gracias por la espera. Especialmente, gracias, mamá, por ser quién más me ha motivado a continuar con los estudios, dándome seguridad en momentos que ha faltado.

Agradezco de igual manera a los docentes que me han guiado y ayudado para la realización de la Tesis. Al profesor Rubén por mostrarme un área de la Ingeniería que me ha cautivado y que me llevó a adentrarme a los campos de Machine Learning, Data Mining e Inteligencia Artificial que me han hecho encontrarme nuevamente con lo que más me gusta, que es la investigación y programación. A Joao y Marcos por formar parte de una comisión que me ha hecho replantear muchas veces mis postulados y conclusiones, indicándome de frente cuando no estaban de acuerdo con mis planteamientos pero también felicitándome cuando lo hacía bien. Gracias también a los profesores de los que fui auxiliar y ayudante, por darme la posibilidad de realizar clases que es otra de mis grandes pasiones.

No puedo evitar mencionar a mi grupo actual de trabajo y también amigos, los chaqueteros irremediables de SK9. Sin ustedes lo más probable es que estaría trabajando en algo que no me satisface. Hay potencial, muchachos. Incluyo acá a mi papá, quien fuera la persona que me atrajo hacia uno de los proyectos más interesantes que he tenido la posibilidad de trabajar, que es la cervecería.

Infinitas gracias a mis amigos! Sean de Santiago o Chillán hacen que mi vida sea completa. Imposible mencionar a todos los grupos, futboleros, carreteros, Oikos, pedaleros, músicos, mechones. La LAD se lleva mención especial ¡Que familion!

Y para terminar lo más importante. Gracias a Kritsye Leiva, mi Kiki, y su familia, que para mí forman parte de mi propia familia. Mi compañera de vida, gran parte de lo que soy es gracias a ti. No hay palabras que puedan expresar todo lo feliz que soy por compartir mi vida contigo. Todo el universo sea para ti.

Tabla de Contenido

1. Introducción

Uno de los desafíos más grandes que posee la Ingeniería Civil es poder asegurar que una estructura se mantenga en condiciones seguras para su operación, ya sea que se trate de obras civiles con fines habitacionales o comerciales o bien de infraestructura pública y/o industrial. Evidentemente, hay situaciones especificas en las que tomar la decisión de que un edificio haya perdido su condición de habitabilidad es una tarea que no presenta una dificultad mayor. Por ejemplo, después del terremoto del 27 de Febrero del 2010, varios edificios en Santiago de Chile vieron gravemente afectada su condición, siendo uno de los casos más emblemáticos el del edificio Emerald que quedó visiblemente inclinado. Por supuesto, en una situación así, pese a que políticamente la decisión de evacuación y desalojo permanente puede ser difícil de tomar, desde el punto de vista ingenieril no existe otra opción hasta que se tomen las medidas necesarias para asegurar sanidad estructural. Sin embargo, la gran mayoría del resto de los edificios de Santiago no tuvo daño visible, por lo que es muy natural cuestionarse si es que acaso la ocurrencia del terremoto haya generado algún tipo de deterioro que no sea posible de observar. A partir de esta inquietud nace la rama de investigación asociada al Monitoreo de Salud Estructural, que estudia metodologías con las cuales poder determinar de manera temprana, automática y en tiempo real, la ocurrencia de anomalías o cambios en la condición estructural, con la finalidad de poder tomar acciones lo antes posible evitando que haya un perjuicio mayor.

El paradigma clásico de un Sistema de Monitoreo de Salud Estructural, o SHM por sus iniciales en inglés, consiste en una cadena de algoritmos y procedimientos, partiendo desde la instalación de sensores en las estructuras, con los que se realiza la medición de alguna variable física como aceleración, tensión y/o inclinación, por nombrar algunas. Estos sensores vienen acompañados de sistemas de adquisición que capturan las mediciones y permiten procesarlas mediante programas computacionales especialmente diseñados para detectar patrones y cambios en el comportamiento normal de las señales. Por tanto, SHM está fuertemente relacionada con las áreas de investigación de *Minería de Datos* y *Aprendizaje de Máquinas*, y en general cualquier técnica de reconocimiento de patrones y control estadístico de procesos. De hecho, gracias al desarrollo de estas aéreas principalmente matemáticas, es que ha sido posible el avance en técnicas de monitoreo en línea.

Si bien el objetivo original de un sistema de monitoreo de salud estructural consiste en la detección de daños estructurales, también pueden existir cambios estructurales que no representen necesariamente un perjuicio a la obra civil, sino más bien una mejora, como sería por ejemplo la instalación de barras de refuerzo, o las mismas obras necesarias para reparar un edificio dañado por un terremoto. Aún así, (Rytter 1993) define cuatro niveles de identificación de daño los que hasta la actualidad se consideran los pilares del monitoreo estructural, y fueron expandidos añadiendo un nivel más en el trabajo de (Worden and Dulieu-Barton 2004):

1. Nivel 1: Determinar si hay presente daño en la estructura.
2. Nivel 2: Determinar la localización del daño.
3. Nivel 3: Determinar el tipo de daño presente.
4. Nivel 3: Cuantificación del nivel o magnitud del daño.
5. Nivel 4: Predicción de la vida útil remanente de la estructura.

Responder cada uno de estos niveles representa un esfuerzo adicional, y siendo los cuatro primeros principalmente dedicados a indicar un diagnóstico estructural, el último nivel busca

predecir comportamiento futuro por lo que es necesario incluir conocimiento de fatiga de materiales, propagación de fallas u otras materias relevantes (J. P. Santos 2014).

Las estrategias de SHM pueden tener muchas variantes. Una primera distinción es determinada por la posibilidad de uso de un enfoque *inverso* o uno *directo*. El enfoque inverso consiste en realizar estimaciones de los modelos paramétricos o analíticos que mejor representen las mediciones. El enfoque directo, en cambio, no calcula modelos sino que simplemente extrae información sensible a los cambios estructurales mediante métodos estadísticos. En la presente tesis se estudiarán algunos métodos utilizando el enfoque directo, el que también recibe el nombre de 'data-driven' o empírico. Dentro de este enfoque, dependiendo de qué estrategia de aprendizaje estadístico se utilice, se genera una segunda distinción mayor, entre los tipos de aprendizaje supervisado y no supervisado. Los algoritmos supervisados se aplican en los casos que se cuente con conjuntos de datos y se conozca a qué clase o condición correspondan. Un ejemplo sería contar con registros de aceleración y saber exactamente a qué condición estructural corresponden, como es el caso de ensayos de laboratorios. Por el contrario, cuando no se conocen las clases, o cuando no hay información a priori sobre las condiciones estructurales, se utilizan las estrategias de aprendizaje que se conocen como no-supervisadas.

La extracción de *features* o características sensibles a los cambios estructurales es uno de los pasos bien estudiados en SHM. El paso siguiente, que consiste en clasificar las características en comportamientos estructurales conocidos o desconocidos, generalmente se lleva a cabo entrenando algoritmos con líneas base de referencia, en las que se asume que la estructura tiene una condición sana o sin cambios. Las características nuevas son analizadas por estos algoritmos entrenados para determinar si el comportamiento sigue sin presentar cambios. Dentro de estos algoritmos se pueden encontrar 'Statistical Process Control', 'Multi-Layer Perceptron Neural Neworks' o 'Support Vector Machines'. Una alternativa a estos métodos que no requiere de un uso de una línea de referencia es el análisis por agrupamiento, que clasifica las características en base a una medida de semejanza. Una de las grandes desventajas de estos algoritmos recae en su costo computacional; sin embargo, se ha mezclado con la transformación de los datos en objetos simbólicos, que son una forma de representar los datos originales de una forma condensada pero manteniendo la información sensible. De esta forma, gracias al pequeño tamaño de los objetos simbólicos, se provee a la clasificación de suficiente eficiencia computacional (J. Santos, Orcesi, et al. 2014).

El principal objetivo del presente trabajo de Tesis consiste en el estudio de algoritmos de agrupamiento y objetos simbólicos a partir de series de aceleración. Generalmente las series de aceleración son utilizadas para realizar identificación de parámetros modales y son éstos los que son usados para el monitoreo de salud estructural. Sin embargo, en este estudio se analiza la capacidad de las series brutas de aceleración para detectar cambios estructurales. Los resultados esperados son obtener una clasificación efectiva de los estados estructurales, y para ello serán aplicados los métodos estudiados en ensayos de laboratorios y en registros reales obtenidos del edificio de la Torre Central de la FCFM.

El informe está organizado por capítulos. El Capítulo 1 corresponde a la presente Introducción. Luego, en el Capítulo 2 se estudia y detallan los métodos asociados a la transformación de las series de aceleración en objetos simbólicos para su posterior clasificación utilizando algoritmos de agrupamiento. Se analizan distintas medidas de distancias entre objetos, diferentes algoritmos de agrupamiento y por último distintos índices de validación de las clasificaciones encontradas. Utilizando registros de ensayos de laboratorio se hace análisis de

sensibilidad con los que se determina cual es la combinación de métodos son los que logra una mejor clasificación y usando esa combinación se aplica la metodología a los registros de la Torre Central. Durante este capítulo la metodología es completamente no supervisada y libre de base de referencia.

En el Capítulo 3 se extiende la forma en que se pueden analizar las series de aceleración, mediante la identificación de modelos autoregresivos. Esto resulta en la extracción de características sensibles a cambios estructurales en un nuevo espacio, en adición a los conocidos espacio del tiempo (aceleración bruta) y espacio de las frecuencias (identificación de frecuencias modales). El espacio generado por los parámetros de los modelos autoregresivos también presenta características que lo hacen útil al momento de realizar un monitoreo de salud estructural. Durante este capítulo se explican los modelos autoregresivos, cómo calcularlos y cuáles son sus propiedades con las que se puede detectar un cambio en el comportamiento estructural. La metodología que se utiliza en este capítulo es no supervisada, pero sí considera el uso de una base de referencia.

En el cuarto capítulo, se realiza una mezcla de lo estudiado en los capítulos dos y tres, buscando aprovechar las ventajas encontradas en cada uno de ellos. Es así como se transforman las series de aceleración en el espacio del tiempo, en series de coeficientes de modelos autoregresivos, los que finalmente son clasificados usando una metodología no supervisada y libre de línea de referencia. Al igual que en los capítulos anteriores, se hace un análisis de sensibilidad con el que finalmente se encuentra una metodología para el estudio de los registros de la Torre Central.

Por último, en el Capítulo 5 se resumen los principales resultados obtenidos en todo el desarrollo de la Tesis, así como también se da una perspectiva para la implementación de lo estudiado en un sistema de monitoreo en línea.

2. Análisis por Objetos Simbólicos

2.1. Introducción.

El presente capítulo tiene por objetivo estudiar una metodología novedosa en el ámbito de detección de cambios estructurales a partir de datos instrumentales. Si bien lo estándar ha sido estudiar la variación de los parámetros modales, la estrategia analizada utiliza directamente las series de aceleración, buscando clasificar o agrupar distintos estados estructurales utilizando una representación a través de objetos simbólicos. Este tipo de representación consiste en un procedimiento de extracción de características sensibles y una posterior fusión de éstas, permitiendo pasar de una serie de variables en el tiempo a una serie de contenido estadístico. Un análisis utilizando este paradigma es un seguimiento o estudio a la variación del contenido estadístico de las señales de aceleración y tiene las ventajas así como también los límites que estos parámetros puedan poseer.

El desarrollo de la teoría detrás de los objetos simbólicos se debe al mejoramiento progresivo de los computadores y su capacidad de cálculo, lo que ha provocado que la creación de grandes bases de datos sea algo rutinario (L Billard and Diday 2002). Una de las formas de trabajar controladamente con la masiva cantidad de datos y extraer información valiosa de ella es categorizarla y representarla usando listas, intervalos, distribuciones, entre otras cosas, lo que es un ejemplo de la generación de datos simbólicos.

En la literatura, varios autores han investigado la aplicabilidad de esta metodología para detectar cambios estructurales. En (Cury, Crémona, and Diday 2010), los autores convierten a objetos simbólicos tanto los datos de aceleración como las propiedades modales de la estructura y aplican algoritmos de agrupamiento para clasificar diferentes comportamientos estructurales, indicando que los parámetros modales entregan resultados más robustos. (J. P. Santos et al. 2013) utiliza datos de sensores estáticos, los que normaliza utilizando PCA para posteriormente realizar un procedimiento en base a objetos simbólicos similar al de (Cury et al, 2010). El mismo autor sigue la misma línea en sus trabajos (J. Santos, Cremona, et al. 2014), (J. Santos, Orcesi, et al. 2014) y (J. P. Santos 2014), en donde transforma datos estáticos en objetos simbólicos utilizando intervalos intercuartiles. (Alves et al. 2015), por su parte, transforma datos de aceleración en objetos simbólicos usando histogramas, los que logra clasificar con resultados positivos tanto en ensayos de laboratorio como en el estudio de las vibraciones de un puente.

La intención de aplicar Analísis por Objetos Simbólicos al problema de detección de cambios estructurales consiste en poder identificar y clasificar distintos estados estructurales. En un caso muy particular en el que se tiene una estructura con dos estados, uno sano y otro dañado, lo que se espera es que las series de aceleración que fueron adquiridas en el estado sano terminen siendo agrupadas en un mismo conjunto, y lo mismo para las series del estado dañado. Se tendrían de esta forma 2 agrupaciones de series de aceleración, representando cada agrupación a un estado estructural distinto. El procedimiento o paradigma de este tipo de análisis es el siguiente:

1. Adquisición de series de aceleración
   En esta etapa se miden la aceleración en distintos puntos de una estructura. El número de sensores puede ser variado así como también el largo de la medición.

2. Fusión/Extracción/Representación en Objetos Simbólicos
   Cada medición de aceleraciones en el tiempo se procesa para transformarla en un objeto simbólico conteniendo la información estadística de las series de aceleración, pero con una dimensionalidad mucho menor. Si bien se habla de fusión, extracción y representación como conceptos distintos, todos comparten el hecho de que en la práctica son algoritmos que transforman un set de datos en otro. Sin embargo, tienen una finalidad distinta. Extracción se refiere principalmente a algoritmos dedicados a la identificación de modelos; Fusión, a algoritmos dedicados a la condensación de los resultados de la extracción para considerar múltiples variables y/o sensores. Representación, por su parte, corresponde a un nivel de abstracción mayor donde se acepta que la información bruta es caracterizada o representada por un set de parámetros que definen el objeto simbólico.
3. Cálculo de Distancias entre objetos simbólicos.
   Cada objeto simbólico representa una medición en el tiempo, y cada objeto puede ser ubicado en un espacio de $\Re^{sxm}$, en el cual es intuitivo pensar en alguna medida de distancia entre objetos. En este paso se calculan dichas distancias.
4. Agrupamiento de objetos simbólicos.
   Se aplica un algoritmo de agrupamiento, en el cuál cada objeto simbólico es asignado a un conjunto, dependiendo de la distribución espacial de todo el set de objetos. Cada conjunto representa un estado estructural distinto.

En general, la entrada de la metodología son las series de aceleración adquiridas por el sistema de monitoreo, mientras que la salida es el número de estados estructurales encontrados o clasificados, junto con una función de asignación que indica a qué conjunto, representando un estado estructural en particular, pertenece cada serie de aceleración. Dicho de otra forma, si se considera el conjunto de series adquiridas, la metodología devuelve una partición de éste conjunto con la propiedad de que todo par de subconjuntos tiene intersección vacía.

Una aclaración importante de realizar tiene que ver con la interpretación de la partición encontrada. Los algoritmos utilizados, al estar basados en series aceleración, son fuertemente influenciados por la amplitud de las vibraciones lo que a su vez es una condición de respuesta frente a la energía entrante al sistema. Eventualmente la clasificación podría arrojar particiones que distinguen entre diferentes tipos de excitaciones, lo que correctamente se puede interpretar como distintos comportamientos estructurales ('estructura moviéndose' vs 'estructura quieta'). Sin embargo, la clasificación entrega poca información acerca de la naturaleza del cambio en el comportamiento estructural, pudiéndose tratar de daños, del régimen de uso de la estructura o de distinta energía de entrada, por nombrar algunas de las posibles causas.

El presente capitulo se desarrolla como sigue: En la Sección 2.2, se detalla extensivamente la transformación de los datos de aceleración en objetos simbólicos, para lo que se utilizaran intervalos intercuartiles e histogramas, generando dos set de datos simbólicos distintos. Luego, en la Sección 2.3, se muestra la forma de extraer información útil de estos conjuntos, utilizando métrica asociada a los dos tipos de datos, intercuartil e histograma, para definir el concepto de distancia entre un par de objetos simbólicos. La forma de clasificar los objetos simbólicos se explica en la Sección 2.4, donde se muestran los algoritmos de agrupamiento destinados a encontrar la mejor partición posible de los datos. Por último, en las Secciones 2.5 y 2.6 se ejecutan las metodologías desarrolladas, aplicadas a una estructura de laboratorio y al edificio de la Torre Central del Campus Beauchef de la Universidad de Chile, respectivamente.

2.2. Representación en Objetos Simbólicos.

Normalmente entre la etapa de adquisición de registros de aceleración y la aplicación de cualquier algoritmo de monitoreo de salud estructural, tanto si este use la extracción de variables modales o alguna otra característica sensible a los cambios estructurales, se realiza un proceso de limpieza de datos. Este puede consistir en, por ejemplo, filtros pasa banda en el espacio de las frecuencia, o filtros RMS en el espacio del tiempo y amplitud de la señal. Cada sistema de adquisición requiere un proceso de limpieza determinado por las características de los sensores y su montaje, así como también de las variables medidas. Durante esta sección se muestra el procedimiento para pasar de una serie de aceleración limpia a una descripción de estos datos usando objetos simbólicos.

Hay variadas formas de convertir los datos en forma a clásica a su representación en objetos simbólicos(Alves et al. 2015). Por ejemplo, las señales se pueden representar como

- Histograma de k-intervalos: $\{a_t\} \rightarrow \{C_1, P_1; C_2, P_2; \dots; C_w, P_w\}$
- Intervalos interX-iles: $\{a_t\} \rightarrow (T_{inf}, T_{sup})$
- Intervalos min/max: $\{a_t\} \rightarrow (\min\{a_t\}, \max\{a_t\})$
- Desviación Estándar: $\{a_t\} \rightarrow \sigma^2(\{a_t\})$

En este trabajo de tesis se consideran los tipos de objetos simbólicos intercuartiles y de histograma para extraer y condensar las series de aceleración. A continuación se detalla el procedimiento para llevar a cabo la representación (J. P. Santos 2014).

Considerar una medición adquirida de largo τ segundos. Si el sistema tiene una tasa de muestreo de Fs [Hz] y posee s canales, la medición puede ser almacenada en su forma clásica en una matriz de dimensiones $(Fs \cdot \tau \times s)$, en la que cada fila es un estado de aceleraciones en el tiempo.

Para el caso de un objeto tipo intercuartil, cada sensor de la medición se transforma en un par intercuartil(T_{inf}^r, T_{sup}^r). Suponiendo que se ordenan ascendentemente los datos asociados al sensor r-ésimo, el valor T_{inf}^r corresponde al dato que tiene el 25% debajo de él. Análogamente, $T_{\sup}^r$ corresponde al valor que tiene el 75% de los datos debajo de él. Es por esto que este tipo de objetos se llama intercuartil, ya que la representación excluye los datos ubicados en los cuartiles inferior y superior, mientras que el intervalo $[T_{inf}^r, T_{sup}^r]$ contiene en su interior al 50% de todos los datos. En este caso, la medición de $(Fs \cdot \tau \times s)$ queda reducida a un objeto simbólico de dimensiones $s \times 2$.

Por otra parte, para el caso de histogramas la creación de objetos simbólicos no es tan directa, ya que él calculo de distancias entre este tipo de objetos requiere que ciertas condiciones se satisfagan, como por ejemplo, que las categorías de los histogramas sean las mismas para todas las mediciones realizadas. Como las categorías en un histograma de datos numéricos son determinadas por los extremos de los intervalos, es necesario especificar cuántos intervalos se considerarán y cuáles son sus extremos respectivos. No obstante, los datos de aceleración tienen un rango extenso, pensando en que hay momentos donde la estructura está prácticamente en reposo absoluto y otros en que está en movimiento extremo (terremoto). Dicho esto, se debe estimar un rango de aceleraciones límites de aceptación $\pm a_{lim}$, dentro de los cuales se generan intervalos equiespaciados, y fuera de ellos se crean dos intervalos extremos, el primero con un rango de valores $(-\infty, -a_{lim})$, y el otro de $(a_{lim}, +\infty)$. De esta forma, las aceleraciones que caen fuera del rango de aceptación son contabilizadas en los intervalos extremos. Todas las

mediciones son procesadas con el mismo número de intervalos y con los mismos extremos de intervalos. Al condensar los datos clásicos en un objeto simbólico usando histogramas, la medición de $(Fs \cdot \tau \times s)$ se reduce a un objeto de $(N \times s)$, donde N es el número de intervalos utilizados. La inclusión de intervalos extremos no se ha visto aplicada en otros artículos pero nace de la realidad en Chile de convivir con movimientos sísmicos de manera cotidiana. La selección de las amplitudes límites puede tener gran incidencia en la sensibilidad final de los algoritmo y es algo a tener consideración al momento de analizar los resultados.

De forma matemáticamente estricta, y según lo estipulado en (J. P. Santos 2014), un objeto simbólico T_i es definido como

$$T_i = [T_i^{(1)} \dots T_i^{(r)} \dots T_i^{(p)}] \tag{2.1}$$

donde cada $T_i^{(r)}$ es un valor estadístico que describe una serie de datos clásicos. En esta notación, (r) hace referencia al $r-$ésimo sensor de un total de p sensores. Además, cuando se utilizan intervalos (no necesariamente intercuartiles), los $T_i^{(r)}$toman la forma

$$T_i^{(r)} = \left(T_{i,inf}^{(r)}, T_{i,sup}^{(r)}\right) \tag{2.2}$$

donde el subíndice 'inf' indica el límite inferior del intervalo, mientras que el subíndice 'sup' indica el límite superior de éste. Análogamente, si se consideran histogramas, los $T_i^{(r)}$ son representados como sigue

$$T_i^{(r)} = (\mathrm{P}_{\mathrm{ik}}^{(\mathrm{r})}; k = 1, \dots, \omega) \tag{2.3}$$

con $P_{ik}^{(r)}$ siendo la frecuencia relativa del intervalo $k-$ésimo, de un total de ω intervalos utilizados. En este caso, la frecuencia relativa se define como la razón entre el número de ocurrencias y el número total de datos de la medición. Cabe destacar que en la Ecuación (2.3) no se explicitaron lo límites de los intervalos utilizados, ya que todos los objetos creados con histogramas utilizan exactamente los mismos, y la definición de distancia para histograma no toma en consideración el ancho ni los extremos de los intervalos.

A modo de ejemplo, en la Figura 2.1 se aplican histogramas a series de aceleración típicas de un edificio, considerando aceleraciones extremas de $a_l = \pm 0.00004[g]$ y 20 intervalos. Si bien el rango de aceleraciones puede parecer pequeño, es efectivamente el rango obtenido para el edifico en un régimen de excitación ambiental pura. Cabe destacar que en esta figura todas las aceleraciones de la serie caen dentro del intervalo. Por otro lado, en la Figura 2.2 se muestra la aplicación de histogramas a las mismas series, pero utilizando un rango incluso más pequeño, con límites $a_l = \pm 0.00002\ [g]$. En este último caso, no todas las aceleraciones caen dentro del intervalo, lo que se puede apreciar claramente en las ocurrencias de contenedores extremos, sobre todo en los sensores N°4 y N°6.

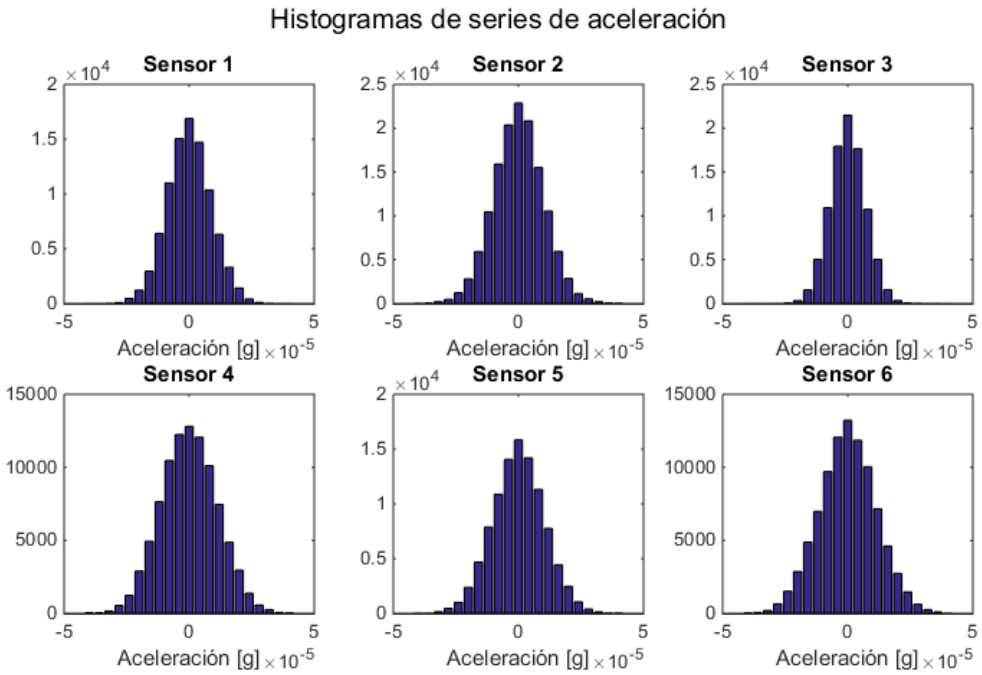


Figura 2.1. Ejemplo de histogramas. $a_l = \pm 0.00004$ [g]

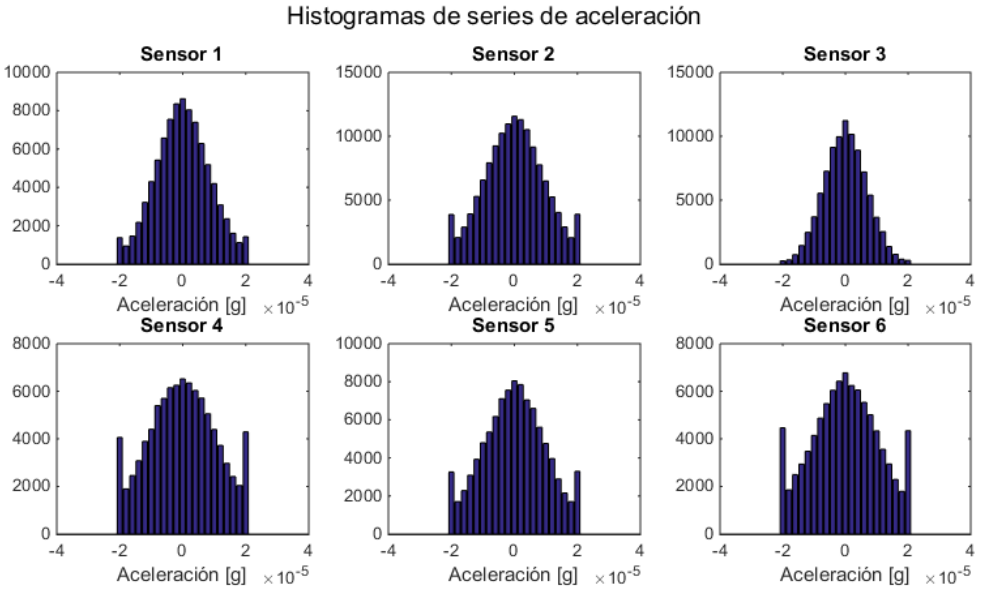


Figura 2.2. Ejemplo de histogramas. $a_l = \pm 0.00002$ [g]

Cabe destacar que el efecto de los intervalos extremos en los algoritmos no puede asegurarse que sea algo positivo o negativo, por lo que se debiera realizar una investigación sobre la sensibilidad de los métodos frente a los límites escogidos.

En cuanto al número de datos en las series de aceleración, (J. P. Santos 2014) hace un estudio de la robustez del análisis por objetos simbólicos dependiendo del largo de las

mediciones, comparando mediciones realizadas cada 5 minutos y cada 30 minutos, indicando que el método es más robusto en las últimas. Sin embargo el análisis llevado a cabo en su trabajo de Tesis es para datos estáticos y generalizar dichos resultados hacia series de aceleración puede no ser una suposición correcta. Estudiar la sensibilidad del método en relación al largo de las mediciones no forma parte de la presente investigación, por lo que se utilizan series de un largo predefinido en el sistema de adquisición, establecido en aproximadamente 15 minutos.

2.3. Distancias.

En la sección anterior se explicó cómo transformar la data clásica a un objeto simbólico cuyas dimensiones dependen del tipo de herramienta estadística que se haya utilizado. Si se considera una instalación instrumental de solo 3 sensores, y se transforman los datos a un objeto simbólico utilizando la desviación estándar de la muestra, entonces este objeto tendría 3 dimensiones. En dicho caso particular, los objetos pueden ser graficados en un espacio tridimensional, lo que permite observar la distribución de los objetos de una forma común e intuitiva para el ojo humano.

En la Figura 2.3, se muestra una distribución en 3D de objetos simbólicos creados a partir de la desviación estándar de 3 sensores. Los datos son generados aleatoriamente por lo que no representan un experimento real, pero es solo para hacer un ejemplo ilustrativo. En esta figura, se puede ver claramente como la mitad de los datos tienen una desviación estándar mayor para el Sensor 1, indicando que existe un comportamiento distinto entre estos dos grupos de objetos simbólicos.

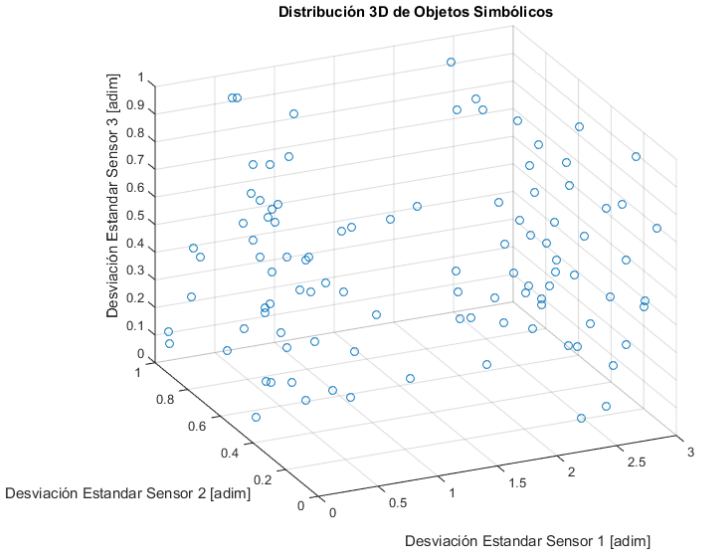


Figura 2.3. Una distribución espacial de objetos simbólicos.

Hay distintas formas de graficar los objetos simbólicos. (J. P. Santos 2014) muestra un ejemplo donde puede representar un objeto de 4D, en un plano 2D, mediante superposición de los objetos. Sin embargo, a medida que las dimensiones de los objetos van aumentando, crear un gráfico de distribución espacial se imposibilita. No obstante, existen herramientas matemáticas desarrolladas para trabajar en espacios de múltiples dimensiones. En particular, interesa conocer la distancia que existe entre un objeto y otro, pues como se vio en el gráfico anterior, una distancia elevada entre objetos indica la ocurrencia de un evento anormal. Lo más intuitivo sería utilizar una medida de distancia euclidiana en un espacio $\Re^n$, lo cual es factible y efectivamente

se hace, pero además se utilizan otros tipos de métricas definidas específicamente para objetos simbólicos, las que otorgan distintos pesos a las propiedades estadísticas, detallándose a continuación.

2.3.1. Distancias para objetos tipo intercuartil

Considerar un set de N objetos simbólicos $\{T_i, \dots, T_N\}$. Cada uno de estos objetos es descrito por p intervalos intercuartiles $T_i^{(r)} = \left(T_{i,inf}^{(r)}, T_{i,sup}^{(r)}\right)$, obtenidos de p sensores instalados en una estructura. Las métricas utilizadas relacionan dos objetos: T_i y T_j. En esta sección se detallan tres tipos distintos de cálculo de distancia entre objetos simbólicos intercuartiles.

1. **Haussdorf Euclideana Estandarizada.**

$$d_{ij} = \left(\sum_{r=1}^{p}\left[\frac{\phi_r(T_i, T_j)}{H_r}\right]^2\right)^{\frac{1}{2}} \tag{2.4}$$

donde

$$\phi_r(T_i, T_j) = \max\left(\left|T_{i,inf}^{(r)} - T_{j,inf}^{(r)}\right|, \left|T_{i,sup}^{(r)} - T_{j,sup}^{(r)}\right|\right) \tag{2.5}$$

es definida como la medida de disimilitud de Haussdorf y H_r^2 es el término de estandarización calculado como

$$H_r^2 = \frac{1}{2N}\sum_{i=1}^{N}\sum_{j=1}^{N}\left[\phi_r(T_i, T_j)\right]^2 \tag{2.6}$$

2. **Ichino-Yaguchi Euclideana Estandarizada**

$$d_{ij} = \left(\frac{1}{p}\sum_{r=1}^{p}\frac{1}{|Y_r|}\left(\phi_r(T_i, T_j)\right)^2\right)^{\frac{1}{2}} \tag{2.7}$$

en este caso, ϕ_r es la medida de disimilitud de Ichino-Yaguchi

$$\begin{aligned}\phi_r(T_i, T_j) &= \left|T_i^{(r)} \oplus T_j^{(r)}\right| - \left|T_i^{(r)} \otimes T_j^{(r)}\right| \\ &+ \Omega\left(2\left|T_i^{(r)} \otimes T_j^{(r)}\right| - \left|T_i^{(r)}\right| - \left|T_j^{(r)}\right|\right)\end{aligned} \tag{2.8}$$

Los operadores Cartesian join$\oplus$, Cartesian meet $\otimes$ y la norma del intervalo $|[\cdot]|$, se definen como se muestra en las Ecuaciones (2.9), (2.10) y (2.11), respectivamente,

$$T_i^{(r)} \oplus T_j^{(r)} = \left[\min\left(T_{i,inf}^{(r)}, T_{j,inf}^{(r)}\right), \max\left(T_{i,sup}^{(r)}, T_{j,sup}^{(r)}\right)\right] \tag{2.9}$$

$$T_i^{(r)} \otimes T_j^{(r)} = \left[\max\left(T_{i,inf}^{(r)}, T_{j,inf}^{(r)}\right), \min\left(T_{i,sup}^{(r)}, T_{j,sup}^{(r)}\right)\right] \tag{2.10}$$

$$\left|T_i^{(r)}\right| = T_{i,sup}^{(r)} - T_{i,inf}^{(r)} \tag{2.11}$$

mientras que el término de estandarización $|Y_r|$ se define como se expresa en la Ecuación (2.12).

$$|Y_r| = \left|\max_s\left(T_{sup}^{(r)}\right) - \min_s\left(T_{inf}^{(r)}\right)\right| \tag{2.12}$$

3. Gowda-Diday

$$d_{ij} = \sum_{r=1}^{p} \phi_r\left(T_i, T_j\right) \tag{2.13}$$

donde

$$\begin{aligned}\phi_r\left(T_i, T_j\right) = &\frac{\left|\left|T_{i,sup}^{(r)} - T_{i,inf}^{(r)}\right| - \left|T_{j,sup}^{(r)} - T_{j,inf}^{(r)}\right|\right|}{k_r} \\ &+ \frac{\left(\left|T_{i,sup}^{(r)} - T_{i,inf}^{(r)}\right| - \left|T_{j,sup}^{(r)} - T_{j,inf}^{(r)}\right| - 2I_r\right)}{k_r} \\ &+ \frac{\left|T_{i,inf}^{(r)} - T_{j,inf}^{(r)}\right|}{|Y_r|}\end{aligned} \tag{2.14}$$

$$k_r = \left|\max\left(T_{i,sup}^{(r)}, T_{j,sup}^{(r)}\right) - \min\left(T_{i,inf}^{(r)}, T_{j,inf}^{(r)}\right)\right| \tag{2.15}$$

$$I_r = \left|\max\left(T_{i,inf}^{(r)}, T_{j,inf}^{(r)}\right) - \min\left(T_{i,sup}^{(r)}, T_{j,sup}^{(r)}\right)\right| \tag{2.16}$$

Los fundamentos matemáticos de estas métricas pueden ser encontrados por separado en (Billard and Diday 2007), (Ichino and Yaguchi 1994), (Gowda and Diday 1991), pero (J. P. Santos 2014) hace un muy buen resumen de estos.

2.3.2. Distancias para objetos tipo histograma

La definición de distancia para histograma es tomada y modificada de (Billard and Diday 2006). Las modificaciones se introducen ya que en la referencia la distancia es usada de una forma mucho más general que el objetivo de esta investigación.

Considerar una serie de mediciones, en nuestro caso de aceleraciones, de $nt \times s$, donde $nt = Fs * \tau$ y s es el numero de sensores, la cual ha sido transformada a un set de objetos simbólicos $\{T_i\}$ usando histogramas. El objeto T_i puede ser descrito como la frecuencia relativa de los contenedores o intervalos utilizados. Suponiendo que las series de cada sensor se transforman usando el mismo número $p$de intervalos, la caracterización de T_i formalmente queda como

$$T_i = \left\{\left(P_{11}, \dots, P_{p1}\right), \dots, \left(P_{1r}, \dots, P_{pr}\right), \dots, \left(P_{1s}, \dots, P_{ps}\right)\right\} \tag{2.17}$$

donde P_{kr} es la frecuencia relativa del intervalo $k-$ésimo asociado al sensor $r-$ésimo. Cabe destacar que el objeto simbólico queda de dimensiones $1 \times (p * s)$, por lo que puede ser escrito como un vector fila, lo que facilita enormemente el cálculo de la distancia categórica para histogramas, la que es definida a continuación.

Considerar un set $\{T_i\}$ de Nobjetos simbólicos de histograma, la distancia categórica entre dos objetos T_i y T_j es definida como (Billard and Diday 2006)

$$d_{ij}^2 = \frac{1}{s}\sum_{a=1}^{s}\sum_{k_a=1}^{p}\left(\sum_{n=1}^{N} P_{nk_a a}\right)^{-1}\left(P_{ik_a a} - P_{jk_a a}\right)^2 \qquad (2.18)$$

En la Tabla 2.1 se muestran valores para 4 objetos de histograma creados usando 5 intervalos y dos sensores. Nuevamente, lo datos son ficticios y son solo para la representación. Notar que los dos primeros objetos muestran una distribución normal de los datos adquiridos, mientras los dos últimos objetos muestran una distribución uniforme. Si bien la fórmula de distancia categórica en la Ecuación (2.18) aparenta ser compleja, cuando los objetos se escriben como en la Tabla 2.1 la operación del cálculo de este tipo de distancia es bastante sencillo. En este ejemplo, las distancias entre los objetos T_1 y T_2, T_1 y T_3, y T_3 y T_4 valen 0.12, 0.43 y 0.05, respectivamente, indicando que entre los objetos de similar distribución la distancia es menor que entre los objetos de distribución distinta. Un ejemplo detallado puede ser encontrado en(Billard and Diday 2006).

Tabla N° 2.1. Objetos simbólicos de histograma.

Sensores →	s_1					s_2				
Objetos ↓	$k_1 = 1$	$k_1 = 2$	$k_1 = 3$	$k_1 = 4$	$k_1 = 5$	$k_2 = 1$	$k_2 = 2$	$k_2 = 3$	$k_2 = 4$	$k_2 = 5$
T_1	0,02	0,21	0,49	0,26	0,02	0,04	0,23	0,46	0,24	0,03
T_2	0,05	0,33	0,42	0,18	0,03	0,02	0,20	0,45	0,27	0,05
T_3	0,19	0,22	0,19	0,19	0,21	0,20	0,21	0,20	0,21	0,18
T_4	0,21	0,18	0,19	0,21	0,21	0,19	0,22	0,18	0,21	0,20
$\sum_{n=1}^{N} P_{nk_a a}$	0,47	0,94	1,29	0,84	0,47	0,46	0,87	1,29	0,92	0,46

2.3.3. Matrices de distancias

Considerando nuevamente la Figura 2.3, si bien con tan solo mirar la distribución se identifican dos conjuntos de puntos, interesa poder mostrar matemáticamente que los puntos conforman estas agrupaciones. Para esto es necesario desarrollar conceptos y herramientas matemáticas que no presenten la subjetividad del ojo humano, lo que además posibilita la automatización del proceso.

Suponiendo que se cuenta con un set de objetos simbólicos $\{T_1, \ldots . T_n\}$, una de las cosas de utilidad es calcular la matriz de distancias D, que es la matriz cuyas entradas d_{ij} son la distancia entre el par de objetos T_i y T_j. Cabe destacar que, por definición, la matriz D es cuadrada de $n \times n$, sus elementos en la diagonal son nulos ya que todos los objetos tienen distancia nula a ellos mismos, y además es simétrica. La matriz D es solo una de las formas de representar la distancia entre cada par de objetos, ya que también puede ser expresada en forma de lista considerando las propiedades ya mencionadas.

El cálculo de la matriz de distancias tiene por objetivo estudiar distintos grupos de objetos y analizar las distancias entre ellos. En el caso de que haya poca distancia entre objetos de un mismo grupo y gran distancia entre objetos de distinto grupo se puede afirmar objetivamente que se trata de cúmulos o agrupamientos óptimos. Ya sea en su forma cuadrada o en lista, la matriz D tiene la ventaja de que libera al usuario del espacio multidimensional, otorgando una nueva posibilidad de visualización usando una escala de grises, como se muestra en la Figura 2.4.

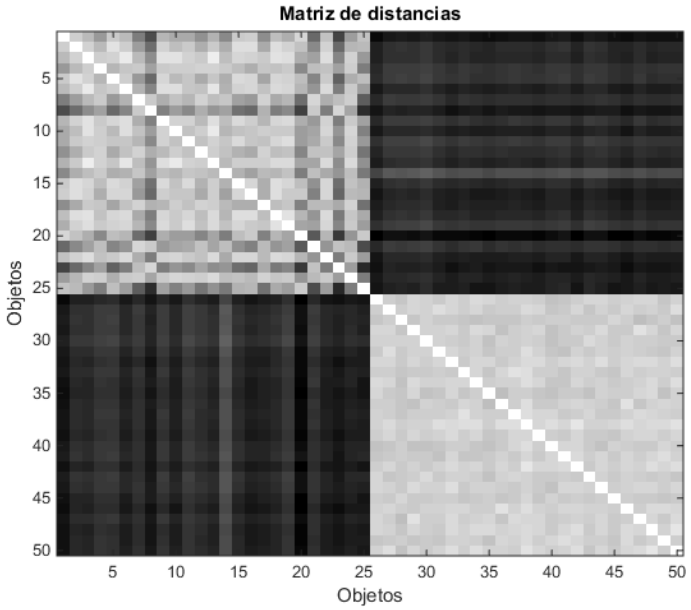


Figura 2.4. Ejemplo de un diagrama de matriz de distancias. La primera mitad de los datos tiene distribución normal, y la segunda mitad una distribución uniforme.

Este tipo de diagrama se genera otorgando un color en la escala de grises a cada entrada de la matriz. La escala va desde 0, asignándole el color blanco a todas las entradas que tengan valor cero (por eso se aprecia una diagonal de color blanco), y el negro a la entrada que tenga el mayor valor. Para valores entremedio se genera una interpolación en escala de grises. Con tan solo mirar estos diagramas, es posible ver que los colores pueden ser agrupados. Ahora bien, como los objetos son creados en orden a medida que las mediciones son adquiridas, es de esperar que la presencia de estados estructurales estuvieran ordenados de la misma manera y no mostrasen una distribución aleatoria. Más aún, si un diagrama muestra comportamiento aleatorio, es factible indicar que no existe una clara diferenciación entre distintos estados estructurales.

En la Figura 2.4, los datos fueron creados aleatoriamente, en los que la primera mitad posee una distribución normal, mientras que la segunda posee una distribución uniforme. Todos los datos fueron procesados mediante histogramas, usando 20 intervalos. Se puede observar que entre los datos de la primera mitad los colores son claros, indicando poca distancia entre estos

objetos, mientras que las distancias de objetos pertenecientes a diferentes mitades es mucho mayor. Esto implica que agrupaciones de colores claros sugieren comportamientos similares, mientras que agrupaciones de colores oscuros indican la posibilidad de comportamientos distintos.

Si bien el análisis mirando los diagramas requiere de un ojo experimentado para sacar conclusiones inmediatas, hay algoritmos especializados en generar agrupaciones óptimas de una forma objetiva, los que se explican en la siguiente sección.

2.4. Clasificación usando algoritmos de agrupamiento

La salida de la sección anterior consiste en una matriz o lista de distancias entre todos los objetos simbólicos del set de datos. Cabe recordar que cada uno de estos objetos representa a un lapso de mediciones establecido por el sistema de adquisición, siendo de aproximadamente 15 minutos en este trabajo. El objetivo del presente segmento es, a partir de la matriz de distancias, generar una partición (subconjuntos) de todos los objetos, en la que cada conjunto de objetos represente un estado estructural distinto, lo que vendría siendo la parte final del análisis por objetos simbólicos.

La agrupación de objetos en distintos subconjuntos, forma parte de los conocidos algoritmos de agrupamiento, cuya finalidad es justamente encontrar una partición que genera los conjuntos más compactos y con la mayor separación entre ellos.

Para desarrollar los algoritmos, primero es necesario definir los conceptos que hay detrás de éstos. Considerar un set de objetos simbólicos $\{T\} = \{T_1, T_2, \dots, T_n\}$. Un clúster C_k se define como un subconjunto de $\{T\}$. Una partición P_K de $\{T\}$ es un conjunto de K grupos con la particularidad de que la unión de todos ellos es el set original, mientras que la intersección de cada par de ellos es vacía. Es decir, cada elemento de $\{T\}$ pertenece a solo un clúster de la partición P_K.

Ahora, considerar una partición $P_K = \{C_1, C_2, \dots C_K\}$ de un set de objetos simbólicos. La distancia total dentro de un clúster se define según la Ecuación (2.19), en la que $c(i)$ es una función de asignación en la que el elemento T_i es asignado al clúster k-esimo. En este punto es necesario hacer una distinción. La regla de asignación c(i) se puede leer como "Los elementos son asignados al clúster más cercano". El resultado de la aplicación de la regla c(i) es finalmente la partición de interés, cuyo resultado se puede interpretar como un vector de índices. En este sentido la regla es un input al algoritmo y el vector de índices es el output. Notar que todos los algoritmos de agrupamiento entregan el vector de índices como salida.

$$WC(C_k) = \frac{1}{2} \sum_{c(i)=k} \sum_{c(j)=k} d_{ij} \qquad (2.19)$$

Siguiendo la misma lógica, la distancia total agregada dentro de grupos, para una partición P_K, corresponde a la suma de todas las distancias $WC(C_k)$, lo que se expresa en la Ecuación (2.20) (Cury, Crémona, and Diday 2010). Intuitivamente, el objetivo de cualquier algoritmo de agrupamiento es minimizar ésta distancia (Hastie, Tibshirani, and Friedman 2011), lo que asegura grupos compactos de poca distancia entre elementos pertenecientes a estos.

$$W(P_K) = \sum_{k=1}^{K} WC(C_k) = \frac{1}{2}\sum_{k=1}^{K}\sum_{c(i)=k}\sum_{c(j)=k} d_{ij} \tag{2.20}$$

A su vez, la distancia total agregada del set de objetos simbólicos, no es más que la suma de todas las distancias existentes en el grafo completo definido por los objetos, lo que matemáticamente se expresa en la Ecuación (2.21)

$$OD(\{T\}) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} d_{ij} \tag{2.21}$$

Por último, la distancia total agregada entre grupos, se define simplemente la distancia total agregada del set de objetos menos la distancia total agregada dentro de grupos. Esta distancia adquiere importancia en los índices de la Sección 2.4.4, y se establece matemáticamente en la Ecuación (2.22)

$$B(P_K) = OD - W(P_K) \tag{2.22}$$

Es importante destacar que la distancia total, según la Ecuación (2.21) no depende de la partición considerada, sino que es una propiedad intrínseca del set de objetos simbólicos, mientras que la distancia total agregada dentro de grupos sí depende de la partición, y es lo que en teoría se desea minimizar. Como un set de objetos simbólicos se trata de un conjunto finito, el número de subconjuntos posibles también es finito lo que asegura la existencia de una partición óptima. No obstante, evaluar cada partición sería de una complejidad numérica muy grande, y es este motivo lo que llevó al desarrollo de los algoritmos de agrupamiento. Como se explica más adelante, los métodos son de naturaleza iterativa y pueden presentar el problema de convergencia a mínimos locales.

A continuación se presentan a modo general los fundamentos de algunos de los algoritmos de agrupamiento que existen en la literatura. Para una descripción más detallada se puede revisar (J. P. Santos 2014), (Alves et al. 2015), (Hastie, Tibshirani, and Friedman 2011)

2.4.1. Algoritmos de agrupamiento.

Algoritmos de nubes dinámicas.

Esta clase de algoritmos busca minimizar $W(P_K)$ realizando una serie de cálculos iterativos. Una interpretación del nombre 'nubes dinámicas' es que los conjuntos o grupos creados pueden ser visualizados como una nube en el espacio multidimensional, y estas van cambiando de forma en cada iteración. Como entrada, reciben el set de objetos simbólicos y el número K de grupos que se cree que existen en los datos. A partir de ahí, se crean K prototipos, o ubicaciones iniciales de las 'nubes' (grupos), y cada objeto es asignado a la nube que tenga más cercana, lo que resulta en la función de asignación $c(i)$. La ubicación de cada nube es calculada nuevamente, considerando la posición de todos los objetos asignados a ellas. Finalmente se calcula una función objetivo de mínimos cuadrados, expresada en la Ecuación (2.23).

$$OF = \frac{1}{2}\sum_{k=1}^{K}\sum_{c(i)=k}\sum_{c(j)=k} d_{ij}^2 \tag{2.23}$$

Si se alcanza un criterio de convergencia, se detiene el algoritmo; en caso contrario, se itera con la nueva posición de las nubes, lo que puede generar una nueva asignación y una disminución en el valor de la función objetivo. La existencia de la solución se fundamenta en el hecho de que los conjuntos de objetos son finitos, pero la convergencia es fuertemente dirigida a un punto de mínimo local, por lo que el algoritmo se ejecuta varias veces con distintos prototipos iniciales, eligiendo el mínimo global de las ejecuciones como la solución óptima.

Algoritmos jerárquicos

Este clase de algoritmos se diferencia de los de nubes dinámicas en que calculan todo el árbol de particiones óptimas para todos los números de grupos posibles (de 1 a K) . Sin embargo, lo hacen de una manera heurística, siguiendo reglas iterativas que minimizan la función objetivo.

Básicamente, hay dos tipos de algoritmos jerárquicos: divisivos y aglomerantes. En el primer tipo, se parte de un solo clúster conteniendo a todos los objetos simbólicos, y en cada iteración se divide un clúster en otros dos, y así sucesivamente hasta llegar a que cada objeto se ubica en un clúster distinto. En cambio, los aglomerantes son justamente lo contrario, partiendo de cada objeto asignado a un clúster distinto, y en cada iteración se agrupan dos grupos en uno solo, hasta llegar a un único clúster. Ambos tipos de algoritmos pueden ser representados en un gráfico de dendrograma, que permite una clara visualización de los procesos y a partir del cual se puede obtener la partición óptima para cualquier número de grupos. Un ejemplo de esto se muestra en la Figura 2.5, sacada de (J. P. Santos 2014), en donde se muestra cómo es posible obtener la partición para 3 grupos, realizando un corte en el dendrograma.

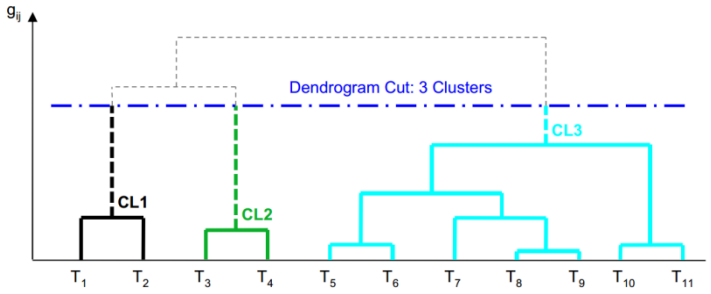


Figura 2.5. Ejemplo de obtención de grupos a partir de un dendrograma (J.P Santos 2014).

2.4.2. Validación de la partición escogida

Uno de los problemas de los algoritmos mencionados, es que necesitan que el usuario indique la cantidad de grupos presentes en los datos, lo que no siempre se puede saber a priori ya que los cambios estructurales o de comportamiento pueden ocurrir en cualquier instante de tiempo. Es por esto que para determinar una partición con un número óptimo de grupos, se evalúan una serie de índices que mediante su minimización o maximización indican el número de grupos que genera la mejor partición de los datos, valorando la separación entre grupos y lo compactos que éstos sean.

Existen diversos tipos de índices de validación, pero los más adecuados para el tipo de análisis desarrollado corresponden a los de 'validación relativa', que son calculados para las K particiones óptimas, recorriendo desde un sólo clúster, hasta Kgrupos. Este tipo de índices solo dependen de las particiones creadas y sus propiedades como la distancia total entre grupos, y no del conocimiento previo que se tenga de las situaciones estructurales ni de una línea de referencia. En (J. P. Santos 2014), se estudiaron cuatro índices de validación relativa: Calinski & Harabasz, Gamma, C* y Global Silhouette, indicando que el primero y el último son los que

arrojan una mejor identificación del número de grupos óptimo. En la presente tesis se utilizan estos índices para evaluar las particiones creadas. A continuación se dan más detalles de estos dos validadores.

Calinski & Harabasz

El índice de Calinski y Harabasz (Calinksi and Harabasz 1974)(J. P. Santos 2014)es una relación entre el número de objetos totales, el número de grupos en la partición analizada, y las distancias dentro y entre grupos. Matemáticamente establecido en la Ecuación (2.24), cuando su valor se maximiza es cuando se presenta la partición idónea para el set de objetos simbólicos estudiados. En esta ecuación, N representa al número total de objetos, t al número de grupos presentes en la particion P_t, y $W(P_t), B(P_t)$ fueron definidas anteriormente como las distancias dentro y entre-clúster, respectivamente.

$$CH(P_t) = \frac{B(P_t)}{W(P_t)} \frac{N-t}{t-1} \tag{2.24}$$

Global Silhouette Index

Para el cálculo de este índice es necesario primero definir varios parámetros. Considerar el ancho $s_i^{(k)}$del objeto simbólico T_i, perteneciente al clúster k-esimo, expresado en la Ecuación (2.25), donde $a_i^{(k)}$ es definido como la distancia promedio entre T_i y los $(M_k - 1)$ demás objetos pertenecientes al clúster k −esimo (Ecuación 2.26), y $b_i^{(k)}$ es la mínima distancia promedio entre el objeto T_i y los objetos pertenecientes a todos los demás grupos (Ecuación 2.27).

$$s_i^{(k)} = \frac{b_i^{(k)} - a_i^{(k)}}{\max\left(a_i^{(k)}, b_i^{(k)}\right)} \tag{2.25}$$

$$a_i^{(k)} = \frac{1}{M_k - 1} \sum_{\substack{j=1 \\ i \neq j}}^{M_k - 1} d_{ij}\ , \quad 1 \leq i \leq M_k \tag{2.26}$$

$$b_i^{(k)} = \min_{\substack{\mathrm{r}=1\ldots\mathrm{K} \\ \mathrm{r}\neq\mathrm{k}}} \left(\frac{1}{M_r} \sum_{j=1}^{M_r} d_{ij} \right), \quad 1 \leq i \leq M_k \tag{2.27}$$

El índice de silueta s_k del clúster C_k, es el promedio del ancho de los objetos pertenecientes al clúster k, mientras que el índice de silueta global SIL de la particion P_t es el promedio de los índices de silueta de sus grupos, tal como se establece en las Ecuaciones (2.28) y (2.29), respectivamente. Al igual que el índice CH, la partición óptima se encuentra cuando la función SIL alcanza su valor máximo.

$$s_k = \frac{1}{M_k} \sum_{i=1}^{k} s_i^{(k)} \tag{2.28}$$

$$SIL(P_t) = \frac{1}{K}\sum_{k=1}^{K} s_k \tag{2.29}$$

2.5. Caso de Estudio: Ensayos de Laboratorio

En la presente sección se estudian los métodos descritos anteriormente, utilizando ensayos de laboratorio bajo condiciones conocidas. En el Laboratorio de Estructuras del Departamento de Ingeniería Civil, el estudiante Pastor Villalpando realizó una serie de ensayos de daño progresivo a una estructura de metal de altura 1.99 [m], distribuidos en 6 niveles la cual estaba anclada a una mesa vibratoria con capacidad de generar vibraciones horizontales al nivel basal (Villalpando et al. 2016). Los niveles, formados por elementos de metal de sección rectangular se comportan como un diafragma rígido, mientras que las columnas son conformadas por barras planas de acero de una sección transversal de 50x3 [mm]. Se instalaron 9 acelerómetros con una tasa de frecuencia de 200 [Hz], ubicados en los distintos pisos y en diferentes direcciones. La Figura 2.6. muestra un modelo de la estructura analizada, así como también la ubicación de los sensores.

Villalpando realizó ensayos de vibración controlada, simulando registros de terremotos y también de ruido blanco, aplicados a distintas condiciones de daño, el que se aplicó mediante una reducción de la sección transversal de una de las columnas ubicadas bajo el segundo nivel, así como también a distintas configuraciones de masa mediante la incorporación de pesos equivalentes al 0.5% de la masa total de la estructura. La aplicación del daño progresivo se puede ver en la Figura 2.7., mientras que las condiciones de daño se resumen en la Tabla 2.2.

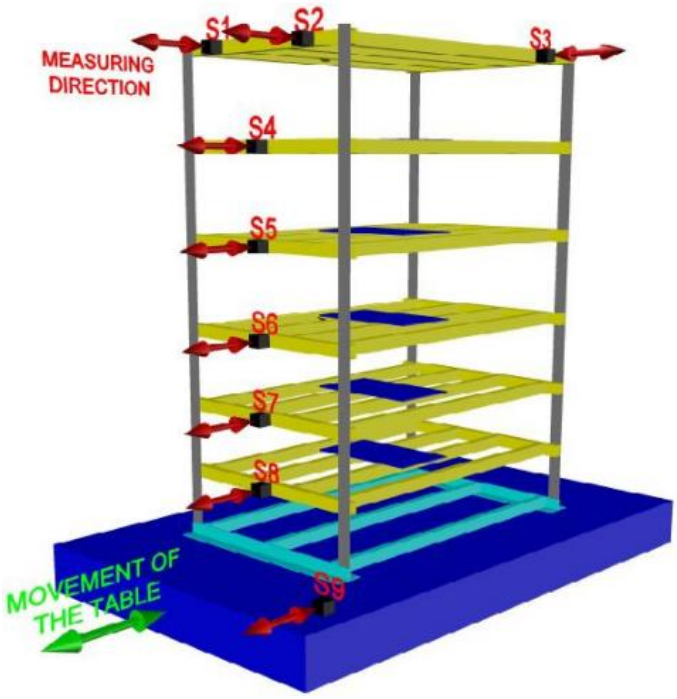


Figura 2.6. Modelo de la estructura analizada y ubicación de los sensores.
(Villalpando et al. 2016)

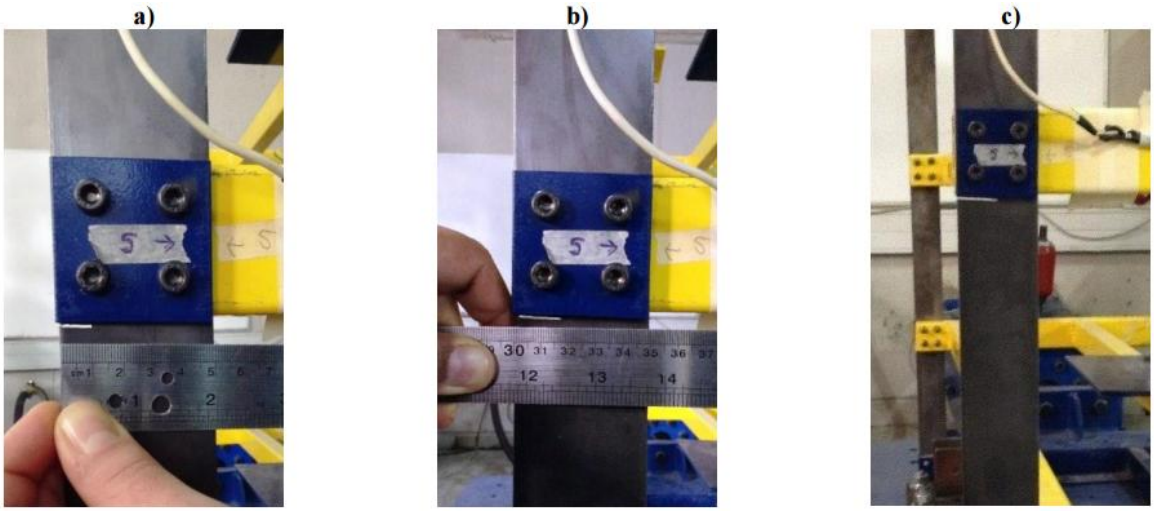


Figura 2.7. Aplicación del daño. a) Corte de 10mm. b) Corte de 17.5mm. c) Corte de 25mm.

(Villalpando et al. 2016)

Tabla 2.2. Condiciones de daños aplicados. Fuente: Modificada de Villalpando 2015.

Test	Condición	Leyenda	Caso N°
Normal	Normal	NORM	1
Normal	Normal + 0.5% masa añadida	NR+M	2
Daño Leve	Reducción del 20% en elemento 5	RE20	3
Daño Leve	Reducción del 20% en elemento 5 + 0.5% masa añadida	R20M	4
Daño Leve	Reducción del 35% en elemento 5	RE35	5
Daño Leve	Reducción del 35% en elemento 5 + 0.5% masa añadida	R35M	6
Daño Leve	Reducción del 50% en elemento 5	RE50	7
DañoLeve	Reducción del 35% en elemento 5 + 0.5% masa añadida	R50M	8
Daño Severo	Reducción del 35% en todas las columnas, nivel 2	RT35	9
Daño Severo	Reducción del 50% en todas las columnas, nivel 2	RT50	10

A cada una de estas condiciones de daño, aplicados en la columna frontal izquierda bajo el piso del sensor 7 (Figura 2.6), se les impuso 3 ruidos coloreados basales de aproximadamente 5 minutos de duración, llamados Ruido0, Ruido2 y Ruido 4, y se registró la respuesta mediante los acelerómetros. Esto resulta en campañas de adquisición de 10 condiciones de daño para las 3 aceleraciones basales distintas.

Se espera que como los ensayos se realizan en un ambiente controlado, la energía de entrada para todas las condiciones debiese ser la misma. Esto permite suponer que las diferencias entre las respuestas y por tanto de las series de aceleraciones adquiridas dependan de las condiciones estructurales más que de la excitación. De esta forma, para distintos casos de daño el contenido estadístico de las señales debe ser distinto y detectable usando algoritmos de objetos simbólicos.

Para validar el resultado del agrupamiento se necesitan numerosos objetos, pues de otra forma se puede llegar a una partición que tenga un objeto por clúster, lo cual no tiene sentido práctico. En estos ensayos, como cada condición de daño tiene una sola respuesta de 5 minutos, se realizaron divisiones de las series de aceleración para obtener 3 objetos por ensayo. Al realizar esta división lo que se está haciendo es otorgar más peso al comportamiento instantáneo de la respuesta, lo cual es contraproducente con los supuestos estadísticos y puede llevar a errores en los resultados. Sin embargo, es la única forma de poder trabajar con el set de ensayos realizados y por tanto hay que tomarlo en consideración al analizar el output de los algoritmos.

En la Figura 2.8. se muestran los registros de aceleración de la mesa para la vibración basal Ruido0. En los nombres de cada gráfico se destaca la condición estructural de daño, la aceleración promedio y la desviación estándar de la señal. Como es de esperar, el promedio de cada señal es nulo, pero la desviación estándar es diferente para cada ensayo, lo que indica que la mesa no tiene un comportamiento único. Además, en los casos RE50, RT35 y RT50 se pueden ver picosde impactos y a su vez una disminución de la desviación estándar, implicando una menor energía de entrada a la estructura. En la Figura 2.9. se grafican los espectros de densidad de potencia de la aceleración de la mesa y la energía de entrada E_i calculada como la integral bajo las curvas de densidad de potencia, confirmando que las excitaciones basales no son idénticas e invalidando la hipótesis de forzante controlada. En dicha figura, el caso 1 corresponde a "NORM", mientras que el décimo está asociado al caso "RT50"; los casos entremedio siguen el mismo orden que la leyenda del gráfico superior.

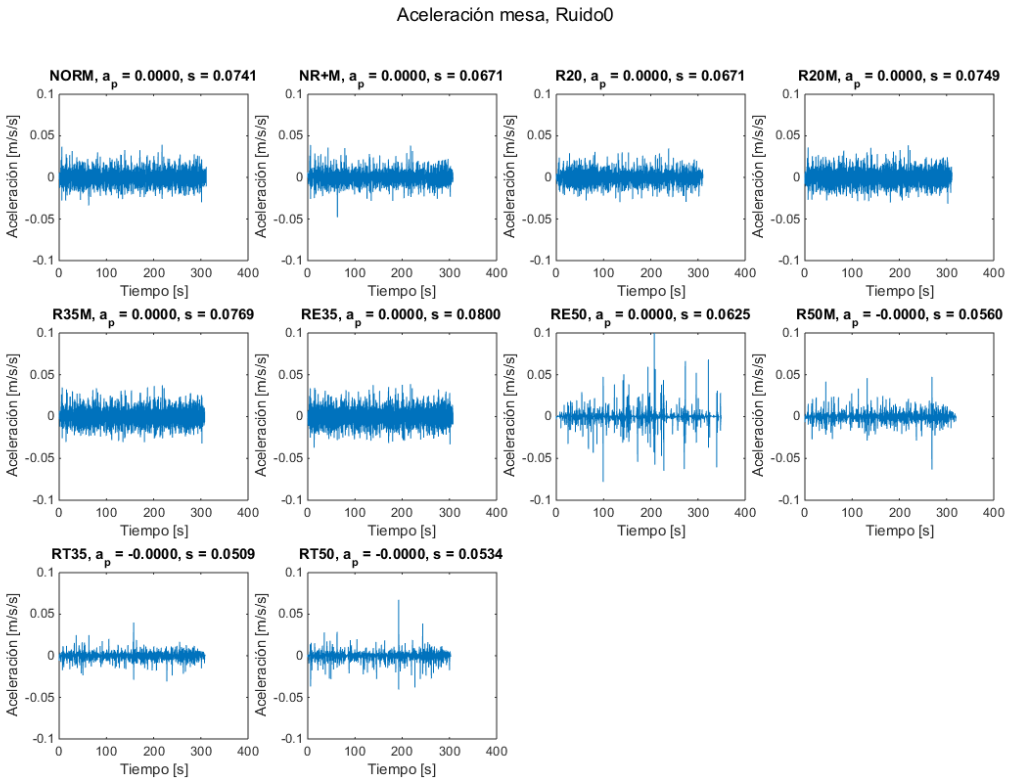


Figura 2.8. Registros de aceleración. Ruido0.

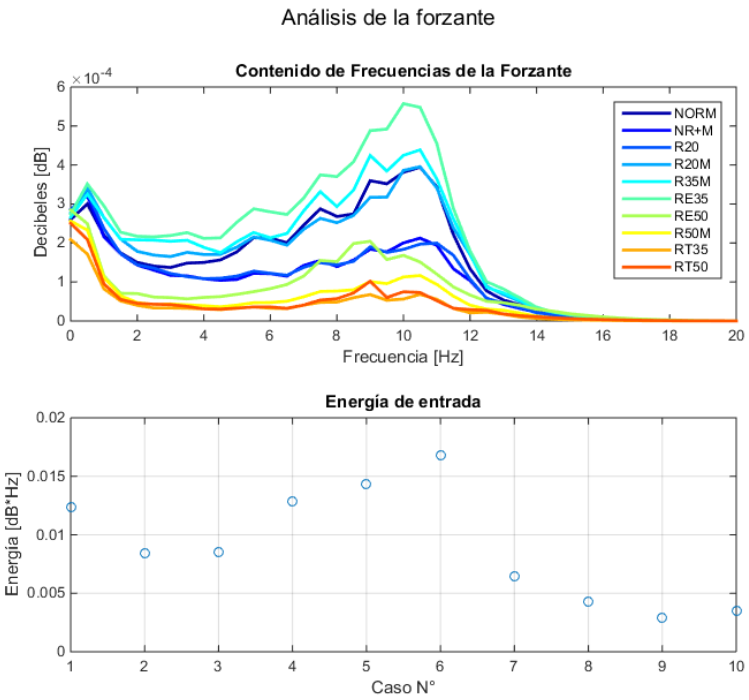


Figura 2.9. Contenido de Frecuencias, Ruido0.

Con cada ensayo de 5 minutos, se generan 3 objetos simbólicos. Entre cada par de ensayos se aplica la metodología expuesta anteriormente para generar las matrices de distancias, esperándose poder apreciar diferencias entre los objetos creados a partir de distintos ensayos. Como son 10 ensayos, existen 45 pares posibles, por lo cual se escogen algunos casos estructurales para no generar tantos gráficos. Los casos estructurales escogidos se muestran en la Tabla 2.3 y los resultados para las matrices de distancias entre los pares posibles se muestran en la Figura 2.10. Claramente hay algunos pares donde es evidente la existencia de una diferencia en el comportamiento estructural del marco estudiado. Sin embargo, para estudiar la naturaleza de esta diferencia se calcula la razón entre las energías de entrada de la Figura 2.9, mediante la fórmula de la Ecuación 2.30. Notar que en este caso, las funciones $\max(\cdot)$ y $\min(\cdot)$ son eligen simplemente el máximo y mínimo entre un par de energías previamente calculadas.

$$R_{i,j} = \frac{\max(E_i, E_j)}{\min(E_i, E_j)} \tag{2.30}$$

Un valor de R_{ij} igual a 1 indica que los ensayos i , $j$tuvieron la misma energía de entrada. Como se puede apreciar en la Figura 2.10, la separación de colores muestra una alta correlación con un valor R alto, lo que se puede interpretar como que en verdad la diferencia entre los objetos simbólicos tiene relación con la diferencia en la energía de entrada más que con la condición estructural. Más aún, cuando existe un valor R cercano a 1, no se pueden apreciar grupos evidentes.

Tabla 2.3. Pares de casos estructurales escogidos.

1	2	3	4	5
RE35 - RE50M	R35M-R50M	R35M-RE35	R20M-R50M	R20M-RE35M
6	**7**	**8**	**9**	**10**
R20M-R35M	NORM-R50M	NORM-RE35M	NORM-R35M	NORM-R20M

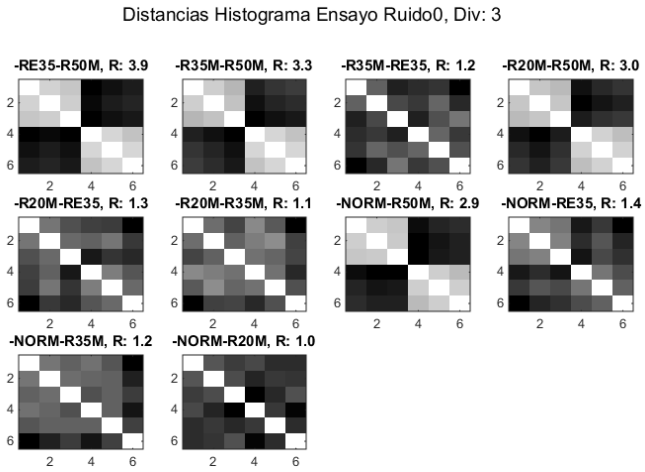


Figura 2.10. Matrices de distancias. Objetos tipo histograma. Ruido0.

Lo expuesto en el párrafo anterior es a partir de una inspección visual de las matrices de distancias. Para realizar un análisis más objetivo de los resultados, se utilizan los algoritmos de agrupamiento por nubes dinámicas y el índice de validación CH con la finalidad de obtener el número de grupos óptimo así como también la asignación de cada uno de los objetos a los clúster, para cada combinación de ensayos probadas. En la Tabla 2.4 se muestran los resultados obtenidos al aplicar esta metodología a las distancias graficadas en la Figura 2.10.

Tabla 2.4. Resultados de agrupamiento para ensayos de Ruido0.

	Combinación N°									
N° Grupos	2	2	3	2	2	3	2	2	3	3
Objetos	Asignación									
Obj 1	2	1	3	2	2	3	2	1	3	2
Obj 2	2	1	3	2	2	2	2	1	1	2
Obj 3	2	1	1	2	2	1	2	1	1	1
Obj 4	1	2	2	1	1	3	1	2	3	3
Obj 5	1	2	2	1	1	2	1	2	2	3
Obj 6	1	2	1	1	1	1	1	2	2	1

Basado en lo expuesto en la Tabla 2.4 se destaca que en el 60% de las combinaciones, el algoritmo pudo detectar correctamente dos comportamientos estructurales distintos, y la asignación de los objetos a los distintos grupos se realizó con 100% de precisión. Por otra parte, el peor de los resultados arrojó 3 grupos y un 33% de asignación correcta. Si bien esto puede parecer una metodología exitosa, hay que destacar que la evidencia indica que el algoritmo está detectando un comportamiento estructural distinto producto de que las excitaciones tienen amplitudes distintas, no debido a una condición estructural distinta. Esta afirmación será sometida a prueba durante la siguiente sección, en la que la amplitud medida por los sensores será multiplicada por un factor de escala en relación a la energía de entrada, calculada a partir del espectro de potencia de la Figura 2.9.

2.5.1. Análisis de Sensibilidad de la Metodología.

En la sección anterior se mostró un procedimiento general con el que consigue clasificar automáticamente en distintos conjuntos a las series de aceleración adquiridas. En particular, se aplicó la metodología considerando objetos obtenidos a partir del cálculo de histograma de las series de aceleración, junto con el método de agrupamiento de nubes dinámicas y posteriormente una validación usando el índice de Calinski & Harabasz. La elección de estos submétodos fue más bien arbitraria por lo que en la presente sección se realiza un análisis de sensibilidad con el que se busca justificar dicha elección.

Considerar los ensayos de Ruido detallados anteriormente. A modo de resumen, estos constan de 3 diferentes excitaciones de ruido coloreado, aplicados a 10 condiciones estructurales distintas. La idea es crear todas las parejas posibles de estados estructurales, hacer una realización de la metodología para alguna combinación de sub-métodos, y finalmente calcular el porcentaje de asignaciones correctas. Lógicamente, un resultado completamente correcto es aquél que entrega dos conjuntos, con la primera mitad de los objetos pertenecientes a un conjunto y la segunda mitad al otro conjunto. No obstante, como los algoritmos no son perfectos, hay algunas asignaciones que son incorrectas. En la Tabla 2.5. se muestra un ejemplo de asignaciones y su porcentaje de exactitud.

Tabla 2.5. Ejemplos de particiones y su exactitud.

Partición	Obj 1	Obj 2	Obj 3	Obj 4	Obj 5	Obj 6	Porcentaje
A1	1	1	1	2	2	2	**100%**
A2	1	2	2	1	2	2	**50%**
A3	1	2	1	2	2	2	**83%**
A4	1	2	3	4	5	6	**33%**

Dentro de las opciones que se estudiarán en relación a los sub-métodos a aplicar, se encuentran los siguientes:

- Tipo de Objeto Simbólico: {Interquartil, Histograma}
- Algoritmo de Agrupamiento: {Jerárquico Aglomerativo, K-Means, Dynamic-Medoids (Nubes Dinámicas)}
- Índice de Validación: {Calinski & Harabasz, Silhouette}
- Factor de Escala: {Sin escalar, Con aplicar escala por energía}

Cabe mencionar que cada algoritmo de agrupamiento tiene asociado una distancia en particular. En este caso, la aplicación del método Jerárquico Aglomerativo está asociada a una distancia Haussdorf-Euclideana, mientras que K-Means y Dynamic-Medoids (Nubes Dinámicas) lo están a distancia Euclideana y Gowda-Diday, respectivamente.

La cantidad de combinaciones posibles de sub-métodos es igual a $2 \cdot 3 \cdot 2 \cdot 2 = 24$, y la cantidad de pares de ensayos es igual a $\binom{10}{2} = 45$. Esto implica que se calculan muchísimas matrices de distancias sobre las cuales se aplican los algoritmos de agrupamiento. No obstante, el resultado que más importa tiene relación con el porcentaje de exactitud, lo que permite generar tablas de resumen que condensan toda la información. Antes de mostrar los resultados, es importante detallar el procedimiento de normalización por energía de entrada:

Factor de Escala según energía de entrada.

Observar nuevamente la Figura 2.9 en la que se muestra la diferencia de energías de entrada para los distintos ensayos. Siguiendo la misma notación, E_i representa la energía de entrada asociada al caso estructural i-ésimo. De esta forma, el factor de escala se define según la Ecuación (2.31), donde $\overline{E}$ es el promedio de las energías.

$$\lambda_i = \sqrt{\frac{\overline{E}}{E_i}} \tag{2.31}$$

Al aplicar este factor de escala a las series de la Figura 2.9, se obtienen los resultados mostrados en la Figura 2.11, en los que se aplica el factor nuevamente a las señales obtenidas en la mesa, con los que se calcula el espectro y finalmente se vuelve a integrar para calcular la energía normalizada. Es notorio que la aplicación logra llevar las energías de entrada a un mismo nivel. Por tanto, aplicando el factor de escala a las mediciones del resto de los sensores, se podría aseverar que hay una mayor igualdad entre las energías de cada ensayo.

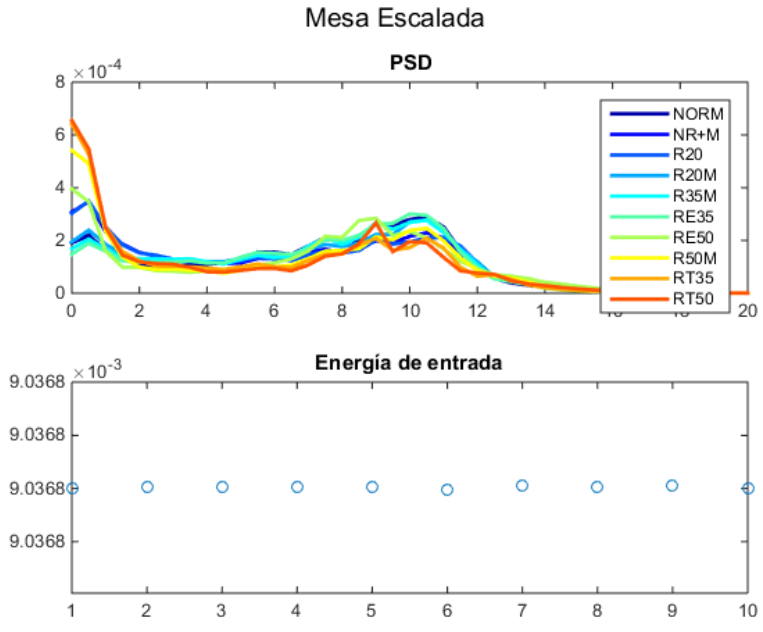


Figura N°2.11. Resultados de la aplicación del factor de escala por energía.

Resultados del análisis de sensibilidad.

Recordando que existen 45 pares de estados estructurales para cada combinación posible de submétodos, los resultados mostrados a continuación son el promedio del porcentaje de exactitud para los pares estructurales. En total son 24 combinaciones, las que fueron aplicadas para cada uno de los tres distintos tipos de excitación de ruido.

Tabla N°2.6. Resultados de análisis de sensibilidad.

Combinación	R0 (%)	R2 (%)	R4 (%)	Combinación	R0 (%)	R2 (%)	R4 (%)
Hier_CH_NoEsc_Hist	87	69	74	Hier_CH_NoEsc_IQ	84	69	70
Kmean_CH_NoEsc_Hist	93	71	77	Kmean_CH_NoEsc_IQ	88	73	69
Dyn_CH_NoEsc_Hist	89	73	80	Dyn_CH_NoEsc_IQ	90	72	76
Hier_SIL_NoEsc_Hist	53	49	47	Hier_SIL_NoEsc_IQ	51	48	45
Kmean_SIL_NoEsc_Hist	51	48	48	Kmean_SIL_NoEsc_IQ	48	47	43
Dyn_SIL_NoEsc_Hist	49	48	47	Dyn_SIL_NoEsc_IQ	48	47	44
Hier_CH_Esc_Hist	64	55	60	Hier_CH_Esc_IQ	60	53	61
Kmean_CH_Esc_Hist	78	62	71	Kmean_CH_Esc_IQ	64	55	53
Dyn_CH_Esc_Hist	76	61	68	Dyn_CH_Esc_IQ	64	59	67
Hier_SIL_Esc_Hist	47	43	42	Hier_SIL_Esc_IQ	43	41	44
Kmean_SIL_Esc_Hist	47	44	45	Kmean_SIL_Esc_IQ	42	42	38
Dyn_SIL_Esc_Hist	47	43	41	Dyn_SIL_Esc_IQ	44	42	44

En la Tabla N°2.6 se muestran los resultados obtenidos para todas las combinaciones, una vez que se calculó el promedio de los porcentajes de exactitud para cada uno de los pares posibles entre estados estructurales. Ahora, si agrupamos los promedios en categorías respecto a los sub-métodos, encontramos los resultados que se muestran en la Tabla 2.7.

Tabla 2.7. Resultados del análisis de sensibilidad.

Histograma (%)			
R0	R2	R4	Promedio
65.1	55.4	58.2	**59.6**

Interquartil (%)			
R0	R2	R4	Promedio
60.4	54.0	54.5	**56.3**

Calinski & Harabasz (%)			
R0	R2	R4	Promedio
78.0	64.3	68.8	**70.4**

Silouette (%)			
R0	R2	R4	Promedio
47.4	45.1	43.9	**45.5**

Sin Factor de Escala por Energía (%)			
R0	R2	R4	Promedio
69.2	59.4	60.0	**62.9**

Con Factor de Escala por Energía (%)			
R0	R2	R4	Promedio
56.3	50.0	52.7	**53.0**

Hierarchical Agglomerative (%)			
R0	R2	R4	Promedio
61.1	53.5	55.2	**56.6**

K-Means (%)			
R0	R2	R4	Promedio
63.8	55.1	55.4	**58.1**

Dynamics Medoids (%)			
R0	R2	R4	Promedio
63.4	55.5	58.4	**59.1**

Selección de Mejores Indicadores (%)			
Histograma, CH, Sin Escala, Dyn-Med			
R0	**R2**	**R4**	**Promedio**
76.4	**61.1**	**67.6**	**68.4**

Discusión de Resultados.

En este punto del análisis es recomendable recordar que el objetivo del estudio de sensibilidad es poder determinar qué combinación de sub-métodos entregan los mejores resultados. Es por esto que se utilizan los promedios de los porcentajes de exactitud como indicadores de la calidad del resultado. En los siguientes párrafos se discuten los resultados y se selecciona la combinación que mostró la mayor exactitud al momento de su aplicación.

En la Tabla 2.7, los resultados son agrupados según cada sub-método de forma aislada. Esto quiere decir que de las 24 combinaciones de la Tabla 2.6, se agrupan todos los resultados que tengan que ver con un sub-método en particular. De esta forma, tanto para objetos simbólicos derivados de histogramas como aquellos derivados de interquartil, por ejemplo, hay una correspondencia de 12 resultados, los que se promedian y resumen en la Tabla 2.7. Esto es análogo para los demás tipos de algoritmos y permite una rápida comparación entre ellos.

Partiendo la discusión por los índices de validación, vemos que CH tiene una exactitud mucho mayor que SIL. Por tanto, CH será utilizado en el resto de la investigación como validador de particiones óptimas. El porqué de este resultado se escapa del alcance de esta tesis, pero cabe destacar que la formulación matemática e implementación en Matlab de CH es muchísimo más sencilla que la de SIL.

En relación a los algoritmos de agrupamiento estudiados, recordemos que cada uno está asociado a una distancia distinta. El algoritmo jerárquico hace uso de la distancia de Gowda-Diday en el caso de objetos tipo histogramas, y la distancia Haussdorf-Euclideana en el caso de

objetos interquartile; K-means utiliza distancia euclideana para ambos tipos de objetos; Dyn-Medoids es análogo al algoritmo jerárquico. La diferencia entre los tipos de objetos hace que sea mejor introducir una nueva tabla con los resultados agrupados considerando esta distinción, lo que se muestra en la Tabla 2.8. Para la creación de dicha tabla, se omitieron los resultados asociados al índice de validación SIL.

Tabla N°2.8. Resultados agrupados por algoritmo de agrupamiento y tipo de objeto.

Jerárquico							
Histograma				IQ			
R0	R2	R4	Prom	R0	R2	R4	Prom
75.5	**62.0**	**66.7**	**68.1**	**72.0**	**61.1**	**65.5**	**66.2**
Kmeans							
Histograma				IQ			
R0	R2	R4	Prom	R0	R2	R4	Prom
85.2	**66.4**	**74.1**	**75.2**	**75.9**	**63.9**	**61.1**	**67.0**
Dyn-Medoids							
Histograma				IQ			
R0	R2	R4	Prom	R0	R2	R4	Prom
82.9	**66.9**	**73.8**	**74.5**	**76.9**	**65.5**	**71.5**	**71.3**

A partir de la Tabla 2.8, se pueden apreciar varios resultados interesantes. Primero que todo, observamos que el método jerárquico aglomerativo es el de menor rendimiento, independientemente del tipo de objeto simbólico utilizado, y además muestra poca diferencia porcentual (menos que un 2%) en su exactitud al comparar los resultados de histograma o interquartil. Por otra parte, entre K-means y Dyn-medoids hay muy poca diferencia porcentual. Analizando el resultado separando por objeto simbólico de histogramas, tomando en consideración que K-Means utiliza distancia Euclideana y a su vez que Dynamic-Medoids utiliza la distancia Gowda-Diday diseñada para objetos de histograma, llama enormemente la atención que el resultado favorezca a K-Means, indicando de que la mencionada distancia no aporta realmente información o tiene un nivel de sensibilidad similar a una distancia Euclideana. Pero cuando aislamos resultados asociados a objetos interquartil, la conclusión es al revés, mostrando una mayor sensibilidad la distancia de Haussdorf-Euclideana. Ahora bien, los objetos derivados de histogramas tienen una dimensionalidad mucho mayor (de un orden de 10 veces más dimensiones), que los de interquartil, y puede que este aumento de dimensionalidad sea el responsable de una mayor sensibilidad, lo que nos lleva a otro punto que vale la pena recalcar: independientemente del método de agrupamiento utilizado, los objetos simbólicos obtenidos a partir del cálculo de histogramas arrojan resultados más exactos que los de tipo interquartil. Esta afirmación era una de las primeras que salía a la vista al observar la Tabla 2.7 original. Efectivamente, el objeto tipo interquartil se puede asociar directamente con la amplitud, mientras que los objetos de histograma además de contener información relacionada con la amplitud, también aporta conocimientos respecto a la distribución de las mediciones.

Por último, el punto más crítico al analizar los resultados corresponde a la comparación entre las series sin aplicación de factor de escala, con las que sí tuvieron la multiplicación. Para hacer más fácil la comprensión, se elabora la Tabla 2.9, en la que se efectúa una nueva selección y agrupación de resultados, esta vez sin incluir ni los asociados a SIL, ni los asociados a

Jerárquico. En esta tabla hay dos resultados primordiales. El primero tiene que ver con que cuando hay una normalización de la energía, los resultados tienden a ser menor precisos. Esto nos indica que la amplitud del movimiento es determinante al momento de aplicar esta metodología ya que simplemente, los cambios en las mediciones producto de un cambio estructural quedan enmascarados por los cambios en la amplitud debido a una mayor energía de entrada. En otras palabras, en el caso de una edificio real con sistema de monitoreo de salud estructural, hay que buscar la forma de asegurar que se está comparando la estructura bajo el mismo régimen de excitación. Si en el párrafo anterior mencionábamos que los objetos de histograma contenían información de amplitud y distribución, y que los interquartil sólo de amplitud, la poca diferencia entre estas dos opciones para el caso sin escalar, favorece a la afirmación de que la amplitud es altamente determinante, haciendo que el aporte del histograma por concepto de distribución sea menor. En cambio, los resultados en el caso normalizado muestran justamente una mayor información otorgada por los objetos tipo histograma, apoyando una vez más el supuesto.

Tabla N°2.9. Diferencias entre series escaladas y las sin escalar.

Histograma Sin Escalar				IQ Sin Escalar			
R0	R2	R4	Prom	R0	R2	R4	Prom
89.7	70.8	77.0	**79.2**	89.7	71.3	71.9	**77.6**
Histograma Escalado				IQ Escalado			
R0	R2	R4	Prom	R0	R2	R4	Prom
72.7	59.4	66.0	**66.0**	62.8	55.7	60.2	**59.6**

Conclusiones

A partir de los resultados y la discusión, se concluye que los métodos más apropiados para el estudio, de acuerdo con el ejemplo estudiado, corresponden a utilizar objetos derivados de histogramas, usar K-Means Euclideano o Dynamic-Medoids con Gowda-Diday como algoritmo de agrupamiento, y finalmente validar y calcular la partición óptima con el índice de Calinski & Harabasz. Cabe destacar que este resultado es válido para la estructura y ensayos estudiados y no debiesen ser generalizados para todo tipo de estructuras y condiciones operacionales.

2.6 Caso de Estudio: Torre Central, FCFM.

En esta sección se expone la aplicación de la metodología de agrupamiento de objetos simbólicos obtenidos desde mediciones brutas de aceleración, aplicado al edificio de la Torre Central de la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile. El objetivo utilizar la teoría detallada en las secciones anteriores, buscando identificar la ocurrencia de cambio estructural debido al terremoto del 27 de Febrero del 2010.

2.6.1. Estructura y adquisición de datos.

La estructura considerada en este estudio ya ha sido analizada anteriormente por memoristas y tesistas del Departamento de Ingeniería Civil de la FCFM. Pablo León hace una muy buena descripción de la instalación en su memoria de Título, la cual se resume en los siguientes párrafos.

La Torre Central es uno de los edificios que componen el Campus Beauchef de la Universidad de Chile. Fue construido en 1962 y actualmente alberga oficinas administrativas. Un set de fotografías del edificio se muestra en la Figura 2.11.

Figura 2.11. Fotografías del edificio Torre Central. ***fuente: Memoria Pablo León.***

Desde su construcción, se han aplicado varias remodelaciones estructurales con el fin de liberar espacios y modificar la usabilidad del edificio. En 1993, por ejemplo, se generaron aperturas en muros interiores para generar accesos a nuevas oficinas. En el 2008, se remodeló el interior del segundo y tercer piso, instalando tabiquería para reacomodar espacios.

La Torre Central se encuentra instrumentada con 8 acelerómetros unixiales, modelo EpiSensor ES-U2 de marca Kinemetrics, y el sistema de adquisición está configurado para generar registros a una tasa de muestreo de 200[Hz], los que son almacenados en archivos que guardan la información cada 15 minutos. Las características y disposición de los sensores se encuentran en la Tabla 2.10. Su ubicación vista en planta se muestra en la Figura N°2.12.

Tabla 2.10. Características de Sensores. (fuente: Pablo León)

Ubicación	Canal	Sensor	Serial	Dirección	Observaciones
Piso 8	1	EPI 4	346	Oeste - Este	-
	2	EPI 5	345	Norte - Sur	-
	3	EPI 6	504	Sur - Norte	Exterior
Piso 3	4	EPI 7	1334	Este - Oeste	-
	5	EPI 8	1336	Sur - Norte	-
	6	EPI 9	1335	Sur - Norte	Exterior
Subterráneo	7	EPI 10	1337	Este - Oeste	-
	8	EPI 12	1339	Sur - Norte	-

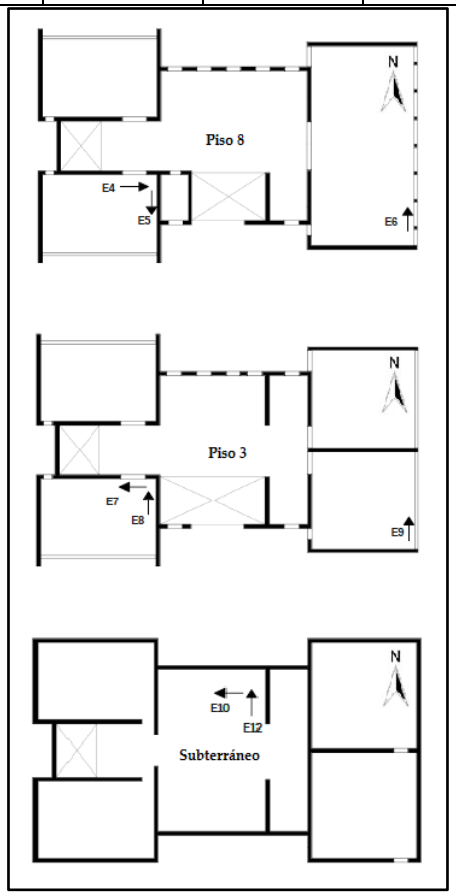


Figura N°2.12. Ubicación de los sensores.

Selección de canales y limpieza de datos.

Los datos disponibles para el estudio de la metodología en la Torre Central corresponden a archivos de medición de aceleración cada 15 minutos para los 31 días de Enero y Marzo del 2010, con excepción de las primeras 10 horas del primero de Marzo. Como es sabido, el 27 de

Febrero ocurrió un terremoto de gran magnitud que afectó a toda la zona central de Chile. Es de esperar que la Torre haya visto modificadas sus propiedades dinámicas producto de esto, lo que es deseado de poder observar en los resultados.

Antes de aplicar la metodología a los registros, es necesario realizar un preprocesamiento de limpieza de datos. Esto consiste en aplicar filtro pasa banda en el espacio de la frecuencia con el que se logra eliminar gran parte del ruido eléctrico y operacional presente en los registros. Sin embargo, no todos los canales tienen la misma calidad y por tanto se deben eliminar los que pese a la aplicación del filtro sigan arrojando datos muy ruidosos. En definitiva, de los 8 canales, solo se utilizan los 4, 5 y 6 que corresponden a los acelerómetros instalados en el piso 3.

2.6.2. Distancias

El cálculo de distancia se realiza considerando la formula de Gowda-Diday, si bien también se aplica el algoritmo K-Means para el análisis. Una de las principales preguntas que vale la pena hacerse es decidir que objetos comparar entre sí. Por ejemplo, se podría utilizar objetos provenientes del mismo día (i.e primer lunes del mes), o comparar absolutamente todos los objetos. Lamentablemente, si aceptamos el supuesto de que la amplitud de las vibraciones medidas es determinante en el resultado, no es recomendable comparar objetos obtenidos de registros provenientes de distinto regímenes de excitación, o en términos prácticos, no es bueno comparar registros nocturnos con muy poca vibración, con aquellos obtenidos durante el edificio en condición operacional.

Para ir ejemplificando lo mencionado en el párrafo anterior, en la Figura 2.12 se muestra una matriz de distancias obtenida para el día Lunes 4 de Enero del 2010. Al ser un día hábil, se puede esperar que la estructura presente evidencias de excitación externa en los horarios de oficina, habitualmente en Chile desde las 8:00 hasta las 19:00, aunque por supuesto que depende del edificio y organización de sus ocupantes. En el caso de la Torre Central, la Figura 2.13 muestra claramente una diferencia entre los objetos obtenidos durante los periodos nocturnos y los diurnos. Es más, parece existir una transición suave entre estos estados, lo que tiene sentido si se piensa que no todas las personas llegan exactamente a la misma hora. Es evidente que si se utiliza un algoritmo de agrupamiento para la matriz de esta figura, los conjuntos entregados realmente van a estar separados por la condicion operacional, obteniendo estados de "estructura quieta" y "estructura en movimiento". Esto es contraproducente con el objetivo de detectar cambios estructurales, pero abre nuevas posibilidades al uso de los registros de aceleración ya que vemos que también se pueden detectar patrones del comportamiento de uso del edificio. Más aún, este resultado es generalizable y se puede llegar a comparar matrices de distancias obtenidas de días distintos. La Figura 2.14 muestra las matrices de distancias obtenidas para 4 semanas de enero de 2010, en las que se omite los valores de los ejes por un tema de espacio. En esta figura, resulta interesante ver como incluso se pueden apreciar claras diferencias entre días de semana y los días de fines de semana. Aún más, la mayoría de los días de semana presentan una cruceta en los objetos correspondientes al horario de colación.

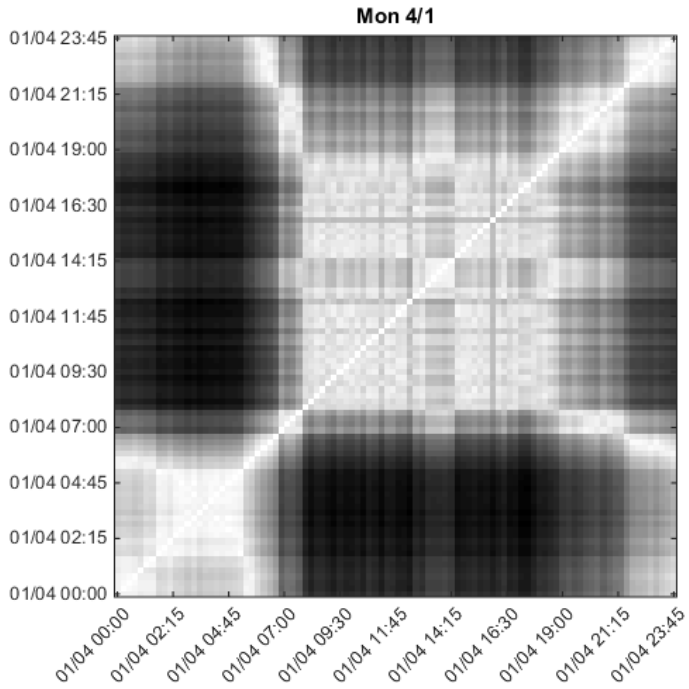


Figura N°2.13. Matriz de distancias para el 4 de Enero del 2010.

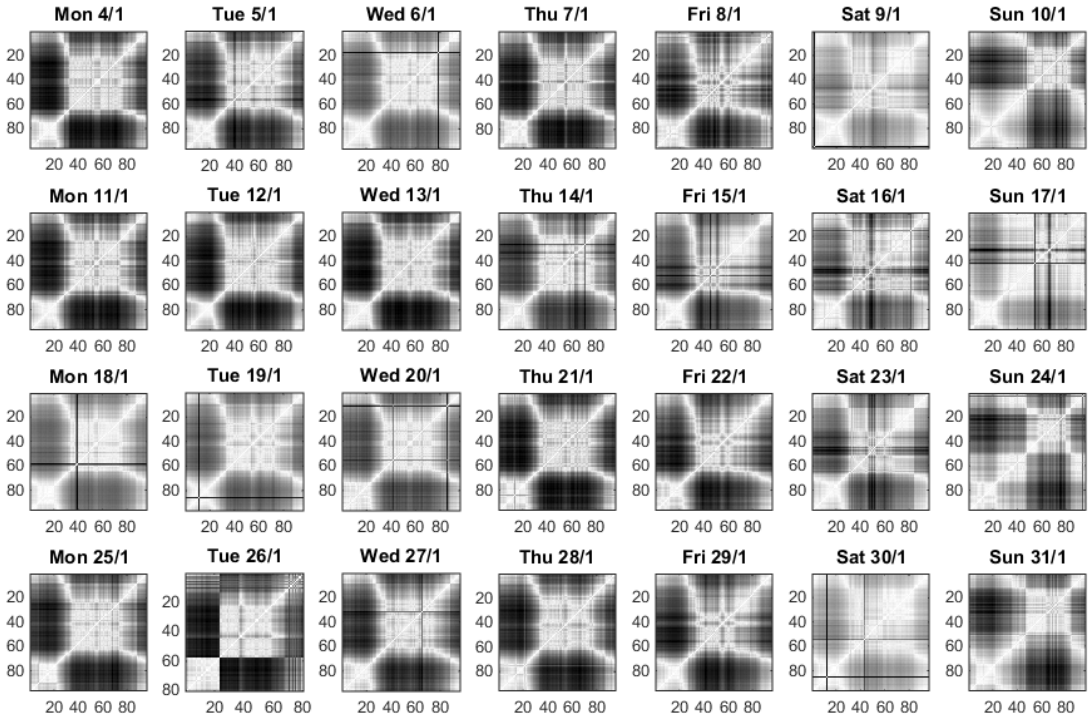
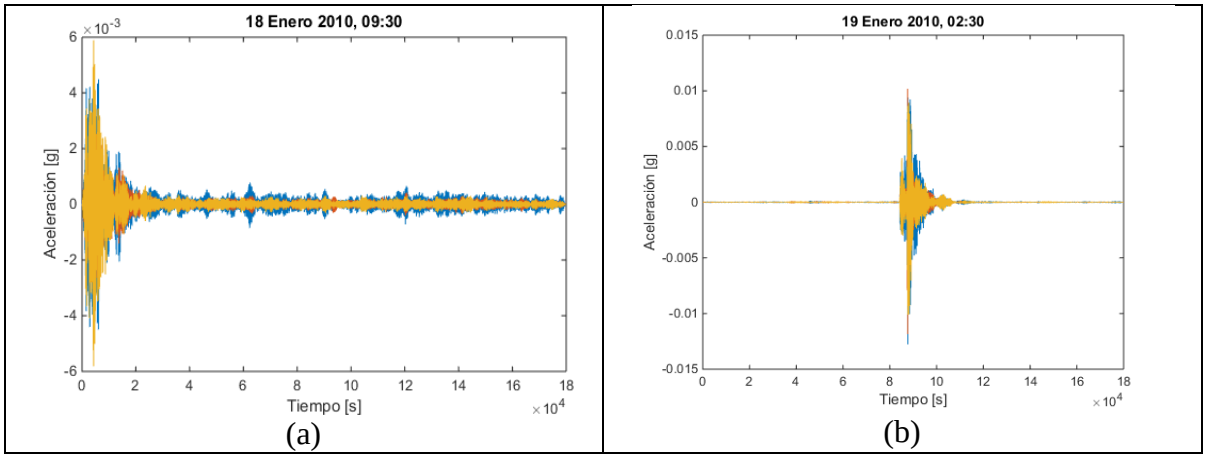


Figura N°2.14. Matrices de distancias para Enero 2010.

Otro punto importante de notar en la Figura 2.14, son las cruces negras que aparecen con un comportamiento aleatorio. Estas corresponden a movimientos sísmicos registrados por los sensores, y es fácil de verificar ya que tan solo basta con mirar las series de aceleración

correspondientes al objeto responsable de la cruz. Para ejemplificar esto, observar la matriz de distancia correspondiente al lunes 18 de Enero. En dicha matriz, se observa una cruz negra cerca del objeto N°40, aproximadamente. Específicamente, ésta corresponde al objeto creado a partir del registro de las 9:30 am, el cual se muestra en la Figura 2.15(a) evidenciando claramente la presencia de un movimiento telúrico en el inicio del registro. En la Figura 2.15(b), se muestra lo mismo para la cruz apreciada en la matriz de distancia del 19 de Enero del 2010, confirmando una vez más lo sensible del método a la amplitud del movimiento. En adelante, llamaremos objetos 'outliers' a los que ocasionan las cruces negras.



(a) (b)

Figura N°2.15. Ejemplos de registros que incluyen un sismo.

Lo mencionado en el párrafo anterior es de suma importancia al momento de utilizar las matrices de distancias como entrada para los algoritmos de agrupamiento, ya que se alejan de lo que sería una matriz ideal. A modo de recordatorio, una matriz de distancias que contenga una separación de estados perfecta es la mostrada en la Figura 2.4, en la que se puede apreciar claramente dos conjuntos de valores de distancias: claro, representando a los objetos cercanos entre sí; y oscuro, a los distanciados entre sí. Por tanto, la ocurrencia de cruces negras indica que un objeto en particular está muy alejado del resto de los demás objetos, y los algoritmos de agrupamiento así lo percibirán, dejándolo en un clúster aislado. A continuación se estudiarán un par de formas con las que se puede trabajar estos casos.

Eliminación de objetos muy distanciados.

Considerar la matriz de distancias de la Figura 2.16 correspondiente al domingo 17 de Enero. En esta figura se pueden apreciar más de una cruz negra. Como los algoritmos de agrupamiento son de orden computacional cuadrático, es conveniente eliminar los objetos outliers antes de la clasificación. El método de eliminación de objetos muy distanciados nace a partir del estudio de la distribución del valor de las distancias. Observar la Figura 2.17 que incluye un gráfico de las distancias de la Figura 2.16 en forma de lista y además el histograma de ellas. Notar que la distribución no es gaussiana, pero que sí es posible apreciar la ocurrencia de anomalías en la serie de distancias. Dichas anomalías son atribuibles a los objetos outliers.

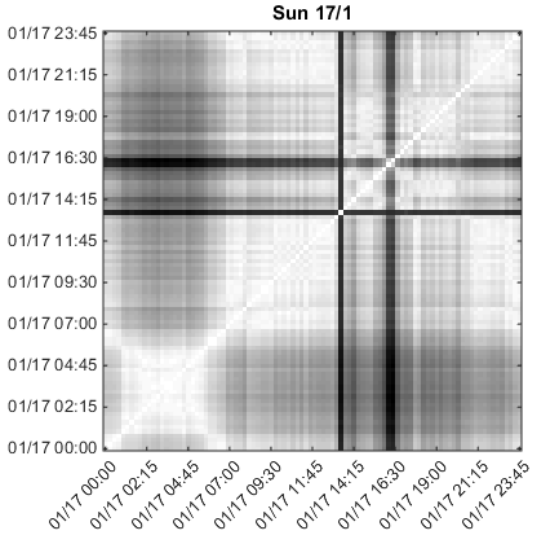


Figura N°2.16. Matriz de distancias del domingo 17 de Enero.

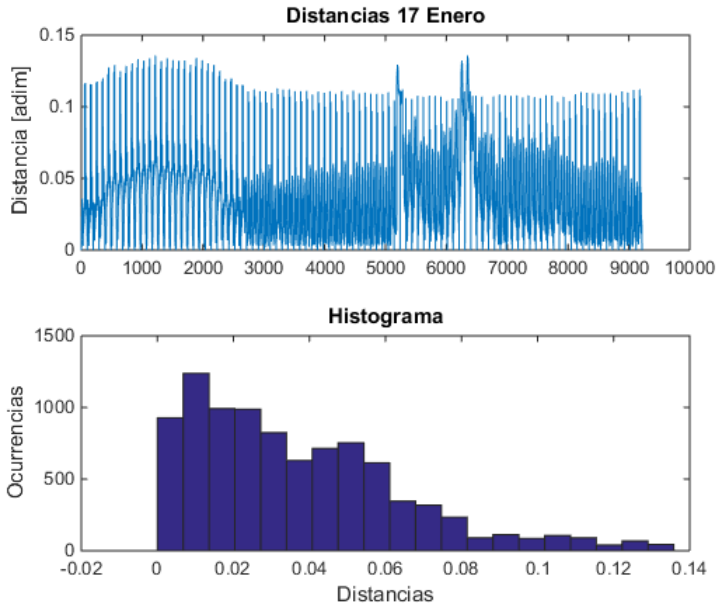


Figura N°2.17. Serie de distancias y su histograma.

Una de las opciones que se pueden considerar hasta este punto, consiste en calcular una función de densidad de probabilidad a partir del histograma, y eliminar del cálculo las distancias que se encuentren sobre cierto valor; por ejemplo, eliminar las distancias que se encuentren sobre el 90% (i.e eliminar el decil superior). Sin embargo, esa estrategia enfrentaría dos dificultades. La primera de ellas es que no todas las matrices presentan objetos outliers, y estipular un porcentaje de aceptación global haría que se eliminaran objetos perfectamente válidos. La otra dificultad, aún mayor, es que el decil superior no tiene porqué pertenecer a los mismos objetos, es decir, se puede dar que se esté eliminando una columna de la matriz a medias. Para resolver este problema, es mejor trabajar con la distancia total de cada objeto, calculada como la suma de las distancias de un objeto hacia todos los demás. Cabe destacar que en el caso de una matriz de distancias, la distancia total de asociada a un objeto no es más que calcular la suma de la columna

correspondiente a dicho objeto. La Figura 2.18 superior contiene la distancia total de cada uno de los 96 objetos asociados al 17 de Enero. En ella se observa la presencia de tres objetos claramente fuera de la tendencia. Más aún, en la Figura 2.18 inferior, se ordenan las distancias de forma ascendente.

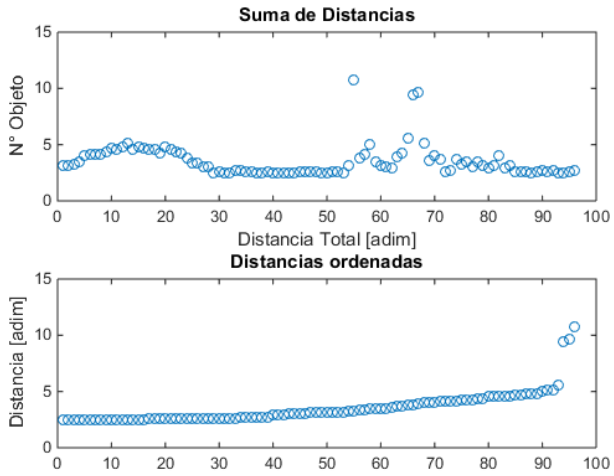


Figura N°2.18. Distancia total de cada objeto.

Nuevamente, la pregunta es cómo seleccionar esos tres objetos que claramente están fuera de la tendencia. El método que se propone consiste en realizar un ajuste bilineal de la serie de distancias ordenadas, y seleccionar los objetos que pertenezcan a la primera parte del ajuste. Estos objetos son los que estarían dentro de la tendencia, y el método tiene la gran ventaja de que el número de objetos a eliminar depende de la calidad de éstos. La Figura 2.19 contiene la aplicación del ajuste bilineal, mientras que la Figura 2.20 muestra la matriz del 17 de Enero una vez que se eliminaron los objetos outliers.

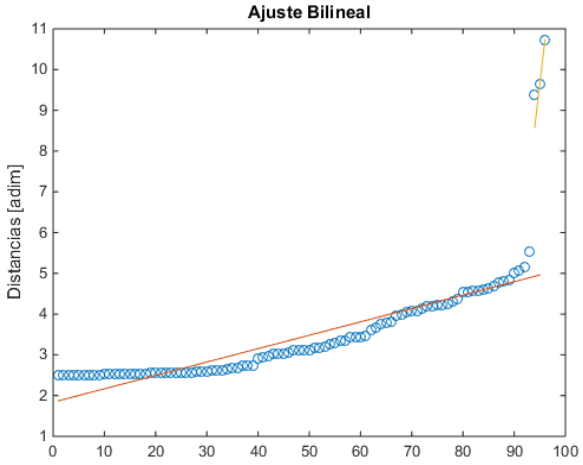


Figura N°2.19. Ajuste bilineal para las distancias agregadas del 17 Enero.

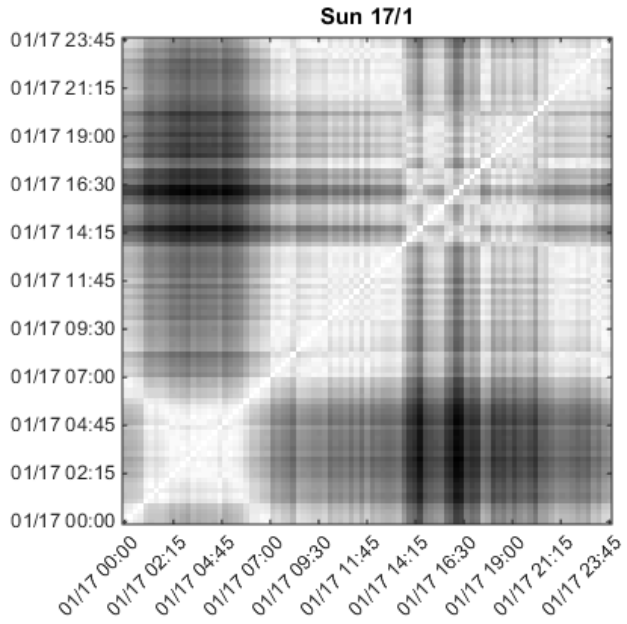


Figura N°2.20. Matriz del 17 de Enero 2010 sin objetos outliers.

Estudio de los límites de intervalos extremos.

En la Sección 2.2, específicamente en las Figuras 2.1 y 2.2, se muestran resultados de histogramas habiendo considerado dos límites distintos para los intervalos extremos. La relación de estos límites con la sensibilidad final del método completo en relación a la clasificación estructural no es clara, pero a continuación se realizan comparaciones cualitativas para distintos límites y se muestra cómo podrían ser utilizados para eliminar el efecto de objetos outliers.

Considerar nuevamente el domingo 17 de Enero, el cual posee tres objetos outliers y que en la sección anterior fueron eliminados por distanciamiento. Ahora, se verá cómo es posible eliminar el efecto del aumento de amplitud, sin necesidad de eliminar los objetos, mediante una selección adecuada de intervalos extremos. En la Figura 2.20 se muestra una comparación de matrices de distancias para el 17 de Enero utilizando cuatro límites distintos para los intervalos extremos. Estos límites corresponden a (a) $a_l = \pm 0.0005g$, (b) $a_l = \pm 0.00005g$, (c) $a_l = \pm 0.000005g$ y (d) $a_l = \pm 0.0000005g$. En (a), se observa claramente la presencia de objetos outliers al existir cruces negras. Sin embargo, a medida que se achica cada vez más el intervalo, se ve que dichas cruces van desapareciendo, hasta llegar al punto de ser imperceptibles en el caso (d). Este resultado es muy llamativo, ya que mantiene la integridad de los registros al no existir la necesidad de eliminar objetos y pareciera ser una alternativa muy viable para trabajar en lugares con alta sismicidad. No obstante, al imponer límites de aceleración tan pequeños puede que se esté aplicando un sesgo el cual en este momento es muy difícil de determinar si existe o no. Sin duda que es un método que requiere de mayor investigación.

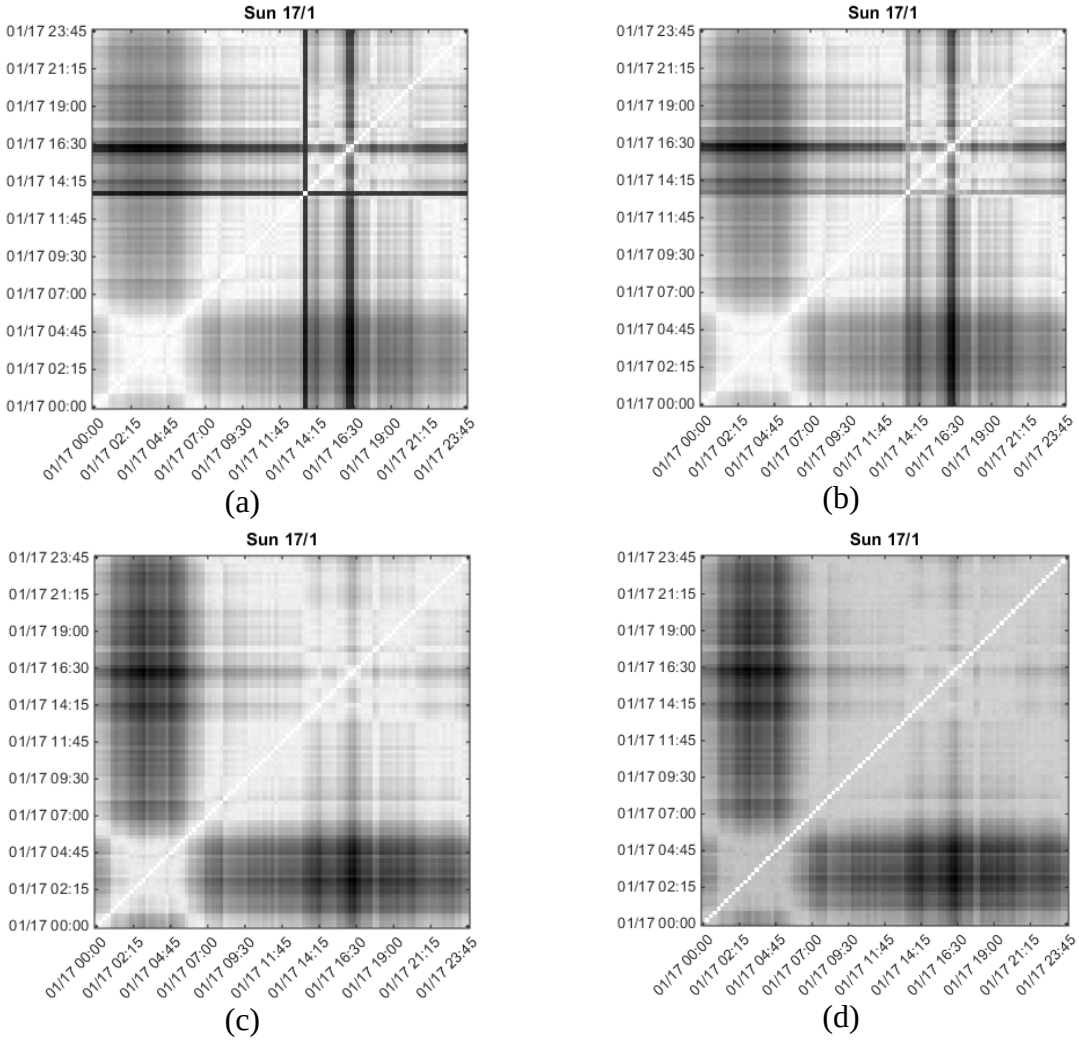


Figura N°2.20. Aplicación de distintos intervalos extremos. (a) $a_l = \pm 0.0005g$, (b) $a_l = \pm 0.00005g$, (c) $a_l = \pm 0.000005g$, (d) $a_l = \pm 0.0000005g$

2.6.3. Agrupamiento.

En este apartado se aplicarán métodos de agrupamiento a distintas matrices de distancias calculadas desde los registros de aceleración. Antes de pasar de lleno a la ejecución de los algoritmos, hay que decidir cuál será el conjunto original de objetos a ser particionado. Considerar la situación de un edificio en constante monitoreo. En el caso de la Torre Central, el sistema genera un registro nuevo cada 15 minutos, lo que en términos de la metodología equivale a un objeto simbólico nuevo cada 15 minutos. Si bien una opción es utilizar todo el universo de objetos como input para los algoritmos de agrupamiento, en el caso de registros de aceleración no es la mejor idea ya que hemos visto como la presencia de distintos regímenes de excitación genera objetos distanciados por la amplitud del movimiento más que por cambios estructurales. Además, la complejidad computacional de utilizar todos los objetos haría que el método sea muy costoso. Otra opción sería comparar los objetos de dos días similares. Un ejemplo de la matriz de distancia de esta alternativa se muestra en la Figura 2.21, en la que se consideraron objetos de dos lunes distintos, uno de Enero y otro de Marzo del 2010. Al observar los resultados de esta figura, se puede asegurar que tampoco es la mejor opción ya que vemos como el efecto de la amplitud es similar en ambos días, y por tanto, es altamente probable que los algoritmos de agrupamiento agrupen en un mismo subconjunto a objetos provenientes de distintos días.

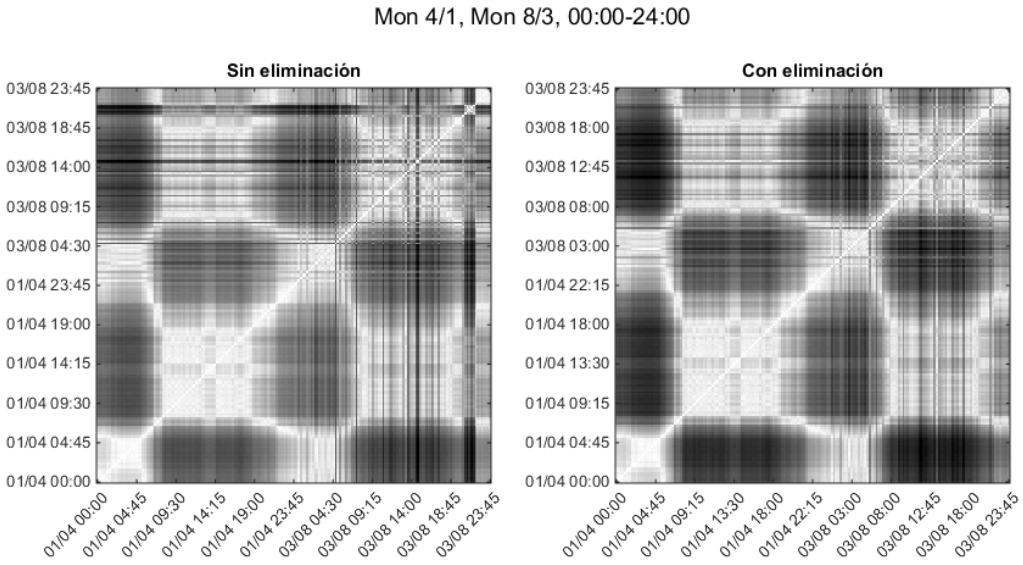


Figura N°2.21. Matriz de distancias utilizando objetos de dos días.

En definitiva, la alternativa que se propone toma en consideración lo anteriormente mencionado, y se basa en la comparación de objetos simbólicos obtenidos a una misma hora y en un mismo día; por ejemplo, los lunes entre las 2:00 y 6:00 AM. Esto permite analizar el comportamiento de la estructura con algún grado de igualdad en el régimen de excitación, aunque obviamente depende de las condiciones ambientales y operacionales que puedan ocurrir durante la noche.

Aplicación del Algoritmo

Se aplican los algoritmos de agrupamiento de Dynamic-Medoids y K-Means a los registros obtenidos los días lunes entre las 2:00 y 6:00 AM, creando pares de días entre el 4 de Enero, versus los lunes restantes de los meses de Enero y Marzo. Se incluyen los lunes de enero para probar si es que hay alguna diferencia en los resultados, buscando probar el algoritmo frente a falsos positivos, y los de Marzo para buscar evidencias de agrupamiento post terremoto del 27F. Además, se incluye la comparación entre el uso o no de la eliminación por distanciamiento y a su vez se aplican dos casos de límites extremos.

Validación de las particiones.

Los algoritmos de agrupamiento se aplican varias veces con un número creciente de grupos, desde 2 hasta 10, y el óptimo se obtiene al maximizar el índice de Calinski-Harabasz. En la Figura 2.22 se muestran los gráficos del índice CH para los registros de Enero vs Marzo, considerando una amplitud extrema de $a_l = \pm 0.0005g$. A partir del número óptimo de grupos se rescata la partición considerada como la que mejor representa el universo de objetos. A esta partición se le calcula el porcentaje de objetos correctamente asignados, ya que se espera que los registros obtenidos en Marzo sean agrupados en un conjunto distinto a los objetos de Enero. Cabe destacar que debido a que el número óptimo de clústers puede ser mayor que dos, los porcentajes de exactitud pueden llegar a ser bastante menores que un 50%.

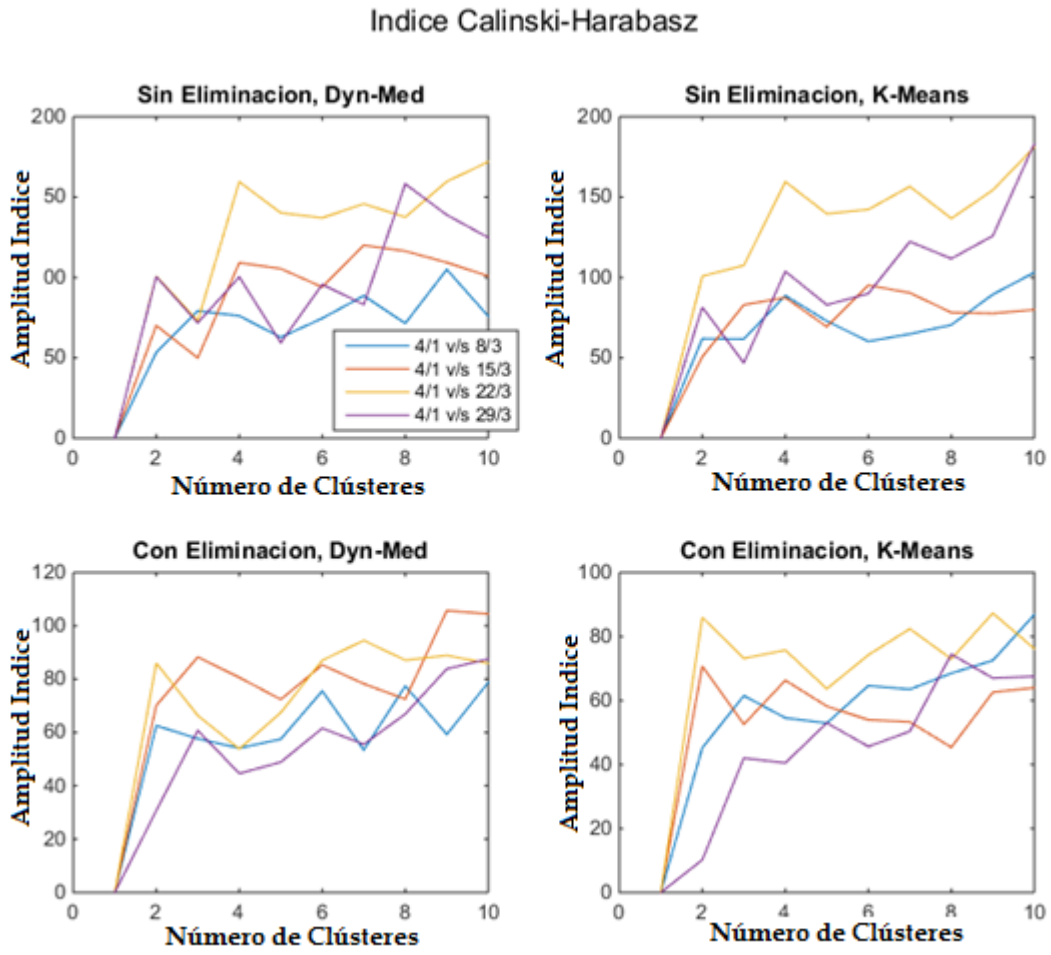


Figura N°2.22. Índice de Calinksi-Harabasz para Enero vs Marzo.

Resultados.

En la tabla 2.11 se muestran los resultados al comparar entre los pares de lunes de Enero del 2010, para una amplitud extrema de $a_l = \pm 0.0005g$. Como es un test de falsos positivos, no se desea encontrar una agrupación correcta sino más bien una mezcla de objetos. La evidencia muestra que este objetivo sí se consigue, ya que a pesar de que se encontró un óptimo de dos grupos en dos ocasiones en el caso del algoritmo K-Means con eliminación de los objetos outliers, el porcentaje total de exactitud no supera el 50%, indicando que estos conjuntos son en verdad una mezcla de objetos de ambos días. Debido a que no hay evidencia de la ocurrencia de un cambio estructural durante Enero del 2010, no se puede hacer un mayor análisis de esta tabla. Sin embargo, la Tabla 2.11 muestra los resultados correspondientes a la misma metodología pero considerando un límite de 0.0000005g, es decir, una amplitud mucho menor. En tal caso, los resultados son análogos, no pudiendo evidenciar una clara diferencia entre los días lunes de Enero. Esto implica que el algoritmo no diferencia entre registros simplemente por el hecho de ser registros distintos, lo cual es un punto a su favor. Aún así, llama la atención de que los algoritmos de agrupamiento tienen un comportamiento contrario en lo que respecta a los límites de los intervalos extremos: Dynamic-Medoids parece tener una mayor sensibilidad para la amplitud pequeña, mientras que K-Means es mejor para amplitudes mayores.

Tabla N°2.11. Resultados Falsos Positivos. Limite 0.0005g

Límite [g]	0.0005	04-ene			
Sin Eliminación	Análisis	11-ene	18-ene	25-ene	Promedio
Dyn-Med	N° de Clúster	7	8	4	6.3
	Porcentaje	25%	22%	41%	29%
K-Means	N° de Clúster	7	3	6	5.3
	Porcentaje	31%	47%	28%	35%
Con Eliminación	Análisis	11-ene	18-ene	25-ene	Promedio
Dyn-Med	N° de Clúster	5	8	5	6.0
	Porcentaje	29%	27%	26%	27%
K-Means	N° de Clúster	2	2	4	2.7
	Porcentaje	54%	58%	33%	48%

Tabla N°2.12. Resultados para lunes de Enero, con un límite de 0.0000005g.

Limite [g]	0.0000005	04-ene			
Sin Eliminación	Análisis	11-ene	18-ene	25-ene	Promedio
Dyn-Med	N° de Clúster	2	6	5	4.3
	Porcentaje	59%	22%	28%	36%
K-Means	N° de Clúster	5	4	6	5.0
	Porcentaje	31%	34%	28%	31%
Con Eliminación	Análisis	11-ene	18-ene	25-ene	Promedio
Dyn-Med	N° de Clúster	2	3	2	2.3
	Porcentaje	56%	42%	54%	51%
K-Means	N° de Clúster	3	4	2	3.0
	Porcentaje	40%	35%	57%	44%

Habiendo pasado el test de falsos positivos con un resultado favorable, ya que no se reporta un cambio de estado entre los objetos pertenecientes a Enero, es tiempo de analizar cómo se comporta la metodología frente a la comparación de objetos provenientes de meses distintos y con la ocurrencia de un evento sísmico mayor entremedio. Es de esperar que el edificio haya

visto modificadas sus propiedades dinámicas producto del terremoto del 27F por lo que se desea encontrar agrupaciones claras entre Enero y Marzo.

Las Tablas 2.13 y 2.14 contienen los resultados para los pares de lunes entre el 4 de Enero con los posibles de encontrar en Marzo. Recordamos que el primero de Marzo del 2010 el sistema de adquisición de datos comenzó a funcionar después de las 11AM por lo que no existen datos entre las 2:00 y 6:00AM. En estas tablas, se confirman algunos resultados anteriores, como que la sensibilidad de los métodos aumenta cuando los límites de los intervalos extremos disminuye. Además, si bien existe un caso en que la eliminación perjudica el resultado, es sólo una excepción, mientras que en el resto de los casos la eliminación por distanciamiento supone una mejora considerable en la exactitud, sobre todo en los casos de límites de intervalos extremos iguales a$a_l = \pm 0.0000005g$. Pero el resultado más importante es que hay indicios de que existe información relacionada al cambio estructural, ya que los porcentajes de exactitud son mucho mayores que los del caso de falsos positivos, llegando a alcanzarse un 91% de asignaciones correctas para la comparación entre el 4 de Enero vs el 22 de Marzo, usando Dynamic-Medoids, eliminación y una amplitud pequeña de los intervalos extremos. Aún así, el porcentaje global no alcanza en ningún caso el 75% , por lo que la metodología dista de ser ideal.

Tabla N°2.13. Resultados para pares de lunes de Enero y Marzo.$a_l = \pm 0.0005g$

Limite [g]	**0.0005**	**04-ene**				
Sin Eliminación	Análisis	08-mar	15-mar	22-mar	29-mar	Promedio
Dyn-Med	N° de Clúster	10	10	10	10	10
	Porcentaje	34%	31%	47%	34%	37%
K-Means	N° de Clúster	10	6	10	10	9
	Porcentaje	38%	63%	50%	38%	47%
Con Eliminación	Análisis	08-mar	15-mar	22-mar	29-mar	Promedio
Dyn-Med	N° de Clúster	7	9	9	10	8.75
	Porcentaje	38%	39%	25%	28%	32%
K-Means	N° de Clúster	10	2	8	8	7
	Porcentaje	34%	86%	38%	41%	50%

Tabla N°2.14. Resultados para pares de lunes de Enero y Marzo. $a_l = \pm 0.0000005g$

Limite [g]	0.0000005	04-ene				
Sin Eliminación	Análisis	08-mar	15-mar	22-mar	29-mar	Promedio
Dyn-Med	N° de Clúster	3	2	7	7	4.75
	Porcentaje	53%	88%	34%	31%	52%
K-Means	N° de Clúster	7	8	5	8	7
	Porcentaje	31%	44%	69%	34%	45%
Con Eliminación	Análisis	08-mar	15-mar	22-mar	29-mar	Promedio
Dyn-Med	N° de Clúster	4	2	2	2	2.5
	Porcentaje	41%	86%	91%	71%	72%
K-Means	N° de Clúster	3	2	5	2	3
	Porcentaje	48%	86%	52%	71%	64%

Para una mayor comparación se elabora la Tabla 2.15, en la que se resumen los resultados más importantes para el estudio de los lunes entre 2:00 y 6:00 AM. En esta tabla, es notoria la mejora introducida tanto por la eliminación por distanciamiento como también por la utilización de límites más pequeños para la creación de los histogramas. Además, es importante notar que Dynamic-Medoids al utilizar la distancia Gowda-Diday que está diseñada como multi-categoría, obtiene beneficios frente a la distancia básica Euclideana utilizada por el algoritmo K-Means.

Tabla N°2.15. Resumen de resultados.

	Limite [g]	0.0005		
Eliminacion	NO		SI	
Enero v/s	Enero	Marzo	Enero	Marzo
Dyn-Med	29%	37%	27%	32%
K-Means	35%	47%	48%	50%
	Limite [g]	0.0000005		
Eliminacion	NO		SI	
Enero v/s	Enero	Marzo	Enero	Marzo
Dyn-Med	36%	52%	27%	**72%**
K-Means	31%	45%	48%	64%

Por último, en la Tabla 2.16 se muestran los resultados al haber aplicado la metodología considerando límites de 0.0000005g y con limpieza de objetos outliers, a todos los días de semana laboral. Si bien la tendencia se mantiene, la diferencia entre Dynamic-Medoids y K-Means se hace mucho menor. En definitiva, con estos resultados se puede afirmar que el análisis usando objetos simbólicos derivados de histogramas de series de aceleración sí arroja

información respecto al estado estructural, pero es necesario mantener prudencia ya que los resultados distan de ser los ideales.

Tabla N°2.16. Resultados considerando cada uno de los días hábiles.

Lunes	Enero	Marzo	Jueves	Enero	Marzo
D-M	45%	72%	D-M	56%	91%
K-M	44%	64%	K-M	61%	84%
Martes	Enero	Marzo	Viernes	Enero	Marzo
D-M	48%	68%	D-M	44%	55%
K-M	41%	85%	K-M	46%	61%
Miércoles	Enero	Marzo	**Promedio**	**Enero**	**Marzo**
D-M	42%	77%	**D-M**	**47%**	**72%**
K-M	43%	61%	**K-M**	**47%**	**71%**

2.7 Conclusiones.

Las conclusión más importante del presente capítulo tiene que ver con el cumplimiento del objetivo planteado al momento de iniciar el estudio. Con todos los resultados establecidos, se asevera que bajo ciertas condiciones y utilizando una combinación adecuada de algoritmos, sí se puede extraer información relacionada con el estado estructural a partir de objetos simbólicos derivados de registros de aceleración.

Dentro de las conclusiones más específicas, se encuentra que cuando la variable medida tiene media 0, los objetos derivados de intervalos interquartil son una representación de la amplitud de la señal, mientras que los objetos derivados de histograma contienen información tanto de la amplitud como también de la distribución de la variable, aportando con mayor sensibilidad pero a expensas de una dimensionalidad mayor.

En lo que respecta a la amplitud de la señal, se concluye que la energía de entrada al sistema es determinante en el algoritmo. Durante los ensayos de laboratorio hubo casos de agrupamiento perfecto pero más relacionado a la diferencia de energía entre los ensayos que a una diferencia en la condición estructural. Cuando se aplicó un factor de escala para normalizar la energía de entrada los resultados fueron mucho más aleatorios. En el caso de series de aceleración obtenidas de la Torre Central, las amplitudes distintas durante el transcurso del día hacen que no sea una opción viable utilizar el universo completo de objetos simbólicos. Los algoritmos de agrupamiento en dicho caso arrojarían resultados de clasificación por la condición de excitación de la estructura, entregando, por ejemplo: "estructura quieta", y "estructura en movimiento". Aún así, comparar estos objetos abre la posibilidad de utilizar los registros de aceleración para realizar otros tipos de estudios. Una posibilidad es usarlos para determinar la sanidad operacional de la estructura: ¿Funcionan todos los ascensores? ¿El aire acondicionado se prende en horarios que no corresponden? Esta forma no convencional de usar los registros de aceleración hace que sea más atractiva la instalación de un sistema de monitoreo, ya que vemos que aún no se descubre toda la gama de información posible de extraer a partir de las señales.

Sobre la ocurrencia de eventos sísmicos durante la adquisición de los registros, se concluye de que éstos se descubren en las matrices de distancias ya que aparecen como cruces negras. Si bien se consideró realizar un pre-procesamiento de los registros para eliminar los eventos sísmicos antes de la aplicación real del algoritmo, los resultados mostraron que los

objetos simbólicos que incluyen sismos pueden ser identificados y eliminados posteriormente. Además, el efecto del aumento de la amplitud por el sismo se puede suavizar considerando límites pequeños para la creación de los histogramas. La combinación de estas dos estrategias resulta en una clasificación más exacta.

En relación a los índices de validación y algoritmos de agrupamiento, se encontró que el índice Calinski-Harabasz es el más adecuado para este análisis y se destaca que es muy simple en su formulación matemática, más intuitivo y de menor complejidad computacional. Por último, se concluye que tanto K-Means como Dynamic-Medoids son los algoritmos de agrupamiento que entregan los mejores resultados y además con salidas muy parecidas, pese a utilizar distancias distintas. La aparición de un cambio estructural se revela como un aumento en el número óptimo de grupos, pero es casi imperceptible. Un mejor indicador sería un índice de secuencialidad de las particiones, ya que los cambios estructurales siempre marcan un antes y un después.

3. Autoregresión y Reconocimiento de Patrones Estadístico.

3.1. Introducción.

En el capítulo anterior se explicó cómo se utilizan algoritmos de análisis de objetos simbólicos para el monitoreo de salud estructural, siguiendo el paradigma de (Hoon Sohn et al. 2000) mediante la extracción de características sensibles y su clasificación en diferentes estados usando técnicas de reconocimiento de patrones. Empezando con series de aceleración limpias, se extrajo parámetros sensitivos a cambios estructurales cumpliendo dos objetivos: servir como una condensación de datos consiguiendo reducir la complejidad numérica asociada y transformar la información original en una que en teoría contiene la mayor información posible acerca del estado estructura. Luego, se utilizaron algoritmos de clasificación por agrupamiento para lograr separar las entradas de aceleración en distintos subconjuntos interpretados como estados estructurales diferentes. En el presente capítulo, se muestra otro procedimiento para la extracción de 'características' y su posterior clasificación usando algoritmos de reconocimiento de patrones.

La base teórica en la que se fundamenta el procedimiento mencionado en el párrafo anterior tiene su origen en los estudios de las series de tiempo y las técnicas disponibles para el modelamiento de una variable en el espacio del tiempo y la predicción futura de ésta a partir de un modelo. Este procedimiento resulta muy intuitivo: por ejemplo, utilizando los datos de habitantes en el mundo podemos ajustar un modelo que prediga la población mundial en los años venideros. Por supuesto, el modelo no es perfecto y año a año veremos que existe una diferencia entre lo predicho y el valor real medido por los censos de las naciones. Sin embargo, mientras el modelo se encuentre relativamente bien ajustado, veremos que el error muestra propiedades estadísticas estables, pero si en algún determinado momento el mundo sufre algún cambio que altere las condiciones globales (por ejemplo, una pandemia), entonces el modelo dejará de predecir correctamente el número de habitantes, lo que se podrá apreciar en un aumento en la desviación estándar del error. Más aún, si se ajusta un nuevo modelo utilizando las nuevas mediciones que consideran la pandemia, las diferencias entre los parámetros que definen éste nuevo modelo y el anterior serán claras. Por lo tanto, utilizando modelos de predicción podemos identificar cambios globales en las condiciones que gobiernan la variable predicha, ya sea mirando las propiedades estadísticas del error residual o actualizando constantemente el modelo hasta la ocurrencia de una diferencia notoria en los parámetros de éste.

Como las series de aceleración de una estructura civil son un caso particular de las series en el espacio del tiempo, se puede aplicar la metodología de modelamiento y predicción mencionada anteriormente con la finalidad de realizar un monitoreo de salud estructural. Además, los modelos son un nivel de abstracción mayor, por lo que es de esperar que tengan algo de información acerca del contenido de frecuencias y que junto con esto disminuyan la influencia del contenido en el tiempo. La técnica de modelamiento de las series de aceleración utilizada en este trabajo de tesis se llama Autoregresión y consiste en la predicción de la aceleración para un instante de tiempo futuro, calculada como una combinación lineal de estados de aceleraciones presentes y pasados. Los coeficientes utilizados en la combinación lineal son calculados a partir de un ajuste del modelo a una serie de aceleración. La diferencia entre la aceleración predicha y una nueva medición es considerada como el parámetro sensible a los cambios estructurales pues, como se dijo anteriormente, este error es considerado estable mientras el sistema no haya cambiado sus propiedades. A su vez, a cada nueva medición de aceleraciones adquirida se le puede ajustar un modelo de autoregresión y realizar un estudio estadístico de los parámetros de los modelos obtenidos. Se trabaja por tanto, con dos 'características' o características sensibles:

los errores residuales de los modelos, y los parámetros o coeficientes de autoregresión. De esta forma, se generan nuevas series en el tiempo con las cuales se realizan los procesos estadísticos de reconocimiento de patrones para determinar la existencia de un cambio en las series originales. Estas técnicas o métodos de reconocimiento de patrones son variadas y se detallan durante el desarrollo del capítulo.

Los modelos autoregresivos se comenzaron a utilizar para el análisis de estructuras civiles a comienzos del 2000 con el trabajo de los investigadores Sohn y Farrar. En (Hoon Sohn et al. 2000) los autores utilizan varios métodos de control estadístico de procesos para realizar un monitoreo de salud de un ejemplo numérico, a partir de los errores residuales obtenidos de un modelo autoregresivo. La idea detrás de dicho estudio, era generar una regla de decisión de 3 output (condición normal, daño leve y daño severo), dependiendo de si la variable sensible, en este caso el error, se salía del control estadístico. En (Hoon Sohn and Farrar 2001) se propone una extensión a los modelos, mediante la incorporación de un input externo (ARX o AutoRegresiveeXogenous), que busca mejorar la sensibilidad de los métodos al intentar modelar la relación con las excitaciones ambientales y además se presenta una idea de localización de daño y toda el procedimiento se aplica a un experimento de laboratorio de 8 GDL. Más adelante y siguiendo la misma metodología, en (H. Sohn, Farrar, and Hunter 2001), los autores tratan de generar un procedimiento de normalización de los inputs mediante el uso de una base de datos de señales de referencia. En (H. Sohn, Czarnecki, and Farrar 2000), se proponen métodos complementarios para la combinación de los datos de varios sensores en una sola señal, y de esta forma obtener la mayor información posible. Ya en (H. Sohn, Farrar, and Hunter 2001) los autores proponen una metodología más desarrollada para la identificación del daño en un bote de patrullaje pudiendo separar correctamente dos estados estructurales. Durante todos los artículos mencionados anteriormente, el enfoque estaba basado principalmente en los errores residuales. Fue en (Hoon Sohn et al. 2001) que se propuso una medida estadística para comparar los parámetros del modelo en vez de los errores, usando la distancia de Mahalanobis y logrando una separación clara de las señales. Así se ha ido desarrollando la teoría, buscando siempre nuevas características sensibles y nuevos métodos estadísticos para lograr su clasificación.

El presente capitulo esta ordenado por secciones. En la 3.2. se presentan los modelos autoregresivos, desde su formulación matemática hasta el cómo son utilizados para la identificación de cambios estructurales. Se detalla también la forma de pre-procesar las señales para poder ser utilizadas en las metodologías, lo que se demuestra usando un ejemplo. En 3.3. se aplica la metodología a unos ensayos de laboratorio; en 3.4 al edificio de la Torre Central del Campus Beauchef y finalmente en 3.5. se destacan las principales conclusiones y resultados.

3.2. Modelos Autoregresivos

3.2.1. Formulación matemática

Consideremos una serie en el espacio del tiempo $\{a_t\} = \{a_1, a_2, a_2, \dots\}$. Un modelo autoregresivo de orden p, $AR(p)$, es aquél que calcula la respuesta presente como una combinación lineal de p estados pasados, según la Ecuación 4.1

$$a_t = \phi_1 a_{t-1} + \phi_2 a_{t-2} + \cdots + \phi_p a_{t-p} + e_t \tag{3.1}$$

los valores $\{\phi_i\}$se llaman coeficientes autoregresivos y son los parámetros que definen el modelo ajustado. El término $e_t$representa el error de no poder ajustar perfectamente la información de la serie, pero en caso de tener un buen modelo se supone que $e_t$posee una distribución normal.

La fórmula 3.1 es la forma intuitiva y matemáticamente sencilla de describir lo que son los modelos autoregresivos, pero para poder trabajar y entender algunos paquetes de software es necesario introducir un nuevo operador, llamado 'Backshift Operator' - B. Este operador tiene la propiedad de devolver un estado pasado al ser aplicado sobre un valor particular de alguna serie, lo que se estipula matemáticamente en la Ecuación 4.2

$$By(t) = y(t-1) \qquad (3.2)$$
$$B^n y(t) = y(t-n)$$

Trabajando algebraicamente la Ecuación 3.1 y aplicando el operador Backshift conseguimos:

$$a_t - \phi_1 a_{t-1} - \phi_2 a_{t-2} - \cdots - \phi_p a_{t-p} = e_t$$
$$a_t - \phi_1 B a_t - \phi_2 B^2 a_t - \cdots - \phi_p B^p a_t = e_t$$

$$\left(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^n\right) a_t = e_t \qquad (3.3)$$

A partir de ésta última ecuación se observa que el modelo autoregresivo $AR(p)$ queda completamente definido por el polinomio característico:

$$\Phi(p) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p \qquad (3.4)$$

Y de esta forma se logra la formulación compacta de los modelos autoregresivos en la Ecuación 3.5. Como se mencionó anteriormente, esta notación adquiere relevancia al trabajar con algunos paquetes de software, lo que será explicado más adelante.

$$\Phi(p) a_t = e_t \qquad (3.5)$$

El ajuste de un modelo de autoregresión a una señal tiene que pasar por una etapa de prediseño en lo que respecta al largo de las señales utilizadas y al largo del ajuste. Por ejemplo, en un sistema de medición de aceleraciones podemos encontrar señales de 15 minutos de duración, por lo que hay que decidir si para ajustar un modelo se utiliza la señal completa o solo una sección de ésta. Esto genera una distinción en los errores que se encuentran después de ajustado el modelo, pudiéndose encontrar errores de ajuste y errores residuales. Los errores de ajuste son aquellos que resultan del proceso mismo de entrenamiento, mientras que los errores residuales son los errores que se van encontrando a medida que se utiliza el modelo para predecir estados futuros y comparándolos con las mediciones. En general, los errores residuales son mayores que los errores de ajuste.

3.2.2. Identificación del modelo

Fundamentos del procedimiento.

La identificación del modelo de autoregresión asociado a una señal de aceleración requiere un procesamiento previo de los datos para poder mantener una complejidad numérica reducida. Como en general la adquisición de datos se hace a una alta frecuencia de muestreo, el orden de los modelos tendría que ser muy alto para poder considerar un intervalo de tiempo de entrenamiento adecuado; en consecuencia, se debe realizar un proceso de decimado en el que se

reduce la tasa de muestreo consiguiendo una reducción en el número de datos de las señales. Sin embargo, para mantener la máxima información posible, primero hay que estudiar el contenido de frecuencias de las señales usando series de Fourier y PSD, y usar esa información para justificar el factor de decimado que permita observar las frecuencias presentes. Por otra parte, debido a que las señales de aceleración son fuertemente influenciadas por el input de energía sobre la estructura, éstas son normalizadas para reducir dicho efecto.

Una vez que las señales han sido pre-procesadas, se continúa con la identificación del modelo autoregresivo. Como se ha mencionado, antes de ajustar el modelo es necesario saber el orden que se utilizará. Contar con dicha información no es trivial pero por el momento asumiremos este dato como conocido y en la próxima sección se comentará cómo obtener con argumentos matemáticos el orden aproximado del modelo.

Supongamos que contamos con una serie $\{a_t\}$ de N datos obtenida a partir de la adquisición de un sensor de aceleración. Se desea ajustar un modelo autoregresivo $AR(p)$ (nótese orden p) , utilizando los N datos como entrenamiento. Los parámetros del modelo de autoregresión son obtenidos mediante las ecuaciones de Yule-Walker. Estas relacionan los coeficientes con la correlación de la señal, obteniendo un sistema de ecuaciones con el que se logra obtener un modelo óptimo para la señal estudiada. En este trabajo de tesis, el método empleado está implementado en Matlab y corresponde a la rutina 'aryule'. Ahora es cuando cobra importancia la Ecuación 3.4 ya que ésta rutina recibe como entrada una señal cualquiera en el espacio del tiempo y el orden del modelo que se está identificando, y devuelve los parámetros ordenados tal como en 3.4. Como el polinomio definido en dicha ecuación considera al estado actual, el primer elemento devuelto por la rutina no debe ser considerado y los valores reales de los coeficientes son los inversos aditivos de los valores retornados. Esta metodología genera $(N-p)$ errores de ajuste y ningún error residual, pues estos últimos se empiezan a generar tan pronto como se comience a utilizar el modelo para predecir estados de aceleración futuros.

Obtención del orden óptimo del modelo.

Saber a ciencia cierta el orden del modelo no es algo trivial. De forma análoga a los sistemas de identificación de parámetros modales, se debiese evaluar de forma incremental el orden y realizar un estudio de convergencia de la solución para encontrar el orden óptimo. Si bien con los algoritmos modales se trabaja con diagramas de estabilización de los parámetros (periodos y amortiguamiento), en el caso de modelos autoregresivos la convergencia se analiza con criterios secundarios que tienen que ver con los errores y con la función de autocorrelación parcial.

Figueiredo et al. (2011) estudiaron cuatro índices de validación de orden de modelos autoregresivos. De ellos, en la presente tesis se estudian solo dos, de los que se dará una breve explicación para dar completitud al presente estudio.

- AIC (Akaike Information Criterion, Akaike 1974)
  Es un valor estadistico que se ha utilizado para calcular la capacidad de generalización de modelos lineales. El valor se calcula según la ecuación a continuación.

$$AIC(p) = Nln(e) + 2p \tag{3.6}$$

  donde $e = SSE/N$ es el promedio de los errores cuadráticos y $N = n - p$ es el número de observaciones usadas en ajustar el modelo. Como es de esperar que los errores

disminuyan al aumentar el orden del modelo, el índice AIC representa un balance entre la capacidad de ajuste y la complejidad del modelo.

- RMS (Root mean squared errors).
  Uno de los índices más sencillos consiste en estudiar la convergencia de los errores según la Ecuación (3.7)

$$RMS(p) = \left(\frac{1}{N}\sum_{i=1}^{N} e_i^2\right)^{\frac{1}{2}} \tag{3.7}$$

En el estudio de Figueiredo se determinó que utilizando los índices no es posible llegar a una solución única con respecto al orden del modelo a utilizar, sin embargo los resultados son coherentes y presentan una variabilidad aceptable. Cabe destacar que tanto AIC como RMS tienen como desventaja que es necesario ajustar progresivamente todos los modelos con un orden creciente para poder hacer el cálculo de los índices.

Ejemplo de obtención del orden del modelo.

Para ejemplificar la metodología detallada en la sección anterior, se utilizará una señal de aceleración obtenida durante un experimento en laboratorio. Los detalles del experimento serán comentados en otro apartado de este capítulo, pero por el momento no son necesarios. De hecho, se verá que el procedimiento general no requiere de información extra en lo que respecta a dimensiones de las estructuras o ubicaciones de sensores.

En la Figura 3.1 se muestra la señal de aceleración que se utilizará para el ejemplo. La unidad de la aceleración no se ha especificado puesto que para el ejemplo no es necesaria. La frecuencia de muestreo es de 200 [Hz] y como el largo de la señal es de aproximadamente 300 segundos, en total hay alrededor de 60000 datos. Para pre-procesar la señal hay que aplicar un decimado, posterior a la realización de un estudio del contenido de frecuencia de la señal. Si bien se está analizando los datos obtenidos de un solo sensor, el análisis de frecuencias se realiza para todos los sensores de manera de unificar el decimado. Este estudio de los periodos presentes en la señal se efectúa con PSD y sus resultados se muestran en la Figura 3.2. Es posible apreciar de este último gráfico que la estructura ensayada posee frecuencias de hasta 12 [Hz], lo que implica que para poder observarlas se debe considerar una frecuencia de muestreo post-decimado de al menos 25[Hz]. En consecuencia, para pasar de 200 a 25 [Hz] se requiere de un factor de decimado igual a 8, y una vez aplicado, la señal posee solo alrededor de 7500 datos. Todo este procedimiento es resumido en la Figura N°3.3, en la que se compara la señal original y la señal que resulta luego de aplicar el decimado y la normalización. Cabe destacar que una de las consecuencias de realizar lo anteriormente estipulado, es que la señal con la que se trabajará ya no tiene unidades de aceleración. Además, si la señal original posee poca amplitud, como ocurre en el caso de realizar mediciones de aceleración en un edificio bajo vibraciones ambientales durante las horas de la madrugada, lo que se hace realmente al normalizar la señal es amplificarla, lo que puede aumentar el nivel de ruido relacionado a las mismas mediciones.

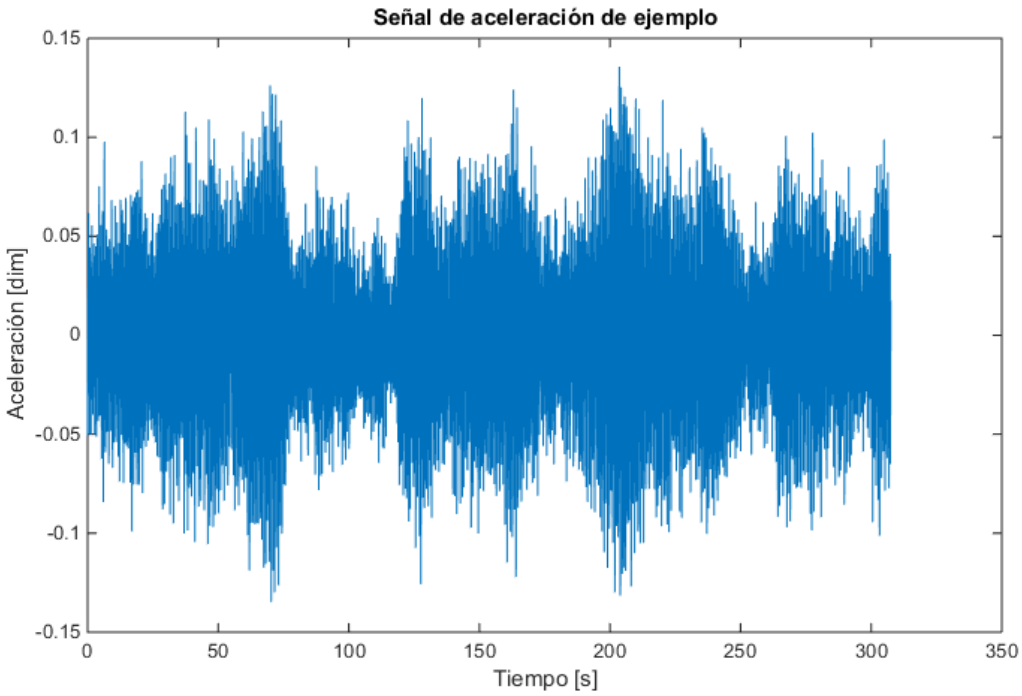


Figura N°3.1. Señal de aceleración de ejemplo.

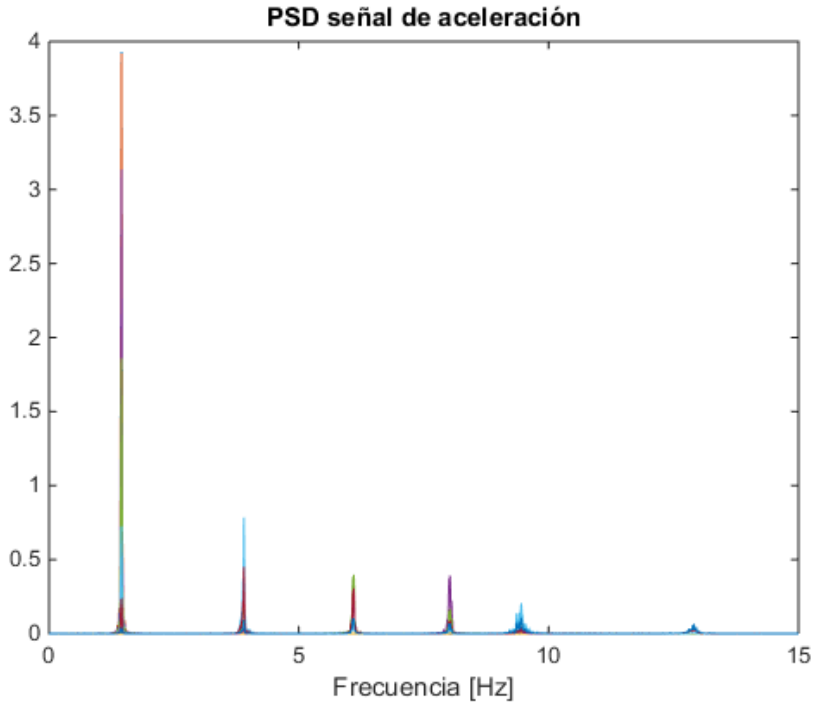


Figura N°3.2. Contenido de frecuencias de la señal de ejemplo.

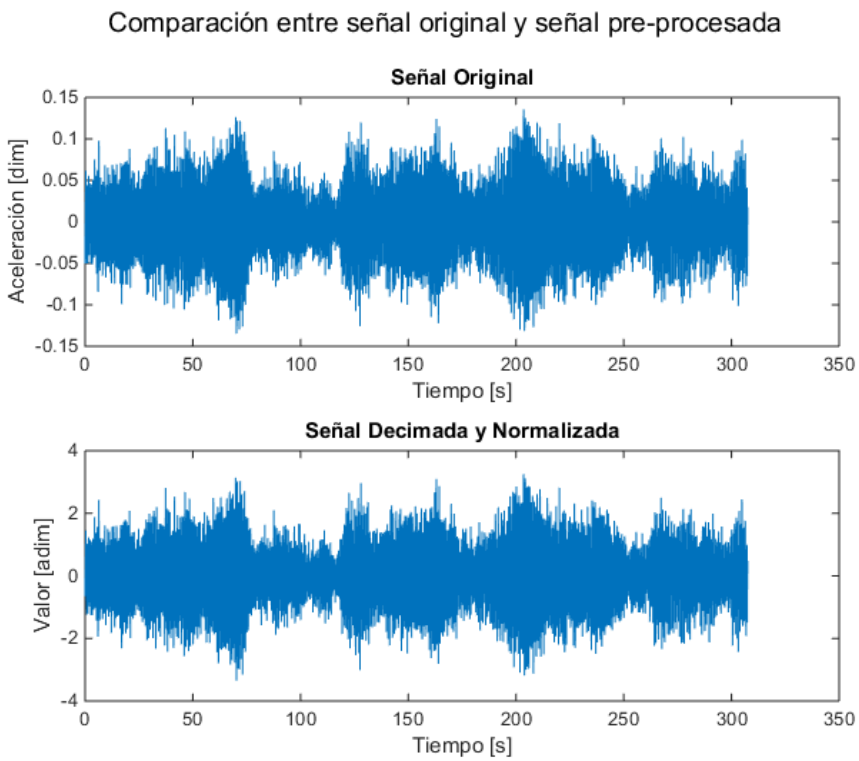


Figura N°3.3. Ejemplo de pre-proceso de señales. Decimado y Normalización.

Ya con las señales pre-procesadas, se puede continuar con el proceso de modelado y predicción. Sin embargo, previamente debe hacerse el estudio del orden óptimo para los modelos autoregresivos. Como se mencionó anteriormente, esto se realiza con los índices AIC y RMS. Para ello, se calculan los modelos incrementando sucesivamente el orden de estos y se almacenan los valores correspondientes a los índices, los que luego se analizan para determinar el orden óptimo. Para ambos índices, el orden optimo está determinado por aquel que minimice las funciones estipuladas en las Ecuaciones 3.6 y 3.7. Es importante notar que RMS puede no tener un mínimo, sin embargo la función AIC tiene al menos uno. Estos resultados se muestran en forma gráfica en la Figura 3.4, donde además de realizar un estudio del orden óptimo de los modelos, también se analiza la relación de estos con el largo de la señal de ajuste. Considerando que la señal cuenta con aproximadamente 7500 datos y una frecuencia de muestreo de 25 [Hz], se analizaron señales de una duración aproximada de 30[s], 60[s], 2[min], 4[min] y 5[min]. Lo obtenido en ésta investigación, es que tanto AIC como RMS muestran una convergencia asíntota, es decir, ningún índice alcanza un mínimo local, y por tanto no se podría asegurar la existencia de un orden de modelo autoregresivo óptimo. Sin embargo, la evidencia muestra que se alcanzaun pseudo-mínimo alrededor de un orden igual a 15, y a partir de ese momento la variación en el valor de los índices es muy pequeña. Es más, si bien el índice AIC muestra valores distintos para los distintos largos de la señal de ajuste, sí se puede apreciar una solución óptima y única entorno al orden 15, en el sentido de que todos los largos de ventana de ajuste muestran la misma tendencia de no tener mucha variabilidad después de un orden igual a 15. Por otra parte, los resultados para RMS son incluso más fuertes, ya que el índice no solo muestra una solución única (el mismo orden cercano a 15, para cada largo de señal de ajuste), sino que además los valores de los errores parecen converger independiente del largo de la señal utilizada. Esto es de suma importancia, ya que el no poder apreciar diferencias notables en los errores de ajuste, indica que

existe un margen bastante amplio al momento de tomar la decisión del largo de señal a utilizar para realizar los ajustes y la creación de los modelos. A modo de resumen, para el ejemplo presente se establece un orden óptimo de modelo igual a 15, y se deja a criterio la elección del largo de señales a utilizar. Esto último dependerá en gran medida del tipo de característica que se extraiga a partir de la utilización de los modelos autoregresivos.

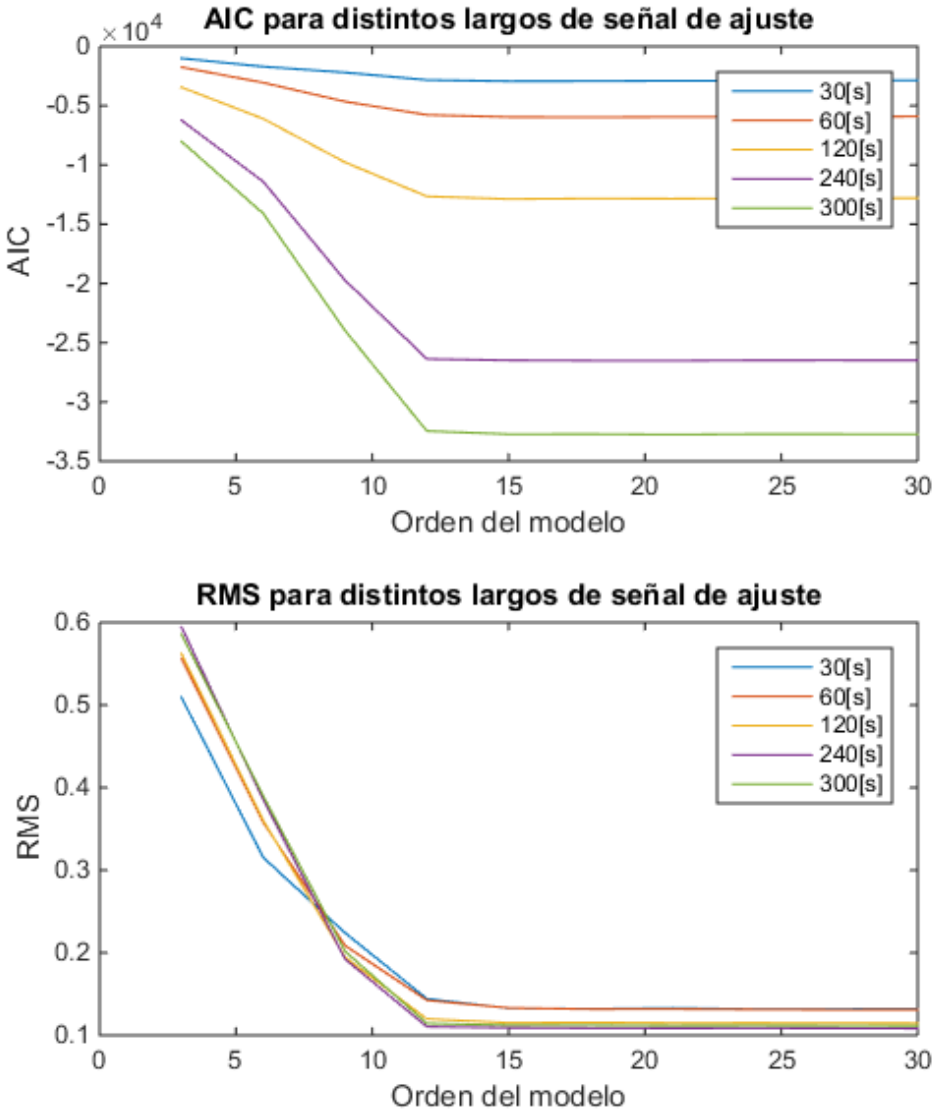


Figura N°3.4. Criterios de elección de orden vs largo de la señal de ajuste.

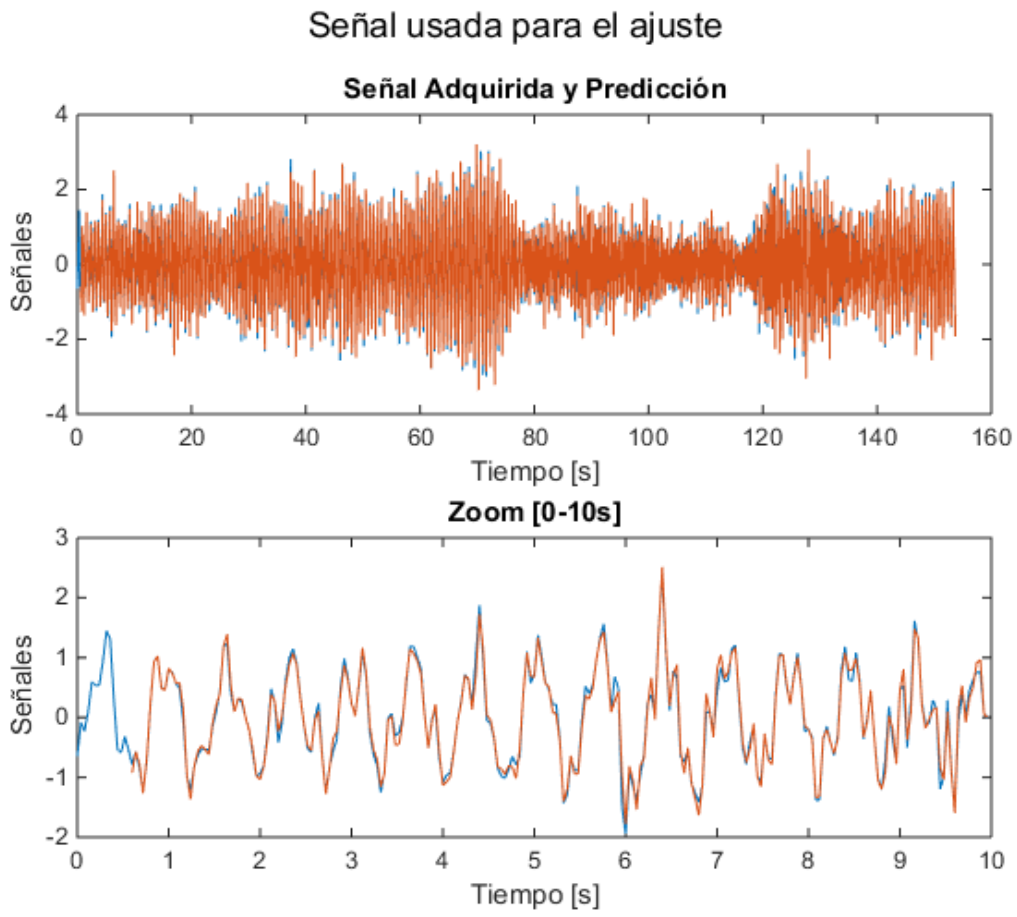


Figura N°3.5. Comparación entre señal adquirida y señal ajustada.

A continuación se ejemplifica las diferencias entre los errores de ajuste y los errores residuales. Recordando que se cuenta con una señal de aproximadamente 5 minutos, ésta se divide en dos mitades de básicamente el mismo largo. La primera mitad se utiliza para entrenar o ajustar el modelo, generando únicamente errores de ajuste, mientras que la segunda mitad se usa para predecir valores futuros de la aceleración y estos valores son comparados con los adquiridos para producir únicamente errores residuales. La Figura 3.5 muestra en un mismo gráfico la señal utilizada para el ajuste y la señal que resulta de dicho ajuste, así como también un zoom a los primeros 10 segundos para poder apreciar de mejor forma los resultados. Si bien es notable la exactitud con que el modelo ajusta los datos, hay que recordar que se está entrenando el modelo con todos los datos de la señal, por lo que no es de extrañar lo certero de éste. No obstante, en la Figura 3.6, los datos utilizados corresponden a la parte de la señal que no fue considerada para el ajuste, por lo tanto se trata de información que cae fuera del set de entrenamiento, y aún así el modelo predice de forma ajustada los datos adquiridos.

El hecho de que los modelos autoregresivos parezcan predecir tan correctamente los valores medidos es una de las piedras angulares para su uso en la identificación de cambios estructurales en la ingeniería civil, ya que es de esperar que un cambio en el comportamiento de las señales no pueda ser representado por un modelo tan bien ajustado a otro estado, generando un error residual mayor.

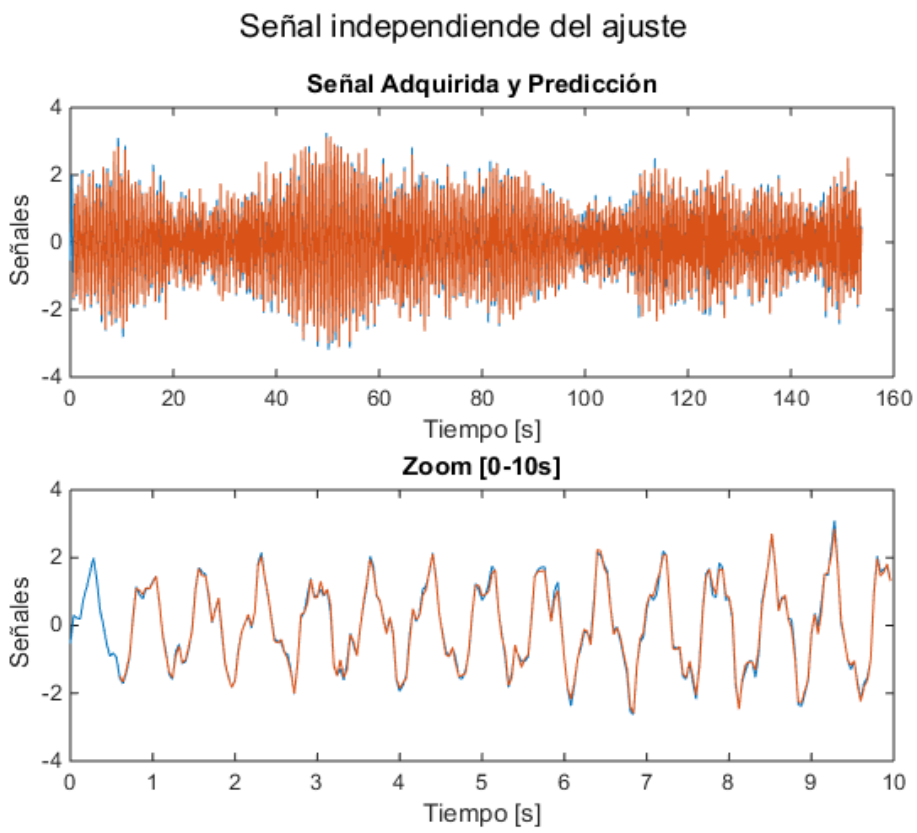


Figura N°3.6. Comparación entre señal adquirida y predicción de una serie nueva.

Para concluir con el ejemplo, en la Figura 3.7 se muestran los gráficos correspondientes a los errores de ajustes, calculados a partir de la diferencia entre el modelo y la señal utilizada para calcularlo, y los errores residuales, calculados como la diferencia entre una señal nueva y la predicción del modelo. Para ambos casos se puede apreciar como el comportamiento de los errores parece seguir una distribución normal, afirmación que a su vez se apoya en los gráficos de histogramas de los errores, las que sin duda dejan ver ésta condición. Cabe destacar que la amplitud de los errores se puede relacionar con la desviación estándar de éstos, y en este caso los errores residuales muestran una mayor desviación que los errores de ajuste, lo que es de esperar en el proceso de modelamiento. Sin embargo la diferencia entre estos valores es tan pequeña que bien puede deberse a una variación estadística. Es importante recordar que la señal no ha cambiado entre la primera y segunda mitad de la Figura 3.7.

En la siguiente sección se detallarán algunas de las metodologías existentes para la detección de cambios en los sistemas estructurales, que utilizando modelos autoregresivos se puede estudiar mediante la incapacidad de un modelo de mantener su exactitud prediciendo el comportamiento de una señal, o mediante el estudio de cambios en el modelo en sí mismo.

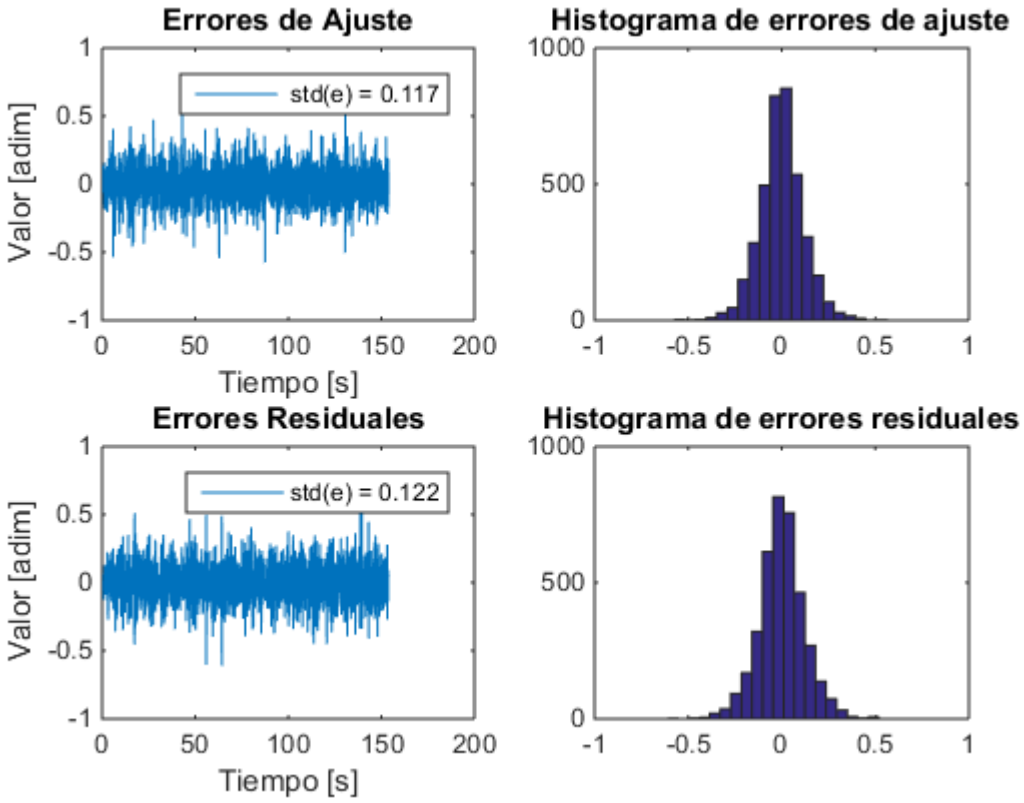


Figura N°3.7. Errores de ajuste y errores residuales y sus histogramas.

3.2.3. Extracción de 'Características' y su clasificación.

Características basados en errores.

Metodología Básica.

Una de las primeras metodologías encontradas en la literatura que utiliza AR(p) para detectar cambios estructurales es la de (Hoon Sohn et al. 2000). En dicho artículo se presentan los fundamentos de AR(p) y se postula una forma de utilización de los errores residuales mediante la detección de outliers. En definitiva, la característica usada es directamente el error residual, y se aplican algoritmos estadísticos para detectar cambios. Formalmente, supongamos que se cuenta con una adquisición de aceleración $\{a_t\}_r$la cual luego del proceso de normalización y decimado tiene un total de N datos. A esta serie se le ajusta un modelo $AR(p)$, generando $(N-p)$ errores de ajuste $\{e_t\}_r$, los que serán utilizados para calcular los límites de control estadístico.

Cuando haya una nueva señal adquirida, se utiliza el modelo calculado previamente para predecir los estados de aceleración, generando una serie de errores residuales $\{e_t\}$. Para monitorear la variación de las características, estas primero son agrupadas en $m$conjuntos no solapados de n elementos, con n usualmente igual a 4 o 5:

$$\begin{matrix} e_{11} & e_{12} & \cdots & e_{1n} \\ e_{21} & e_{22} & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ e_{m1} & e_{m2} & \cdots & e_{mn} \end{matrix} \qquad (3.9)$$

A continuación, se calcula el promedio y la desviación estándar de cada uno de los m subgrupos:

$$\overline{X_i} = mean(\{e_{i1}, \dots, e_{in}\}) \tag{3.10}$$

$$S_i = std(\{e_{i1}, \dots, e_{in}\}) \tag{3.11}$$

Finalmente, se construye un gráfico de control trazando una línea central en el promedio de los $\overline{X_i}$, y dos líneas horizontales adicionales correspondiendo a los límites de control superior e inferior, versus el N° de subgrupo. Las líneas de control se definen como:

$$UCL = CL + Z_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}}, \qquad LCL = \ CL - Z_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}}$$
$$CL = mean(\overline{X_i}) \tag{3.12}$$

El valor $Z_{\frac{\alpha}{2}}$ representa el cuantil $\alpha/2$ de la distribución normal estándar. La varianza S^2es estimada promediando la varianza $S_i$de todos los subgrupos:

$$S^2 = mean(S_i) \tag{3.13}$$

Si el sistema experiencia algún cambio estructural es muy probable que sea evidenciado por un número elevado de observaciones fuera de los límites de control, y cuando esto ocurre al valor se le llama 'outlier'. El monitoreo estructural se realiza graficando nuevos valores $\overline{X_i}$ y S_i junto con los valores de control previamente calculados. En este momento cabe destacar que la metodología presentada, es considerada un tipo de algoritmo supervisado ya que se calculan los límites de control con respecto a una señal de referencia.

Autoregresión con Entrada Externa

En (Sohn et al. 2001), los autores proponen una extensión a la metodología anterior, mediante la inclusión de información extra referente a entrada de excitación externa a los que pueda verse sometida una estructura. Efectivamente, los modelos AR(p) solo modelan el comportamiento de la variable consigo misma, dejando de lado la relación de las vibraciones con las forzantes; es decir, ignoran el efecto causa-efecto. Los modelos autoregresivos que toman en consideración la información de agentes externos se llaman ARX (AutoregresivewitheXogenous input), y realizan un modelado lineal similar a los modelos AR(p) más simples, pero incorporando un vector de valores reconocidos como input.

El procedimiento parte por la normalización y decimado de todas las señales, al igual que la metodología anterior. Luego se construye un modelo AR(p) usando una señal de aceleración de referencia$x(t)$, el cual se especifica a continuación para poder usar la notación más adelante:

$$x(t) = \sum_{j=1}^{p} \phi_{xj}\, x(t-j) + \ e_x(t) \tag{3.14}$$

A partir de esta ecuación se desprenden supuestos muy importantes. Primero, que el modelo AR(p) no es capaz de describir perfectamente los datos; segundo, que ésta incapacidad se produce en parte por la falta de modelación entrada-salida, y por último que los valores asociados a $e_x(t)$ en verdad corresponden a una representación del input mismo. Por lo tanto, al realizar este procedimiento se están identificando dos componentes: el modelo AR(p) que relaciona las variables consigo mismas, y el input del sistema (o al menos una representación de este). Como aún falta poder identificar el comportamiento del sistema en relación a la forzante, se aplica una segunda etapa de identificación, esta vez asumiendo que $e_x(t)$corresponde a la entrada a la que se vio forzada la estructura durante la adquisición de la señal $x(t)$. Esto queda estipulado formalmente en la siguiente ecuación:

$$x(t) = \sum_{i=1}^{a} \alpha_i\,(t-i) + \sum_{j=0}^{b} \beta_j\, e_x(t-j) + \epsilon_x(t) \tag{3.15}$$

En la Ecuación 3.15 los parámetros α_i expresan la relación de la variable x consigo misma, y los coeficientes β_i guardan la relación entrada/salida de la estructura. En este caso, el modelo ARX(a,b) tampoco es capaz de representar perfectamente los datos, generando un error residual $\epsilon_x(t)$. En principio, se esperaría que $e_x(t) > \epsilon_x(t)$ debido a un aumento en la complejidad del modelo, pero esto no necesariamente es así ya que el cálculo del modelo AR(p) pudo arrojar un modelo sobreajustado. No obstante, en teoría el uso de un modelo más complejo arrojaría resultados más sensibles a los cambios en la estructura. De esta forma, al comparar los modelos AR(p) y ARX(a,b), lo único que se puede aseverar es que el último corresponde a uno más completo, y que si se tiene información del input (obtenido de adquisición o identificado mediante modelo AR), es deseable utilizar el modelo ARX para predecir y calcular una característica con un nivel de sensibilidad superior.

Utilizando una nueva señal $y(t)$ obtenida de una condición estructural desconocida, se repite la Ecuación (3.14), identificando el input presente durante la adquisición de esta nueva señal:

$$y(t) = \sum_{j=1}^{p} \phi_{yj}\, y(t-j) + \; e_y(t) \tag{3.16}$$

Luego, se estudia la capacidaddel modelo ARX(a,b) calculado en (3.15) para reproducir la relación entrada/salida entre $y(t)$ y $e_y(t)$:

$$\epsilon_y(t) = y(t) - \sum_{i=1}^{a} \alpha_i y(t-i) - \sum_{j=0}^{b} \beta_j\, e_y(t-j) \tag{3.17}$$

Notar que en esta parte, el modelo ARX(a,b) que representa el comportamiento estructural frente a forzantes externas ya fue calculado e identificado usando una señal de referencia. Si este modelo no fuera un buen candidato para representar el comportamiento de la estructura durante la adquisición de nuevas señales, se espera que exista un cambio significativo en la distribución de

probabilidad de $\epsilon_y(t)$, lo que se podría apreciar al comparar las desviaciones estándar de $\epsilon_x(t)$ y $\epsilon_y(t)$.

Finalmente, en (Sohn and Farrar 2001) se define la razón $\frac{\sigma(\epsilon_y)}{\sigma(\epsilon_x)}$como la característica sensible al cambio estructural. Si esta razón toma valores más grandes que algún umbral de control se asume que el sistema ha sufrido algún tipo de cambio estructural. Sin embargo, para establecer el valor del umbral, se deben adquirir datos desde distintas condiciones operacionales y estimar la distribución de probabilidad de la función $\frac{\sigma(\epsilon_y)}{\sigma(\epsilon_x)}$. En (Sohn and Farrar 2001) se menciona que una simulación Monte Carlo puede servir para estos propósitos.

Características' Basados en coeficientes.

Estudio del comportamiento de los coeficientes de modelos autoregresivos.

En (Figueiredo et al. 2011; Omenzetter and Brownjohn 2006) se estudió la capacidad de los coeficientes como características sensibles a los cambios estructurales, obteniendo que se generan cambios en las amplitudes de los parámetros ante la presencia de daño o algún otro tipo de modificación a las propiedades del sistema. En el artículo mencionado, se pone el énfasis en la comparación del orden de los modelos, por lo que se hace poca mención a la forma de calcular los parámetros para un set de datos en particular. Por ejemplo, dada una señal de aceleración de 7000 datos, ¿Con cuántos puntos se ajusta el modelo? Una opción sería utilizar toda la señal y obtener solo un set de coeficientes AR; otra, considerar ventanas móviles y generar una serie de parámetros en el tiempo. Esta última forma hace que para cada señal exista un conjunto de coeficientes AR que describen su comportamiento, y por tanto lo intuitivo es estudiar la distribución de estas nuevas series y ver de qué forma pueden expresar los cambios estructurales.

Para el estudio de las series de parámetros AR en el tiempo, se utilizarán los experimentos de laboratorio ya mencionados anteriormente. La idea es poder determinar si los parámetros AR sirven como característica sensible a los cambios estructurales, y para ello se analizarán tres condiciones estructurales distintas, ensayadas con dos excitaciones de ruido diferentes. El orden del modelo AR utilizado es de 15 tal como fue mostrado en secciones anteriores, y la serie de aceleración considerada corresponde a la del sensor N°1 (ubicado en el piso superior). La forma de crear las series de parámetros AR es mediante la utilización de ventanas móviles de 60 segundos de duración (1500 datos) que son desplazadas cada una muestra. Dependiendo del largo de la señal original (de aproximadamente 7500 puntos), el uso de ventanas generará aproximadamente 6000 conjuntos de parámetros, es decir, para cada uno de los 15 coeficientes AR, hay una serie de aproximadamente 6000 puntos. En la Figura 3.8 se muestran series de tiempo de los coeficientes N°1, 3, 6 y 10, para tres estados estructurales (Condición Normal, reducción 20% en una columna, y reducción del 50% en una columna). Cada una de las señales mencionadas contiene la condición en su totalidad (no son mitad sana, mitad dañada, sino que o bien toda la señal es de la condición sana, o toda es de la condición dañada). Además, los gráficos se agrupan según la excitación impuesta, siendo el grupo de la izquierda el correspondiente al Ruido0 mientras que el de la derecha corresponde al Ruido4. De esta figura se puede observar que la variabilidad de todos los parámetros en términos de amplitud es considerable (Justificable: σ^2), indicando que los éstos no están fuertemente ligados con las propiedades dinámicas del sistema, al menos al ser analizados sin el uso de Vector-AR. Además, esta variabilidad aparentemente es la misma independiente de la condición estructural. Por otra parte sí es posible apreciar un comportamiento distinto en las series de los parámetros para el caso RE50 durante el

Ruido4. En definitiva este gráfico no aporta mucha información encuanto a sensiblidad a los cambios estructurales. Siguiendo con el análisis, la f3.9 muestra los histogramas de las series de parámetros AR. En ella, se logran ver diferencias más claras en la distribución de ocurrencias de los parámetros, para una misma excitación. Si se observa la primera fila, se ve que la distribución para el caso RE50 es bastante distinto que para los otros dos casos, mientras que mantiene una estructura similar al comparar la misma condición (RE50), pero con distinta excitación. Esto último podría indicar capacidad del algoritmo para detectar cambios estructurales con robustez suficiente para diferenciarlos de distintas vibraciones ambientales. Por último, para concluir el estudio del comportamiento de los parámetros AR, en la Figura 3.10 se muestra las funciones de autocorrelación de las series, usando el procedimiento para su cálculo detallado en (J.P. Santos 2014). Ya en esta figura se puede apreciar claramente un comportamiento distinto para el caso de daño más severo, RE50, mientras que para las condiciones normal y RE20 las diferencias no son notorias. La diferencia se aprecia en que los valores de ACF para el caso de daño severo son todos positivos y de una amplitud mayor que la de los casos sin daño y de daño leve, en los que los valores son más bien oscilantes en torno al cero.

De éste estudio se puede concluir que los parámetros AR efectivamente expresan cierta sensibilidad a los cambios estructurales. Lo que falta es encontrar la forma más eficiente de extraer ésta información sensible y poder clasificarla en distintos estados estructurales. Una metodología posible es mediante la distancia de Mahalanobis, que es una métrica multivariable que considera la distribución de las variables al momento de hacer el cálculo. Otra forma es calcular el espectro de frecuencias de los parámetros y utilizar la distancia Cosh, diseñada específicamente para el espacio de las frecuencias. Un procedimiento novedoso sería usar la teoría de objetos simbólicos explicada en el capítulo 3, y clasificar los estados estructurales utilizando la métrica relacionada con el cálculo de distancias entre histogramas (categorical distance), aplicado a la series de coeficientes AR. Las primeras dos metodologías mencionadas, encontradas en la literatura, son explicadas a continuación, mientras que la tercera, propuesta en la presente Tesis, se explica y aplica en el próximo capítulo.

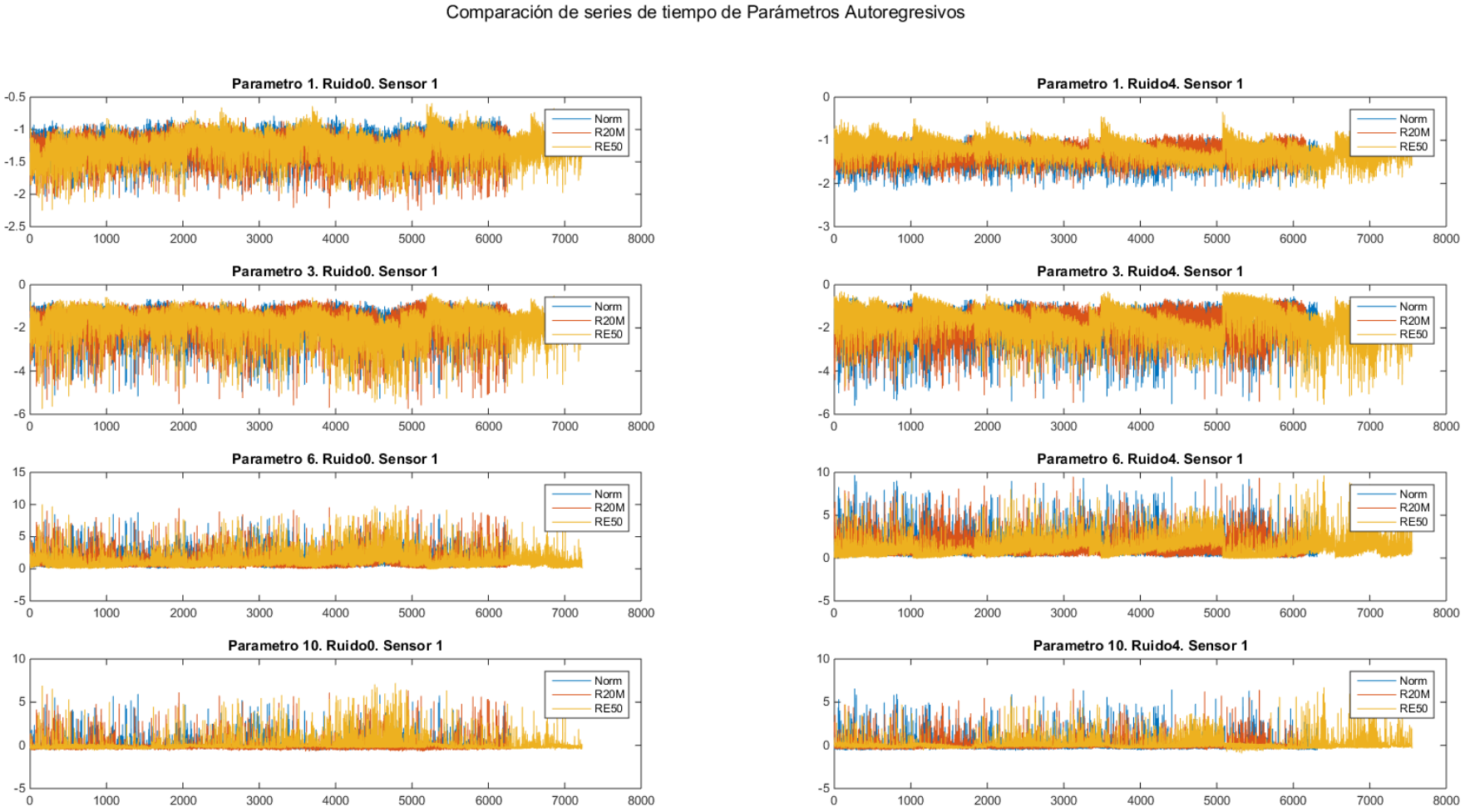


Figura N°3.8. Series de parámetros AR en el tiempo.

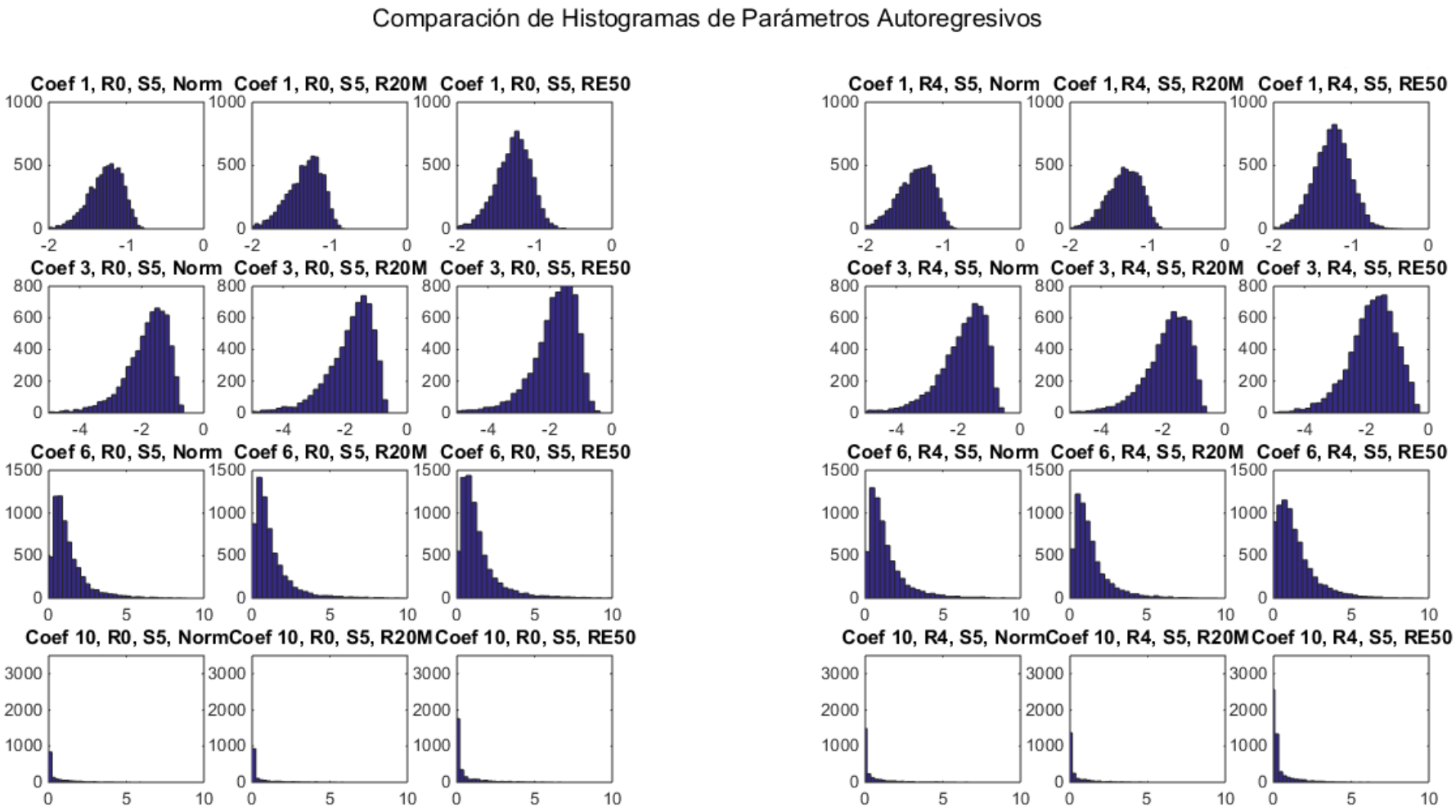


Figura N°3.9. Histogramas de series de parámetros AR.

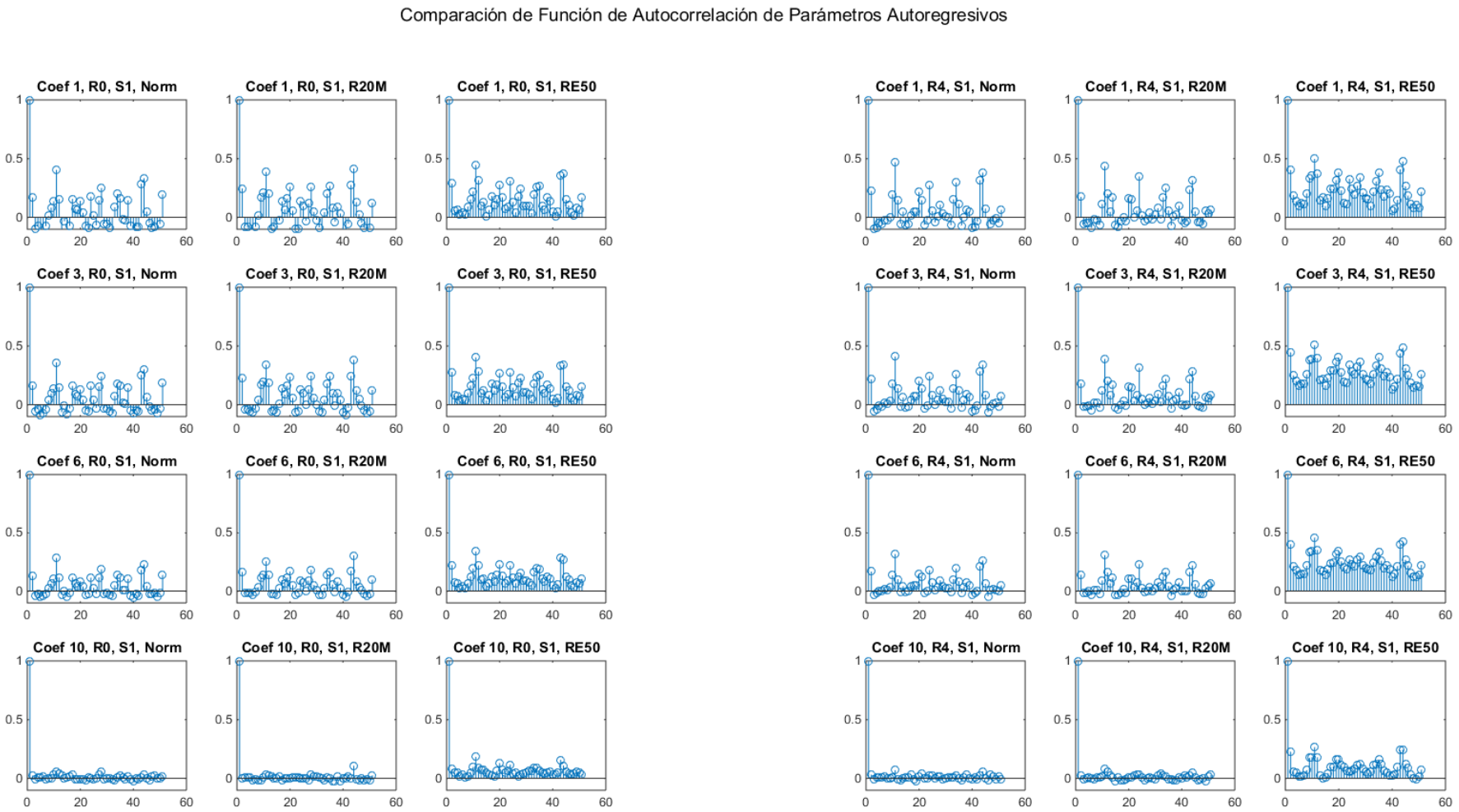


Figura N°3.10. Función de autocorrelación de parámetros AR.

Distancia de Mahalanobis.

La clasificación de estados estructurales mediante el uso de la distancia de Mahalanobis está fundamentada en la detección de valores atípicos de las variables bajo estudio. A estos valores se les llama 'outliers', y tienen la característica de ser llamativamente distintos al resto del conjunto de datos y por tanto se supone que fueron generados por un otro mecanismo o comportamiento estructural (H. Sohn, Worden, and Farrar 2002).

En el caso de un proceso de una sola variable, una de las formas más sencillas propuestas para la evaluación de outliers consiste en el cálculo de una desviación estadística:

$$z_i = \frac{(x_i - \bar{x})}{s} \tag{3.18}$$

donde x_i es la medición evaluada como potencial outlier y $\bar{x}$ junto con s son el promedio y desviación estándar de los valores anteriores o de un estado de referencia. La forma de evaluar si z_i es un outlier recae en su comparación con un umbral de decisión.

La extensión a procesos multivariables en el cual existen n observaciones de p variables, se realiza mediante la distancia de Mahalanobis dada por la Ecuación 3.19

$$D_i = (X_i - \bar{X})^T S^{-1} (X_i - \bar{X}) \tag{3.19}$$

donde X_i es el vector evaluado, $\bar{X}$ es el vector promedio de las variables y S es la matriz de covarianza del proceso. Nuevamente, la determinación de si X_i es un outlier se realiza mediante la comparación con un umbral.

En el caso de monitoreo estructural usando modelos $AR(p)$, se considera que los p coeficientes de los modelos son las variables del proceso, calculados en subventanas de la señal de aceleración obtenida de algún sensor. Además, se requiere un estado base con el cual poder calcular la distribución de los parámetros bajo una condición estructural normal y así obtener la matriz de covarianza.

3.3. Aplicación

En esta sección se aplican las metodologías anteriormente mencionadas para el estudio de los ensayos de laboratorios presentados en el capitulo anterior.

3.3.1 Metodología: Outliers de errores residuales

A modo de resumen, la metodología de detección de valores atípicos considera una señal de referencia con la cual se construyen los límites de control estadístico y además se calcula el modelo AR(p) base. Esta señal debe ser una con la estructura en un estado o condición normal, y por lo tanto, corresponde a los ensayos sin aplicación de daño ni adición de masa ni aplicación de temperatura. En esta parte, se usan todos los datos de referencia para entrenar e identificar el modelo. A su vez, se recuerda que se aplica un decimado para reducir la tasa de muestreo a 25[Hz], todas las señales se normalizan con respecto a su media y desviación estándar y el orden del modelo es p=15.

El procedimiento se describe a continuación:

1. Calcular un modelo $AR(p)$ para la señal de referencia de n datos. Esto devuelve p coeficientes del modelo y $(n-p)$ errores de ajuste.
2. Se calculan los límites de control usando los $(n-p)$ errores. Para ello, los errores son subdivididos en grupos de 4, a los que se les calcula la media y desviación estándar, generando $(n-p)/4$ valores de $\overline{x_i}$ y s_i. Finalmente se calculan los LCL, CL, y UCL a partir de los valores $\overline{x_i}$ y s_i, utilizando las Ecuaciones 3.9 - 3.12.
3. Para una nueva señal, se calculan errores residuales usando el modelo $AR(p)$ obtenido en el punto 1. Nuevamente se dividen en grupos de 4 y se calcula la media de cada grupo, finalmente comparándolos con los LCL y UCL obtenidos en el punto 2.
4. Se cuenta el N° de outliers presentes en la nueva señal.

En la Figura 3.11 se muestra el gráfico de X-bar para la señal de referencia. Considerando que dicha serie de aceleración cuenta con aproximadamente 7500 datos, la agrupación divide los datos en cerca de 1800 valores de X-bar. En esta figura se puede apreciar la existencia de algunos valores atípicos lo cual es esperable debido a la forma en que se construyen los límites basado en el supuesto de normalidad de la distribución de errores. En la Figura 3.12, se muestra el gráfico x-bar para un ensayo realizado con un 50% de reducción de sección en una columna. La presencia de outlierses clara, pudiéndose notar muchísimos valores fuera de los rangos. Estos gráficos se pueden replicar para todos los sensores y para todos los casos de daño o condiciones estructurales. Como el fin último es contar el número de valores atípicos presentes, tiene mayor utilidad hacer un gráfico de barras con la cantidad de outliers. Estos resultados se muestran en la Figura 3.13 para el caso del sensor N°1, y se pueden realizar notables aseveraciones al respecto. Por una parte, prácticamente todos los casos estructurales con presencia de daño arrojaron un mayor número de outliers, comprobando que la metodología es efectiva en la detección de cambios estructurales, pero no solo eso, sino que en términos globales a medida que el cambio estructural es más severo, el numero de outliers también aumenta, significando que el procedimiento podría ser utilizado como un indicador de la magnitud del cambio. A su vez, los resultados para el sensor N°5 (ubicado cerca al lugar de la aplicación del daño), son análogos a los del sensor N°1, pero con la salvedad de que el número de outliers es mayor a los del sensor 1 en todas las condiciones estructurales, indicando que además se podría usar esta información para localizar los cambios estructurales.

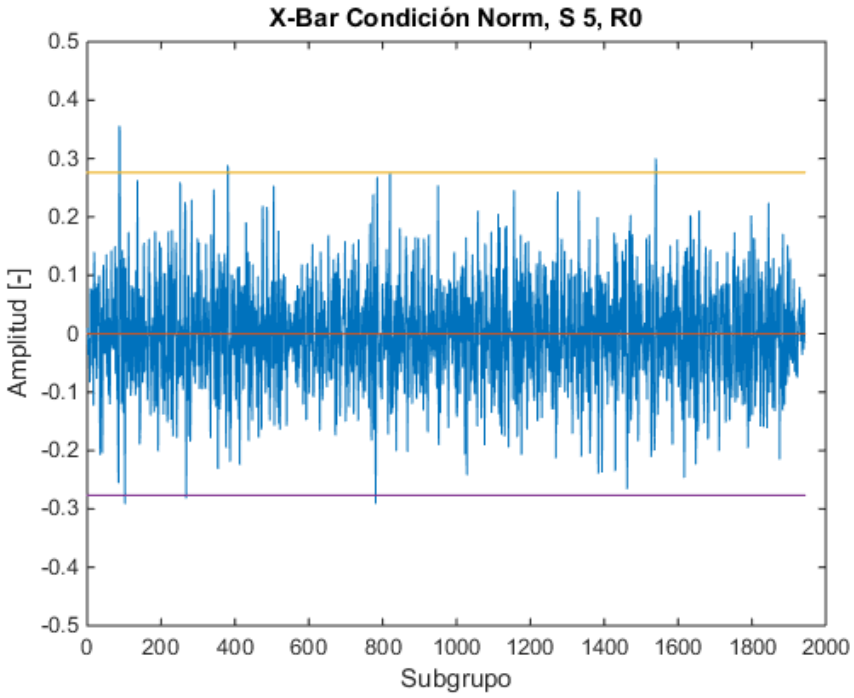


Figura N°3.11 X-Bar de la señal de referencia, con los límites superior e inferior.

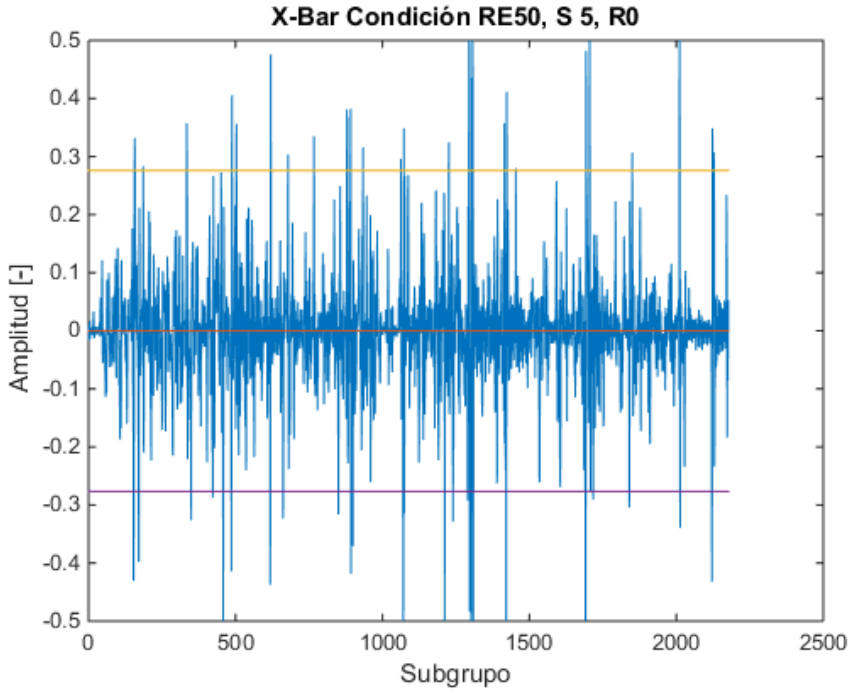


Figura N°3.12. X-Bar para una reducción del 50% en una columna.

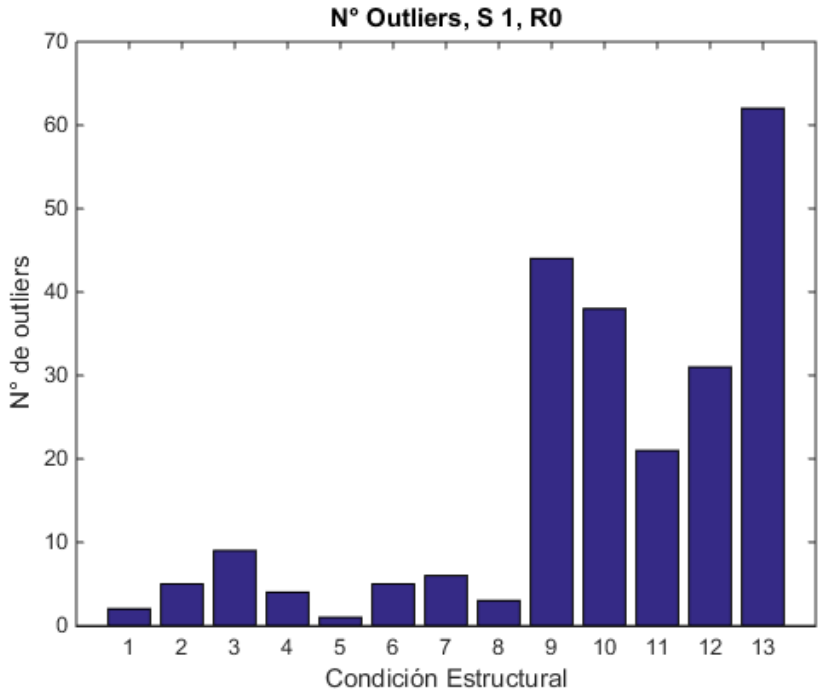


Figura N°3.13. N° de outliers según condición estructural. Sensor 1.

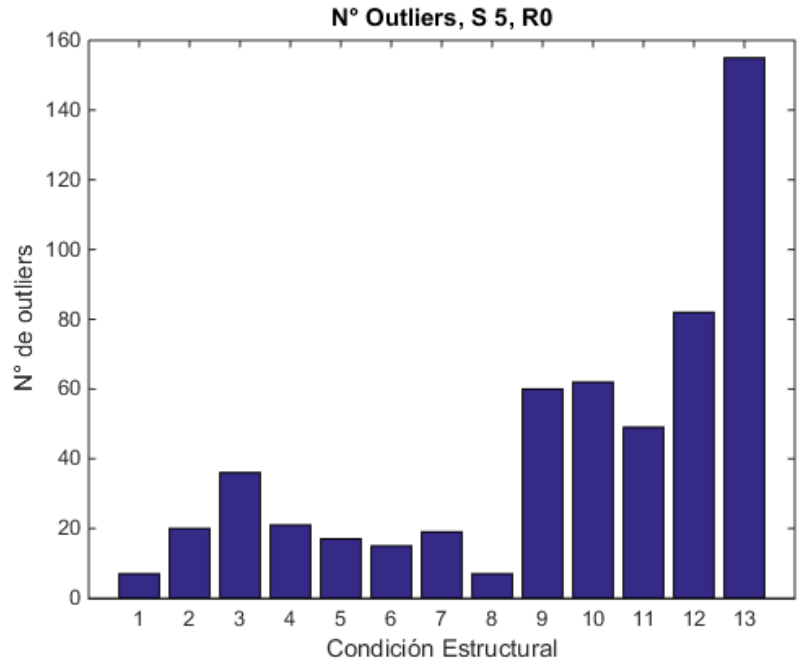


Figura N°3.14. N° de outliers según condición estructural. Sensor 5.

3.3.2. Metodología: Distancia de Mahalanobis.

A diferencia de la metodología basada en errores residuales, el procedimiento usando la distancia de Mahalanobis considera la creación de ventanas solapadas en las series de aceleración para el cálculo de los parámetros de los modelos AR(p). Por otra parte, los métodos comparten la

necesidad de una base o línea de referencia, que corresponde a la condición estructural sin daño. Para comprobar que la metodología es efectiva en distinguir estados estructurales distintos, pero pasando las pruebas de falsos positivos, la matriz de covarianza con la que se calcula la distancia considera solo la mitad de la señal de aceleración de la estructura sana. Así, en los gráficos de resultados se podrá verificar si el algoritmo es capaz de detectar los cambios estructurales o tan solo detecta valores fuera de los de entrenamiento.

El procedimiento es como sigue:

1. Creación de ventanas solapadas de una señal de aceleración de referencia. A cada una de estas ventanas se calculan los parámetros del modelo AR(p). Estos datos se utilizan como base de datos de coeficientes de referencia. El largo de las ventanas es de 60[s]. Este paso se realiza solo hasta la mitad de la señal de referencia.
2. Luego, para cada señal nueva, se concatena la señal de referencia al comienzo y luego se crean ventanas solapadas y calculan los coeficientes AR(p). Con estos coeficientes se calcula la distancia de Mahalanobis, considerando la base de datos creada en el punto 1. Este cálculo se realiza usando la función Matlab 'mahal'.
3. Se grafican los valores obtenidos para las distancias.

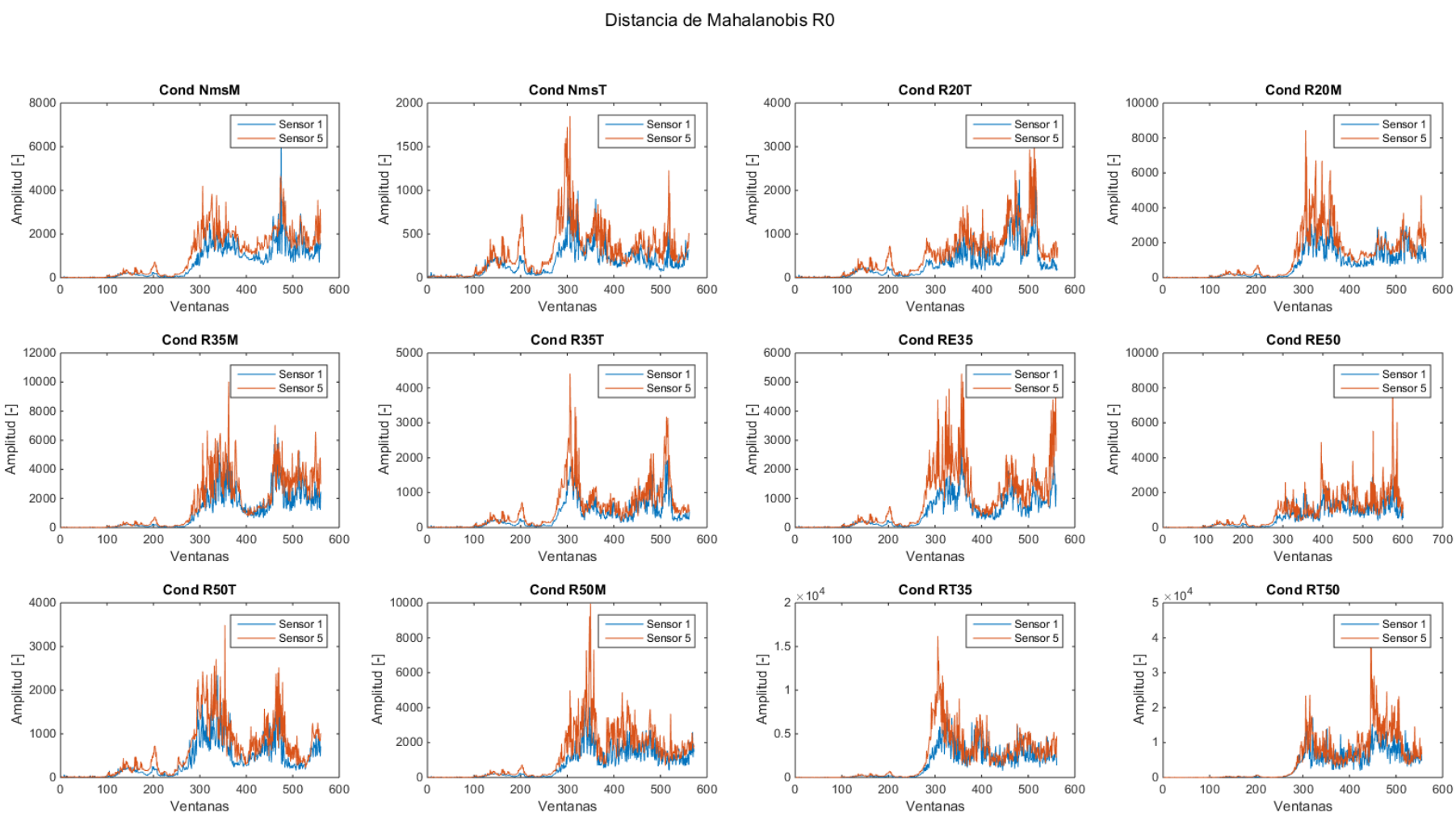


Figura N°3.15. Distancias de Mahalanobis.

La Figura 3.15. contiene los resultados de la aplicación de la metodología de cálculo de distancias de Mahalanobis aplicado a los ensayos de laboratorio. En ella, se muestran las distancias tanto para el sensor 1 como para el sensor 5, y para cada condición estructural. Como se mencionó anteriormente, solo la primera mitad de la señal de referencia se utilizó para la construcción de la base de datos, por lo que en los gráficos de la Figura 3.15, el primer cuarto de los resultados corresponden a las distancias de los parámetros que coinciden con la base de datos. El segundo cuarto son los parámetros de la señal de referencia no considerados en la base de datos y los últimos dos cuartos corresponden a la serie de aceleración de un estado estructural distinto. Los resultados son notorios. Primero, cabe destacar que la distancia de Mahalanobis tiene gran capacidad de detectar resultados dentro del comportamiento de entrenamiento, lo que es posible de observar analizando los primeros 100 datos (valores de distancia) de cada gráfico, cuyos valores son muy cercanos a cero. Además, pese a que existe un aumento en las distancias para la mitad de la señal de referencia que no está incluida en la base de datos, el incremento es apenas perceptible en la mayoría de gráficos, indicando que el procedimiento pasa la prueba de falsos positivos. Por otra parte, todas las mitades correspondientes a casos estructurales con daño muestran un aumento notorio en la distancia de Mahalanobis, lo que confirma la capacidad de diferenciar estados estructurales distintos. Otro punto interesante de notar, es que la distancias generadas a partir del sensor 5 son en general mayores que la del sensor 1, lo que podría significar que el método es más sensible mientras más cerca se esté del daño o cambio, pudiendo ser extendido al problema de localización. Sin embargo, faltan más datos para poder hacer una aseveración con tantas implicancias. Por último, pareciera que la distancia de Mahalanobis tiene una alta correlación con el daño aplicado, lo que es apreciable en los últimos dos gráficos, RT35 y RT50, en los que la distancia alcanza sus valores máximos. Aún así, la clasificación final en estados estructurales está basada en la definición de un umbral de decisión el cual no es claro donde ubicarlo sin saber a priori los resultados.

3.4. Aplicación en Torre Central.

En la presente sección se estudia el comportamiento de los modelos autoregresivos como indicadores de cambios estructurales, aplicado a los registros de aceleración de la Torre Central de la FCFM, cuyas características fueron presentadas en el cápitulo anterior. Recordar que se cuenta con registros cada 15 minutos, los que fueron filtrados en el espacio de las frecuencias con un filtro pasa banda el que considera el rango de frecuencias del edificio. Además, se seleccionaron los tres sensores con la mejor calidad. La base de datos de registros considera los días de Enero y Marzo del 2010. La ocurrencia del terremoto del 27 de Febrero del 2010 generó cambios estructurales en el edificio, los que se esperan puedan ser evidenciados con las metodologías presentadas.

3.4.1. Cálculo del orden del modelo autoregresivo.

El primer paso de la metodología, independiente del tipo de feature que se vaya a extraer, consiste en la estimación del orden de los modelos autoregresivos que se irán ajustando. Este paso es idéntico al realizado para los ensayos de laboratorio. Previamente se hizo un análisis de Fourier para la limpieza de los datos, obteniendo que el edificio muestra frecuencias modales de hasta 12Hz, por lo que el decimado de los registros considera un factor igual a 8, pasando de una frecuencia de muestreo de 200Hz a una de 25Hz.

Considerando una señal de referencia aleatoria, se realiza el análisis de AIC y RMS. La Figura 3.16 muestra el resultado para la señal del 1 de Enero a las 1745 hrs. Se muestra además la diferencia habiendo normalizado o no la señal. En esta figura, se observa que tanto AIC como RMS muestran un decaimiento inicial fuerte. Al igual que en el caso de laboratorio, no se aprecia un mínimo para ninguno de los índices, pero sí una clara tendencia a la estabilización, la cual parece ser mas lenta en el caso de la señal normalizada.

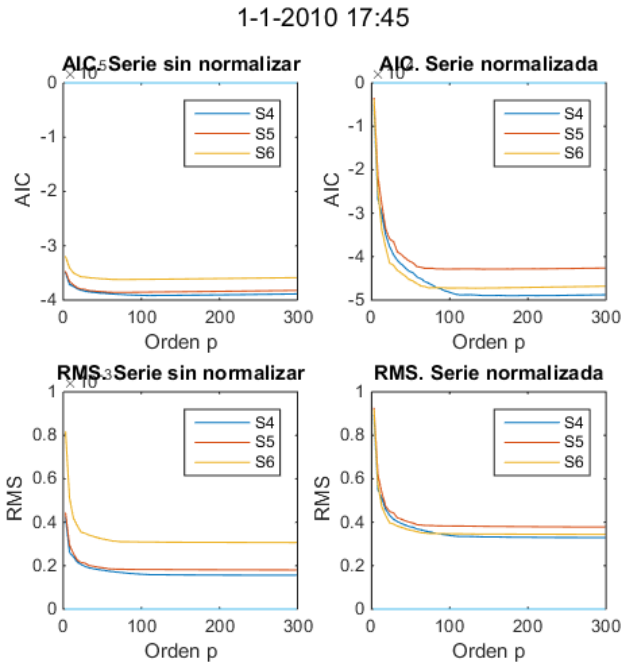


Figura N°3.16. Índices AIC y RMS para la Torre Central. Enero 2010.

La Figura 3.17 muestra los resultados análogos correspondientes a una señal obtenida en Marzo. Los resultados son casi idénticos a los de la Figura 3.16 en términos de la estabilización pero llama la atención de que el valor de los índices tiende a números distintos.

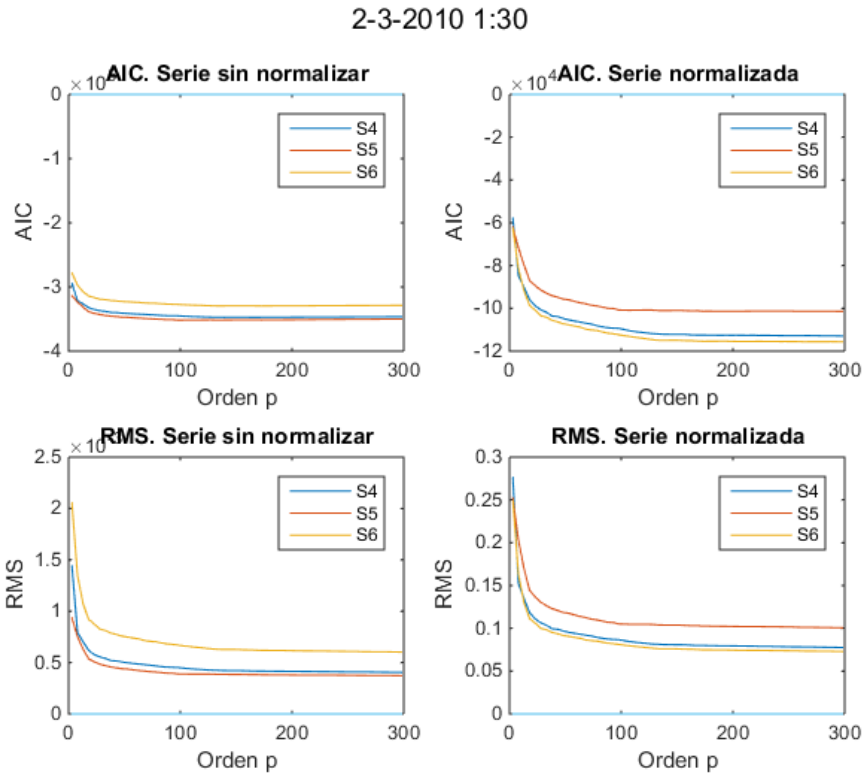


Figura N°3.17. Índice AIC y RMS. Torre Central, Marzo.

A partir de las Figuras 3.16 y 3.17, el orden de los modelos autoregresivos escogido es de $p = 50$. Esta elección se basa en la selección del valor codo,y si bien no corresponde al mínimo, se acepta que la contribución de un orden mayor a 50 no es considerable. Además, el uso de un orden más grande es caro computacionalmente y se corre el riesgo de sobreajustar las señales. La Figura 3.18 muestra un ejemplo de la señal medida post decimado y la señal ajustada, así como también los errores residuales encontrados y su histograma. Se aprecia que el ajuste es de muy buena calidad.

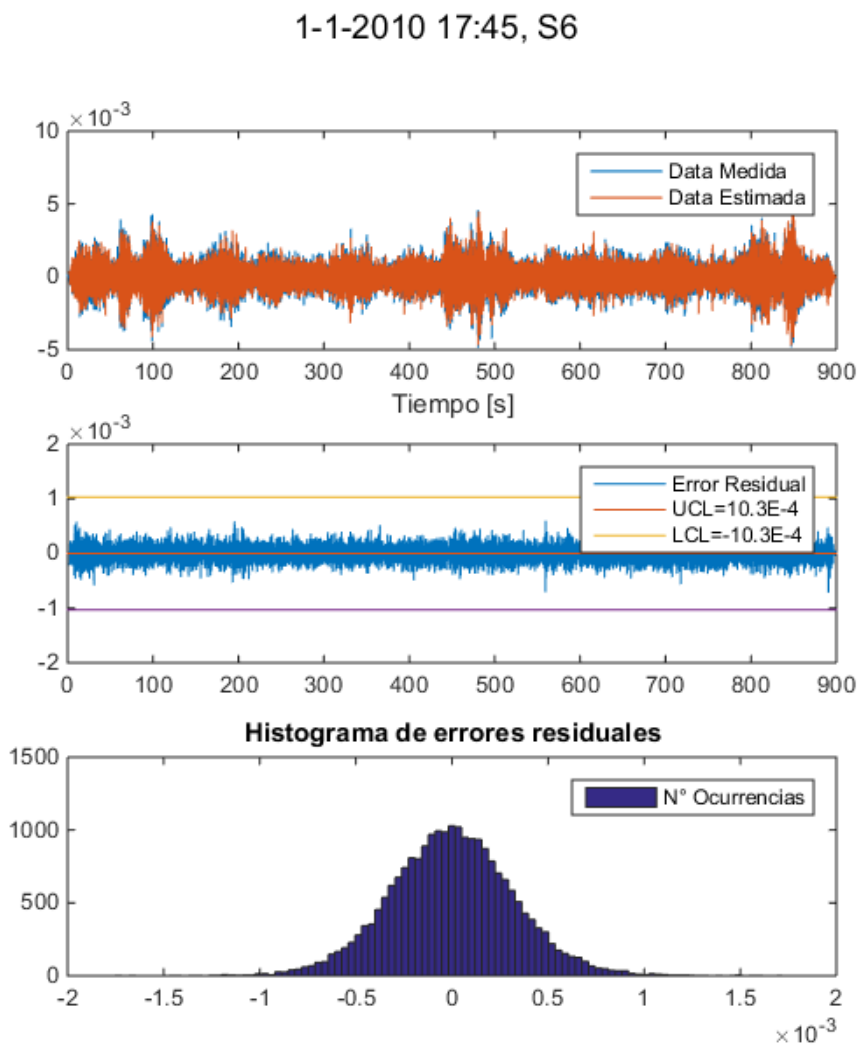


Figura N°3.18. Ajuste de una señal de referencia y sus errores residuales.

3.4.2. Metodología: Outliers de errores residuales.

A modo de recuerdo, esta metodología consiste en ajustar un modelo AR a una señal de referencia, y utilizar dicho modelo para realizar predicciones de las señales futuras. A partir de las predicciones y las señales reales, se calculan los errores residuales y se realiza un control estadístico de éstos. El número de errores que caigan fuera de los limites de control se utiliza como indicador de un cambio.

En el caso de registros obtenidos en una estructura real, nuevamente nos vemos enfrentados al problema de lidiar con distintas amplitudes a lo largo del día. Esto es incluso más problemático en esta metodología ya que se requiere el uso de una señal de referencia, lo que implica distintos resultados si consideramos una señal obtenida durante la noche versus una registrada en plena condición operacional. Para ejemplificarlo, se estudiarán las diferencias entre ambos casos.

La Figura 3.19 muestra los resultados del n°de errores fuera de los límites de control de la primera semana de Enero del 2010. La señal de referencia con la que se obtuvo el modelo

autoregresivo y los límites de control (UCL y LCL) corresponde al 4 de Enero a las 02:00. En esta figura se aprecia claramente como la amplitud de la señal en las horas laborales influye enormemente en el número de errores outliers. De hecho, para el 9 y 10 de Enero, que corresponden a los días sábado y domingo, casi no existen errores fuera de los límites de control.

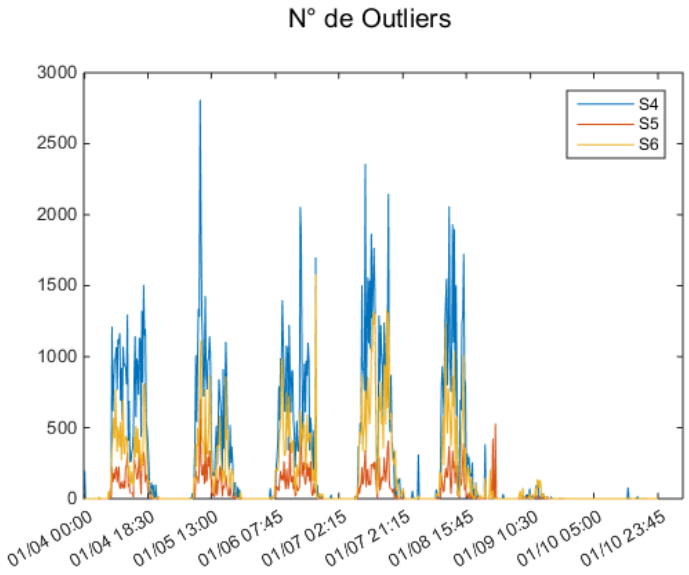


Figura N°3.19. Outliers primera semana de Enero 2010. Referencia: 4 Enero, 2AM

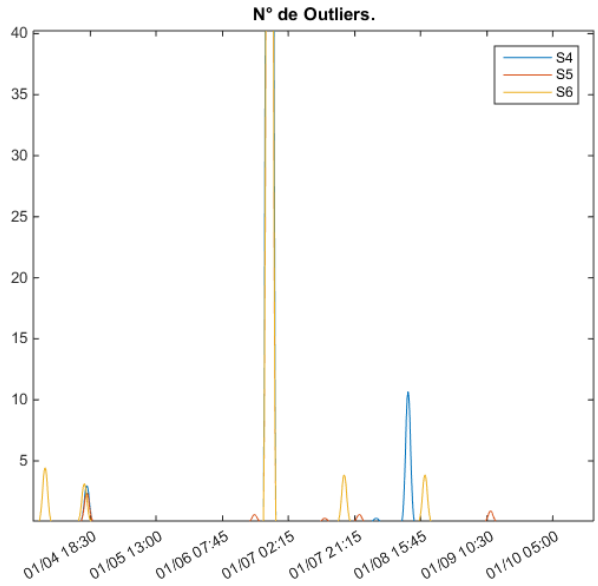


Figura N°3.20. Outliers primera semana de Enero 2010. Referencia: 4 Enero, Mediodía

La Figura 3.20, muestra lo obtenido para la misma de semana pero considerando la señal del 4 de Enero a las 12:00 del mediodía como referencia para el cálculo del modelo y los límites de

control. En esta figura se evidencia que la cantidad de outliers disminuyó notablemente, siendo igual a cero en una gran cantidad de registros.

En las Figuras 3.19 y 3.20 hemos visto que la señal de referencia es de gran importancia a la hora de aplicar la metodología de errores fuera de límites de control. Sin embargo, pese a que los resultados son sumamente distintos dependiendo de la señal escogida, cabe destacar que se podría hacer un análisis análogo al mencionado en el capítulo anterior: utilizar una señal de referencia para condiciones operacionales similares. En este caso, la señal de las 2AM sólo debiesemos considerarla para analizar registros obtenidos durante la noche. Otra opción es normalizar las señales diviendolas por su desviación estándar, en cuyo caso se obtienen resultados muy distintos pero reveladores. Para ejemplificar esto y mostrar una comparación de registros antes y post terremoto, se aplica lo mencionado a registros obtenidos de tres lunes de Enero (4, 11 y 18 de Enero), y tres lunes de Marzo (8, 15, 22 de Marzo), entre las 2 y las 6 AM. Los resultados para esta selección de registros se pueden observar en la Figura 3.21, que no considera normalización de las señales y utiliza el registro del 4 de Enero a las 2AM como referencia. En dicha figura, los resultados son notables, evidenciando una clara diferencia entre los outliers de los lunes de Enero y los de Marzo.

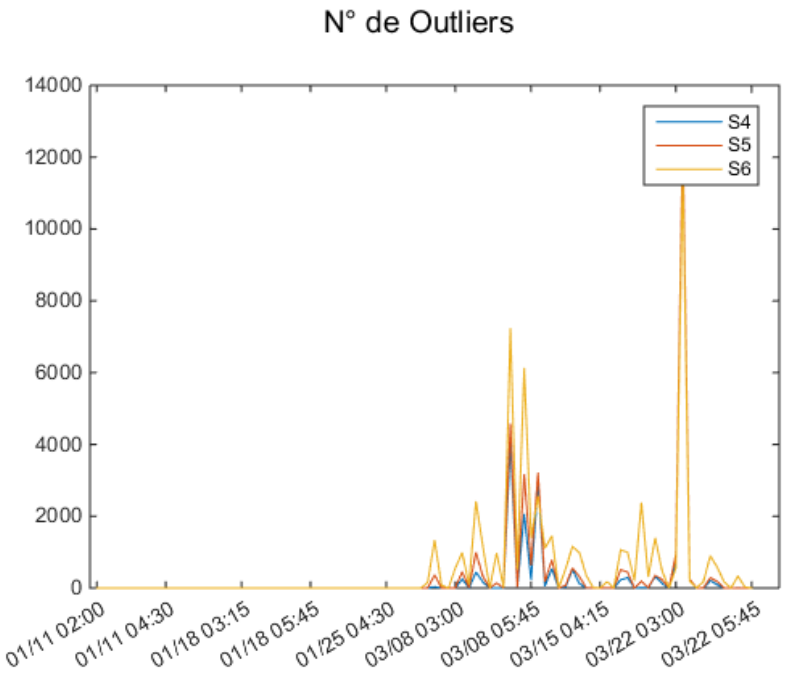


Figura N°3.21. Outliers. Lunes de Enero y Marzo 2010. Ref: 4/1 2AM.

Para cerrar este análisis de outliers, en la Figura 3.22 se muestran los resultados para los mismos registros de la Figura 3.20, pero esta vez considerando el registro del 4 de Enero a las 12 del mediodía, y con normalización de los registros. En dicha figura se aprecia que pese a ser de una condición operacional distinta, existe una clara diferencia entre los resultados pre y post terremotos. Además, el comportamiento de los outliers durante los dias de Enero es sustancialmente distinto, existiendo un gran número de errores fuera de control para cada registro, incluso en aquellos obtenidos durante la misma condición operacional que la señal de referencia.

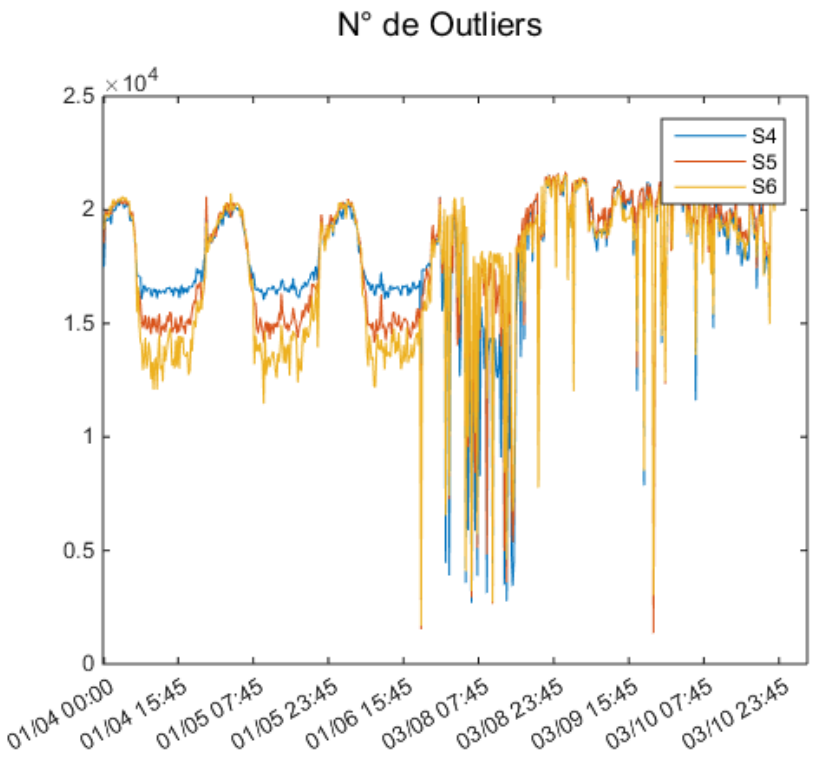


Figura N°3.22. Outliers Enero y Marzo. Registros normalizados. Ref 4/01 12:00 mediodía.

3.4.3. Metodología: Distancia de Mahalanobis.

La metodología que hace uso de la distancia de Mahalanobis necesita el cálculo de series de parámetros AR a partir de ventanas móviles de los registros de aceleración. En el capítulo 4, para el análisis de los ensayos de laboratorio se realiza un estudio de cómo se deben generar las series en relación al largo de las ventanas que se debiesen considerar. En dicho capítulo el análisis está basado en la convergencia del RMS de los errores residuales en función del largo de la ventana y además se realiza un pequeño análisis de la convergencia de la estádistica de los parámetros AR. La conclusión es que es muy dificil justificar un largo de ventana en terminos de la convergencia del promedio o desviación estándar de los parámetros, ya que no existe una tendencia clara de convergencia. El valor RMS sí presenta convergencia (Figura 3.1), y con ello se justifica la selección del largo de ventana, pero de una forma global. Cabe destacar que a pesar de que los parámetros muestran una relativa dependencia al largo de la ventana, dicha selección no influye en gran medida en los resultados finales y lo importante es mantener la selección constante para todos los registros: de tal forma, si existe un sesgo, este será aplicado a todos los resultados por igual.

Creación de las series de parámetros AR.

Considerar un registro de aceleración de Ndatos después del proceso de decimado. El orden de los modelos ajustados es igual a $p = 50$. El tamaño de las ventanas traslapadas tiene un largo igual a $100p$. En total, a partir de un registro de aceleración se genera una serie de 2500 valores para cada uno de los 50 parámetros de los modelos AR.

Cálculo de distancias.

El cálculo de distancias de Mahalanobis considera el uso de una señal de referencia o un set de observaciones de parámetros. Es lo que se consideraría un set de observaciones en estado

"normal". Lo mostrado anteriormente con el estudio de Outliers, indica que probablemente existan diferencias al considerar una señal de referencia durante la noche versus una obtenida durante el día. La fórmula para calcular la distancia se explicitó en la Ecuación 3.19, en la que *S* representa la matriz con el conjunto de observaciones de parámetros AR que se utiliza como referencia.

Distancias: Relación con la señal de referencia.

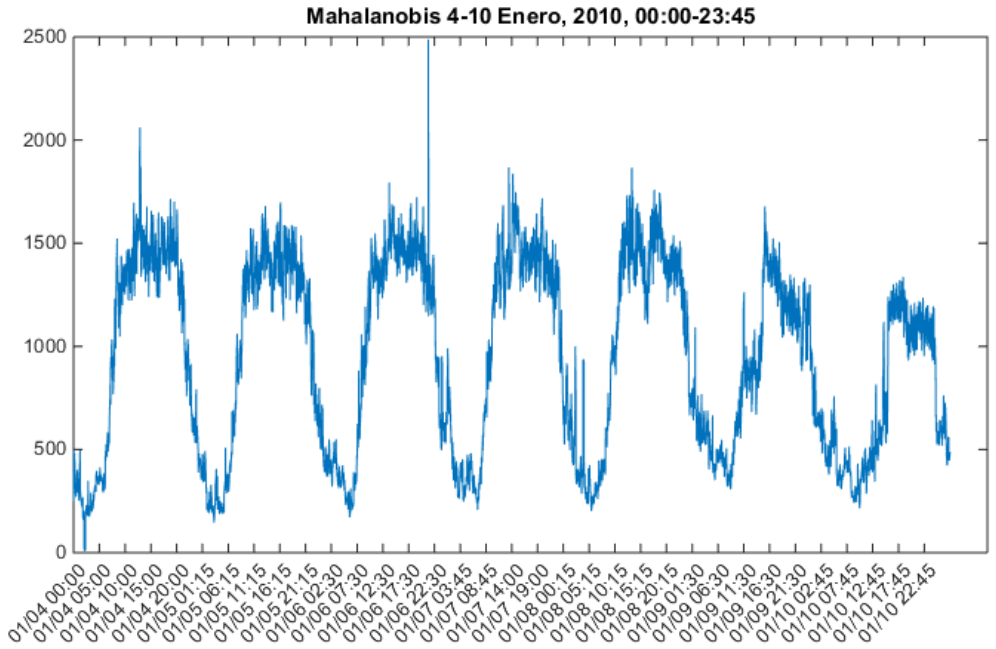


Figura N°3.23. Distancia de Mahalanobis. Primera semana de Enero. Referencia 4/01 2AM.

En la Figura 3.23 se muestra el resultado de la aplicación de la distancia de Mahalanobis para los todos los registros de la primera semana de Enero 2010, considerando la señal del 4 de Enero a las 2AM como referencia. Se observa claramente una periodicidad producto de los cambios en la condicion operacional del edificio, aumentando el valor de la distancia para los registros correspondientes a las horas laborales. Es muy llamativo notar que el aumento de la distancia también se produce en días de fines de semana. Por otra parte, la Figura 3.24 muestra los mismos resultados pero considerando como referencia la señal del 4 de Enero a las 12 del mediodía. En este caso también se puede observar una periodicidad pero de forma contraria en comparación con la Figura 3.23, existiendo un aumento de la distancia de Mahalanobis durante las horas nocturnas. Además, la duración del aumento es mucho menor que en el caso del uso de una señal de referencia nocturna. Una posibilidad es utilizar como referencia un conjunto de observaciones provenientes tanto de la noche como del día. Esto se muestra en la Figura 3.25, donde si bien aún es apreciable la periodicidad, el aumento del valor de la distancia es muchísimo menor. Cabe destacar que los días con el aumento más apreciable corresponden al sábado y domingo, lo cual es atribuible a que la señal de referencia de las 12 del mediodía del lunes no representa bien a los días de fin de semana, a pesar de que sí lo hace para el resto de días laborales.

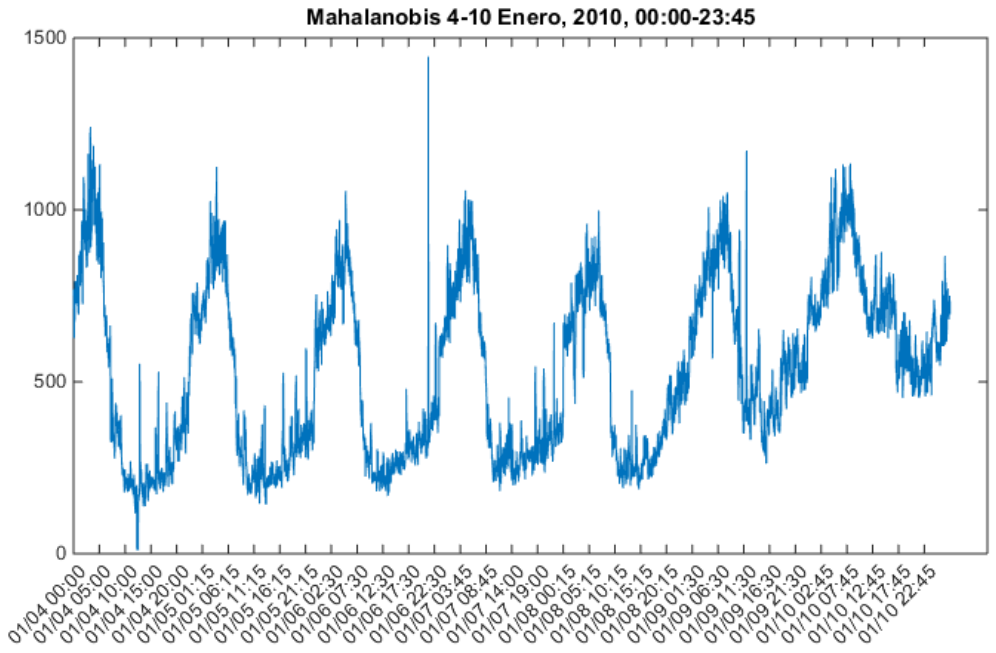


Figura N°3.24. Mahalanobis. Primera semana de Enero. Referencia 4/01 12 mediodía.

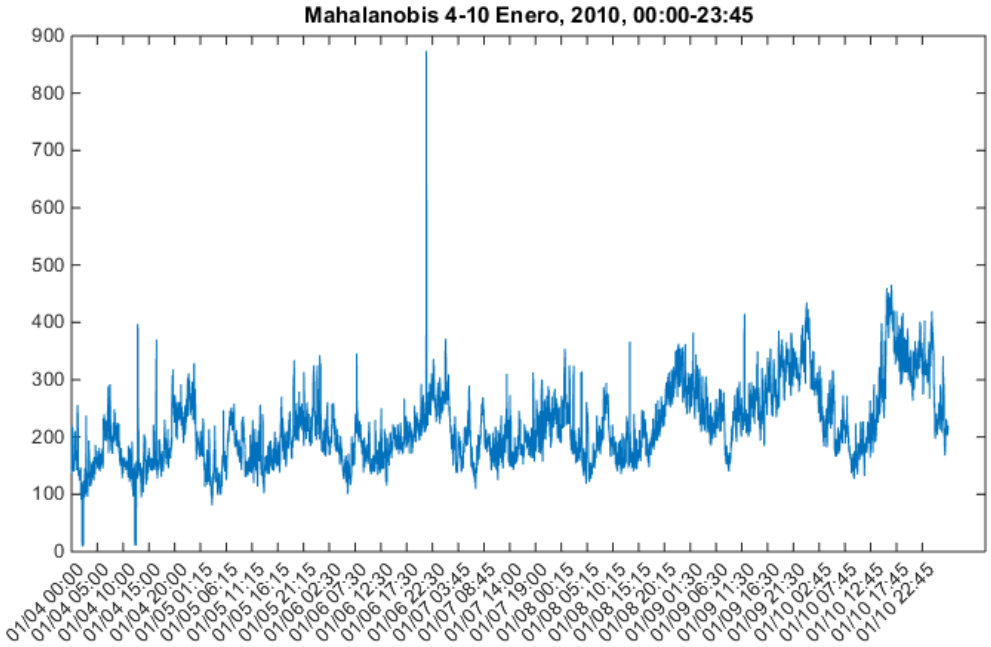


Figura N°3.25. Mahalanobis. Primera semana de Enero. Referencia 4/01 a las 2AM y 12Mediodia.

Distancias: Comparación de registros de Enero y Marzo 2010.

A continuación se muestra la aplicación de la metodología de cálculo de distancias de Mahalanobis para la comparación de registros obtenidos antes y después del terremoto del 27F. En la Figura 3.26 se muestran los resultados usando como referencia el 4/01 a las 2AM y 12 del mediodía. En este caso la diferencia entre los registros obtenidos en Enero versus los obtenidos en Marzo es notoria, indicando sin lugar a dudas la ocurrencia de un cambio estructural mayor. Uno de los puntos importantes a destacar es que los resultados son consistentes sin importar la

hora en que hayan sido obtenidos los registros de Marzo. Aún así, es conveniente conservar la metodología de comparar registros adquiridos durante la misma condición operacional, lo que se muestra en la Figura 3.27, considerando solo registros obtenidos entre las 2 y 6AM. Los resultados son idénticos en el sentido de que existe una clara diferencia pre y post terremoto.

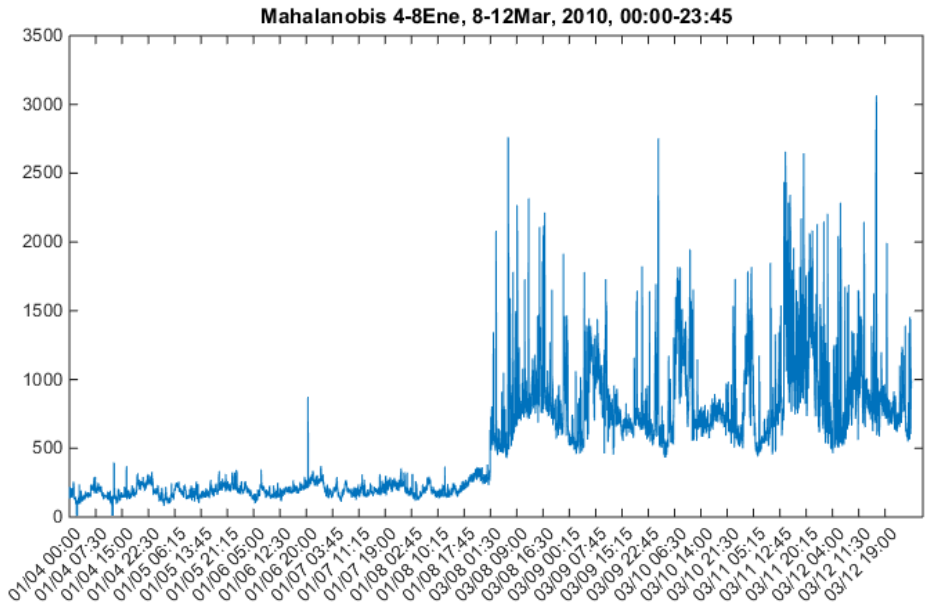


Figura N°3.26 1ra semana de Enero vs 1ra semana de Marzo.

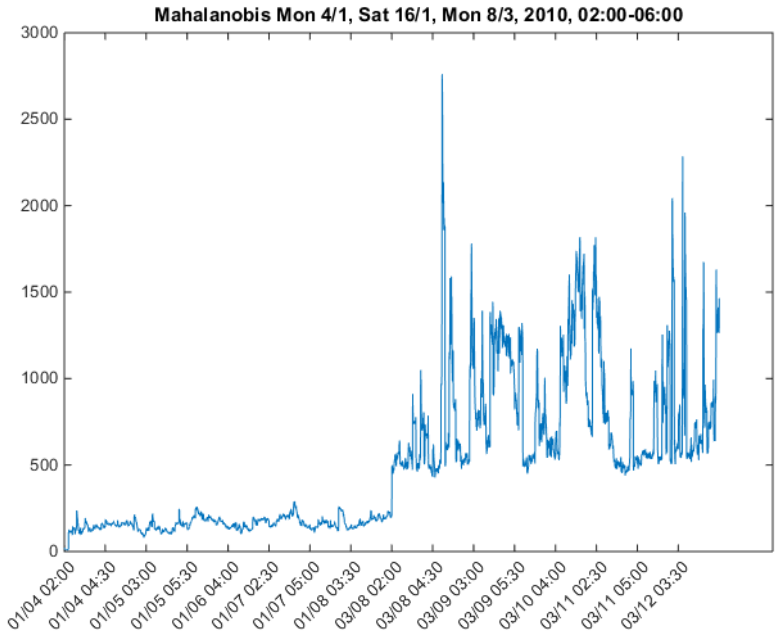


Figura N°3.27. Mahalanobis. 1ra semana de Enero vs 1ra semana de Marzo entre 2 y 6AM

3.5. Conclusiones

En esta sección, se parte por destacar que cualquier aplicación de metodologías que consideren modelos autoregresivos necesita de dos pasos previos. El primero de ellos consiste en un proceso de decimado con el que se reduce la complejidad númerica al reducir las señales de

aceleración. La selección del factor de decimado, se justifica a partir de un análisis PSD de las estructuras analizadas con la finalidad de mantener el contenido de frecuencias de los registros originales. El segundo paso previo consiste en realizar una estimación del orden de los modelos con los que se realizarán los ajustes. En este sentido, el cálculo del orden óptimo no genera resultados únicos al no percibir la ocurrencia de mínimos globales. Sin embargo tanto AIC como RMS muestran tendencias decrecientes claras que permiten la selección de un orden con el que se alcance un nivel aceptable de ajuste. Una vez seleccionado el orden de los modelos, se destaca que una muy buena calidad de ajuste se logra tanto para periodos estacionarios (vibraciones ambientales y/o operaciones), como para periodos de excitación sísmica, lo que es fundamentado en la estabilidad de los errores de ajuste y residuales. Estos resultados son validados tanto para ensayos de laboratorio como para registros de la Torre Central.

En el caso del uso de características basadas en las propiedades de los errores residuales, se encontró que este tipo de feature son efectivas en encontrar diferencias entre distintas condiciones estructurales. En el caso de ensayos de laboratorio, incluso se encontró que existe mayor cantidad de errores fuera de control para los casos estructurales con mayor nivel de daño. Esto permitiría usar la metodología en un nivel superior de información acerca de la salud estructural al ser un indicador de la magnitud del cambio estructural. En el caso de registros de una estructura real, se encontró que la cantidad de errores outliers depende de la señal que se haya utilizado para calcular el modelo autoregresivo con el que se calculan las predicciones y respectivos errores. Si la señal de referencia es de la noche, hay un aumento considerable de errores para los registros durante el día, mientras que si la señal de referencia es del día, el aumento es prácticamente imperceptible. No obstante, el comportamiento de los errores es mucho más estable si se normalizan las señales por su desviación estándar. Además, si se comparan registros de Enero y Marzo, las diferencias son claras sin importar si la señal es normalizada o no. Aún así, se recomienda la comparación de registros obtenidos durante la misma condición operacional. En lo que respecta al uso de la distancia de Mahalanobis, en el caso de ensayos de laboratorio también se encontraron diferencias claras entre distintas condiciones estructurales, y un aumento en el valor de la distancia para condiciones con mayor nivel de daño. Para registros de la Torre Central, el resultado fue análogo a los de Outliers, en el sentido de que se depende de la señal de referencia. Sin embargo, como en esta metodología se pueden considerar observaciones provenientes de distintos registros, al mezclar observaciones nocturnas y de día los resultados presentan menor variación. Al comparar resultados de registros obtenidos en Enero y Marzo, la diferencia es clara, indicando la ocurrencia de un cambio estructural mayor. Nuevamente, se recomienda la comparación de registros obtenidos durante la misma condición operacional para evitar cualquier tipo de sesgo introducido por distintos niveles de excitación.

Como punto final, se concluye que tanto los errores residuales como la distancia de Mahalanobis presentan información sensible referente al estado estructural y podrían ser utilizados en un sistema de monitoreo de salud estructural, pero poseen el inconveniente de necesitar una señal de referencia la cual podría ser obtenida de un registro que ya contiene un cambio estructural.

Capítulo 4. Agrupamiento de parámetros AR(p)

4.1. Introducción.

El presente capítulo contiene la última metodología de monitoreo de salud estructural de sistemas civiles o mecánicos estudiada en este trabajo de Tesis, la cual consiste en la clasificación de distintas condiciones estructurales mediante la utilización de algoritmos de agrupamiento sobre series de tiempo de parámetros de modelos autoregresivos. Esta nueva metodología propuesta mezcla los estudios realizados en los capítulos 2 y 3, buscando aprovechar las ventajas que poseen esos métodos. A modo de recordatorio, en el tercer capítulo se estudió el uso de algoritmos de agrupamiento sobre objetos simbólicos obtenidos a partir de series de aceleración. Los fundamentos de dicha metodología consistían en que el contenido estadístico de las series de aceleración, representado mediante objetos simbólicos, podía servir como característica sensible a los cambios estructurales y finalmente ser clasificado en distintos conjuntos de objetos, los que caracterizarían diferentes estados estructurales presentes en las series de aceleración adquiridas. Por otra parte, en el cuarto capítulo se estudió la capacidad de predicción e identificación de sistemas estructurales utilizando modelos autoregresivos, con los supuestos de que al modelar y predecir series de aceleración de un sistema sin cambios estructurales, tanto los errores residuales como los parámetros de los modelos debieran mostrar un comportamiento estable.

Los resultados arrojados en los capítulos anteriores muestran que los algoritmos de agrupamiento funcionan muy bien al momento de clasificar 'características' de comportamiento similar, sin embargo, los objetos simbólicos extraídos directamente de las series de aceleración son demasiado sensibles a la energía de entrada, haciéndolos poco robustos ya que la clasificación puede arrojar como resultado "estructura en movimiento" y "estructura quieta" dependiendo de la amplitud de las series. Por otra parte, la metodología de autoregresión mostró gran capacidad de ajuste e identificación de modelos, incluso pasando la prueba de falsos positivos, pero está limitada en lo relacionado con la clasificación de estados estructurales ya que siempre se necesita de una línea base de referencia. La motivación y objetivo general del presente capítulo es, por tanto, utilizar la capacidad de clasificación de los algoritmos de agrupamiento de forma no supervisada, en conjunto con la robustez de los modelos autoregresivos para identificar el comportamiento de las series de aceleración, obteniendo una metodología única y completa para el monitoreo de salud estructural. Además, dentro de los objetivos específicos del capítulo se encuentra determinar cuáles son los parámetros que hacen que la metodología entregue la mayor sensibilidad en lo que respecta a cambios estructurales, pero sin separar estados de distintas condiciones operacionales o ambientales. Para cumplir con estos objetivos, se establecen estudios del comportamiento del algoritmo frente a distintos sensores, energías de entrada y diferentes coeficientes de los modelos autoregresivos. Finalmente, al igual que en los capítulos anteriores, toda la teoría y metodología propuesta se prueba utilizando ensayos de laboratorio y series de aceleración adquiridas en el edificio de la Torre Central del campus Beauchef de la Universidad de Chile.

El presente capítulo esta ordenado por secciones. La primera corresponde a la introducción descrita en estos párrafos. Luego, en la Sección 4.2. se detalla la teoría utilizada en el algoritmo propuesto. Si bien el contenido ya fue desarrollado en los capítulos anteriores, en esta sección se vuelven a mencionar, pero de forma resumida, los métodos usados para que éste capítulo sea auto contenido. En seguida, en la Sección 4.3. se encuentra el estudio de sensibilidad para determinar los sensores y coeficientes autoregresivos que muestran ser más adecuados para el algoritmo propuesto y entregan mayor información, así como también las comparaciones

llevadas a cabo para poner a prueba la capacidad de la metodología y no generar distinciones entre diferentes condiciones operacionales. En el apartado 4.4, se encuentra la aplicación de la metodología tanto para los ensayos de laboratorio y la Torre Central, en las Secciones 4.4.1 y 4.4.2, respectivamente. Por último, las conclusiones más importantes y los resultados destacados se resumen en la sección 4.5.

4.2. Fundamentos teóricos.

El procedimiento general del algoritmo propuesto en éste capítulo considera dos etapas de extracción de 'características' y otra de clasificación, y comienza por la transformación de las series de aceleración en series de parámetros AR(p) utilizando ventanas móviles. Este primer proceso ya es de por sí una etapa de extracción de características sensibles a los cambios estructurales ya que en teoría los modelos autoregresivos representan el contenido de frecuencias de la señal y ya en (Peeters 2000) se demostró que los modelos VAR son equivalentes a las propiedades modales. Luego, estas series nuevas de coeficientes AR(p) son utilizadas como input para el algoritmo de objetos simbólicos, que una vez más condensa las series en variables de menor complejidad numérica pero extrayendo la información sensible a los cambios estructurales, la cual está relacionada con la distribución estadística de las series. Finalmente, en la tercera etapa, los objetos simbólicos son agrupados usando algoritmos de agrupamiento, obteniendo conjuntos que caracterizan los distintos estados estructurales presentes (J. P. Santos 2014). En las siguientes secciones, se detalla el procedimiento para cada una de las etapas necesarias para llevar a cabo la metodología completa de monitoreo de salud estructural.

4.2.1. Series de tiempo de parámetros AR(p)

Transformación a series de parámetros AR(p)

Esta primera etapa parte del supuesto de que se conoce el orden óptimo de los modelos autoregresivos que se calcularán. Como se explicó en el Capítulo 3, este valor puede obtenerse utilizando los índices AIC y/o RMS. Con el orden de los modelos, se puede proceder a la primera extracción de características, que en este caso corresponden a los coeficientes de modelos autoregresivos ajustados a ventanas móviles en las señales de aceleración.

Considerar una señal de aceleración $\{a_t\}$ con N datos luego del proceso de decimado. La idea es generar una serie en el tiempo de parámetros $AR(p)$, para lo cual se utilizan ventanas móviles las cuales pueden ser traslapadas o no, y a cada una de dichas ventanas se le ajusta un modelo autoregresivo. De esta forma, si cada ventana móvil de $n$datos queda representada por un conjunto de $p$parámetros AR, y las ventanas se generan cada un dato, la transformación genera un resultado de $p$series de $(N-n+1)$ datos. En esta parte es interesante notar que el valor de n es escogido por el usuario, y más importante, dependiendo del valor de $n$la transformación puede actuar como una reducción o una amplificación de la información, lo que en definitiva puede generar series con mayor o menor sensibilidad a cambios estructurales. En efecto, el resultado de la primera etapa genera $p(N-n+1)$datos, lo que para $n < (N+1) - N/p$ resulta en

$$p(N-n+1) > p\left(N - N - 1 + \frac{N}{p} + 1\right) > N \qquad (4.1)$$

es decir, en un número mayor de datos que el original. Hay que ser cuidadoso con este procedimiento, ya que el aumento de datos no equivale a un aumento en la sensibilidad y podría introducir series carentes de información o significado estructural. Siguiendo la misma línea en relación a los valores posibles de n, si se ocupa $n = (N+1) - N/p$ las series de parámetros $AR(p)$ tendrán un número de datos igual a

$$N - \left[(N+1) - \frac{N}{p}\right] + 1 = N - N - 1 + \frac{N}{p} + 1 = N/p \qquad (4.2)$$

lo que equivale a ocupar ventanas no solapadas de un largo igual al orden de los modelos; es decir, usando ventanas no solapadas la transformación solo genera una reducción de la información ya que el largo de las ventanas no puede ser menor o igual al orden de los modelos. Esta discusión no se pone a prueba en la presente investigación, pero es importante tenerla en consideración al momento de analizar los resultados.

Antes de realizar la transformación a series de parámetros AR(p), las señales de aceleración pueden pasar por un proceso de normalización estadística según la Ecuación 4.3.

$$\{\bar{a_t}\} = \frac{\left(\{a_t\} - \overline{\{a_t\}}\right)}{\sigma(\{a_t\})} \tag{4.3}$$

El proceso de normalización anterior tiene por finalidad hacer que las señales no se vean influenciadas por la energía de entrada, pero manteniendo la información relacionada con las frecuencias y por tanto con los coeficientes de autoregresión. La decisión que hay que tomar tiene relación con el momento en que se realiza el proceso. Esto puede ser: normalizar a toda la señal, a cada ventana o combinar ambas propuestas. En este estudio se sigue la última opción, teniendo la intención de evitar generar diferencias producto de la amplitud en cada ventana.

Una ventaja que no ha sido mencionada aún sobre la presente metodología, es que permite realizar el estudio independiente del número de sensores que haya instalados y que realicen la adquisición de datos, lo que era uno de los puntos débiles en el capítulo 3. Esto se debe a que los algoritmos de agrupamiento y objetos simbólicos están diseñados para poder trabajar en un entorno de múltiples variables. Aún así, cabe mencionar que no todos los sensores aportan información valiosa relacionada con los cambios estructurales, como sería el caso de un sensor ubicado en un nodo de las formas modales predominantes, o que si se aplica la metodología incluyendo todos los sensores se puede enmascarar los cambios que detecte algún sensor en particular. Es más, analizando estas señales cada una por sí sola, se podría investigar la localización del cambio estructural.

Resumiendo, a partir de las mediciones de s sensores, la transformación a series de parámetros $AR(p)$ arroja como resultado un total de $s \cdot p$ series, las que serán nuevamente transformadas a objetos simbólicos para su posterior clasificación. Cabe destacar que es posible aplicar la metodología por separado a cada una de las series mencionadas.

Características de las series de tiempo de parámetros AR(p).

En la sección anterior se mencionó que el largo de las ventanas influye en el resultado de la transformación a series de tiempo de parámetros $AR(p)$. Uno de los objetivos de la presente sección es determinar la forma en que influye en número de datos en las ventanas y así poder determinar un tamaño adecuado de éstas para mejorar el rendimiento de la metodología. Para responder a esta pregunta, se utilizarán los ensayos de laboratorio que ya han sido expuestos en capítulos pasados.

Como la metodología considera un análisis de ventanas móviles con las que se calculan las series de tiempo de parámetros AR(p), lo que interesa es poder determinar el largo de ventanas que haga que el resultado muestre algún tipo de convergencia. Lógicamente, el procedimiento necesita aumentar gradualmente el tamaño de las ventanas hasta alcanzar algún criterio de convergencia útil. La búsqueda de este criterio no es algo trivial, y para intentar responder la pregunta hay que recordar que el análisis por objeto simbólico está basado en las

propiedades estadísticas de las series de entrada, en particular en los histogramas y/o intervalos interquartiles, por lo que lo ideal es que las series de parámetros $AR(p)$ muestren un comportamiento adecuado en dichas propiedades. Uno de los criterios que se usan en esta sección corresponde a una extensión del método RMS usado en el capítulo 3 para determinar el orden óptimo de los modelos de autoregresión, con la diferencia de que en esta aplicación se deja fijo el orden, mientras que en cada iteración se aumenta el tamaño de las ventanas.

Considerar que una de las señales de aceleración tiene N datos. El método ajusta un modelo autoregresivo de orden $p$utilizando una ventana ubicada al inicio con $n = kp$ datos. Notar que el cálculo de los coeficientes no se realiza con el total de datos sino con una fracción de estos. Luego, se predicen los estados de aceleración para todo el largo de la señal original a partir del modelo ajustado y se calculan los errores residuales. Finalmente se obtiene el RMS considerando estos $(N - p)$ errores, de los cuales $(n - p)$ son de ajuste y $(N - n)$ son residuales. Este procedimiento se aplica iterativamente aumentando el valor de k, lo que se analiza utilizando un gráfico. Lo que se espera es lo siguiente: en cada iteración se generan errores de ajuste y residuales, pero a medida que se aumenta el valor de k, aumenta el número de errores de ajuste y disminuye el de errores residuales. Como se mostró en el capítulo anterior, los errores de ajuste tienen una menor amplitud que los errores residuales, por lo que es esperable que el procedimiento planteado presente algún tipo de convergencia, cómo se ejemplificará a continuación.

A modo de recuerdo, los ensayos que se consideran son series de aceleración adquiridas para 13 condiciones estructurales, 3 distintos registros de excitación de ruido blanco, y cada señal tiene aproximadamente 7000 datos luego del proceso de decimado. Se aplicó el procedimiento de convergencia para 4 condiciones estructurales y para los 3 casos de excitación de ruido, esperando que de existir convergencia ésta sea independiente de los casos de excitación y de la condición estructural. Se estudiaron largos de ventana que van desde $2p$ hasta $300p$. Los resultados se muestran en la Figura 4.1, los que son bastante esclarecedores al respecto. En dicha figura, a pesar de que no se alcanza un mínimo, sí se logra apreciar una tendencia asíntota a un valor de convergencia que no depende ni del caso estructural ni de la excitación. Esta tendencia no debe ser considerada como haber encontrado un óptimo, sino más bien como que se llega a un punto en el que un mayor largo de ventana no aporta información relevante, puesto que los modelos ya ajustan suficientemente bien los datos originales. La selección del largo no tiene un criterio claro y si bien se podría intentar encontrar un valor codo mediante un ajuste bilineal del RMS, en esta sección se escoge un orden localizado en la parte estable del RMS. Esto ocurre alrededor de un largo de ventana igual a 100 veces el orden del modelo, lo que considerando una frecuencia de muestreo de 25Hz, y el orden igual a 15, resulta en un largo adecuado de 60 segundos, que era justamente lo utilizado en el capítulo 3, y uno de los resultados secundarios al momento de realizar el estudio de índices de validación de modelos autoregresivos. En adición a lo anterior, la Figura 4.2. arroja más información ya que es posible ver diferencias en el valor RMS para distintas condiciones estructurales. Esto tiene que ver con la capacidad de los modelos autoregresivos de ajustar series no lineales (con daño), y llama mucho la atención de que en este caso se ajuste de mejor forma la serie dañada que la no dañada.

Si bien no forma parte de los objetivos de esta sección, en las Figuras N°4.2, 4.3 y 4.4 se estudia el uso de RMS como característica sensible al daño.

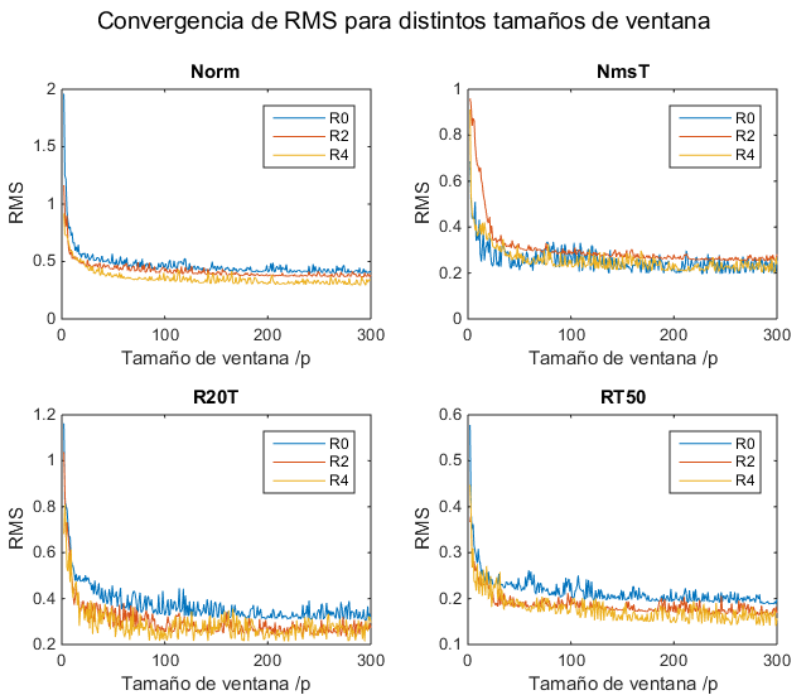


Figura N°4.1. Convergencia de RMS para distintos tamaños de ventana.

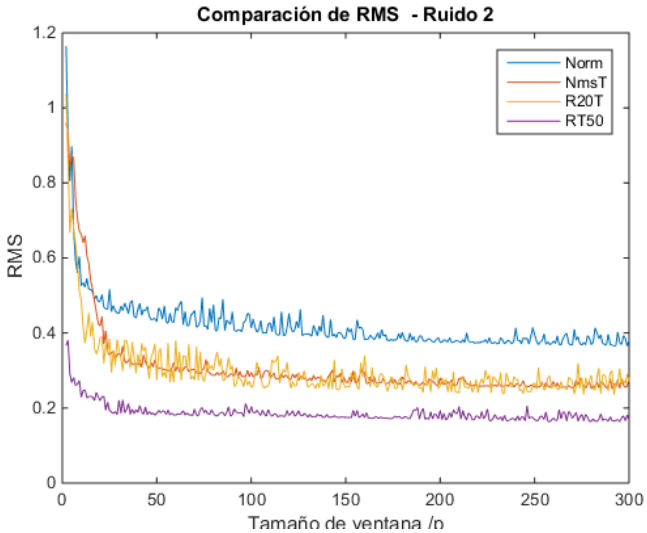


Figura N°4.2. Comparación de RMS para distintas condiciones estructurales. Ruido 2.

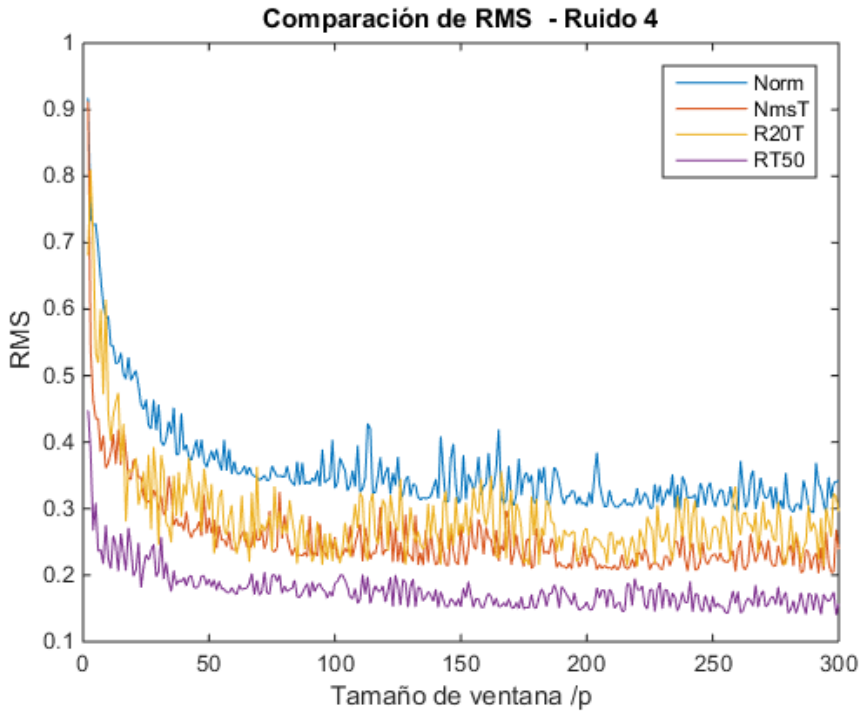


Figura N°4.3. Comparación de RMS para distintas condiciones estructurales. Ruido 4.

La Figura 4.2 compara los resultados obtenidos de RMS para distintas condiciones estructurales pero compartiendo la misma excitación de Ruido 2; la Figura 4.3, en tanto, considera la excitación Ruido 4. En ambas figuras se puede ver una clara distinción entre las distintas condiciones estructurales Normal y RT50, mientras que los estados de daño leve, Normal más masa y Reducción 20%, parecen tener el mismo comportamiento. Uno de los hechos más llamativos, es que RMS parece ser coherente en el sentido de que a mayor cambio estructural, se aprecia una mayor diferencia entre los valores obtenidos.

Por último, para cerrar el estudio de la convergencia de RMS y su sorpresivo comportamiento como característica sensible a los cambios estructurales, en la Figura 4.4 se comparan los resultados para distintas excitaciones y dos estados estructurales diferentes. Nuevamente, se obtuvo que los casos de daño muestran una clara separación incluso para excitaciones distintas. Es decir, RMS puede ser una 'feature' robusta en términos de distintas condiciones operacionales.

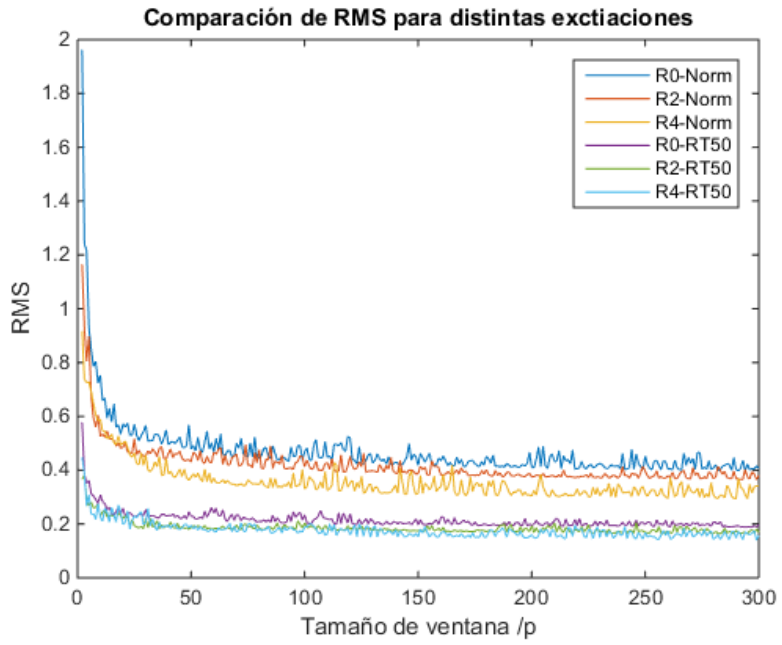


Figura N°4.4. Comparación de RMS. Distintas excitaciones.

Hasta el momento, podemos concluir que el largo de ventana a utilizar debe tener una cantidad de datos de alrededor de 100 veces el orden de los modelos autoregresivos ajustados para conseguir un valor de RMS estable. Sin embargo no hemos estudiado en profundidad el comportamiento de los parámetros o coeficientes del modelo AR frente a distintos largos de ventana. Esto se realizará mediante el estudio del promedio, desviación estándar e histograma de los coeficientes obtenidos para cada ventana, en iteraciones sucesivas con un largo de ventanas creciente, según el algoritmo siguiente:

1. Ajustar modelo AR(p) para una ventana de largo $n = kp$
2. Mover ventana, recorriendo la señal original para tener las p series de coeficientes AR(p).
3. Calcular propiedades estadísticas: promedio, desviación estándar e histograma.
4. Aumentar el largo de ventana e iterar.

Los resultados en las Figuras 4.5 y 4.6 consideran la excitación de Ruido 0 y la condición estructural Normal, y si bien no son muy alentadores ya que se ve mucha variabilidad en algunos parámetros y una alta dependencia en relación al largo de la ventana, la mayoría muestra un comportamiento estable en torno a un valor. Es particular, desde los coeficientes N°3 al N°8, el promedio depende casi linealmente con el largo de las ventanas. Notar que esto no es necesariamente algo negativo ya que la metodología real considera un largo de ventanas fijo lo que hace que los resultados sean coherentes (comparamos resultados obtenidos siempre con el mismo largo de ventanas). Sin embargo, lo que se puede apreciar en la Figura 4.5 es que el cálculo del modelo AR(p) no identifica fielmente las propiedades del sistema sino que simplemente es un ajuste a las señales de aceleración, ya que de ser así hubiésemos tenido un comportamiento mucho más estable o al menos haber mostrado convergencia. Para apoyar lo anterior vemos en más detalle la Figura 4.6, que muestra la desviación estándar de los coeficientes en función del largo de la ventana, de 10 a 300 veces el orden de los modelos. En dicha figura, apreciamos que la variabilidad es crítica en los coeficientes desde el N°3 al N°12

inclusive, los que muestran un crecimiento lineal a medida que aumentan los largos de las ventanas. Por otro lado, en estas Figuras, 4.5 y 4.6, se añaden los resultados de las propiedades estadísticas de las series luego de aplicar un filtro de media móvil con 5 coeficientes, lo que es una forma de suavizar los resultados y disminuir el efecto de coeficientes que se encuentren muy lejos del promedio. En cuanto al promedio, no se puede apreciar una diferencia notable, pero en relación a la desviación estándar, la variabilidad obtenida es mucho menor luego de aplicar el filtro.

Los resultados de la metodología para histogramas por razones de espacio no se pueden mostrar de la misma forma que las figuras anteriores, por lo que decidió mostrar los resultados solo para los coeficientes N°1, N°7 y N°15, por encontrarse al inicio, mitad y final del conjunto de parámetros. De forma análoga a lo anterior, se grafican resultados con y sin filtro media móvil de 5 coeficientes, desde la Figura 4.7 hasta la 4.12.

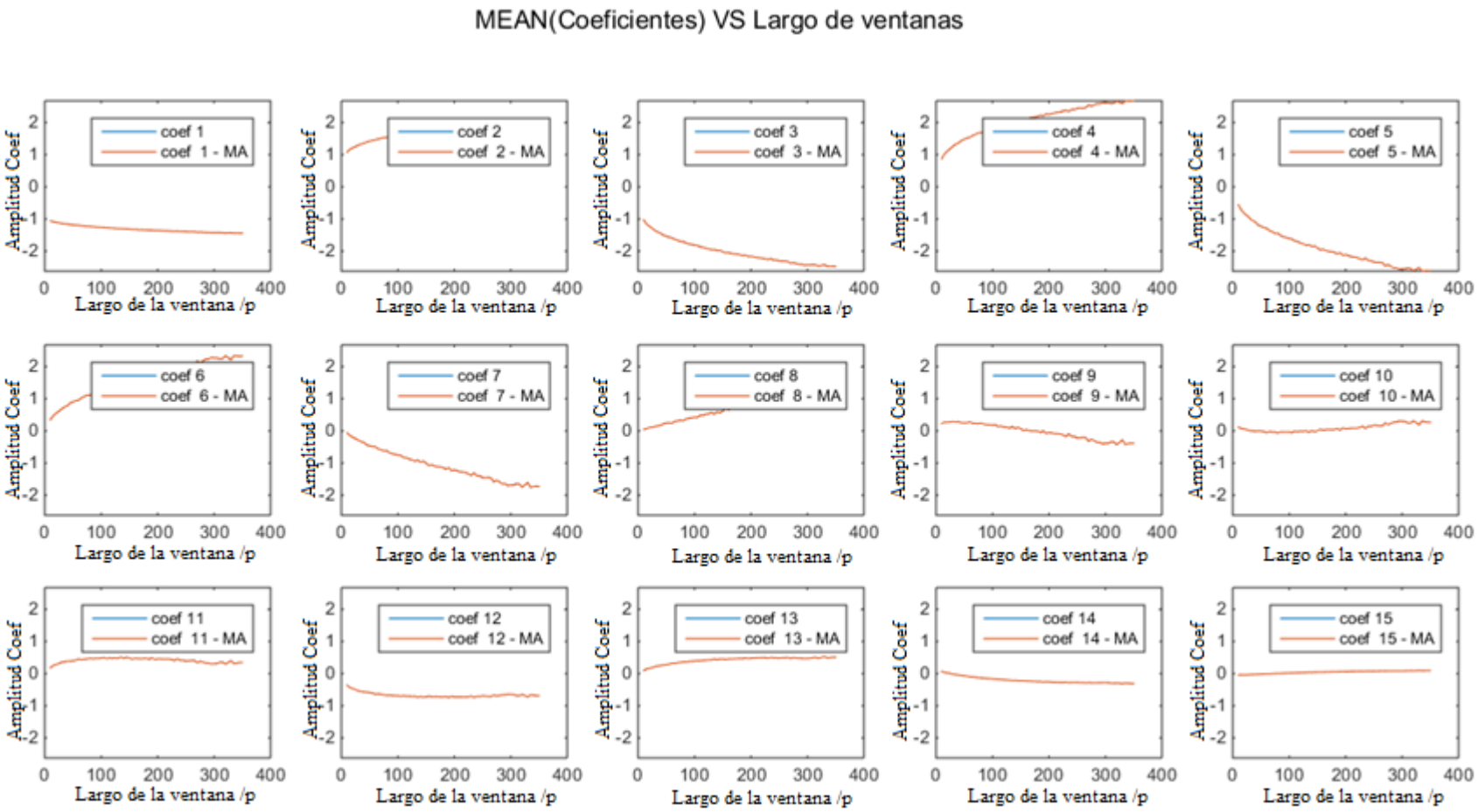


Figura N°4.5. Promedio de coeficientes autoregresivos en función del largo de ventana.

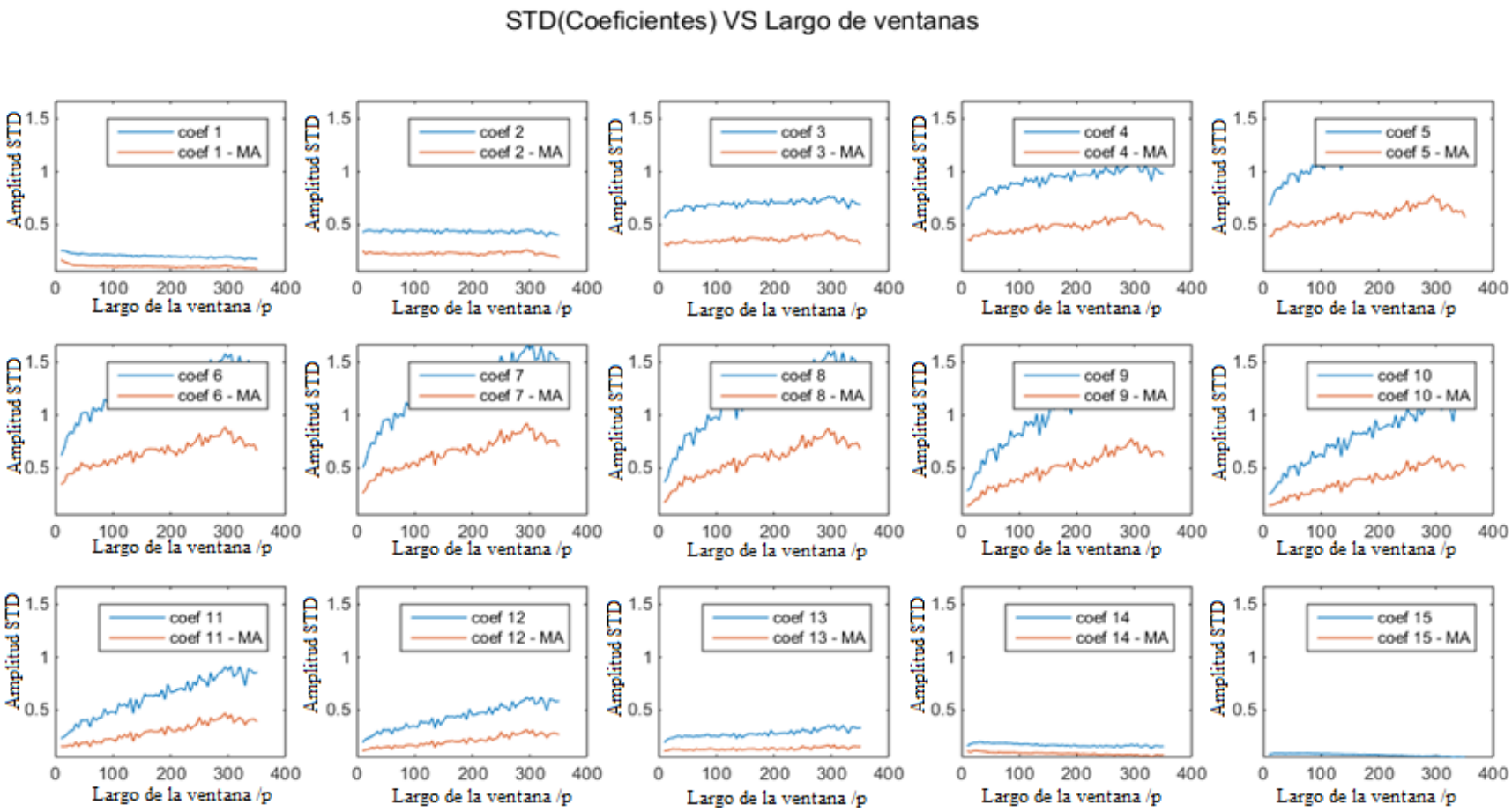


Figura N°4.6 Desviación Estándar de coeficientes autoregresivos en función del largo de ventanas.

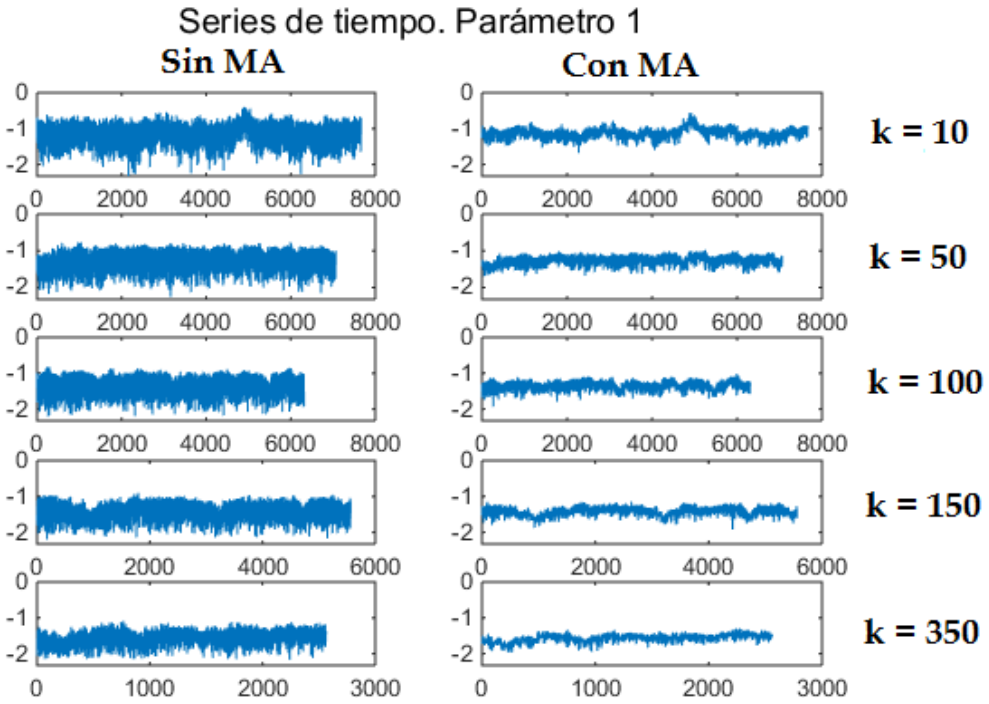


Figura N°4.7. Series de tiempo. Parámetro 1.

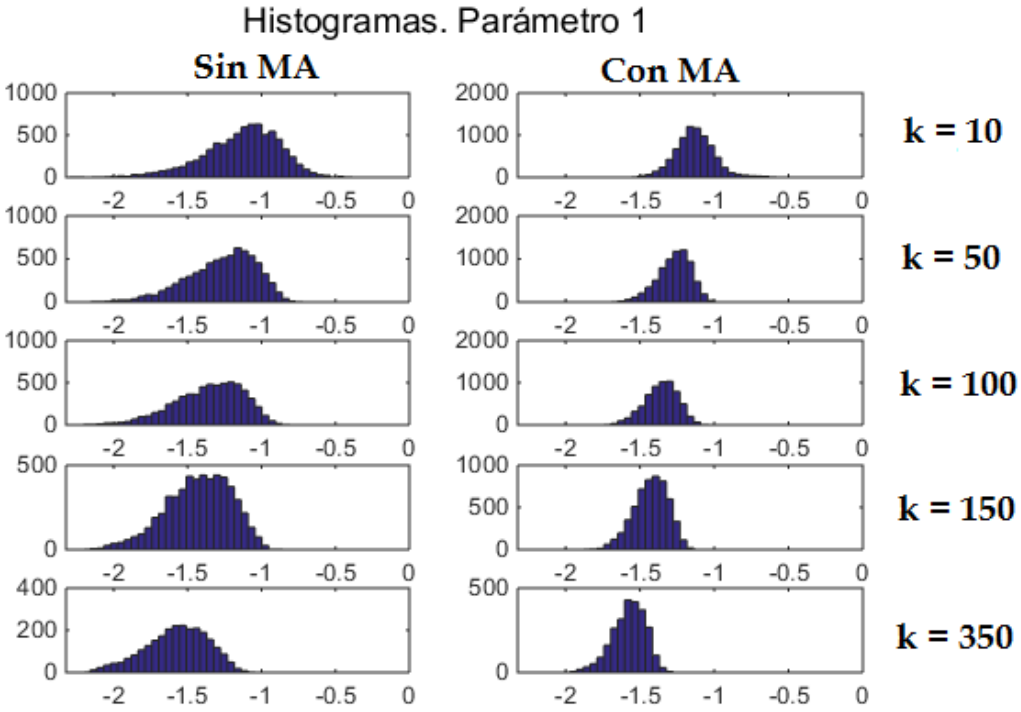


Figura N°4.8. Histogramas. Parámetro 1.

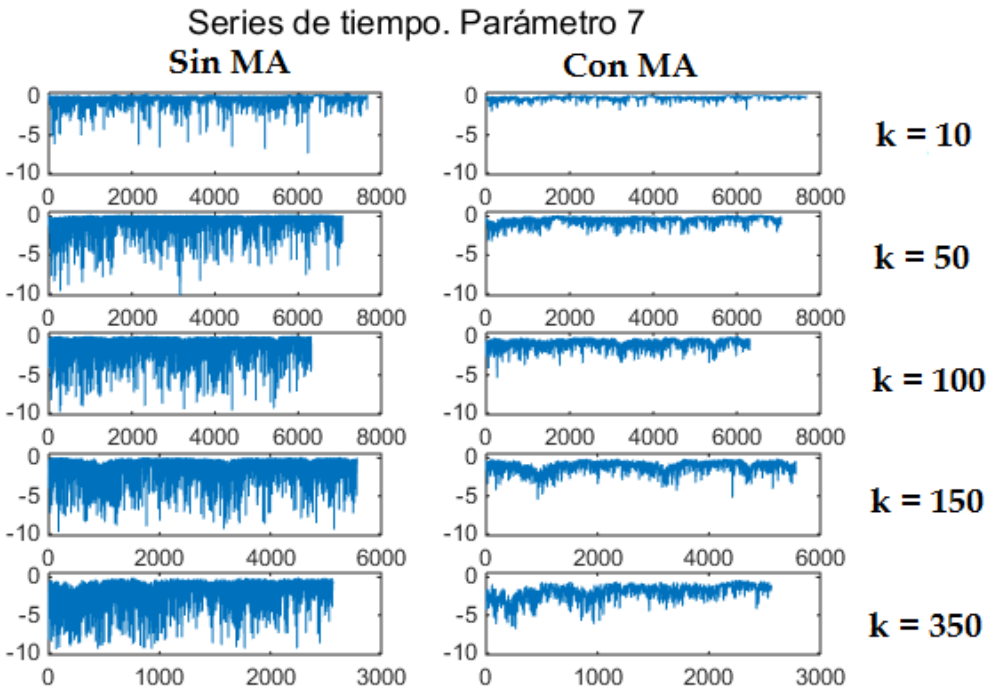


Figura N°4.9. Series de tiempo. Parámetro 7.

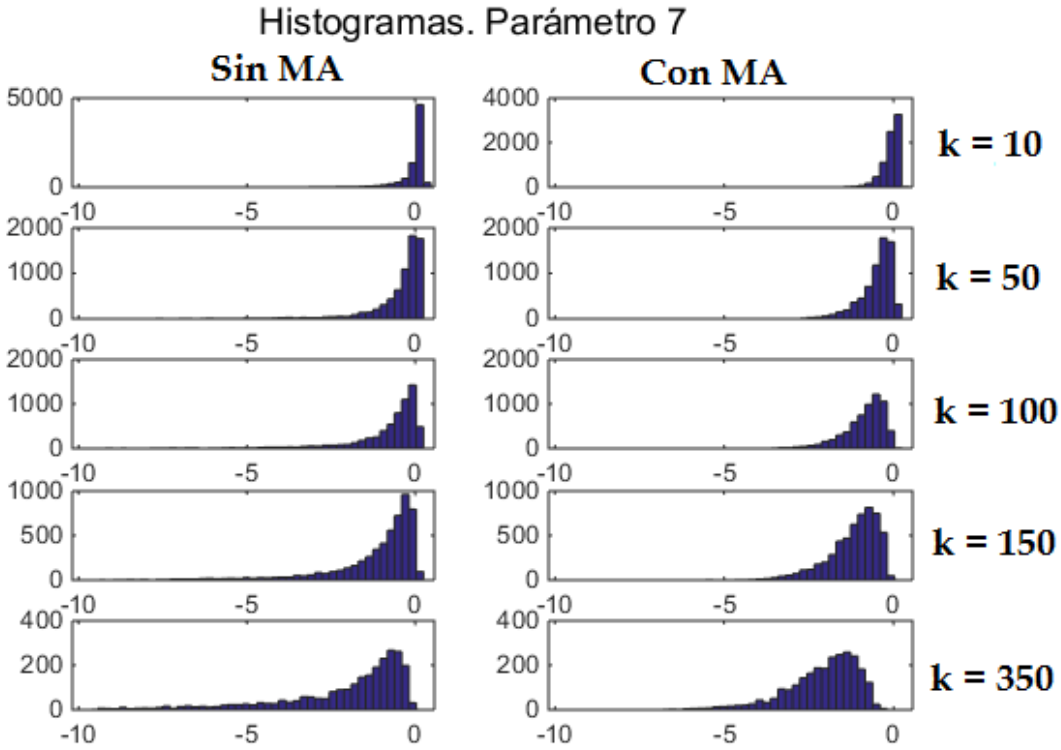


Figura N°4.10. Histogramas. Parámetro 7.

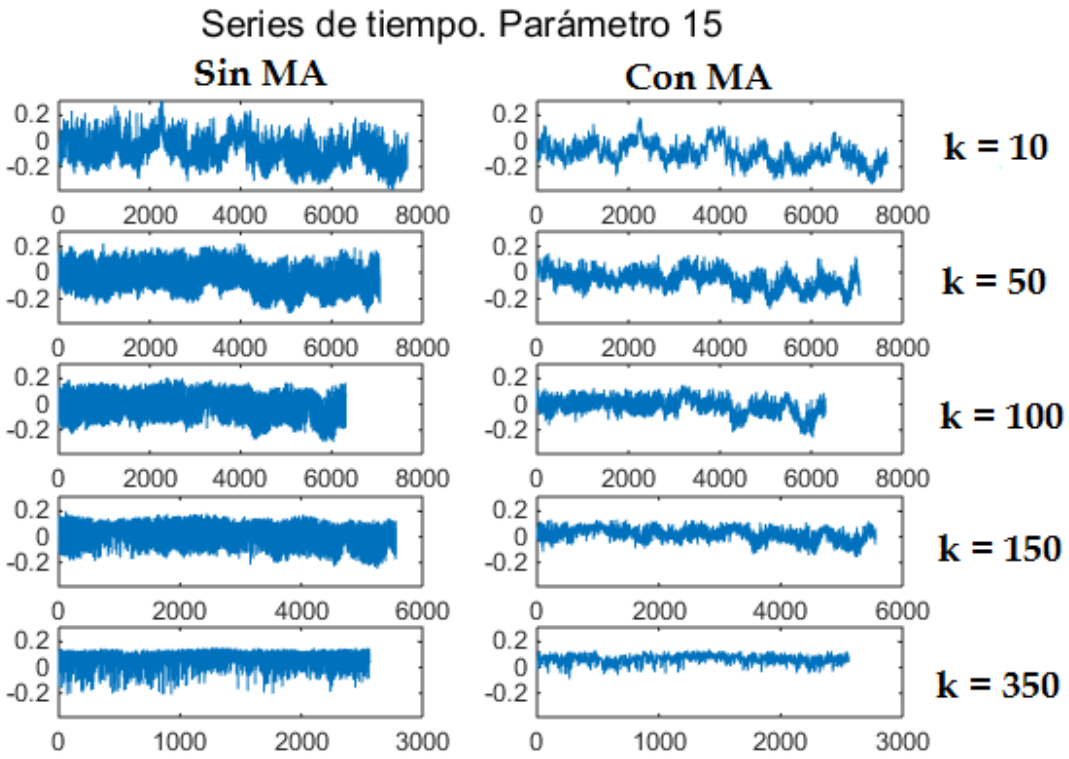


Figura N°4.11. Series de tiempo. Parámetro 15.

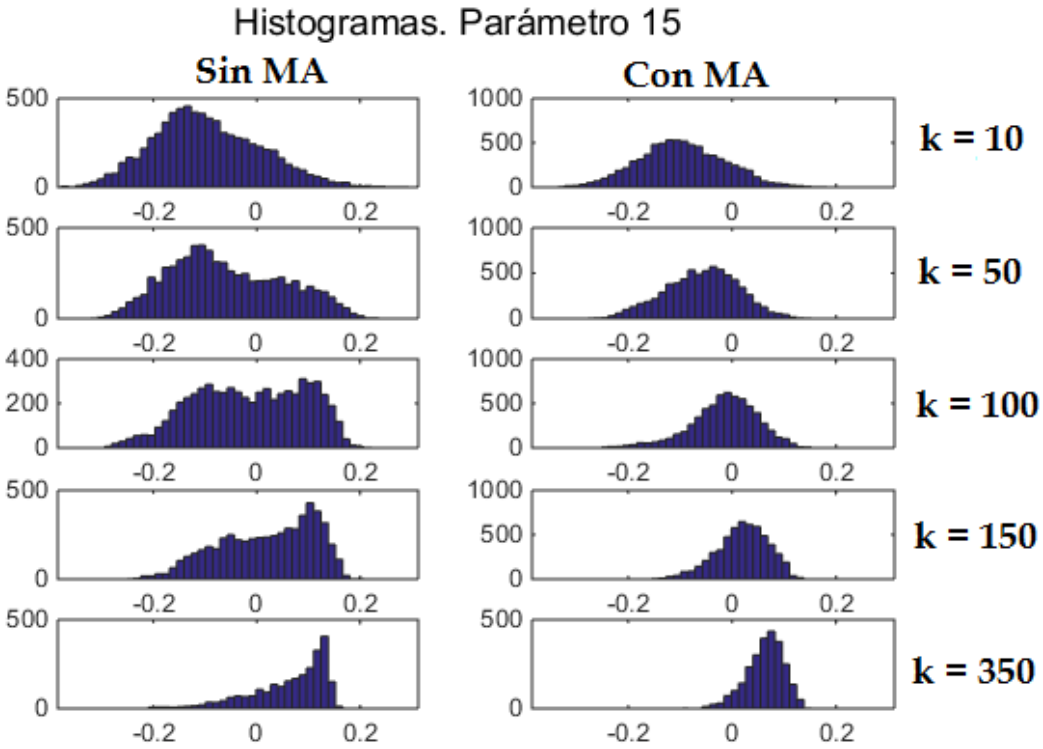


Figura N°4.12. Histogramas. Parámetro 15.

En las figuras anteriores, se pueden observar varios resultados interesantes. Por una parte, a partir de las series de tiempo se concluye que la amplitud de los coeficientes no sigue un

comportamiento coherente en relación al largo de las ventanas, ya que el coeficiente N°7 aumenta su valor a medida que las ventanas crecen. Este comportamiento ya había sido observado en los gráficos de promedio y desviación estándar. Además, con el filtro moving average se obtienen valores mucho más estables y parecen ser más adecuados para la metodología que se utilizará para la generación de objetos simbólicos. No obstante, se aprecia dependencia de la amplitud en el tiempo, lo que generará objetos outliers si se utilizan ventanas de tiempo pequeñas para la generación de estos. A su vez, a partir de los gráficos de histograma se observa la tendencia creciente o decreciente de los valores (peaks no son estables), mismo resultado que se había notado en los gráficos de promedio de parámetros, y además el filtro media móvil genera series con un histograma con una forma mucho más parecida a la campana de Gauss.

En resumen, el estudio del comportamiento de los parámetros AR(p) con respecto al largo de ventana a utilizar arrojó resultados con gran variabilidad. Los coeficientes no parecen tener una marcada convergencia hacia algún valor, pero usados en conjunto sí se aprecia una convergencia en términos del RMS. Además, la alta variabilidad mostrada se puede reducir considerablemente utilizando filtros de media móvil. Para la metodología siguiente, en la que se transformarán estas series de coeficientes AR(p) en objetos simbólicos, se usarán ventanas de tiempo de un largo igual a 100 veces el orden de los modelos ajustados, y se aplicará un filtro media móvil de 5 coeficientes.

4.2.2. Creación de objetos simbólicos a partir de parámetros AR(p).

Al igual que en la sección anterior, mucha de la teoría que se utiliza en ésta ya fue explicada en el capítulo correspondiente al estudio de la clasificación usando objetos simbólicos, por lo que no se hará una explicación exhaustiva de los términos ni ecuaciones. No obstante, se detallarán los métodos usados para mantener el capítulo auto contenido.

Considerar que se tiene un conjunto de (ps) series de parámetros $AR(p)$, obtenidos a partir de la transformación de los datos adquiridos por s sensores. En esta sección, veremos cómo transformar estas series en objetos simbólicos, para lo que utilizaremos dos procedimientos: por histogramas y por intervalos interquartiles.

Transformación a objetos simbólicos de histogramas.

Uno de los requerimientos para la transformación a objetos simbólicos usando histogramas, es que las categorías de estos sean iguales para todas las mediciones. Realmente, lo que se necesita es establecer un rango de valores y un número de categorías para cada serie analizada, no para todo el set de datos. No obstante, generar esta diferencia hace que el proceso sea más complejo numéricamente por lo que en este caso solo se utilizará el mismo rango y el mismo número de categorías para cada serie. Formulando lo anterior matemáticamente, considerar (p*s) series de parámetros AR(p), cada una con un total de U datos. Esta series pueden estar condensando, por ejemplo, una adquisición de 15 minutos en algún edificio instrumentado, o un ensayo de laboratorio de 5 minutos. El objeto simbólico T_i que representa estas series queda definido como indica la ecuación 4.4. En dicha ecuación la notación (r) hace referencia a la r −ésima serie de parámetros $AR(p)$.

$$T_i = [T_i^{(1)} \dots T_i^{(r)} \dots T_i^{(ps)}] \tag{4.4}$$

En el caso de un histograma, los $T_i^{(r)}$ tienen la forma de la ecuación 4.5

$$T_i^{(r)} = (\mathrm{P}_{\mathrm{ik}}^{(\mathrm{r})}; k = 1, \dots, \omega) \tag{4.5}$$

donde $\mathrm{P}_{\mathrm{ik}}^{(\mathrm{r})}$ es la frecuencia relativa del intervalo k −ésimo, de un total de ω intervalos utilizados. Notar que en este caso, hay un total de (ps) histogramas, por lo tanto, cada T_i tiene una dimensión igual a $(p \cdot s \cdot \omega)$. En este punto cabe recalcar que la definición de objetos usando histogramas no toma en consideración los límites de las categorías utilizadas, ya que todos los objetos utilizan exactamente las mismas categorías y además, la definición de distancia entre objetos simbólicos de histograma no considera el ancho ni los extremos de los intervalos.

Transformación a objetos simbólicos usando intervalos intercuartil.

La ecuación 4.4 que define la notación de un objeto simbólico genérico también es válida cuando se trata de objetos tipo intercuartil. En este caso, como estamos usando objetos representados por intervalos, los $T_i^{(r)}$ toman la forma

$$T_i^{(r)} = \left(T_{i,inf}^{(r)}, T_{i,sup}^{(r)}\right) \tag{4.6}$$

donde el subíndice 'inf' indica el límite inferior del intervalo, mientras que el subíndice 'sup' indica el límite superior de éste. Recordemos que para intervalos interquartile, $T_{i,inf}^{(r)}$ es el valor que se encuentra sobre el 25% de todos los datos de la serie r −ésima de parámetros, mientras que $T_{i,sup}^{(r)}$ es el valor que se encuentra sobre el 75%.

4.2.3. Distancias utilizadas.

Objetos interquartil.

Considerar un set de N objetos simbólicos $\{T_i, \dots, T_N\}$. Cada uno de estos objetos es descrito por ps intervalos intercuartiles $T_i^{(r)} = \left(T_{i,inf}^{(r)}, T_{i,sup}^{(r)}\right)$, obtenidos de ps series de parámetros $AR(p)$. Para calcular las distancias entre dos objetos T_i y T_j, utilizaremos la distancia Haussdorf Euclideana Estandarizada, que se detalla en la ecuación 4.7

$$d_{ij} = \left(\sum_{r=1}^{p}\left[\frac{\phi_{\mathrm{r}}(\mathrm{T_i}, \mathrm{T_j})}{\mathrm{H_r}}\right]^2\right)^{\frac{1}{2}} \tag{4.7}$$

donde

$$\phi_{\mathrm{r}}(\mathrm{T_i}, \mathrm{T_j}) = \max\left(\left|T_{i,inf}^{(r)} - T_{j,inf}^{(r)}\right|, \left|T_{i,sup}^{(r)} - T_{j,sup}^{(r)}\right|\right) \tag{4.8}$$

es definida como la medida de disimilitud de Haussdorf y H_r^2 es el término de estandarización calculado como

$$H_r^2 = \frac{1}{2N}\sum_{i=1}^{N}\sum_{j=1}^{N}\left[\phi_r(T_i, T_j)\right]^2 \tag{4.9}$$

Objetos Histograma.

Considerar un set de N objetos simbólicos $\{T_i, \dots, T_N\}$. El objeto T_i puede ser descrito como la frecuencia relativa de los contenedores o intervalos utilizados. La caracterización matemática de T_i se expresa en la ecuación 4.10

$$T_i = \{(P_{111}, \dots, P_{11\omega}), \dots, (\dots, P_{\phi lo}, \dots), \dots, (P_{1s1}, \dots, P_{ps\omega})\} \tag{4.10}$$

donde $P_{\phi lo}$ es la frecuencia relativa del intervalo o −ésimo asociado a la serie del parámetro ϕ del sensor l-ésimo, de un total de ω intervalos para cada una de las (ps) series. Para calcular la distancia usamos la definición de distancia categórica, descrita por (Billard & Diday 2006).

$$d_{ij}^2 = \frac{1}{ps} \sum_{a=1}^{ps} \sum_{k_a=1}^{\omega} \left(\sum_{n=1}^{N} P_{nk_a a} \right)^{-1} \left(P_{ik_a a} - P_{jk_a a} \right)^2 \tag{4.11}$$

4.2.4. Agrupamiento.

Para ambos tipos de objetos se utilizaron los métodos K-means usando distancia euclideana, y el método jerárquico aglomerativo usando la distancia de Haussdorf-Euclideana en el caso de objetos intercuartil, y distancia categórica en el caso de histogramas. La selección de la partición óptima se hace mediante el criterio de Calisnki Haranbasz. Recordar que estas metodologías ya fueron estudiadas en el capítulo 2, obteniendo que el índice Calisnki-Harabasz es el que entrega mejores resultados para este estudio y además es el más sencillo en su formulación matemática y al momento de su programación. Por otra parte, en el caso de series de aceleración se obtuvo que tanto Dynamic Cloud como K-means arrojaron resultados similares al momento de realizar agrupamiento de objetos obtenidos de histogramas, en contraste con el algoritmo Jerárquico que no obtuvo tan buenos resultados. Sin embargo, al tratarse de objetos de otra naturaleza, en el presente capítulo se vuelve a analizar la distancia intercuartil para verificar los resultados del capítulo 2.

4.3. Análisis de sensibilidad.

En esta sección se estudia el comportamiento de la metodología propuesta frente a decisiones en lo que respecta al uso de distintos algoritmos para agrupar y creación de objetos simbólicos, y además, se busca determinar cuáles son los parámetros de modelos autoregresivos que muestran una mayor sensibilidad a los cambios estructurales. Por otra parte, se analizan los resultados obtenidos al seleccionar sólo algunos de los sensores disponibles.

4.3.1. Metodología de análisis.

En esta sección se analizan los resultados obtenidos al estudiar los 13 casos estructurales del ensayo de laboratorio ya presentado en capítulos y secciones anteriores.

En términos globales, la metodología entrega una solución óptima mediante la selección del número de grupos que maximiza el índice de Calinski Harabasz y la partición asociada a esa cantidad de conjuntos. Para ello, los algoritmos de agrupamiento utilizan una matriz de distancias entre todos los objetos simbólicos a partir de la cuál generan las particiones cuyas características fueron mencionadas en el capítulo 2. Los objetos simbólicos, por su parte, son creados usando la metodología descrita en la sección anterior. Sin embargo, cabe destacar que para poder contar con varios objetos simbólicos a los que aplicar los algoritmos de agrupamiento, es necesario dividir los ensayos en ventanas no solapadas. Esto último aumentará la influencia de fenómenos instantáneos o de corta duración lo cual no es deseable ya que se pierde robustez pero es la única forma de poder proceder con los ensayos disponibles. Las soluciones encontradas son evaluadas dependiendo si separan correctamente los dos estados estructurales comparados, lo que equivale a decir que entregan un número óptimo de grupos igual a 2 (considerando que comparan solo dos condiciones por vez), y además, que los objetos asociados a las ventanas creadas desde una misma condición estructural sean asignados al mismo clúster. Esta es una condición objetiva que permitirá comparar los resultados sin necesidad de observar las matrices de distancias.

El procedimiento general es como sigue:

Primero, se crearán objetos simbólicos a partir de ventanas de las series de parámetros AR. La idea es comparar los objetos creados desde el ensayo en condición normal, con los otros 12 casos. Esto genera 13 gráficos de distancias entre objetos simbólicos (incluyendo la comparación del caso base consigo mismo). Un ejemplo de estos gráficos se muestra en la Figura 4.13, calculado para ensayos de Ruido0, en la que se generaron 3 ventanas para cada serie de parámetros AR mientras que se consideraron todos los parámetros pero sólo el sensor N°3 (mirar el título superior del gráfico). Fijarse que el gráfico muestra matrices de 6x6, siendo los primeros 3 objetos pertenecientes a la serie de condición base, mientras que los últimos 3 pertenecen a la serie con la cual se está realizando la comparación, lo que genera el título de cada matriz. La cantidad de combinaciones posibles hace que sea inadecuado mostrar todos los gráficos de distancias generados; además, estos gráficos sólo son un insumo para el siguiente paso, que es el de calcular la partición óptima para distintos números de grupos. Esta etapa es realizada utilizando objetos basados en histograma.

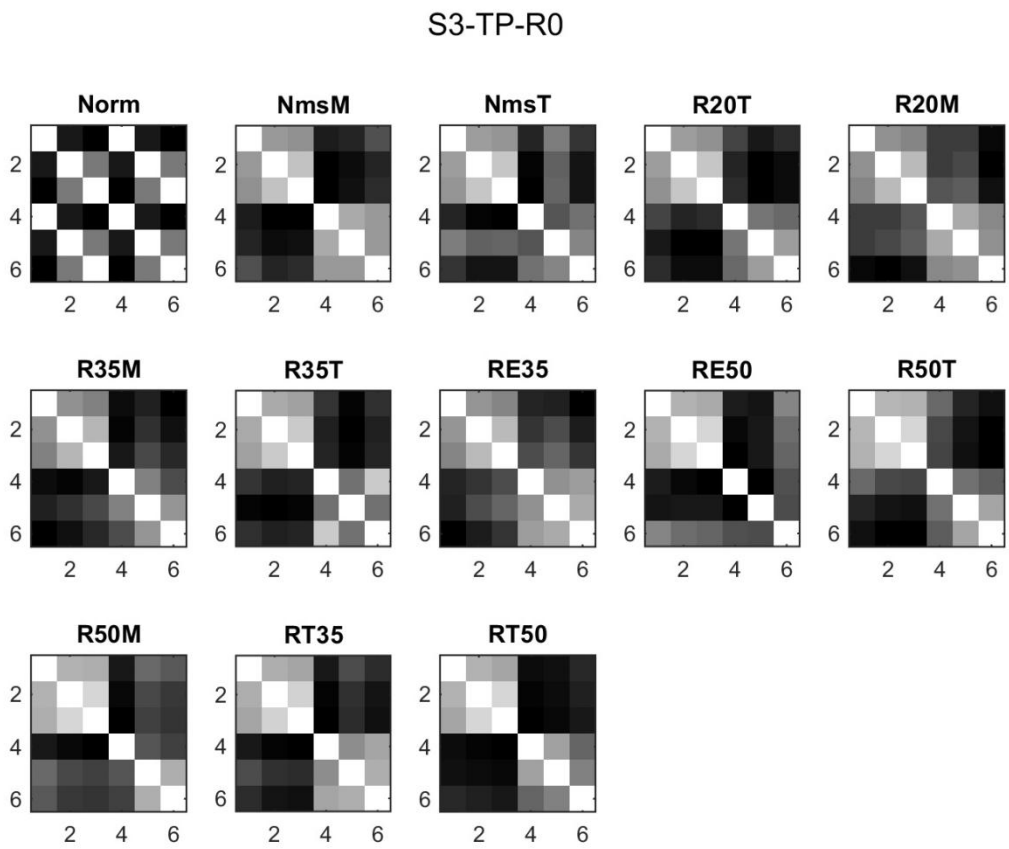


Figura N°4.13 Ejemplo de gráficos de distancias.

El segundo paso, consiste en que partiendo de las matrices de distancias, se utilizan algoritmos de agrupamiento para calcular la partición óptima para distintos números de grupos. El output es un arreglo de índices que asigna cada ventana a un clúster en particular. Luego, se evalúa el índice CH para determinar el número de grupos presentes en el set de datos. Esto queda ejemplificado en la Figura 4.14, obtenida para un análisis realizado considerando solo los tres primeros parámetros y sólo los tres primeros sensores. Recordemos que éstos gráficos son de valor de CH v/s número de grupos, y que el número de grupos óptimo está asociado con el valor máximo de CH. El primer gráfico se puede ignorar ya que es una comparación de una serie consigo misma, lo que provoca la aparición de grupos idénticos, haciendo que el cálculo del índice CH sea imposible de realizar. Esto es debido a que la distancia "within-cluster" es igual a cero en este caso, lo que ocurre también cuando cada objeto es asignado a un clúster diferente, por lo que el número de grupos igual a 6 tampoco se muestra en los gráficos. En esta etapa, como estamos comparando solo dos estados estructurales, esperamos que el resultado indique un número de grupos presentes igual a 2; por supuesto, esto es el caso ideal y como ya podemos observar de la Figura 4.14, no siempre obtendremos este resultado. Nuevamente, el número de gráficos asociados a las combinaciones posibles es demasiado grande como para poder mostrarlos todos, y además tampoco es de utilidad pues lo que realmente consideraremos como output es lo del siguiente paso.

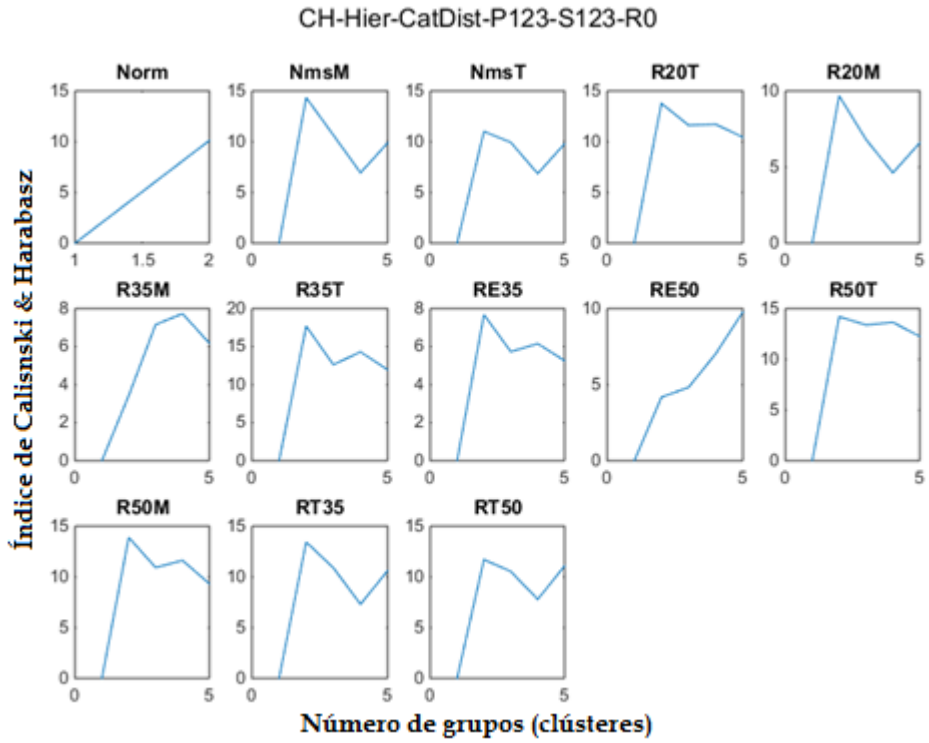


Figura N°4.14. Ejemplo de índices de validación de grupos.

Finalmente, considerando el número de grupos presentes indicado por el índice CH para cada comparación, recuperamos la partición asociada a dicho número. Sabiendo que en el caso de estos ensayos los tres primeros objetos corresponden a una condición normal y los tres objetos finales están asociados a una condición con daño aplicado, esperamos que la partición óptima sea aquella que asigna los tres primeros objetos a un clúster, y los otros tres objetos a otro clúster. Con esto en mente, calculamos el porcentaje de objetos correctamente asignados para cada combinación, y por último evaluamos la performance global. Un ejemplo se muestra en la Tabla N°4.1, calculada para la misma combinación de parámetros y sensores de la Figura 4.14. Notamos que en este caso, el algoritmo tuvo una muy buena performance global.

	K-Means. Histograma. P123-S123-R0											
Condición	NmsM	NmsT	R20T	R20M	R35M	R35T	RE35	RE50	R50T	R50M	RT35	RT50
Objeto1	1	2	1	2	3	2	2	5	2	1	1	1
Objeto2	1	2	1	2	1	2	2	1	2	1	1	1
Objeto3	1	2	1	2	1	2	2	1	2	1	1	1
Objeto4	2	1	2	1	4	1	1	2	1	2	2	2
Objeto5	2	1	2	1	2	1	1	3	1	2	2	2
Objeto6	2	1	2	1	2	1	1	4	1	2	2	2
(%)	100	100	100	100	67	100	100	50	100	100	100	100
Overall	93%											

Tabla N°4.1. Ejemplo de resultados de partición óptima para una combinación en particular.

4.3.2. Resultados Análisis de Sensibilidad.

A lo largo de este apartado se mostrarán y analizarán los resultados obtenidos para el análisis de sensibilidad de la metodología. Principalmente, la información será resumida en forma de tablas que contienen el porcentaje de asignaciones correctas para cada combinación. De esta forma, podremos analizar el comportamiento de cada uno de los sensores y de cada uno de los parámetros autoregresivos frente a cada una de las condiciones estructurales.

Sensibilidad Sensores.

Las Tablas 4.2. y 4.3 contienen los resultados obtenidos al aplicar la metodología integrando al análisis la totalidad de parámetros autoregresivos, pero un solo sensor a la vez. Esto permite estudiar el comportamiento por separada de cada uno de los sensores. En particular, las dos tablas mencionadas resumen la información para todos los casos estructurales de la aceleración basal llamada Ruido0, cuyas propiedades globales ya fueron estudiadas en el capítulo 2.

En análisis de los resultados debe partir dejando bien en claro cuál era el resultado esperado o la forma en que se evaluará la performance del procedimiento propuesto. En este sentido, se definió como output de la metodología una partición de asignación óptima, y esta se evalúa calculando el porcentaje de asignaciones correctas. Una asignación del 100% quiere decir que los tres primeros objetos, calculados a partir de la primera señal, fueron asignados a un mismo clúster, mientras que los otros tres objetos fueron asignados a un segundo clúster. Esto es el caso de éxito total, pero no nos sirve mucho para evaluar la performance negativa. La pregunta que hay que responder tiene relación sobre cuál es el porcentaje con el que hay que ser categórico y decir que, ya sea producto de la metodología, los sensores o parámetros, no hay un aporte de información al respecto de las condiciones estructurales presentes. Este porcentaje depende de cuántas condiciones estructurales se estén analizando, y en este caso, cada análisis consta de solo dos de ellas, por lo que en un caso completamente aleatorio, esperamos un 50% de asignaciones correctas. En el caso límite, la peor performance posible para este estudio es de un 33%, lo que ocurre cuando cada objeto es asignado a un clúster distinto. Incluso en tal caso, 2 objetos de los 6 totales habrían sido asignados correctamente. De esta forma, solo los resultados que estén sobre el 50% arrojan información aceptable sobre la presencia de distintos grupos y por tanto, distintas condiciones estructurales.

Ensayos: Ruido 0

K-Means. Euclidean Distance. Todos los Parametros. **R0.**

Sens	N+M	N+T	R20T	R20M	R35M	R35T	RE35	RE50	R50T	R50M	RT35	RT50	**Prom**
S1	100%	100%	100%	100%	67%	100%	33%	50%	50%	100%	50%	83%	**78%**
S2	100%	100%	83%	67%	67%	100%	33%	67%	67%	100%	50%	33%	**72%**
S3	100%	50%	100%	100%	100%	50%	100%	50%	50%	50%	100%	33%	**74%**
S4	100%	17%	83%	67%	67%	100%	50%	50%	67%	100%	100%	100%	**75%**
S5	100%	100%	83%	100%	67%	100%	67%	33%	83%	100%	100%	100%	**86%**
S6	50%	100%	100%	33%	33%	100%	33%	100%	100%	100%	100%	100%	**79%**
S7	67%	100%	100%	67%	67%	100%	33%	50%	100%	100%	100%	100%	**82%**
S8	100%	100%	100%	33%	67%	100%	67%	100%	100%	100%	100%	50%	**85%**
Prom	**90%**	**83%**	**94%**	**71%**	**67%**	**94%**	**52%**	**63%**	**77%**	**94%**	**88%**	**75%**	**79%**
	84%				69%				83%				

Tabla 4.2. Resultados Sensibilidad. K-Means. Euclidean Distance. Todos los Parámetros. R0.

Hierarchical Agglomerative. Categorical Distance. Todos los Parámetros. R0													
Sens	N+M	NmsT	R20T	R20M	R35M	R35T	RE35	RE50	R50T	R50M	RT35	RT50	**Global**
S1	100%	83%	83%	100%	67%	100%	33%	83%	67%	100%	50%	50%	**76%**
S2	100%	83%	83%	50%	67%	100%	33%	67%	67%	100%	50%	50%	**71%**
S3	100%	50%	100%	100%	100%	50%	100%	50%	50%	33%	100%	50%	**74%**
S4	100%	50%	83%	50%	67%	50%	67%	50%	50%	100%	100%	100%	**72%**
S5	100%	83%	83%	83%	50%	100%	33%	50%	83%	100%	100%	100%	**81%**
S6	50%	100%	50%	33%	33%	100%	33%	83%	100%	100%	50%	100%	**69%**
S7	67%	50%	67%	67%	33%	100%	33%	100%	100%	100%	100%	100%	**76%**
S8	100%	67%	67%	33%	50%	100%	67%	50%	100%	100%	100%	50%	**74%**
Prom	**90%**	**71%**	**77%**	**65%**	**58%**	**88%**	**50%**	**67%**	**77%**	**92%**	**81%**	**75%**	**74%**
	76%				66%				81%				

Tabla 4.3. Resultados Sensibilidad. Jerárquica Aglomerativa. Todos los Parámetros. R0.

Partamos el análisis mirando la Tabla 4.2. En este caso estamos estudiando el algoritmo de agrupamiento K-means, utilizando una definición de distancia euclideana. Lo primero que llama la atención, es que hay un caso estructural cuyo resultado promedio está muy cerca al caso aleatorio del 50%, correspondiente al de RE35. De todas formas, el promedio total de efectividad, considerando los resultados de todos los sensores y todos los casos estructurales es de alrededor de un 80%, lo que es obviamente una mejora con respecto al caso aleatorio. Otro punto que llama la atención, es que pareciera que la gravedad del daño influye en el resultado del procedimiento. Esto se observa al comparar el promedio de los 4 primeros casos estructurales, asociados a un daño leve, con el promedio de los últimos 4 casos, asociados a un daño severo. La implicancias de esto son que la metodología es coherente con el nivel de daño y si bien no se puede inferir la gravedad del daño a partir de la salida de ésta, ya que simplemente entrega una función de asignación óptima, sí se puede afirmar que a un mayor nivel de daño la probabilidad de una asignación correcta también aumenta. Por último, comparando los dos algoritmos de agrupamiento utilizados, de la Tabla 4.3, se observa que el de K-means obtuvo una performance levemente superior, sin embargo, en términos computacionales, el método jerárquico es considerablemente más rápido principalmente por dos motivos: es determinístico, en el sentido de que no necesita múltiples ejecuciones para evitar caer en un mínimo local, y además, calcula las distintas particiones posibles a partir de un mismo árbol jerárquico, posibilitando usar el mismo árbol para un número distinto de grupos. A su vez, los resultados de ambos algoritmos de agrupamiento es coherente: la efectividad promedio para cada caso estructural son muy parecidas.

A modo de resumen, se destaca que la performance obtuvo resultados prometedores considerando que la información usada es solo la aportada por cada sensor por separado. Todos los resultados estuvieron por sobre el 50%, lo que en términos coloquiales podemos expresar como que "todos los sensores tienen algo que decir al respecto".

Ensayo: Ruido 2

K-Means. Euclidean Distance. Todos los Parametros. **R2.**													
Sensor	NmsM	NmsT	R20T	R20M	R35M	R35T	RE35	RE50	R50T	R50M	RT35	RT50	Global
S1	33%	67%	50%	67%	83%	33%	83%	67%	67%	50%	67%	100%	**64%**
S2	33%	33%	33%	67%	83%	33%	83%	67%	50%	50%	67%	100%	**58%**
S3	33%	33%	50%	83%	67%	33%	83%	50%	33%	50%	50%	83%	**54%**
S4	33%	33%	67%	67%	67%	33%	67%	50%	50%	50%	50%	100%	**56%**
S5	33%	33%	67%	67%	67%	33%	67%	50%	67%	67%	100%	100%	**63%**
S6	67%	67%	50%	83%	100%	67%	100%	33%	100%	67%	67%	100%	**75%**
S7	50%	67%	83%	33%	50%	50%	33%	67%	67%	50%	33%	100%	**57%**
S8	33%	67%	67%	67%	100%	100%	100%	100%	100%	100%	100%	100%	**86%**
Prom	**40%**	**50%**	**58%**	**67%**	**77%**	**48%**	**77%**	**60%**	**67%**	**60%**	**67%**	**98%**	**64%**
	54%				66%				73%				

Tabla 4.4. Resultados Sensibilidad. K-Means. Euclidean Distance. Todos los Parámetros. R2.

Hierarchical. Categorical Distance. Todos los Parametros. R2.													
Sensor	NmsM	NmsT	R20T	R20M	R35M	R35T	RE35	RE50	R50T	R50M	RT35	RT50	**Global**
S1	67%	33%	67%	67%	83%	50%	83%	50%	67%	67%	67%	100%	**67%**
S2	50%	33%	67%	67%	83%	50%	83%	50%	67%	67%	67%	100%	**65%**
S3	33%	33%	33%	83%	67%	33%	83%	50%	17%	67%	67%	50%	**51%**
S4	50%	33%	67%	67%	67%	33%	50%	50%	50%	67%	67%	100%	**58%**
S5	33%	33%	67%	33%	67%	67%	67%	50%	50%	67%	67%	100%	**58%**
S6	67%	50%	67%	83%	100%	67%	100%	50%	50%	67%	67%	100%	**72%**
S7	50%	67%	83%	50%	50%	33%	50%	50%	33%	50%	50%	100%	**56%**
S8	33%	33%	67%	33%	33%	33%	100%	83%	83%	67%	100%	100%	**64%**
Prom	**48%**	**40%**	**65%**	**60%**	**69%**	**46%**	**77%**	**54%**	**52%**	**65%**	**69%**	**94%**	**61%**
	53%				62%				70%				

Tabla 4.5. Resultados Sensibilidad. Jerárquica Aglomerativa. Todos los Parámetros. R2.

Los resultados obtenidos a partir de la aceleración basal Ruido2, son bastante menos prometedores que los de Ruido0. Efectivamente, en términos globales, se tiene una efectividad de alrededor de 63% en contraste con el 80% obtenido para Ruido0. Claramente, el output de la metodología depende en gran medida de la energía entrante al sistema y a los posibles efectos instantáneos. Recordemos que los objetos simbólicos considerados tienen un alcance de alrededor de un minuto lo que se postula como un largo muy inferior al ideal, y estos resultados pueden ser considerados como un apoyo a esta teoría. Un punto que destaca es la presencia de tres casos estructurales en los que la performance fue peor o igual al resultado aleatorio del 50%, correspondientes a NmsM, NmsT y R35T. Esto representa un contraste muy fuerte con lo obtenido para Ruido0, ya que estos casos habían tenido unos resultados mucho mayores. Aún así, recordemos que se está analizando solo a partir de la información que entrega un sensor por vez, por lo que la metodología podría ser fácilmente mejorada incluyendo más sensores. Esto último se ve apoyado en el hecho de que, incluso considerando la información de un solo sensor, los resultados globales para cada uno de ellos entregan porcentajes sobre el 50%. Es decir, en promedio, todos los sensores entregan información relacionada con la presencia de distintas condiciones estructurales. Otro resultado importante y distinto al obtenido para Ruido0, es que en este caso, la condición estructural que obtuvo la mejor performance fue la asociada al daño más

severo. El resto de los resultados destacados para Ruido0 también se cumplen para esta aceleración basal, siendo el más importante el relacionado con la sensibilidad de la metodología con respecto al nivel de daño, ya que el promedio de efectividad para los casos de daño leve es notablemente menor que para los casos de daño severo. Es más, la diferencia entre estos promedios es mucho mayor que para el caso de Ruido0. Además, K-means es levemente superior a costa de un costo numérico mayor, y la efectividad es similar para cada caso estructural comparando ambos métodos de agrupamiento.

Ensayos: Ruido 4

K-Means. Euclidean Distance. Todos los Parámetros. **R4.**													
Sensor	NmsM	NmsT	R20T	R20M	R35M	R35T	RE35	RE50	R50T	R50M	RT35	RT50	**Global**
S1	67%	83%	67%	67%	67%	67%	100%	33%	100%	67%	67%	67%	**71%**
S2	67%	67%	67%	67%	100%	67%	100%	50%	100%	67%	67%	67%	**74%**
S3	33%	100%	67%	50%	100%	50%	100%	33%	100%	100%	100%	100%	**78%**
S4	33%	100%	50%	100%	100%	83%	100%	33%	100%	100%	83%	100%	**82%**
S5	67%	83%	50%	100%	100%	33%	100%	33%	100%	67%	83%	100%	**76%**
S6	67%	67%	50%	100%	67%	67%	100%	50%	50%	50%	50%	100%	**68%**
S7	67%	33%	50%	100%	100%	50%	100%	50%	100%	67%	100%	50%	**72%**
S8	83%	100%	67%	100%	100%	83%	100%	33%	100%	100%	100%	100%	**89%**
Prom	**60%**	**79%**	**58%**	**85%**	**92%**	**63%**	**100%**	**40%**	**94%**	**77%**	**81%**	**85%**	**76%**
	71%				74%				84%				

Tabla 4.6. Resultados Sensibilidad. K-Means. Euclidean Distance. Todos los Parámetros. R4.

Hierarchical. Categorical Distance. Todos los Parámetros. R4.													
Sensor	NmsM	NmsT	R20T	R20M	R35M	R35T	RE35	RE50	R50T	R50M	RT35	RT50	**Global**
S1	67%	83%	50%	67%	67%	67%	100%	50%	100%	67%	67%	50%	**69%**
S2	67%	67%	67%	50%	100%	67%	100%	50%	100%	67%	67%	67%	**72%**
S3	33%	100%	67%	50%	100%	67%	100%	50%	100%	100%	100%	100%	**81%**
S4	83%	83%	50%	100%	100%	50%	100%	50%	100%	100%	83%	100%	**83%**
S5	50%	50%	50%	100%	100%	83%	100%	50%	100%	50%	67%	100%	**75%**
S6	67%	67%	67%	67%	67%	50%	67%	50%	50%	67%	67%	100%	**65%**
S7	67%	33%	50%	100%	100%	50%	100%	50%	100%	50%	100%	50%	**71%**
S8	83%	100%	83%	83%	100%	100%	67%	50%	100%	100%	100%	100%	**89%**
Prom	**65%**	**73%**	**60%**	**77%**	**92%**	**67%**	**92%**	**50%**	**94%**	**75%**	**81%**	**83%**	**76%**
	69%				75%				83%				

Tabla 4.7. Resultados Sensibilidad. Jerárquica Aglomerativa. Todos los Parámetros. R4.

En lo que respecta a la aceleración basal llamada Ruido4, casi todas las conclusiones referentes a Ruido0 también se cumplen. La efectividad promedio total se encuentra entre las dos aceleraciones basales anteriores, siendo igual a un 76%. No obstante, llama la atención que el caso estructural RE35, sea el que obtuvo el mejor resultado incluso igual a un 100% (todos los sensores fueron capaces de determinar la existencia de dos condiciones distintas), mientras que fue el de peor resultado para Ruido0. Cabe destacar que en este caso, ambos algoritmos de agrupamiento tuvieron el mismo desempeño, pero se recuerda que K-means tiene un costo computacional varios órdenes de magnitud mayor.

Sensibilidad Parámetros.

En este apartado, analizaremos los resultados obtenidos para cuando se consideraron la totalidad de sensores existentes, pero solo un parámetro AR a la vez. Una de las cosas que esperamos es que los resultados sean más prometedores, puesto que como mostramos en la sección anterior, todos los sensores tienen información con la que aportar. Sin embargo, se entiende que hay una balanza entre la información que aportan los sensores y la que aportan los parámetros, por lo que esta mejora puede ser menor que lo esperado.

En la Tabla 4.8 y 4.9 se muestran los resultados obtenidos al estudiar la aceleración basal Ruido0. Lo primero que cabe destacar es que ningún parámetro AR obtuvo resultados menores al 70%, lo que resulta en una mejora considerable. Aún más, hay algunos parámetros que entregaron particiones correctas en un 100% para casi todos los casos estructurales. También se destaca que las condiciones estructurales R35T y RT35 fueron correctamente particionadas para cada uno de los parámetros. Esto es un indicador claro de que conviene considerar la información de la mayor cantidad posible de sensores, y reforzamos esta afirmación destacando el resultado global promedio, de alrededor de un 84%. Por otra parte, todos los parámetros parecen rendir de forma pareja; es decir, no se aprecia alguna tendencia en cuánto a relacionar parámetros AR viejos (los más alejados del instante presente) con daños locales, o a relacionar parámetros AR jóvenes con daños globales. Aún así, no se puede afirmar que no exista una relación o analogía entre los parámetros AR viejos con los modos de vibración superiores. Llama la atención que algunos parámetros tengan un rendimiento cercano al 100%, pero sacar alguna conclusión al respecto sería demasiado arriesgado ya que faltarían más ensayos para poder generalizar estos resultados.

Un punto destacable, es que de acuerdo a lo observado en el análisis de los sensores, a mayor nivel de daño hubo un mayor porcentaje de asignaciones correctas, lo que indica que los coeficientes AR son una 'feature' coherente con el nivel del cambio estructural. Sin embargo se recalca que esta información no es posible de utilizar a nivel de usuario en un formato de análisis no supervisado. En cuanto a la comparación entre los algoritmos de agrupamiento usados, resalta el hecho de que el comportamiento es muy similar, siendo los promedios globales de cada parámetro muy parecidos entre los dos métodos de agrupamiento, lo que también se desprende para los resultados promedio de cada caso estructural. Las únicas excepciones corresponden al parámetro N°7, que para K-means resultó ser el de mejor resultado, mientras que para el método jerárquico fue el de peor efectividad; y la condición estructural R35M que tuvo una diferencia de un 20% entre ambos algoritmos. Por último, se mantiene la tendencia de que K-Means con distancia euclideana obtiene mejor rendimiento a expensas de un tiempo de cómputo mayor

En términos globales, el uso de la información que aportan todos los sensores mejora considerablemente los resultados, siempre y cuando los sensores no tengan algún problema de medición. Más adelante se mostrará el caso en el que se ocupan todos los parámetros y todos los sensores.

K-Means. Euclidean Distance. Todos los Sensores. **R0**													
Param	NmsM	NmsT	R20T	R20M	R35M	R35T	RE35	RE50	R50T	R50M	RT35	RT50	**Global**
P1	100%	100%	100%	100%	100%	100%	67%	83%	100%	100%	100%	83%	**94%**
P2	100%	100%	100%	67%	33%	100%	33%	33%	100%	100%	100%	100%	**81%**
P3	100%	50%	100%	67%	33%	100%	33%	100%	100%	100%	100%	100%	**82%**
P4	100%	67%	67%	100%	100%	100%	100%	50%	67%	33%	100%	100%	**82%**
P5	67%	50%	50%	33%	33%	100%	33%	100%	100%	100%	100%	100%	**72%**
P6	83%	33%	33%	100%	100%	100%	100%	83%	83%	33%	100%	100%	**79%**
P7	100%	100%	100%	100%	100%	100%	67%	100%	100%	100%	100%	100%	**97%**
P8	100%	67%	100%	100%	100%	100%	100%	67%	100%	100%	100%	100%	**94%**
P9	100%	100%	100%	100%	100%	100%	100%	83%	100%	67%	100%	100%	**96%**
P10	100%	100%	100%	100%	67%	100%	33%	50%	100%	100%	100%	100%	**88%**
P11	100%	100%	100%	100%	100%	100%	50%	100%	100%	67%	100%	100%	**93%**
P12	100%	83%	67%	33%	33%	100%	33%	50%	100%	100%	100%	100%	**75%**
P13	100%	100%	67%	33%	67%	100%	33%	100%	50%	100%	100%	100%	**79%**
P14	100%	67%	100%	67%	100%	100%	100%	100%	67%	100%	100%	50%	**88%**
P15	100%	100%	100%	100%	100%	100%	50%	67%	100%	100%	100%	100%	**93%**
Prom	**97%**	**81%**	**86%**	**80%**	**78%**	**100%**	**62%**	**78%**	**91%**	**87%**	**100%**	**96%**	**86%**
	86%				80%				93%				

Tabla 4.8. Resultados Sensibilidad. K-Means. Todos los Sensores. R0.

Hierarchical Agglomerative. Categorical Distance. Todos los Sensores. R0													
Param	NmsM	NmsT	R20T	R20M	R35M	R35T	RE35	RE50	R50T	R50M	RT35	RT50	**Prom**
P1	100%	100%	100%	83%	83%	100%	100%	83%	100%	100%	100%	83%	**94%**
P2	100%	83%	100%	33%	33%	100%	33%	50%	100%	100%	100%	100%	**78%**
P3	100%	67%	100%	67%	33%	100%	33%	100%	100%	100%	100%	100%	**83%**
P4	67%	33%	67%	100%	100%	100%	100%	83%	67%	50%	100%	100%	**81%**
P5	50%	50%	67%	33%	33%	100%	33%	83%	83%	100%	100%	100%	**69%**
P6	83%	33%	83%	100%	100%	100%	100%	50%	67%	50%	100%	100%	**81%**
P7	83%	50%	67%	67%	33%	100%	33%	83%	67%	67%	100%	100%	**71%**
P8	100%	33%	100%	100%	33%	100%	33%	83%	100%	100%	100%	100%	**82%**
P9	100%	100%	100%	100%	67%	100%	100%	83%	100%	67%	100%	100%	**93%**
P10	100%	67%	100%	67%	33%	100%	33%	50%	100%	100%	100%	100%	**79%**
P11	100%	100%	100%	33%	50%	100%	33%	67%	100%	67%	100%	100%	**79%**
P12	100%	83%	50%	33%	33%	100%	33%	50%	83%	100%	100%	100%	**72%**
P13	100%	100%	50%	33%	33%	67%	33%	100%	50%	100%	100%	100%	**72%**
P14	67%	67%	100%	100%	100%	100%	100%	100%	100%	83%	100%	50%	**89%**
P15	100%	100%	50%	67%	100%	100%	50%	50%	100%	83%	100%	100%	**83%**
Prom	**90%**	**71%**	**82%**	**68%**	**58%**	**98%**	**57%**	**74%**	**88%**	**84%**	**100%**	**96%**	**80%**
	78%				72%				92%				

Tabla 4.9. Resultados Sensibilidad. Jerárquica Aglomerativa. Todos los Sensores. R0.

K-Means. Euclidean Distance. Todos los Sensores. **R2**													
Param	NmsM	NmsT	R20T	R20M	R35M	R35T	RE35	RE50	R50T	R50M	RT35	RT50	**Prom**
P1	67%	33%	33%	50%	67%	33%	50%	50%	50%	67%	67%	50%	**51%**
P2	33%	33%	67%	67%	83%	33%	67%	50%	50%	33%	50%	50%	**51%**
P3	67%	50%	67%	83%	67%	33%	50%	67%	100%	100%	100%	100%	**74%**
P4	50%	33%	33%	67%	83%	50%	33%	50%	67%	67%	33%	100%	**56%**
P5	67%	83%	67%	17%	67%	33%	83%	67%	67%	100%	100%	100%	**71%**
P6	67%	50%	33%	67%	67%	33%	83%	100%	100%	67%	100%	100%	**72%**
P7	67%	67%	50%	67%	67%	33%	83%	67%	67%	67%	100%	100%	**69%**
P8	33%	33%	50%	33%	67%	33%	83%	100%	100%	33%	100%	100%	**64%**
P9	50%	67%	50%	67%	67%	67%	83%	67%	67%	67%	50%	100%	**67%**
P10	50%	50%	33%	83%	67%	33%	83%	100%	67%	50%	100%	100%	**68%**
P11	67%	67%	50%	67%	67%	33%	50%	50%	100%	67%	100%	100%	**68%**
P12	50%	33%	50%	83%	67%	33%	83%	50%	50%	33%	67%	67%	**56%**
P13	33%	50%	33%	67%	67%	33%	33%	100%	100%	100%	100%	100%	**68%**
P14	33%	33%	33%	33%	50%	33%	50%	100%	67%	50%	33%	100%	**51%**
P15	67%	50%	67%	67%	67%	67%	83%	67%	100%	67%	50%	50%	**67%**
Prom	**53%**	**49%**	**48%**	**61%**	**68%**	**39%**	**67%**	**72%**	**77%**	**64%**	**77%**	**88%**	**64%**
	53%				62%				76%				

Tabla 4.10. Resultados Sensibilidad. K-Means. Todos los Sensores. R2.

Hierarchical Agglomerative. Categorical Distance. Todos los Sensores. R2													
Param	NmsM	NmsT	R20T	R20M	R35M	R35T	RE35	RE50	R50T	R50M	RT35	RT50	**Global**
P1	67%	50%	50%	33%	83%	67%	50%	50%	17%	67%	67%	33%	**53%**
P2	33%	33%	33%	67%	83%	33%	67%	50%	67%	50%	67%	50%	**53%**
P3	67%	33%	33%	33%	50%	67%	100%	50%	67%	50%	100%	100%	**63%**
P4	33%	33%	33%	67%	50%	33%	50%	100%	67%	67%	50%	100%	**57%**
P5	67%	33%	33%	50%	50%	33%	83%	83%	50%	83%	100%	100%	**64%**
P6	33%	33%	50%	67%	50%	33%	83%	67%	100%	50%	100%	100%	**64%**
P7	33%	33%	50%	50%	67%	33%	83%	83%	50%	33%	100%	100%	**60%**
P8	33%	33%	50%	33%	67%	33%	83%	100%	67%	33%	100%	100%	**61%**
P9	33%	67%	50%	67%	67%	50%	83%	67%	67%	67%	100%	100%	**68%**
P10	33%	50%	50%	83%	67%	33%	83%	100%	67%	50%	100%	100%	**68%**
P11	33%	67%	50%	50%	50%	33%	50%	50%	100%	50%	100%	100%	**61%**
P12	50%	33%	67%	83%	83%	50%	83%	33%	17%	33%	67%	50%	**54%**
P13	33%	33%	33%	83%	50%	33%	50%	100%	100%	100%	100%	100%	**68%**
P14	33%	33%	50%	33%	50%	33%	50%	100%	33%	67%	17%	100%	**50%**
P15	67%	67%	33%	67%	50%	67%	67%	50%	50%	67%	67%	100%	**63%**
Prom	**43%**	**42%**	**44%**	**58%**	**61%**	**42%**	**71%**	**72%**	**61%**	**58%**	**82%**	**89%**	**60%**
	47%				62%				73%				

Tabla 4.11. Resultados Sensibilidad. Jerárquica Aglomerativa. Todos los Sensores. R2.

Los resultados para las otras aceleraciones basales arrojan básicamente las mismas conclusiones que para el estudio de sensibilidad de sensores. Esto es, para la aceleración basal Ruido2, los resultados no son tan prometedores, rondando el 64% en el resultado promedio global. En este caso en particular hubo bastantes situaciones en las que el promedio por condición estructural no alcanzó el 50%, sobretodo en condiciones estructurales de daño leve. En este sentido, pese a que hemos visto que Ruido2 es una excitación conflictiva, también se mantiene la tendencia a una mejor asignación a medida que el daño progresa. Por otra parte, la coherencia de resultados entre los algoritmos de agrupamiento fue incluso mejor que para la de Ruido0, mostrando una correlación casi perfecta al no existir las excepciones mencionadas para dicho caso. Todos estos comentarios también son válidos para la aceleración Ruido4, y cabe mencionar que al igual que en el estudio de sensores, esta aceleración basal obtuvo resultados entremedio de Ruido2 y Ruido0.

K-Means. Euclidean Distance. Todos los Sensores. **R4**													
Param	NmsM	NmsT	R20T	R20M	R35M	R35T	RE35	RE50	R50T	R50M	RT35	RT50	**Global**
P1	67%	50%	50%	67%	100%	50%	67%	50%	100%	50%	100%	33%	**65%**
P2	50%	67%	100%	100%	100%	67%	100%	50%	100%	50%	100%	100%	**82%**
P3	67%	100%	100%	100%	100%	67%	100%	50%	100%	100%	100%	100%	**90%**
P4	67%	100%	83%	100%	100%	67%	100%	83%	100%	100%	100%	100%	**92%**
P5	33%	83%	83%	67%	67%	83%	100%	83%	100%	100%	100%	100%	**83%**
P6	83%	100%	67%	100%	100%	67%	100%	17%	100%	100%	100%	100%	**86%**
P7	67%	67%	67%	67%	67%	83%	100%	50%	100%	67%	100%	100%	**78%**
P8	67%	100%	83%	100%	100%	67%	100%	50%	100%	100%	100%	100%	**89%**
P9	50%	67%	83%	67%	67%	83%	100%	50%	100%	83%	100%	100%	**79%**
P10	100%	67%	100%	100%	100%	67%	100%	33%	100%	100%	100%	100%	**89%**
P11	17%	67%	67%	67%	100%	67%	100%	50%	100%	50%	100%	100%	**74%**
P12	33%	50%	50%	100%	100%	67%	100%	50%	100%	50%	100%	100%	**75%**
P13	50%	100%	83%	100%	100%	83%	100%	33%	100%	100%	100%	100%	**87%**
P14	83%	100%	67%	100%	100%	67%	100%	100%	100%	100%	100%	100%	**93%**
P15	67%	50%	33%	67%	100%	50%	100%	50%	50%	50%	67%	50%	**61%**
Prom	**60%**	**78%**	**74%**	**87%**	**93%**	**69%**	**98%**	**53%**	**97%**	**80%**	**98%**	**92%**	**82%**
	75%				**78%**				92%				

Tabla 4.12. Resultados Sensibilidad. K-Means. Todos los Sensores. R4.

Hierarchical Agglomerative. Categorical Distance. Todos los Sensores. R4													
Param	NmsM	NmsT	R20T	R20M	R35M	R35T	RE35	RE50	R50T	R50M	RT35	RT50	**Global**
P1	67%	67%	33%	50%	100%	50%	50%	50%	100%	50%	100%	50%	**64%**
P2	67%	67%	100%	100%	100%	50%	100%	50%	100%	50%	100%	100%	**82%**
P3	33%	100%	100%	100%	100%	83%	100%	50%	100%	100%	100%	100%	**89%**
P4	67%	100%	67%	50%	100%	83%	100%	50%	50%	100%	100%	100%	**81%**
P5	67%	83%	83%	67%	67%	83%	100%	50%	100%	100%	100%	100%	**83%**
P6	67%	100%	83%	50%	100%	83%	100%	50%	100%	100%	100%	100%	**86%**
P7	67%	50%	50%	67%	67%	83%	67%	50%	100%	67%	100%	100%	**72%**
P8	50%	100%	83%	100%	100%	83%	100%	50%	100%	100%	100%	100%	**89%**
P9	50%	67%	83%	67%	67%	83%	100%	50%	100%	67%	100%	100%	**78%**
P10	50%	67%	50%	100%	100%	83%	100%	50%	100%	67%	100%	100%	**81%**
P11	67%	67%	67%	67%	100%	50%	100%	50%	100%	67%	100%	100%	**78%**
P12	67%	67%	67%	100%	100%	50%	100%	50%	100%	67%	100%	100%	**81%**
P13	50%	100%	50%	100%	100%	83%	100%	50%	100%	100%	100%	100%	**86%**
P14	67%	100%	33%	67%	100%	67%	100%	50%	100%	50%	100%	100%	**78%**
P15	67%	50%	50%	50%	100%	33%	100%	50%	50%	50%	50%	17%	**56%**
Prom	**60%**	**79%**	**67%**	**76%**	**93%**	**70%**	**94%**	**50%**	**93%**	**76%**	**97%**	**91%**	**79%**
	70%				77%				89%				

Tabla 4.13. Resultados Sensibilidad. Jerárquica Aglomerativa. Todos los Sensores. R4.

4.3.3. Resumen de resultados de sensibilidad

Durante la presente sección 4.3 estudiamos el comportamiento en solitario de sensores y parámetros y podemos resumir los resultados más importantes.

1. Se logran resultados superiores al 50%. Incluso un 100% de clasificación correcta en algunos casos.
2. La excitación base afecta bastante el resultado de la salida. Puede deberse a efectos instantáneos producto de objetos simbólicos de poca duración.
3. A mayor nivel de daño, mayor probabilidad de lograr una asignación correcta.
4. Todos los sensores y todos los parámetros aportan información.
5. La performance de los algoritmos de agrupamiento es levemente superior en K-means, pero a un costo computacional mayor.
6. En general, se logra coherencia entre los resultados, con muy pocas excepciones.

En la siguiente sección, vamos a revisar resultados al utilizar la información de todos los sensores y todos los parámetros. Además, se verán los resultados globales al usar objetos simbólicos tipo intercuartil, y la posibilidad de mejorar la metodología considerando el filtro Moving Average mencionado en la sección anterior.

4.4. Resultados casos de estudio.

Los análisis efectuados en esta sección, consideran la información aportada por todos los sensores y todos los parámetros. Se busca comparar los resultados que entregan los objetos tipo histograma e intercuartil, así como también verificar si es que el filtro media móvil que se utilizó para eliminar outliers realmente genera un beneficio para el algoritmo.

4.4.1. Ensayos de laboratorio.

Se consideran los ensayos correspondientes a las tres aceleraciones basales. Para aumentar el número de objetos, las señales se dividen en tres ventanas y las series de parámetros AR consideran modelos de orden 15 a partir de la obtención de orden óptimo calculada en las secciones anteriores.Los resultados se presentan en tablas con los correspondientes porcentajes de asignaciones correctas.

Los resultados para la aceleración Ruido0 se presentan en la Tabla 4.14. Esta aceleración era la que mostraba las asignaciones más correctas hasta el momento, y lo obtenido considerando todos los sensores y todos los parámetros confirman que para este caso en particular, la metodología funciona de forma muy positiva, logrando una clasificación con un 100% de efectividad en el caso de agrupamiento k-means usando histogramas. Se destaca la obtención de este resultado considerando objetos simbólicos de corta duración, lo que permite pensar que la utilización de la metodología en casos más generales puede ser un hecho real. En cuanto a objetos obtenidos por intervalos intercuartil, su performance es inferior a los obtenidos por histograma. Además, si bien el filtro MA muestra una leve mejora de los resultados, ésta es marginal

Ruido 0								
	Histograma				***Interquartil***			
	No MA		Con MA		No MA		Con MA	
	KM	JA	KM	JA	KM	JA	KM	JA
NmsM	100%	100%	100%	100%	83%	100%	83%	100%
NmsT	100%	100%	100%	100%	100%	100%	100%	100%
R20T	100%	100%	100%	100%	83%	100%	83%	100%
R20M	100%	50%	100%	100%	67%	67%	67%	67%
R35M	100%	83%	100%	67%	67%	67%	67%	67%
R35T	100%	100%	100%	100%	100%	100%	100%	100%
RE35	100%	50%	100%	50%	33%	33%	33%	33%
RE50	100%	83%	100%	83%	17%	50%	83%	83%
R50T	100%	100%	100%	100%	100%	100%	100%	83%
R50M	100%	100%	100%	100%	100%	100%	100%	100%
RT35	100%	100%	100%	100%	100%	100%	100%	100%
RT50	100%	100%	100%	100%	100%	100%	100%	100%
Prom	**100%**	**89%**	**100%**	**92%**	**79%**	**85%**	**85%**	**86%**

Tabla 4.14. Resultados para Ruido0.

Ruido 2								
	Histograma				***Interquartil***			
	No MA		Con MA		No MA		Con MA	
	KM	JA	KM	JA	KM	JA	KM	JA
NmsM	33%	50%	33%	50%	67%	50%	67%	50%
NmsT	50%	67%	50%	33%	50%	67%	50%	33%
R20T	33%	50%	67%	50%	67%	67%	67%	67%
R20M	83%	83%	83%	83%	67%	67%	33%	67%
R35M	67%	83%	67%	67%	67%	67%	67%	67%
R35T	33%	33%	33%	33%	67%	50%	33%	33%
RE35	83%	83%	83%	83%	67%	50%	67%	50%
RE50	100%	100%	100%	100%	67%	50%	67%	50%
R50T	100%	100%	100%	100%	50%	67%	67%	50%
R50M	100%	67%	50%	67%	33%	50%	33%	50%
RT35	100%	100%	100%	100%	100%	100%	100%	100%
RT50	100%	100%	100%	100%	100%	100%	100%	100%
Prom	**74%**	**76%**	**72%**	**72%**	**67%**	**65%**	**63%**	**60%**

Tabla 4.15. Resultados Ruido2.

En relación a la aceleración basal Ruido2, se mantiene la tendencia de ser un caso cuya performance es mucho menor que el Ruido0. Se destaca que, por primera vez, el agrupamiento jerárquico tuvo mejor resultado que el de K-means, lo cual es un hecho notable considerando el mayor costo computacional de éste último. Además, se debe remarcar que pese a ser una aceleración basal que ha mostrado ser complicada para el algoritmo, los dos casos estructurales más severos fueron clasificados correctamente al 100%. Nuevamente, los objetos intercuartil muestran porcentajes menores que los de histograma, con la destacable excepción de las

condicione estructurales más severas, que como se mencionó, fueron clasificadas correctamente. Por último, para esta aceleración basal, el filtro MA no aporta en nada y de hecho, perjudica los resultados.

Lo remarcado para aceleración basal Ruido0 también se cumple para el Ruido4. Encontramos una mejora sustantiva en comparación a lo obtenido usando un sensor o parámetro a la vez, logrando una eficiencia promedio de hasta un 92% en el caso de k-means con filtro MA. Aún así, el filtro realmente produce una ganancia marginal y como vimos en el caso de Ruido2 incluso puede llegar a perjudicar los resultados por lo que no se recomienda su uso. Al igual que en los casos anteriores, los objetos histogramas son más sensibles a clasificar correctamente las condiciones estructurales, y se alcanza un 100% de clasificación correcta para los casos extremos, independiente del tipo de objeto utilizado.

A modo de resumen, en esta sección hemos visto que pese a usar objetos simbólicos de corta duración, la metodología ha mostrado tener una performance prometedora, logrando clasificar las condiciones estructurales más severas con una eficiencia del 100%. Además, podemos afirmar que no debiese ser utilizado un filtro MA ya que su comportamiento es variable e incluso perjudica los resultados en algunos casos. De los dos tipos de objetos considerados, los obtenidos a partir de histogramas muestran una sensibilidad mayor, lo que es esperable dado que son objetos de una complejidad mayor y por tanto contribuyen con mayor información relacionada con las series de tiempo. Por último, K-means es el algoritmo que entrega una mejor asignación, pero con costos computacionales mayores. Dentro de las posibles mejoras que se pueden observar de momento, se encuentra la de eliminación de objetos outliers, que describen el comportamiento instantáneo de la estructura frente a alguna excitación de características notablemente distintas a lo que consideramos ruido blanco o excitación ambiental "normal". Esto es posible de conseguir analizando objetos cuya distancia al resto supere un umbral de aceptación, lo que se realizaría en un paso antes de realizar el algoritmo de agrupamiento, pero resulta más atractivo eliminar los objetos que se encuentren en grupos cuya cardinalidad, considerablemente menor a la del resto de grupos. Otra posible mejora es la de aplicar una restricción a las asignaciones posibles, ya que no debiesen mezclarse objetos obtenidos en distintos momentos temporales. Esto es intuitivo en el sentido de que el cambio estructural marca un antes y un después, sin embargo imponer esta restricción en el algoritmo de agrupamiento parece ser demasiado fuerte, implicando que para buscar una solución de esta forma se tendría que hacer un análisis por fuerza bruta, lo cual es muy costoso computacionalmente para conjuntos de objetos muy grandes.

Ruido 4								
	Histograma				Interquartil			
	No MA		Con MA		No MA		Con MA	
	KM	JA	KM	JA	KM	JA	KM	JA
NmsM	67%	67%	67%	67%	33%	33%	67%	33%
NmsT	100%	100%	100%	100%	83%	100%	67%	100%
R20T	100%	67%	100%	67%	67%	83%	67%	83%
R20M	100%	100%	100%	100%	67%	100%	67%	100%
R35M	100%	100%	100%	100%	67%	100%	67%	100%
R35T	67%	83%	83%	83%	83%	83%	83%	83%
RE35	100%	100%	100%	100%	100%	100%	100%	100%
RE50	33%	50%	50%	50%	17%	50%	33%	50%
R50T	100%	100%	100%	100%	100%	100%	100%	100%
R50M	100%	100%	100%	100%	67%	100%	67%	50%
RT35	100%	100%	100%	100%	100%	100%	100%	100%
RT50	100%	100%	100%	100%	100%	100%	100%	100%
Prom	**89%**	**89%**	**92%**	**89%**	**74%**	**87%**	**76%**	**83%**

Tabla 4.16. Resultados Ruido4.

4.4.2. Aplicación en Torre Central.

Las características del sistema de adquisición en la Torre Central fueron descritas en el capítulo 2. A modo de recordatorio, se tienen registros de aceleración de tres sensores ubicados en el piso 3, adquiridos cada 15 minutos. En particular, se cuentan con registros obtenidos durante Enero y Marzo del 2010, periodo que incluye la ocurrencia del terremoto del 27 de Febrero, con una magnitud tal que se tiene certeza de que existió cambios estructurales. La idea de esta sección es aplicar la metodología a estos registros para lograr clasificar los registros pre y post terremotos.

Series de parámetros AR.

Se utilizan las mismas series que las calculadas en el capítulo 3 para el estudio de las distancias de Mahalanobis. Esto consideró el uso de ventanas traslapadas para cada registro, a las que se le ajustó un modelo autoregresivo de orden $p = 50$, y con cuyos parámetros se crearon las series de parámetros. Cada registro de 180000 datos de aceleración, luego de pasar por el proceso de decimado, se transforma en una serie de aproximadamente 2500 observaciones de parámetros. Como cada uno de los tres sensores aporta con 50 coeficientes, las series tienen dimensiones de aproximadamente 2500x150.

Cálculo de distancias.

El cálculo de distancias considera el uso de histogramas para la transformación de las series AR en objetos simbólicos. Esto genera la primera gran dificultad de la metodología, ya que para el cálculo de histogramas se necesita considerar límites de intervalos extremos y además un número determinado de contenedores. En el capítulo 2 esto no fue un problema, principalmente porque las series de aceleración tienen media nula, lo que permite utilizar el mismo rango de forma global. En cambio, las series de parámetros AR tienen una media variable que depende del número del coeficiente estudiado. Para ejemplificar esto, la Figura 4.15 muestra el promedio y desviación estándard de los coeficientes AR(p) para el registro del sensor N°3 durante el 4 de Enero del 2010 a las 2AM. En esta figura se puede apreciar que el valor de los coeficientes es

bastante variable, y tan solo los últimos parecen tener media nula. Esto implica que si se aplican exactamente los mismos intervalos al momento de calcular los histogramas para distintos parámetros, los objetos simbólicos podrían resultar siendo una mala representación de las series. Por ejemplo, considerar un rango desde -1 hasta 1, con una cantidad de contenedores igual a 50. En el caso del parámetro 50, la totalidad de las observaciones caen dentro de dicho rango; sin embargo, para el parámetro 18 casi la todas las observaciones caen en el rango extremo desde 4 hasta infinito, lo cual es observable en la Figura 4.16. Se podría pensar que una solución a este problema es simplemente aumentar los límites de los intervalos extremos e incrementar la cantidad de contenedores, pero esto no necesariamente contribuye ya que en el caso de los parámetros con promedio cercano a 0, existirían muchos contenedores con cero ocurrencias.

Pese lo expuesto en el párrafo anterior, una opción viable es estudiar el comportamiento global de cada uno de los parámetros, buscando encontrar el rango en que normalmente están contenidos. Una de las ventajas del análisis por objetos simbólicos es que cada una de las series analizadas puede utilizar distintos contenedores para el cálculo de los histogramas. Sin embargo, esto representa un grado de uso de señales de referencias a partir de las cuales se calcula el rango de los parámetros.

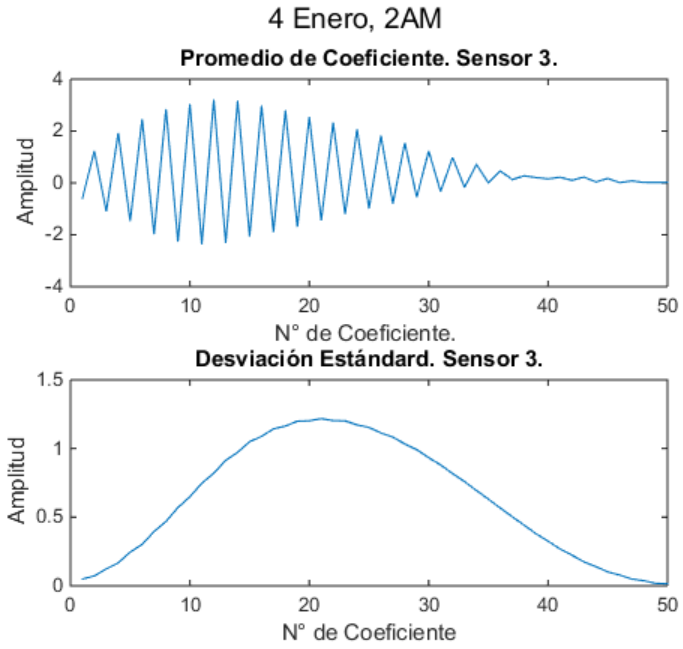


Figura N°4.15. Promedio y Desviación Estándar de los Coeficientes AR.

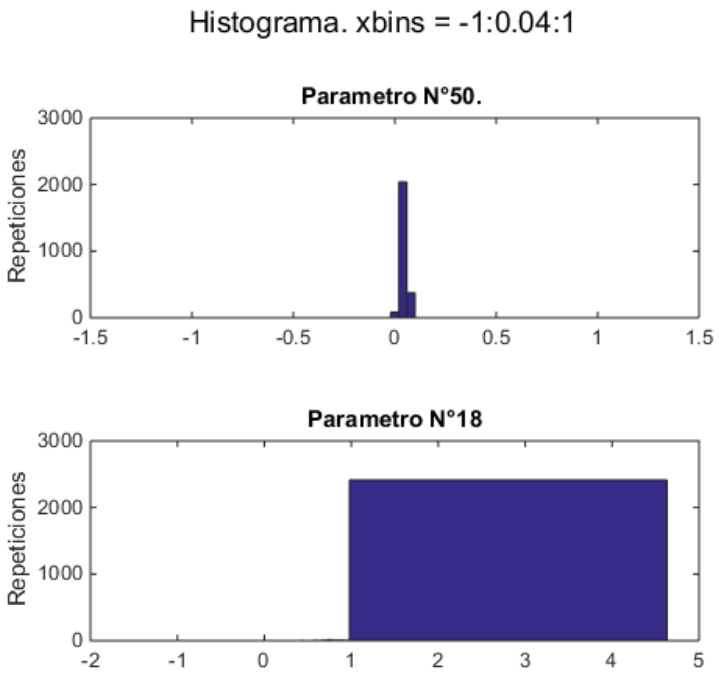


Figura N°4.16. Ejemplo de problema al usar mismos contenedores.

En definitiva, para la creación de las matrices de distancia se utilizarán los dos enfoques mencionados. (1) Considerar un rango fijo global para todos los parámetros, (2) analizar un registro en particular con el que se calculen los valores máximos y mínimos, a partir de los cuales se generen los rangos y categorías para la construcción de los histogramas. Se compararán los resultados finales para determinar cuál de las dos estrategias entrega los mejores resultados.

Metodología de comparación.

Al igual que en el capítulo 2, cada registro será transformado en un objeto simbólico, los que serán clasificados utilizando un algoritmo de agrupamiento. En particular, en esta sección se usará K-Means ya que anteriormente se mostró que entregan resultados similares con Dynamic-Medoids.

La estrategia de comparación de los resultados consiste en clasificar objetos provenientes de la misma condición operacional, entre las 2 y 6AM. Primero se analizarán los resultados al considerar objetos obtenidos de registros de Enero, buscando caracterizar cómo se comporta la metodología al comparar objetos de una misma condición estructural. Finalmente se incluirá la comparación entre objetos de Enero y Marzo, con la idea de encontrar una clasificación que agrupe objetos de distintos meses en distintos conjuntos, como es de esperar debido a la ocurrencia del terremoto en Febrero.

Una de las cosas importantes de recordar es la forma en que se valida el resultado final. Considerar que a partir un set de objetos simbólicos, los algoritmos de agrupamiento arrojan una partición considerada óptima, la que puede tener varios grupos. Esto quiere decir que pese a que se esperaría agrupación perfecta en dos conjuntos representando dos condiciones estructurales distintas, en la práctica es muy difícil obtener este resultado. Por tanto, como siempre se va a estar analizando objetos de distintos días, incluso provenientes de la misma condición estructural, se utilizará como indicador de la calidad del resultado a la cantidad de objetos correctamente

asignados a los conjuntos definidos por los registros de cada día. Por ejemplo, considerar la Tabla 4.17 en la que se muestra un resultado típico de un algoritmo de agrupamiento, donde se indica el grupo al que fue asignado cada objeto. Por ejemplo, según la Tabla 4.17, el décimo objeto fue asignado al grupo n°3, mientras que el quinto objeto al grupo 6. Notar que en dicha tabla, el número óptimo de grupos correspondería a 6, y que el número de objetos correctamente asignado sería de 8, equivalente a un 40% de exactitud.

	Día 1									
Objeto	O1	O2	O3	O4	O5	O6	O7	O8	O9	O10
Asignación	1	1	1	2	6	4	4	5	5	3
	Día 2									
Objeto	O11	O12	O13	O14	O15	O16	O17	O18	O19	O20
Asignación	3	3	3	3	2	5	3	2	1	6

Tabla N°4.17. Resultado típico de un algoritmo de agrupamiento.

Lo que se espera, finalmente, es que el indicador de exactitud de las particiones aumente cuando se comparen objetos de distintos casos estructurales.

La aplicación de los algoritmos se realiza juntando series de parámetros AR de pares de días. Estos días se establecen en la Tabla 4.18.

Test 1	**4/01 & 11/01 (2-6AM)**
Test 2	4/01 & 18/01 (2-6AM)
Test 3	4/01 & 25/01 (2-6AM)
Test 4	4/01 & 8/03 (2-6AM)
Test 5	4/01 & 15/03 (2-6AM)
Test 6	4/01 & 22/03 (2-6AM)
Test 7	4/01 & 29/03 (2-6AM)

Tabla N°4.18. Combinaciones de días.

Como se puede apreciar, los primeros tres test corresponden a combinaciones de registros entre lunes de Enero, lo que corresponde a registros obtenidos durante la misma condición estructural basándose en el hecho de que aún no ocurría el terremoto del 27F. Los últimos cuatro test contemplan registros de Enero y Marzo, por lo que es de esperar que exista algún tipo de clasificación en los resultados. Además, como ya ha sido mencionado, solo se está tomando en consideración los registros correspondientes a las horas nocturnas, con lo que se deja de lado cualquier variación por la excitación o cambios en el uso del edificio post-terremoto.

Resultados.

Recordando que el indicador de la performance de la clasificación corresponde al porcentaje de objetos correctamente asignados, en las Figuras 4.19 y 4.20 se muestra un resumen al haber aplicado la metodología a cada parámetro por separado.

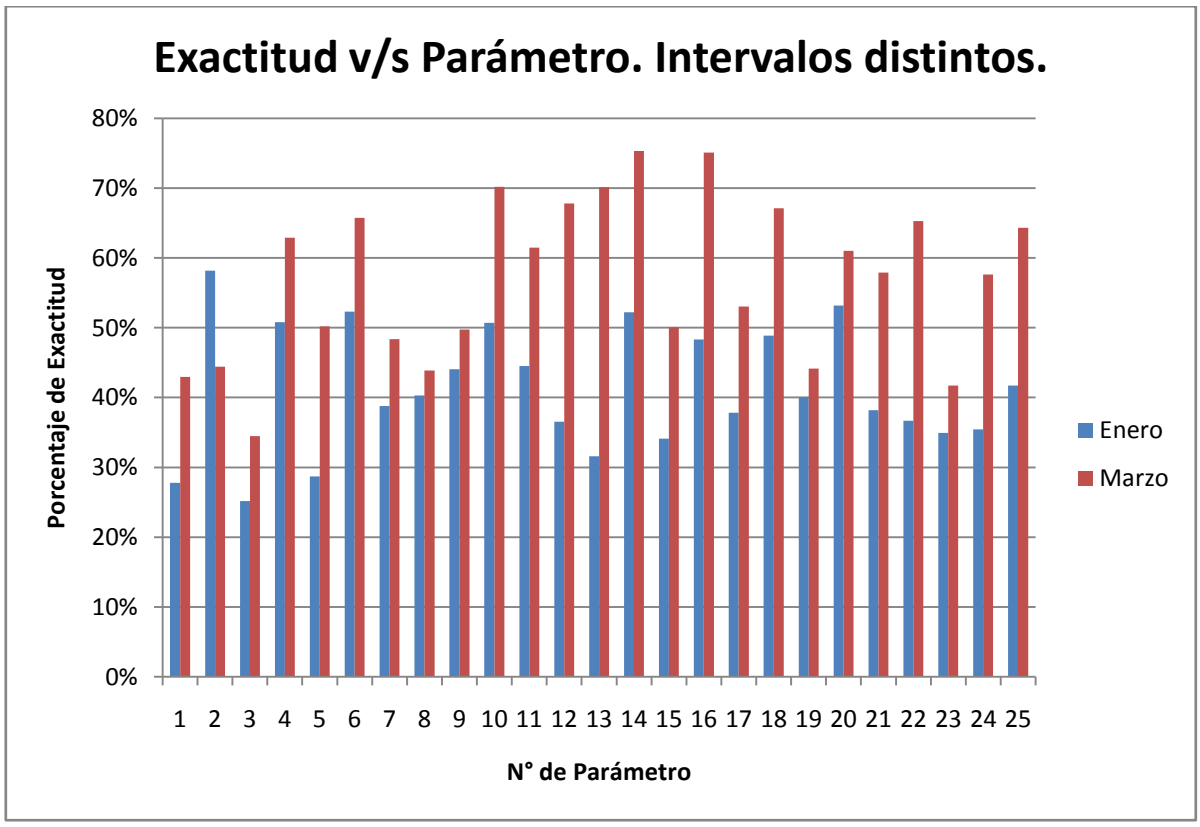


Figura N°4.19. Resumen de exactitud de asignaciones para los 25 primeros coeficientes.

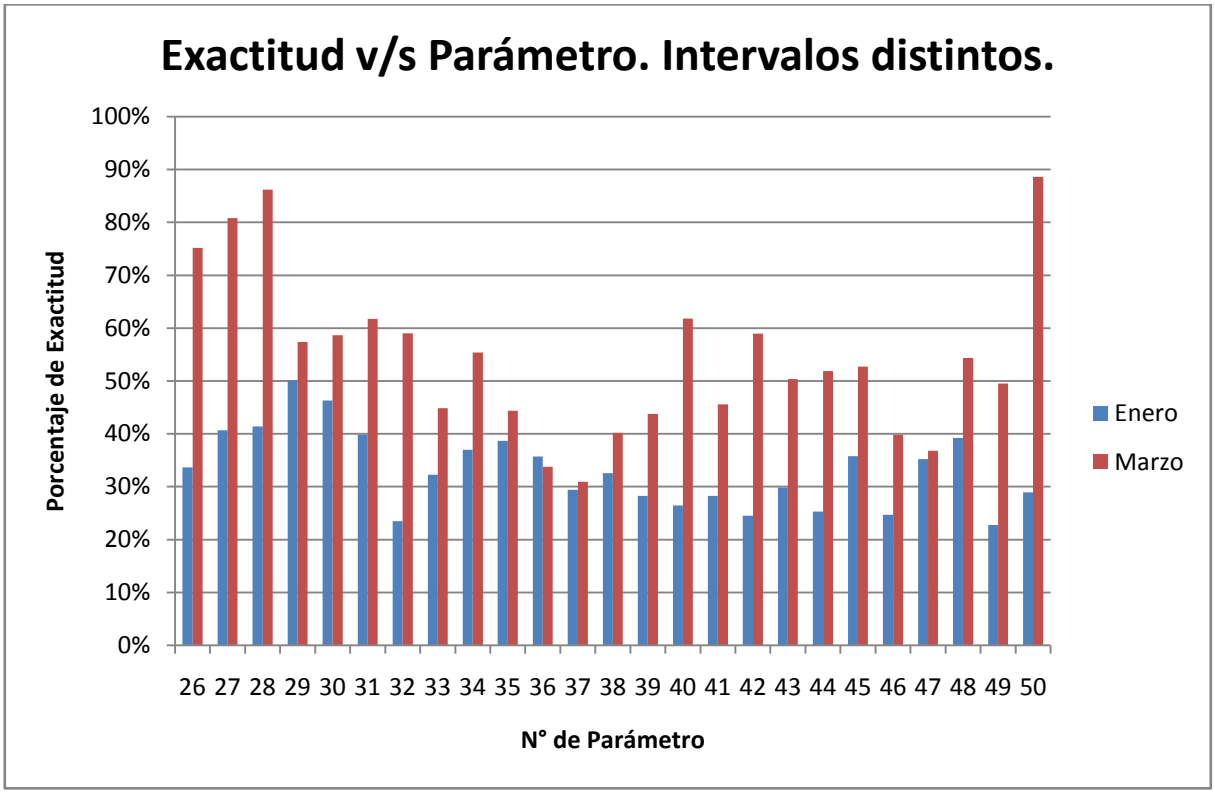


Figura N°4.20. Resumen de exactitud de asignaciones para los últimos 25 coeficientes.

Las Figuras 4.19 y 4.20 se crearon a partir de las Tablas A.1 y A.2 incluidas en el Anexo A, las que contienen los resultados para los 7 test y para cada parámetro, habiendo considerado rangos distintos para el cálculo de los histogramas de cada uno de los parámetros. A modo de resumen, las 4.19 y 4.20 son el promedio de los resultados globales para todos los parámetros y los test correspondientes a combinaciones de Enero-Enero, y Enero-Marzo. En estas figuras se aprecia que cuando se incluyen en el análisis registros provenientes de enero y marzo, hay un aumento considerable en la exactitud de la clasificación. Esto quiere decir que el algoritmo de agrupamiento utilizado es capaz de percibir diferencias entre los objetos de Enero y Marzo, aumentando la cantidad de registros correctamente asignados. Sin embargo, en promedio, ningún parámetro logra clasificar con 100% de exactitud, lo que implica que la metodología no funciona correctamente como clasificador, pero sí como característica sensible a los cambios estructurales. En otras palabras, no se puede crear una partición que represente estados estructurales a partir de registros de aceleración utilizando series AR, pero sí se puede generar un indicador con el que se permita comparar registros obtenidos en momentos distintos, a partir del cual se podría probar otro algoritmo clasificador, o idear una nueva estrategia para encontrar una agrupación de estados estructurales.

En las Tablas A.3 y A.4 del anexo, se encuentran los resultados considerando distintos intervalos. Dichos resultados muestran básicamente la misma tendencia que la indicada en la Figuras 4.19 y 4.20. Para poder comparar con mayor facilidad, considerar la Tabla 4.19, que contiene un resumen de los promedios para las dos posibilidades de intervalos.

		Exactitud	
	Días comparados	Intervalos Distintos	Intervalos Idénticos
Test 1	4/01 & 11/01	37%	43%
Test 2	4/01 & 18/01	37%	41%
Test 3	4/01 & 25/01	38%	43%
Test 4	4/01 & 8/03	56%	54%
Test 5	4/01 & 15/03	57%	57%
Test 6	4/01 & 22/03	53%	52%
Test 7	4/01 & 29/03	56%	49%
Resumen			
	4/01 v/s	Intervalos Distintos	Intervalos Idénticos
	Enero	**37%**	**42%**
	Marzo	**56%**	**53%**

Tabla N°4.19. Resultados promedios.

Recordando que el porcentaje de exactitud indica la cantidad de registros correctamente asignados, al observar la Tabla 4.19 se ven evidencias claras de que cuando se comparan series de parámetros AR obtenidas durante el mes de Enero, el algoritmo de agrupamiento no es capaz de detectar diferencias entre los objetos simbólicos asociados a dichas series, consiguiendo alrededor de un 40% de exactitud. Esto quiere decir que las asignaciones mezcla los registros y el resultado es más o menos aleatorio. En cambio, el porcentaje de exactitud en el caso de comparar series de Enero y Marzo aumenta en un 15% llegando a ser de alrededor de 55%. Si se toma en

consideración el hecho de que una asignación completamente aleatoria dos estados está marcada por un porcentaje del 50%, es destacable que para registros de Enero-Enero el resultado esté por debajo de este límite, mientras que para registros de Enero-Marzo se esté sobre el 50%.

En relación al uso de intervalos distintos o idénticos para la creación de los histogramas, se aprecia que el uso de intervalos distintos, calculados a partir de un estudio de los rangos típicos para cada parámetro, hace que la diferencia entre los resultados antes y después del terremoto sea de casi un 20% de mayor exactitud, en comparación con la diferencia de un 10% para intervalos idénticos para todos los parámetros. Esto quiere decir que si se utiliza el porcentaje de exactitud como un indicador de ocurrencia de cambios estructurales, el considerar intervalos distintos amplifica por un factor igual a dos la sensibilidad.

Por último, otro resultado destacable es que existen algunos parámetros que muestran una sensibilidad superior al resto. En particular, el último de los coeficientes logró clasificar correctamente al 100% de los registros, indicando la existencia de dos estados estructurales, para dos de los 4 test Enero-Marzo.

4.5. Conclusiones.

Uno de los resultados más llamativos del presente capítulo se encontró al analizar el valor RMS para distintas condiciones estructurales en los ensayos de laboratorio. Llamativamente, el RMS mostró una disminución en las condiciones con daño aplicado, lo que implica que en dichos casos los modelos autoregresivos ajustaban de mejor forma los registros de aceleración. Es muy difícil encontrar una explicación a este hecho, que incluso se mantuvo para distintas excitaciones basales. La dificultad recae en que no hay una clara relación entre los modos de vibración de la estructura y los coeficientes de los modelos AR. Un posible estudio futuro sería filtrar la respuesta de la estructura para eliminar el contenido de modos superiores, y con eso estudiar nuevamente el orden óptimo y el valor de los coeficientes, buscando encontrar la relación entre modos y parámetros AR.

Un segundo resultado importante de destacar, tiene relación con el largo de las ventanas con el cual realizar los ajuste de los modelos para el cálculo de los parámetros AR. Se estudiaron distintos largos de ventana para encontrar alguna convergencia, lo cual se pudo evidenciar pero solo en el caso de RMS, que alcanza rápidamente un valor estable, al contrario de lo que ocurrió con el promedio y desviación estándar de los parámetros. Esto implica que el valor de los parámetros y su variabilidad están fuertemente influenciadas por el tamaño de la ventana de ajuste. No obstante, manteniendo un mismo largo para todo el análisis se logra que el sesgo introducido sea parejo.

En los resultados obtenidos para ensayos de laboratorio se destaca que se logran porcentajes de exactitud sobre el 50% y en algunos condiciones estructurales se alcanza el 100%. En general, se aprecia que a mayor nivel de daño hay mayor probabilidad de lograr una asignación correcta. Sin embargo, como los algoritmos de agrupamiento son solo de clasificación, no es factible utilizarlos para medir la gravedad de los cambios estructurales. Además, se observó que lo ideal es incluir en el análisis tantos sensores como sea posible, eliminando aquellos que posean problemas evidentes, lo cual también es válido en relación a los parámetros AR, si bien se recomienda que los parámetros AR sean analizados por separado ya que hay algunos que muestran una sensibilidad mucho mayor que otros.

Para el estudio de la metodología utilizando registros reales, se obtuvo que es muy difícil y poco probable que la clasificación entregue una separación perfecta de estados estructurales.

Sin embargo, se mostró que es posible utilizar el resultado de la clasificación para calcular un indicador de exactitud con el cual se puede realizar una aseveración acerca del la ocurrencia de un cambio estructural, ya que los resultados durante un periodo sin cambios muestran un comportamiento aleatorio, mientras que cuando existe algún cambio los resultados tienden a agrupar los registros en los conjuntos reales. A su vez, nuevamente se encontró que algunos parámetros tienen una sensibilidad mucho mayor que otros.

En general, se puede decir que los resultados de laboratorio fueron mejores que los de la Torre Central. Esto puede deberse a que los registros de los ensayos de laboratorio fueron cortos, lo que derivó en poder realizar el análisis de agrupamiento con pocos objetos y por tanto, con pocas posibilidades de agrupamiento. Además, no hay certeza respecto al nivel del cambio estructural producto del terremoto, y el sistema de adquisición de la Torre Central contaba con menos sensores. Estos puntos hacen meditar sobre el hecho de que aún se pueden hacer mejores pruebas, teniendo mayor seguridad sobre las condiciones estructurales y con un sistemas de adquisición más completo, a partir de las cuales los resultados pueden mejorar.

Por último, existen mejoras que se pueden introducir a la metodología para obtener asignaciones más exactas. Una de ellas es considerar la normalización de las series de aceleración al momento de realizar los cálculos de los parámetros AR. En los ensayos de laboratorio esto se realizó para poder llevar las series a un nivel semejante de energía, pero en los registros reales no se aplicó una normalización ya que se compararon series obtenidas para una misma condición operacional y de excitación. Además, se puede avanzar en la metodología para realizar el cálculo de los histogramas de los parámetros AR. Por ejemplo, si se concatenan dos registros seguidos, se pueden tener mayor numero de ventanas, lo que se traduce en mayor observaciones de parámetros y con ello poder utilizar un número de contenedores mayor. Otra opción es hacer un estudio sobre la distribución misma de los parámetros, y así no solo usar una amplitud distinta para cada parámetro sino también distintos números de intervalos. Sin embargo, como fue mencionado, esto implica que haya que utilizar algunas señales como referencia, lo que resulta en un algoritmo con línea base de referencia.

A modo de extensión, una metodología sin línea base de referencia, sería hacer uso de los errores residuales y/o valor RMS de las ventanas utilizadas para generar las series de parámetros AR. Por una parte, los errores tienen un comportamiento mucho más estable que los parámetros mismos, y cuentan con la ventaja de que tienen media igual a cero, lo que permitiría hacer uso de intervalos idénticos de forma global. El valor RMS, además, ya se mostró que muestra un llamativo comportamiento como feature, la que no depende de una línea de referencia y que podría ser clasificada fácilmente por algún algoritmo de agrupamiento.

5. Conclusiones.

En esta sección es prudente recordar que el objetivo principal de la tesis consiste en ser capaz de identificar la ocurrencia de un cambio estructural, respondiendo al primer nivel de SHM mencionado en la introducción a la tesis. Es decir, el alcance de la metodología consiste en obtener diferencias entre los resultados aplicados a un conjunto que contenga registros de un solo estado estructural, versus los resultados aplicados a conjuntos conteniendo registros de más de un estado estructural.

La conclusión más importante del capítulo 2, relacionado al uso de objetos simbólicos a partir de los registros brutos de aceleración, es que sí se puede extraer información acerca del estado estructural mediante una combinación adecuada de los algoritmos, pero que es muy difícil lograr clasificar perfectamente los registros en los conjuntos definidos por las condiciones estructurales presentes en el set de datos. Específicamente, se aprecia una clara diferencia entre los resultados cuando se comparan objetos de la misma condición estructural versus los resultados cuando se comparan objetos de distintos estados. En relación a la gran influencia que tiene la energía de entrada al sistema sobre el resultado, se concluye que para detectar cambios en el comportamiento estructural, siempre es mejor comparar registros obtenidos durante el mismo régimen de excitación. Además, de los algoritmos utilizados se obtuvo que el uso de histograma aporta con mayor información que el uso de intervalos intercuartiles pero con una dimensionalidad mayor. Junto con esto, en cuanto a algoritmos de agrupamiento, se observaron comportamientos similares entre K-Means (distancia euclideana) y Dyamic-Medoids (distancia categórica). De todas formas, si se quiere extraer la mayor información posible, se recomienda el uso de la distancia categórica ya que su formulación está especialmente diseñada para histogramas. El algoritmo jerárquico, por su parte, mostró ser el de menor rendimiento. Otro punto importante del capítulo, fue el desarrollo de dos métodos para la limpieza de objetos outliers, con los que se consiguió mejorar sustancialmente la sensibilidad de la metodología.

En el capítulo 3, que estudia el comportamiento de algunos indicadores de cambio estructural basado en el ajuste de modelos autoregresivos, se concluye que estas características, al utilizar señales de referencia, aumentan notablemente la sensibilidad a los cambios. Es importante hacer la distinción de que en este caso, la línea base de referencia se utiliza al momento de extraer las características y no en el algoritmo de clasificación. El resultado, por tanto, está influenciado fuertemente por la selección de la referencia. Aún así, la sensibilidad es tan grande que se podrían clasificar perfectamente los estados presentes. En específico, se concluye que la obtención de un orden óptimo para los modelos autoregresivos no es trivial, ya que los índices AIC y RMS no muestran un mínimo claro, pero ya que tienen un decaimiento rápido y posterior comportamiento estable, se logra un muy buen ajuste de las señales incluso para un orden aproximado. Otro punto destacable, es que tanto el número de outliers como la distancia de Mahalanobis, parecen tener un comportamiento creciente con respecto al nivel de daño introducido, y eso abre la puerta a que sean indicadores para responder a un nivel superior de identificación de daño, relacionado con la magnitud del cambio estructural.

En relación al capítulo 4, que aplica la metodología de agrupamiento sobre series de parámetros AR, los resultados son análogos al capítulo 2, es decir, se logra identificar la ocurrencia de un cambio estructural, pero en promedio no se consigue una clasificación perfecta de los registros en los conjuntos reales. Pese a lo anterior, sí se pudo observar la existencia de parámetros autoregresivos que poseen una sensibilidad mayor que el resto y en algunos casos particulares, se obtuvo una clasificación con 100% de exactitud al comparar registros de distintas condiciones estructurales. Esto es un punto a favor al uso de la metodología de parámetros AR,

ya que los resultados del capítulo 2, que se aplica directamente sobre las series de aceleración, son muy similares a los obtenidos en el capítulo 4, que necesita varios pasos y procedimientos extra. Además, se han identificados algunas mejoras posibles que aumentarían la sensibilidad de la metodología. Por ejemplo, se podrían normalizar las señales de aceleración al momento de realizar los cálculos de parámetros AR, o se podrían concatenar registros y así obtener objetos simbólicos más robustos y con mayor números de ventanas e intervalos.

Otro punto destacable del capítulo 4, es que se encontró una nueva feature basada en el valor RMS al momento de ajustar los modelos autoregresivos. Se encontró que el valor RMS disminuye cuando se introduce un cambio estructural. Pese a no haber claridad en el porqué ocurre esta disminución, sería posible utilizarlo y de forma no supervisada y libre de referencia. Otra posibilidad consiste en utilizar los errores residuales mismos, ya que poseen un comportamiento mucho más estable y con media cero, lo que permite realizar el cálculo de histogramas de una forma mucho más genérica.

En definitiva, se concluye que el uso de modelos autoregresivos para ajustar series de aceleración sí aporta mayor información sensible respecto al estado estructural y aún quedan muchas posibilidades de estudio y desarrollo.

6. Bibliografía.

Alves, Vinicius et al. 2015. "Novelty Detection for SHM Using Raw Acceleration Measurements." *Structural Control and Health Monitoring* 22(9): 1193–1207. http://doi.wiley.com/10.1002/stc.1741 (October 28, 2015).

Billard, L, and E Diday. 2002. "Symbolic Data Analysis : Definitions and Examples."

———. 2006. 57 Merrill-Palmer Quarterly *Symbolic Data Analysis: Conceptual Statistics and Data Mining*. eds. Paolo (University of Pavia) Giudici and Geof (Colorado State University) Givens. Wiley.

Billard, Lynne, and Edwin Diday. 2007. *Symbolic Data Analysis: Conceptual Statistics and Data Mining - Lynne Billard, Edwin Diday*. Wiley. http://www.wiley.com/WileyCDA/WileyTitle/productCd-0470090162.html (October 28, 2015).

Calinksi, and Harabasz. 1974. "A Dendrite Method for Cluster Analysis." *Communications in Statistics* (3:1): 1–27. http://dx.doi.org/10.1080/03610927408827101.

Cury, Alexandre, Christian Crémona, and Edwin Diday. 2010. "Application of Symbolic Data Analysis for Structural mCury, Alexandre, Christian Crémona, and Edwin Diday. 2010. 'Application of Symbolic Data Analysis for Structural Modification Assessment.' Engineering Structures 32(3): 762–75. http://dx.doi.org/10.101." *Engineering Structures* 32(3): 762–75. http://dx.doi.org/10.1016/j.engstruct.2009.12.004.

Figueiredo, Eloi et al. 2011. "Influence of the Autoregressive Model Order on Damage Detection." *Computer-Aided Civil and Infrastructure Engineering* 26: 225–38.

Gowda, Kc, and E Diday. 1991. "Symbolic Agrupamiento Using a New Dissimilarity Measure." *Pattern Recognition* 24(6): 567–78. http://www.sciencedirect.com/science/article/pii/003132039190022W.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2011. 1 Elements *The Elements of Statistical Learning*. http://www.springer.com/us/book/9780387848570.

Ichino, M, and H Yaguchi. 1994. "Generalized Minkowski Metrics for Mixed Feature-Type Data Analysis." *IEEE Transactions on Systems Man and Cybernetics* 24(4): 698–708.

Omenzetter, Piotr, and James Mark William Brownjohn. 2006. "Application of Time Series Analysis for Bridge Monitoring." *Smart Materials and Structures* 15: 129–38.

Peeters, Bart. 2000. Status: Published "System Identification and Damage Detection in Civil Engineering." https://lirias.kuleuven.be/handle/123456789/212304.

Rytter, Anders. 1993. *Vibrational Based Inspection of Civil Engineering Structures*.

Santos, João, Christian Cremona, et al. 2014. "Static-Based Early-Damage Detection Using Symbolic Data Analysis and Unsupervised Learning Methods." *Frontiers of Structural and Civil Engineering* 9(1): 1–16.

Santos, João, André D. Orcesi, Christian Crémona, and Paulo Silveira. 2014. "Baseline-Free Real-Time Assessment of Structural Changes." *Structure and Infrastructure Engineering: Maintenance, Management, Life-Cycle Design and Performance* 00(January 2015): 37–41. http://dx.doi.org/10.1080/15732479.2013.858169.

Santos, João Pedro. 2014. "Smart Structural Health Monitoring Techniques for Novelty

Identification in Civil Engineering Structures." Universidade de Lisboa.

Santos, João Pedro, Christian Crémona, André D. Orcesi, and Paulo Silveira. 2013. "Multivariate Statistical Analysis for Early Damage Detection." *Engineering Structures* 56: 273–85. http://linkinghub.elsevier.com/retrieve/pii/S0141029613002459.

Sohn, H., Jerry Czarnecki, and Charles R. Farrar. 2000. "Structural Health Monitoring Using Statistical Process Control." *Structural Engineering* (November): 1356–63.

Sohn, H., C. R. Farrar, and N. F. Hunter. 2001. "Data Normalization Issue for Vibration-Based Structural Health Monitoring." *Proceedings of the International Modal Analysis Conference - IMAC* 1: 432–37. http://www.scopus.com/inward/record.url?eid=2-s2.0-0035052073&partnerID=tZOtx3y1.

Sohn, H., K. Worden, and C. R. Farrar. 2002. "Statistical Damage Classification Under Changing Environmental and Operational Conditions." *Journal of Intelligent Material Systems and Structures* 13(9): 561–74.

Sohn, Hoon, and Charles R Farrar. 2001. "Damage Diagnosis Using Time Series Analysis of Vibration Signals." *Smart Materials and Structures* 10: 446–51.

Sohn, Hoon, Charles R. Farrar, Norman F. Hunter, and Keith Worden. 2001. "Structural Health Monitoring Using Statistical Pattern Recognition Techniques." *Journal of Dynamic Systems, Measurement, and Control* 123(4): 706.

Sohn, Hoon, Michael L Fugate, Charles R Farrar, and Los Alamos. 2000. "Damage Diagnosis Using Statistical Process Control."

Villalpando, Pastor, Viviana Meruane, Rubén Boroschek, and Marcos Orchard. 2016. "Damage Location by Maximum Entropy Method on a Civil Structure." In *Dynamics of Civil Structures, ,* 105–16.

Worden, K., and J. M. Dulieu-Barton. 2004. "An Overview of Intelligent Fault Detection in Systems and Structures." *Structural Health Monitoring* 3(1): 85–98. http://shm.sagepub.com/cgi/doi/10.1177/1475921704041866.

Anexo A

04-ene	Combinado con						
Parámetro	11-ene	18-ene	25-ene	08-mar	15-mar	22-mar	29-mar
1	20%	23%	40%	46%	33%	44%	48%
2	54%	70%	50%	43%	36%	36%	63%
3	29%	30%	17%	50%	30%	31%	28%
4	41%	59%	52%	50%	61%	73%	68%
5	23%	33%	30%	50%	39%	67%	45%
6	40%	54%	63%	70%	57%	67%	70%
7	38%	56%	23%	55%	45%	33%	61%
8	53%	41%	27%	39%	59%	36%	41%
9	33%	56%	43%	42%	70%	44%	43%
10	60%	45%	47%	96%	54%	38%	93%
11	50%	54%	30%	96%	50%	50%	50%
12	30%	26%	53%	71%	88%	63%	50%
13	27%	45%	23%	88%	57%	42%	93%
14	53%	50%	53%	88%	81%	52%	80%
15	27%	36%	40%	88%	30%	46%	36%
16	50%	58%	37%	88%	90%	42%	80%
17	33%	43%	37%	39%	47%	44%	82%
18	60%	50%	37%	46%	90%	56%	77%
19	50%	43%	27%	38%	50%	46%	43%
20	62%	41%	57%	41%	77%	50%	77%
21	47%	41%	27%	41%	57%	58%	76%
22	21%	35%	53%	36%	97%	65%	63%
23	45%	20%	40%	43%	43%	42%	38%
24	36%	31%	39%	29%	97%	52%	53%
25	38%	33%	54%	93%	30%	62%	72%

Tabla A.1. Resultados de análisis para los primeros 25 parámetros. Intervalos Distintos.

04-ene	Combinado con						
Parámetro	11-ene	18-ene	25-ene	08-mar	15-mar	22-mar	29-mar
26	24%	23%	54%	45%	97%	59%	100%
27	37%	26%	59%	100%	97%	58%	69%
28	45%	18%	62%	52%	97%	96%	100%
29	35%	54%	62%	70%	47%	58%	55%
30	45%	39%	54%	45%	60%	93%	37%
31	36%	30%	53%	60%	57%	73%	57%
32	19%	35%	17%	33%	83%	83%	37%
33	25%	30%	41%	43%	33%	38%	64%
34	30%	31%	50%	30%	83%	75%	33%
35	32%	32%	52%	46%	37%	31%	64%
36	18%	30%	59%	31%	33%	41%	30%
37	33%	30%	25%	30%	29%	37%	28%
38	40%	39%	19%	53%	38%	33%	37%
39	32%	30%	23%	68%	36%	41%	30%
40	21%	40%	18%	54%	100%	64%	30%
41	33%	30%	21%	61%	46%	39%	37%
42	18%	31%	25%	54%	88%	41%	53%
43	32%	33%	24%	54%	54%	64%	30%
44	29%	27%	21%	61%	50%	52%	45%
45	62%	22%	24%	50%	60%	61%	40%
46	19%	33%	22%	45%	30%	48%	36%
47	40%	30%	36%	40%	31%	36%	40%
48	36%	57%	25%	58%	35%	25%	100%
49	20%	27%	21%	59%	25%	72%	41%
50	46%	16%	25%	100%	63%	92%	100%

Tabla A.2. Resultados de análisis para los últimos 25 parámetros. Intervalos Distintos.

04-ene	Intervalos Idénticos. Combinado con						
Parametro	11-ene	18-ene	25-ene	08-mar	15-mar	22-mar	29-mar
1	29%	21%	25%	33%	33%	27%	32%
2	29%	40%	32%	29%	36%	28%	30%
3	26%	18%	18%	64%	30%	46%	38%
4	97%	97%	97%	57%	61%	61%	59%
5	69%	25%	50%	43%	39%	41%	64%
6	97%	97%	97%	57%	57%	59%	57%
7	48%	57%	53%	56%	45%	59%	48%
8	97%	97%	97%	62%	59%	55%	53%
9	39%	59%	60%	84%	70%	37%	21%
10	70%	96%	96%	50%	54%	46%	47%
11	30%	30%	30%	58%	50%	44%	33%
12	31%	47%	25%	77%	88%	70%	40%
13	23%	46%	32%	31%	57%	73%	37%
14	67%	47%	52%	81%	81%	38%	53%
15	30%	54%	47%	29%	30%	38%	50%
16	52%	23%	44%	81%	90%	52%	63%
17	43%	44%	30%	48%	47%	24%	29%
18	28%	37%	24%	82%	90%	85%	67%
19	70%	19%	60%	57%	50%	50%	31%
20	62%	48%	58%	86%	77%	73%	53%
21	31%	65%	53%	40%	57%	58%	52%
22	67%	38%	62%	43%	97%	93%	77%
23	63%	62%	55%	64%	43%	31%	55%
24	54%	67%	54%	90%	97%	96%	70%
25	25%	40%	29%	78%	30%	81%	66%

Tabla A.3. Resultados para los primeros 25 parámetros. Intervalos Idénticos.

04-ene	Intervalos Idénticos. Combinado con						
Parametro	11-ene	18-ene	25-ene	08-mar	15-mar	22-mar	29-mar
26	52%	61%	45%	50%	97%	96%	100%
27	27%	69%	59%	52%	97%	62%	72%
28	66%	30%	62%	97%	97%	96%	30%
29	21%	48%	37%	54%	47%	76%	48%
30	24%	41%	42%	69%	60%	96%	60%
31	17%	17%	23%	47%	57%	31%	43%
32	33%	41%	52%	62%	83%	57%	60%
33	37%	26%	50%	68%	33%	60%	71%
34	24%	41%	52%	60%	83%	71%	67%
35	27%	21%	56%	30%	37%	67%	28%
36	33%	26%	28%	63%	33%	31%	23%
37	30%	30%	37%	50%	29%	39%	27%
38	28%	26%	19%	30%	38%	35%	23%
39	32%	23%	35%	57%	36%	33%	37%
40	32%	38%	36%	43%	100%	83%	53%
41	24%	18%	21%	33%	46%	38%	40%
42	35%	21%	28%	30%	88%	36%	40%
43	56%	37%	19%	33%	54%	36%	47%
44	23%	18%	21%	38%	50%	46%	43%
45	56%	32%	26%	33%	60%	27%	54%
46	54%	30%	20%	27%	30%	32%	40%
47	52%	20%	19%	27%	31%	28%	45%
48	30%	21%	22%	30%	35%	25%	30%
49	17%	23%	27%	31%	25%	25%	25%
50	39%	22%	26%	100%	63%	30%	97%

Tabla A.. Resultados para los últimos 25 parámetros. Intervalos Idénticos.