



Universidad de Concepción
Facultad de Ciencias Físicas y Matemáticas
Magíster en Ciencias mención Física

**TESIS PARA OPTAR AL GRADO DE MAGISTER
EN CIENCIAS CON MENCIÓN EN FÍSICA.**

**MULTIFRACTALIDAD
EN SERIES DE TIEMPO DE $PM_{2.5}$
SOBRE SANTIAGO DE CHILE.**

Rodrigo Alfredo Valenzuela Melivilu.

Tutor: Dr. Félix Borotto Cháves
Departamento de Física
Facultad de Ciencias Físicas y Matemáticas
Universidad de Concepción

Cotutor interno: Dr. Rodrigo Abarca del Río
Departamento de Geofísica
Facultad de Ciencias Físicas y Matemáticas
Universidad de Concepción

Cotutor externo: Dr. Giovanni Salini Calderón
Departamento de Matemática y Física Aplicadas
Universidad Católica de la Santísima Concepción

2016

Tabla de Contenido

Lista de figuras	VII
Lista de tablas	IX
Resumen	x
Agradecimientos	XI
1. Introducción a la Contaminación Atmosférica	1
1.1. Contaminación atmosférica	1
1.2. Tipos de fuentes Contaminantes	2
1.3. Contaminación Atmosférica: Efectos en la Salud Humana.	4
1.4. Normativa Asociadas a la Contaminación Atmosférica	5
1.5. Bases de datos disponibles en el País en estaciones EMRP	7
1.5.1. Zona Norte	8
1.5.2. Zona Sur	9
1.5.3. Región Metropolitana de Santiago de Chile	10
1.6. Contaminación atmosférica en Santiago de Chile	11
1.7. Procedimientos en casos extremos en Santiago de Chile.	13

2. Series de tiempo	15
2.1. Estadística aplicada al análisis de Series de Tiempo	15
2.1.1. Estacionaridad	17
2.1.2. Correlación	17
2.2. Regresión Lineal y tendencia	19
2.2.1. Significancia Estadística de la Tendencia	20
2.3. Análisis de Fourier	21
2.3.1. Densidad Espectral	22
2.3.2. Relación entre densidad espectral y períodograma	22
2.3.3. Intervalo de confianza de la densidad espectral	24
2.3.4. Coherencia	26
2.3.5. Significancia estadística de la coherencia	27
2.4. Transformada de Ondeletas	29
2.4.1. Elección de la escala s	32
2.4.2. Cono de Influencia	32
2.4.3. Reconstrucción de señales y filtrado	33
2.4.4. Espectro de Ondeletas o Potencia espectral de Ondeletas	33
2.4.5. Espectro Teórico y Nivel de Significancia	34
2.4.6. Espectro cruzado de ondeletas	35
2.4.7. Coherencia de ondeletas	36
3. Multifractalidad en Series de Tiempo	37
3.1. Fractales.	37
3.1.1. Ejemplos de fractales.	38

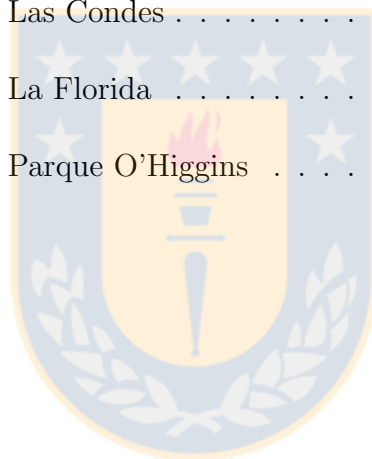
3.1.2. Dimensión Fractal	40
3.2. Formalismo Multifractal	42
3.2.1. Cálculo del Espectro Multifractal	44
3.2.2. Relación entre $\tau(q)$ y $h(q)$	46
3.3. Series de Tiempo y formalismo multifractal.	48
3.3.1. Persistencia, Correlaciones de largo y corto plazo en series de tiempo.	48
3.3.2. Otros ejemplos.	52
4. Análisis y Resultados	56
4.1. Base de Datos	56
4.2. Metodología	57
4.3. Análisis Estadístico y Tendencias	59
4.4. Ondeletas	62
4.5. Multifractalidad	63
5. Conclusión y Discusión	69
A. Transformadas de Ondeletas otras comunas	71
B. Trabajos Presentados	74
Bibliografía	78



Índice de figuras

1.1. Plaza de Tiananmén	2
1.2. Tipos de Fuentes de Contaminación	2
1.3. Diámetros Característicos	3
1.4. Estructura del sistema respiratorio	4
1.5. Ciudad de Santiago de Chile	12
3.1. Ejemplos de Fractales	39
3.2. Atractor de Lorenz	40
3.3. Modelo Bifractal con N=32768 iteraciones	51
3.4. Resultados teóricos y del método MF DFA	52
3.5. Señales Sintéticas	53
3.6. Resultados Señales Sintéticas	53
3.7. Series de tiempo multifractal de camino aleatorio	54
3.8. Espectro multifractal con truncamiento	55
4.1. Mapa del Área metropolitana de Santiago de Chile	57
4.2. Series de tiempo horarias de Concentración de PM _{2,5} en Santiago de Chile	58
4.3. Medias diarias, mensuales y anuales	61

4.4. Evolución de medias invernales y estivales	61
4.5. Análisis de ondeletas en Pudahuel	64
4.6. Espectro Global de Ondeletas	65
4.7. Análisis de Pudahuel 1	66
4.8. Análisis de Pudahuel 2	66
4.9. Análisis de Las Condes	67
4.10. Análisis de La Florida	68
4.11. Análisis de Parque O’Higgins	68
A.1. Análisis de ondeletas en Las Condes	71
A.2. Análisis de ondeletas en La Florida	72
A.3. Análisis de ondeletas en Parque O’Higgins	73



Índice de tablas

1.1. Estaciones de Monitoreo	7
1.2. Zona Norte	8
1.3. Zona Sur	9
1.4. Región Metropolitana	11
1.5. Medidas frente a episodios críticos de contaminación	14
4.1. Coordenadas lat-long	56
4.2. Medias y Desviaciones Estándar (std)	59
4.3. Tendencias y resultados del test de Mann-Kendall	59
4.4. Medias Invernales y Estivales	59

Resumen

Las herramientas desarrolladas por el análisis de series de tiempo, junto con el estudio de sistemas dinámicos y caos permiten estudiar una gran cantidad de fenómenos, ya sean físicos, biológicos o químicos en sus ramas más fundamentales hasta aplicaciones en la industria, medicina, economía. Una de estas aplicaciones se encuentra centrada en estudio de gases y aerosoles.

En particular el problema de la contaminación del aire debido a factores antropógenos como la industria, el tráfico vehicular y el crecimiento de la población urbana han hecho de Santiago de Chile una ciudad saturada, por especies como PM_{10} y $PM_{2.5}$.

En el Capítulo 1, se hará una introducción al concepto de contaminación atmosférica, las fuentes de emisión existentes, la normativa asociada, la situación de las estaciones de monitoreo en diferentes regiones del país además de la región metropolitana y los procedimientos para enfrentar los episodios críticos de contaminación.

A continuación en el Capítulo 2 se presentan los conceptos estadísticos básicos asociados al estudio de series de tiempo, para luego entrar en el análisis espectral de Fourier y finalmente en el análisis multi resolución utilizando ondeletas.

El Capítulo 3 introduce el tema de la fractalidad y multifractalidad de estructuras complejas y como éstas subyacen en las series de tiempo, así como cuales son las variables que permiten caracterizar dichos sistemas.

El Capítulo 4 muestra los resultados al aplicar estas técnicas al estudio de series de contaminación para finalmente en el Capítulo 5 mostrar las conclusiones.

Agradecimientos

Quiero agradecer a mis padres Luis y Edith por darme la oportunidad de acceder a la educación superior y apoyarme aunque no sin reparos en el estudio de la física. A mis hermanos Alejandro, Cesar y Claudia por cuidar de mi desde pequeño. A mi novia Julie quien me ha impulsado a ser un mejor hombre ayudándome a plantearme metas que en algún momento de mi vida pensé que eran imposibles. A mis amigos y todos aquellos que estuvieron en los momentos buenos y malos. Por creer en mí, aún cuando yo no podía hacerlo. Su amor y cariño me ayudaron a permanecer con ustedes salvándome la vida.

También quiero agradecer a mi profesor guía, Dr. Félix Borotto, por su invitación a trabajar con él en el mundo de sistemas complejos. Por su apoyo y comprensión en los momentos de duda tanto a nivel académico como personal. Por abrirme las puertas a ejercer la docencia a nivel superior y en paralelo hacer mis estudios de magíster.

Y a mis cotutores: al Profesor, Dr. Rodrigo Abarca del Río, por empujarme a seguir adelante. Aunque para avanzar en más de alguna ocasión tuve que retroceder hasta obtener los resultados correctos.

Al profesor, Dr. Hernán Astudillo, quien siempre aportó con nuevas ideas para abordar el problema.

Al profesor, Dr. Claudio Faúndez, con quien encontré un nuevo amigo con el cual hablar de fútbol y la vida, siempre teniendo una palabra de ánimo, un consejo y una buena broma que en más de alguna ocasión me sacaron una carcajada.

Al Profesor, Dr. Giovanni Salini, quien me introdujo en el tema de la contaminación atmosférica ayudándome a aplicar las herramientas aprendidas a un tema contingente e importante para el país.

Esta tesis contó con el apoyo parcial del Departamento de Física y Departamento de Geofísica de la Universidad de Concepción en conjunto con el Departamento de Matemáticas y Física Aplicadas de la Universidad Católica de la Santísima Concepción, Concepción, Chile. Se agradece a esta última el apoyo económico a través del proyecto DIN 444713(13/2016), Universidad Católica de la Santísima Concepción.

Capítulo 1

Introducción a la Contaminación Atmosférica

1.1. Contaminación atmosférica

Entenderemos por contaminación atmosférica a la liberación de una sustancia al medio ambiente en cantidades suficientemente grandes para producir daños a nivel ambiental, material y humano, por acción directa o indirecta del hombre. Las concentraciones de dicha sustancia varían con las condiciones atmosféricas llegando a niveles que pueden ser perjudiciales para la salud de las personas que ahí viven. La concentración de material particulado depende de los tipos de fuentes emisoras las cuales pueden ser naturales, como por ejemplo erupciones volcánicas, incendios forestales naturales por rayos. Otras fuentes de emisión son las artificiales (antropógenas) debido a la actividad humana, como por ejemplo quemas agrícolas, calefacción de hogares, tubos de escape de vehículos que ocupan combustibles fósiles, la actividad minera etc. Debemos notar que durante miles de años la atmósfera era capaz de adaptarse a los efectos generados por la contaminación natural, sin embargo, a partir de la revolución industrial en la segunda mitad del siglo XVIII el factor antropógeno ha influenciado a la atmósfera al punto de generar un cambio climático global.

Producto de la gran cantidad de fuentes contaminantes, grandes urbes del mundo se ven sometidas a condiciones donde la concentración de material particulado llega a niveles críticos trayendo una serie de problemas desde afecciones a la salud de la población hasta pérdidas económicas por ausentismo laboral. Un ejemplo de una ciudad sometida a diferentes concentraciones de contaminantes atmosféricos puede ser vista en la figura 1.1 que muestra la plaza de Tiananmén en la ciudad de Beijing, China donde a) muestra un episodio crítico de contaminación y b) condiciones normales.

Chile tampoco es ajeno a esta situación, siendo muchas las ciudades del país que presentan situaciones críticas, por ejemplo Antofagasta, Calama, Tocopilla, Talca, Chillan, Temuco y otras. En específico la ciudad que ha sido prioridad para el gobierno es la ciudad de Santiago de Chile, debido a los graves problemas generados durante las décadas de 1970

1.2. TIPOS DE FUENTES CONTAMINANTES

y 1980, además que es la capital del país y una de las ciudades más pobladas de América Latina [1, 2].

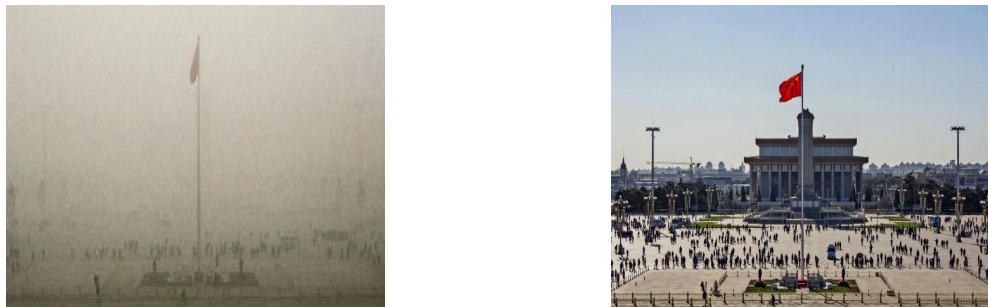


Figura 1.1: **Plaza de Tiananmén:** A la izquierda, episodio crítico de contaminación en Diciembre del 2015, a la derecha condiciones normales un mes antes. Fuente: <https://goo.gl/7S7zta>.

1.2. Tipos de fuentes Contaminantes

Las fuentes de contaminación se dividen en:

- I) **Fuentes fijas:** Corresponde a fuentes tales como las industrias, quemas agrícolas, leña residencial.
- II) **Fuentes móviles:** Corresponde a fuentes tales como automóviles particulares, transporte público, camiones, aviones, etc.



(a) Central termoeléctrica en Tocopilla.



(b) Vehículo con problemas de emisión de gases.

Figura 1.2: **Tipos de Fuentes de Contaminación:** a) Central termoeléctrica de Tocopilla, ejemplo típico de fuente fija. Fuente: <https://goo.gl/Gp3UIx>, b) Automóvil con problemas de emisión, ejemplo de fuente móvil. Fuente: <https://goo.gl/RzakH0>.

Los agentes contaminantes emitidos por estas fuentes se clasifican en:

- I) **Agentes contaminantes primarios:** Son sustancias emitidas por fuentes contaminantes y vertidas directamente a la atmósfera, siendo las más comunes el monóxido de carbono (CO), los hidrocarburos (NO_x), los óxidos sulfurosos (SO_2), dióxido de carbono (CO_2), compuestos orgánicos volátiles (COV), el material particulado *PM*, en general la mayoría de los hidrocarburos y las partículas y otros que se presentan más raramente. Los contaminantes primarios conservan su identidad desde que son emitidos hasta que son medidos por medio de algún instrumento.
- II) **Agentes contaminantes secundarios:** Estos se forman en el aire a partir de reacciones químicas y fotoquímicas entre los contaminantes primarios y los compuestos que están en la atmósfera. Entre ellos NO_2 , O_3 (ozono), SO_3 , HNO_3 , H_2SO_4 , la mayoría de los iones NO_3^- y SO_4^{2-} y los nitro-peróxidos de acilo conocidos por la sigla inglesa PAN (Peroxyacetyl nitrate), que es un potente irritante de los ojos y de la respiración en el smog fotoquímico.

Entre los elementos contaminantes de especial interés están los aerosoles, consistentes de material particulado muy fino menor a $10\mu\text{m}$, que contempla PM_{10} y $\text{PM}_{2.5}$ los cuáles están constituidos por un 38 % de polvo en el ambiente, 11 % de polvo antropógeno, 10 % nitrato, 8 % de amonio, 8 % sulfato, 8 % carbono orgánico, 7 % carbono vegetal, 7 % cloro y 3 % de otros elementos, según CENMA [3]. Estos agentes son los que crean mayores problemas en la salud humana ya que fácilmente llegan hasta los pulmones. En el PM_{10} distinguimos una fracción gruesa, aquella con un diámetro ϕ , $2.5\mu\text{m} < \phi < 10\mu\text{m}$ y una fracción fina $\phi < 2.5\mu\text{m}$ que corresponde al $\text{PM}_{2.5}$. La siguiente figura muestra las proporciones entre los tamaños de PM_{10} y $\text{PM}_{2.5}$ con respecto al diámetro del cabello humano.

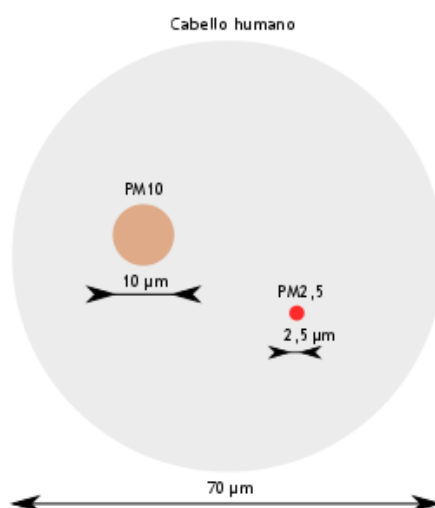
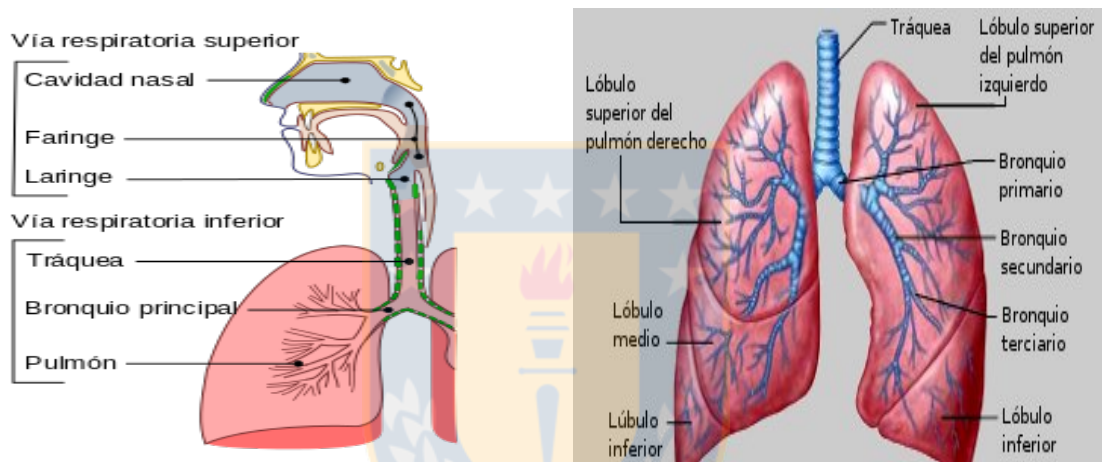


Figura 1.3: **Diámetros característicos:** Material particulado PM_{10} y $\text{PM}_{2.5}$ en comparación con el cabello humano. Fuente: <https://goo.gl/3A0qSJ>.

1.3. Contaminación Atmosférica: Efectos en la Salud Humana.

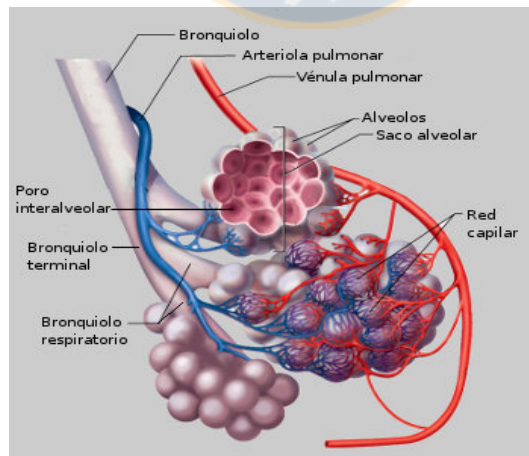
Existe en el mundo un seguimiento en el comportamiento de las sustancias químicas presentes en las masas de aire contaminado, ya que muchas de sus sustancias tiene características cancerígenas o mutagénicas, afectando por ende la salud de la población.

Esto se debe al alto grado de penetración y permanencia que ellas tienen en el sistema respiratorio. Las partículas de diámetro ϕ con un tamaño de $5\mu m$ se depositan en las vías respiratorias superiores (nariz), traquea y laringe, mientras que las de diámetro inferior tienen la probabilidad de depositarse en los bronquios y alvéolos (Ver figura 1.3).



(a) Tracto respiratorio superior e inferior.

(b) Tracto respiratorio inferior.



(c) Bronquios y Alveolos.

Figura 1.4: **Estructura del sistema respiratorio:** En el caso de la fracción gruesa de material particulado esta queda retenida en el tracto respiratorio superior. En cambio la fracción fina es capaz de llegar a la estructuras más pequeñas del tracto respiratorio inferior. Fuente: a) <https://goo.gl/bsvubI>, b) <https://goo.gl/mRBCsm>.

En el caso de Santiago de Chile, se han identificado los siguientes efectos debido el aumento de la fracción fina de material particulado $PM_{2.5}$:

- I) Aumento en el número de atenciones de emergencia por enfermedades cardio respiratorias en niños. [4]
- II) Mortalidad en pacientes con enfermedad pulmonar obstructiva crónica y enfermedad cardiovascular especialmente en ancianos. [5]
- III) Exacerbación de los síntomas e incremento de casos de asma.
- IV) Aumento de riesgo de infartos al miocardio.
- V) Inflamación pulmonar, inflamación sistémica, disfunciones endoteliales y vasculares, desarrollo de aterosclerosis, incremento en la incidencia de infecciones y cáncer respiratorio.

1.4. Normativa Asociadas a la Contaminación Atmosférica

De acuerdo a la constitución política de Chile, el gobierno chileno tiene la obligación de garantizar el derecho a la vida e integridad física (Artículo 19-1) y el derecho a vivir en un medio ambiente libre de contaminación (Artículo 19-8). En Chile existe una norma primaria para el $PM_{2.5}$ y PM_{10} respectivamente, las cuales pasaremos a resumir a continuación. Dichas normas se basan en los dictámenes de la Agencia de Protección Ambiental de Estados Unidos en 1997 (<https://www.epa.gov/laws-regulations>) y la Organización Mundial de la Salud en el 2005 [6] (en inglés US-EPA y WHO respectivamente).

Norma primaria $PM_{2.5}$

- I) Mediante decreto ley N° 39955, a partir del 01 de Enero del 2012 comienza a regir la norma primaria de calidad del aire para material particulado fino $PM_{2.5}$ es veinte micro-gramos por metro cúbico ($20 \mu g/m^3$), como concentración anual, y cincuenta micro-gramos por metro cúbico ($50 \mu g/m^3$), como concentración de 24 horas.
- II) Dicha norma es sobrepasada cuando los promedios diarios registrados durante un año, son mayores a $50 \mu g/m^3$, en cualquier Estación de Monitoreo con Representatividad Poblacional (**EMRP**) capaz de medir agentes PM. (Es decir, la estación de monitoreo debe estar en un área urbana) y cuando el promedio tri-anual de las concentraciones anuales es mayor a $20 \mu g/m^3$ en cualquier estación EMRP.

1.4. NORMATIVA ASOCIADAS A LA CONTAMINACIÓN ATMOSFÉRICA

III) Se definen tres situaciones de emergencia ambiental las cuales siguen el siguiente orden:

- Alerta: cuando los niveles de concentración se encuentran en $80 \mu\text{g}/\text{m}^3$ y $109 \mu\text{g}/\text{m}^3$
- Pre-emergencia: cuando los niveles de concentración se encuentran en $110 \mu\text{g}/\text{m}^3$ y $169 \mu\text{g}/\text{m}^3$
- Emergencia: cuando los niveles de concentración son mayores o iguales a $170 \mu\text{g}/\text{m}^3$

Las concentraciones para establecer las situaciones de emergencia serán obtenidas a partir de un pronóstico de calidad del aire utilizando algún modelo o bien a partir de las mediciones provenientes de alguna de las estaciones de monitoreo de calidad del aire calificadas como EMRP.

Norma primaria PM₁₀

- I) A partir de Junio del 2013, donde se derogó el Artículo 59 de 1998, se establece que la norma primaria de calidad del aire para material respirable PM₁₀ es ciento cincuenta micro-gramos por metro cubico ($150 \mu\text{g}/\text{m}^3$) como concentración a las 24 horas.
- II) Dicha norma es sobrepasada cuando los promedios diarios registrados durante un año, son mayores a $150 \mu\text{g}/\text{m}^3$, en cualquier estación de monitoreo calificada como EMRP y cuando estas estaciones registran un número de más 7 días seguidos con mediciones sobre los $150 \mu\text{g}/\text{m}^3$.

III) Se define tres situaciones de emergencia ambiental las cuales siguen el siguiente orden:

- Alerta: cuando los niveles de concentración se encuentran en $195 \mu\text{g}/\text{m}^3$ y $239 \mu\text{g}/\text{m}^3$
- Pre-emergencia: cuando los niveles de concentración se encuentran en $240 \mu\text{g}/\text{m}^3$ y $329 \mu\text{g}/\text{m}^3$
- Emergencia: cuando los niveles de concentración son mayores o iguales a $330 \mu\text{g}/\text{m}^3$

Las concentraciones para establecer las situaciones de emergencia serán obtenidas a partir de un pronóstico de calidad del aire utilizando algún modelo o bien a partir de las mediciones provenientes de alguna de las estaciones de monitoreo de calidad del aire calificadas como EMRP.

La normativa vigente también estipula los instrumentos de medición de concentraciones ambientales de contaminantes atmosféricos a usar. Estos deben encontrarse en la

lista de los métodos denominados de referencia y equivalencia publicado por la Agencia de Protección Ambiental de Estados Unidos (EPA), También pueden ser aquellas que cuentan con certificación de alguno de los países miembros de la Unión Europea (UE) los que implementan las directrices del Comité Europeo para Estandarizaciones o por último aquellos que cuenten con la certificación de cumplimiento de los estándares de calidad exigidos por el país de origen, siendo acreditado por alguna entidad competente de este.

Para definir los lugares de emplazamiento de los las estaciones de monitoreo EMRP, se deberán emplear los siguientes criterios:

- I) Cantidad de población expuesta en la zona de estudio
- II) Valores absolutos de concentraciones de material particulado respirable PM_{10} (lo cual incluye al $PM_{2.5}$) y tendencias históricas de dichos valores.
- III) Composición química del material particulado respirable al que está expuesto la población en términos de su toxicidad.
- IV) Presencia de desarrollos industriales significativos, volumen del parque automotor existente u otras actividades que generen emisiones de este contaminante o sus precursores sobre las zonas de estudio.

1.5. Bases de datos disponibles en el País en estaciones EMRP

La Tabla 1.1 muestra la situación actual respecto al número de estaciones EMRP por región. Se indica cuantas de ellas monitorean PM_{10} y $PM_{2.5}$. Sin embargo no todas estas estaciones se encuentran en línea y en funcionamiento como lo indica la cuarta fila. Esta información es entregada por el sitio web del Sistema Nacional de Calidad del Aire (SINCA) [7], entidad que depende del Ministerio del Medio Ambiente, MMA.

Región	XV	I	II	III	IV	V	SAM	VI	VII	VIII	IX	XIV	X	XI	XII
#EMRP	1	1	29	26	17	35	13	15	8	34	4	6	2	6	1
# $PM_{2.5}$	1	1	10	2	2	9	13	3	6	12	3	2	3	2	1
# PM_{10}	0	0	25	14	14	29	13	12	7	26	4	6	6	2	0
#EMRP en línea	1	1	11	2	6	22	12	7	5	28	3	1	3	2	1

Tabla 1.1: **Estaciones de Monitoreo:** Se puede ver que el número de estaciones de monitoreo no coincide con el número de estaciones en línea. Fuente: Elaboración propia.

1.5.1. Zona Norte

A continuación haremos un balance de algunas estaciones de monitoreo ubicadas en el país. Dichas tablas entregarán información sobre la región, comuna, número de estaciones de monitoreo y parámetros medidos.

Región	Comuna	# EMRP	Parámetros	
			M	C
XV	Arica	1	X	✓
I	Alto Hospicio	1	✓	✓
II	Antofagasta	1	X	✓
	Calama	4	✓	✓
	Sierra gorda	1	✓	✓
	Tocopilla	4	✓	✓
III	Copiapo	1	X	✓
	Huasco	1	✓	✓
IV	Andacollo	3	✓	✓
	Coquimbo	2	✓	✓
	Salamanca	1	✓	✓
V	Catemu	2	✓	✓
	Concón	4	✓	✓
	Hijuelas	1	✓	✓
	Panqueque	1	✓	✓
	Puchuncaví	5	✓	✓
	Quilpue	1	X	✓
	Quintero	5	✓	✓
	Valparaíso	1	X	✓
	Viña del mar	1	✓	✓

Tabla 1.2: **Zona Norte:** La letra **M** denota parámetros meteorológicos, mientras que la letra **C**, denota parámetros contaminantes. Se observa que las estaciones de Antofagasta, Copiapo, Quilpue y Valparaíso no detectan parámetros meteorológicos. Fuente: Elaboración propia.

En la segunda región cabe destacar que en la comuna de Calama solo tres estaciones de monitoreo se encuentran registrando valores de $PM_{2.5}$ y PM_{10} a partir del año 2012, hasta el día de hoy. La estación de monitoreo del Hospital del Cobre que aparece en línea no tiene ningún dato a libre disposición de $PM_{2.5}$, mientras que para PM_{10} se indica que existen datos entre los años 2002 a 2009. Sin embargo, estos no están en línea.

En la comuna de Tocopilla de las 4 estaciones de monitoreo, solo Gendarmería registra PM_{10} y $PM_{2.5}$ desde el año 2011 hasta el día de hoy. Por otro lado, en el caso de la estación de Gobernación solo registra PM_{10} , donde supuestamente se mide desde el año 2004 al día de hoy. Sin embargo, no existen datos entre el año 2004 al 2011.

Esta situación se repite también en las comunas de Concón, Puchuncaví y Quintero en la quinta región. Cabe destacar también que existen estaciones de monitoreo como Viña

1.5. BASES DE DATOS DISPONIBLES EN EL PAÍS EN ESTACIONES EMRP

del Mar que registran datos desde el año 2004 hasta el día de hoy. Sin embargo, dichos registros tienen pérdidas importantes de datos. Por ejemplo, desde Abril hasta Diciembre de 2010. A continuación mostraremos la situación en estaciones de monitoreo al sur de Chile.

1.5.2. Zona Sur

Región	Comuna	# EMRP	Parámetros	
			M	C
VI	Machali	3	✓	✓
	Rancagua	2	✓	✓
	Rengo	1	✓	✓
	San Fernando	1	✓	✓
VII	Curicó	1	✓	✓
	Linares	1	✓	✓
	Maule	1	✓	✓
	Talca	2	✓	✓
VIII	Cabrero	2	✓	✓
	Chiguayante	2	✓	✓
	Chillan	2	✓	✓
	Concepcion	1	✓	✓
	Coronel	7	✓	✓
	Curanilahue	1	✗	✓
	Hualpen	4	✗	✓
	Los Angeles	2	✓	✓
	Quillon	2	✗	✓
	Ranquil	1	✓	✓
IX	Talcahuano	6	✗	✓
	Padre Las Casas II	1	✓	✓
	Las Encinas II	1	✓	✓
	Museo Ferroviario	1	✓	✓
XIV	Valdivia	1	✓	✓
X	Chaiten	1	✓	✓
	Osorno	1	✓	✓
	Puerto Montt	1	✓	✓
XI	Coyhaique	2	✓	✓
XII	Punta Arenas	1	✗	✓

Tabla 1.3: **Zona Sur**: La letra **M** denota parámetros meteorológicos, mientras que la letra **C**, denota parámetros contaminantes. Se observa que las estaciones de Curanilahue, Hualpen, Quillon, Talcahuano y Punta Arenas no miden variables meteorológicas. Fuente: Elaboración propia.

La situación al sur del país muestra una gran cantidad de estaciones de monitoreo

1.5. BASES DE DATOS DISPONIBLES EN EL PAÍS EN ESTACIONES EMRP

en la octava región. Es por ello que de la tabla 1.3 centraremos nuestra atención en esta región. En la comuna de Chiguayante, la estación de monitoreo de Punteras, cuenta con registros de $PM_{2.5}$ diarios entre el año 2009 y hasta el día de hoy. Sin embargo, se necesita que el muestreo sea horario, para una mejor estadística y estudio de la evolución de dicho contaminante. Para PM_{10} existen registros horarios desde el año 2010 hasta el día de hoy, con un gap importante sin datos de cuatro meses entre Noviembre del 2011 hasta Marzo del 2012. La estación meteorológica de Chiguayante solo registra a partir de Agosto del 2009 midiendo $PM_{2.5}$ y PM_{10} hasta Septiembre del mismo año.

Para la comuna de Chillan el escenario es muy similar. La estación de monitoreo INIA, tiene supuestamente datos desde el año 2008 hasta el 2016 de $PM_{2.5}$. Sin embargo los registros horarios tienen muchos gap, más aún, una ausencia completa de datos entre Abril del 2010 hasta Junio del 2012. Desde Noviembre del 2013 hasta septiembre del 2016 la estación registra datos y pareciera seguir estando operativa. Similar situación se observa para los registros de PM_{10} . Por otro lado la estación de monitoreo de Puren en la misma comuna tiene registros horarios tanto de $PM_{2.5}$ y PM_{10} desde el 2012 hasta el día de hoy.

Para Concepción la única estación de monitoreo en línea se encuentra en Kingston College. Los registros horarios de PM_{10} se encuentran desde el año 2009 hasta Septiembre del 2016, mientras que para $PM_{2.5}$ los registros horarios solo se encuentran desde Mayo del 2015 hasta Septiembre del 2016, siendo que entró en operación también el 2009.

Hemos visto que la situación de mal mantenimiento de las estaciones de monitoreo se da en varias ciudades a lo largo del país. Dicha situación ha sido denunciada previamente por G. Salini [8], donde se hace hincapié en la necesidad de que el Ministerio de Medio Ambiente (MMA), mantenga bases de datos horarias con un registro de datos con mínimas perdidas, en cada ciudad del país que sea declarada como latente o saturada. De esta forma, con una buena base de datos es posible diseñar y poner en marcha un modelo de pronóstico para la calidad del aire en las diferentes zonas del país bajo riesgo por material particulado.

1.5.3. Región Metropolitana de Santiago de Chile

La tabla 1.4 muestra las estaciones EMRP, operativas en cada comuna. De estas estaciones de monitoreo, Test de Prueba se mantuvo operativa durante 20 días entre el 01 al 21 de Abril del 2016. Cerro Navia comenzó a registrar PM_{10} el año 2001, mientras Puente Alto y Talagante el año 2008. Las demás estaciones poseen registros de PM_{10} desde 1997. Para $PM_{2.5}$ las estaciones de monitoreo de Las Condes, La Florida, Parque O'Higgins y Pudahuel tienen registros desde el año 2000, mientras las demás estaciones de monitoreo partieron el año 2008. La estación de monitoreo de Quilicura I partió midiendo tanto $PM_{2.5}$ y PM_{10} el año 2008 y estuvo en operaciones hasta Mayo del 2016. La estación de monitoreo de Quilicura comenzó a medir ambos contaminantes el 2016 siendo la más nueva en funcionamiento. Los datos medidos por estas estaciones pueden ser horarios, diarios o mensuales.

Región Metropolitana	Comuna	# EMRP	Parámetros	
			M	C
	Cerrillos	1	✓	✓
	Cerro Navia	1	✓	✓
	EL Bosque	1	✓	✓
	Independencia	1	✓	✓
	La Florida	1	✓	✓
	Las Condes	1	✓	✓
	Pudahuel	1	✓	✓
	Puente Alto	1	✓	✓
	Quilicura	2	✓	✓
	Parque O'Higgins	1	✓	✓
	Test Prueba	1	✗	✓
	Talagante	1	✓	✓

Tabla 1.4: **Región Metropolitana** :La letra **M** denota parámetros meteorológicos, mientras que la letra **C**, denota parámetros contaminantes.Solo la estación Test de Prueba no detecta parámetros meteorológicos. Fuente: Elaboración propia.

1.6. Contaminación atmosférica en Santiago de Chile

La ciudad de Santiago de Chile, posee una población que bordea el 43.8% del país, según el censo desarrollado el año 2012 por el INE, en una superficie de 135 km^2 . Situada en el centro del país y a una altura de 520 msnm (metros sobre el nivel del mar), aproximadamente, esta rodeada de dos sistemas montañosos, la cordillera de los Andes al Este y la cordillera de la Costa al Oeste, dada esta conformación Santiago se encuentra en lo que se conoce como una cuenca. Dicha cuenca tiene una longitud de 80 km , en dirección Norte-Sur y 35 km de ancho Este-Oeste. Esta se encuentra delimitada en el norte por el cordón montañoso de Chacabuco y por el sur por los cerros de Paine, la cual constituye una unidad orográfica de gran importancia.

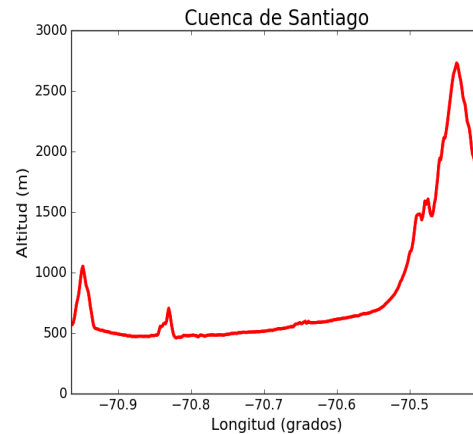
Su clima es de tipo mediterráneo, siendo los meses de Diciembre a Febrero calurosos y secos, alcanzando una temperatura de $35 \text{ }^\circ\text{C}$ a la sombra. La altitud por sobre el nivel del mar crece de Poniente a Oriente.

Debido a su topografía y a las condiciones meteorológicas imperantes, especialmente en la temporada otoño-invierno, se observa por lo general una mala dispersión tanto horizontal como vertical de los contaminantes emitidos por las diversas fuentes existentes en la ciudad. Además existe una tendencia al aumento de la densidad poblacional, lo cual se asocia a un aumento de las fuentes fijas y móviles. Prueba de ello es el aumento del parque automotor y la actividad en el área de la construcción.

La inversión térmica es el fenómeno meteorológico que obstaculiza la adecuada dispersión de los contaminantes. Esta consiste en que se asienta una capa de aire cálido por sobre masas de aire de menor temperatura, impidiendo así el flujo del aire en dirección vertical (corriente ascendente), estabilizando las condiciones e impidiendo, por tanto, la dispersión vertical de los contaminantes. Existen dos tipos de inversión térmica que corresponden a:



(a) Vista Satelital Región Metropolitana de Santiago.



(b) Relieve de la cuenca de Santiago

Figura 1.5: **Ciudad de Santiago de Chile:** a) Vista Satelital que incluye todas las comunas del gran Santiago. Fuente: Google Maps. b) El relieve de la Cuenca de Santiago fue efectuado entre las coordenadas $-33,460;-70,964$ y $-33,424;-70,400$. Fuente: Elaboración propia.

- I) **Inversión Térmica de Subsistencia:** Es aquel fenómeno que se presenta por el predominio de altas presiones en el Pacífico (anticiclón) lo cual va acompañado por el descenso del aire sobre la región metropolitana, el cual al comprimirse se calienta rompiendo la tendencia del gradiente típico de temperatura. Se forma así una capa de inversión térmica cuya base se sitúa a altitudes menores a los 700 (*msnm*).
- II) **Inversión Térmica radiativa:** Es aquel fenómeno que se produce por el enfriamiento de la tierra en períodos fríos especialmente noches de invierno, afectando a las masas de aire inmediatamente contiguas a ella, que quedan debajo de las capas más cálidas.

Por otro lado en la región de Santiago se dan dos tipos de condiciones sinópticas ligadas directamente con los episodios de alta contaminación atmosférica, a las cuales se les denominan tipo A, tipo BPF (Baja Pre-Frontal) y mixta.

- I) **Tipo A:** Estos ocurren por la entrada de masas de aire de altas presiones y altas temperaturas en altura (dorsal cálida), en la troposfera media y sobre la zona central reemplazando a una vaguada en altura (que es lo contrario a una dorsal, área de bajas presiones y temperaturas), cuyo eje presenta normalmente una orientación noroeste sureste. La formación de una vaguada costera se produce a niveles bajos, entre el centro de altas presiones fría al otro lado de la cordillera de los Andes (alta fría migratoria) y un centro de altas presiones sobre el pacífico subtropical, el cual está sincronizado a las condiciones de altura mencionada más arriba. Son estos dos sistemas de altas presiones los que obstaculizan el desplazamiento de la dorsal y la vaguada, las cuales sólo pueden moverse en la dirección norte-sur. Aún más esta

situación produce un viento superficial de este a oeste, provocando que los máximos de concentraciones de PM_{10} se den hacia el sector poniente de la cuenca de Santiago (Pudahuel, Cerro Navia, Lo Prado). Durante los días que dura este tipo de episodio se pueden apreciar cielos despejados con una alta variación de temperatura, días cálidos y noches heladas.

- II) **Tipo BPF:** Estos ocurren por la irrupción de una vaguada desde la costa hacia el interior y el flujo de aire ascendente en la troposfera media, el cual acompaña a un sistema frontal débil que va perdiendo energía a medida que se acerca a la cuenca de Santiago. En este tipo de episodio se puede apreciar un centro de bajas presiones al este de la cordillera y una alta fría migratoria desplazada sobre el océano Atlántico, esta condición está asociada, usualmente, a una cobertura nubosa prefrontal del tipo media y alta con bajo factor de ventilación.
- III) **Mixtos:** Este tipo de episodios, ocurren cuando se alternan los episodios de tipo A y tipo BPF en períodos intermedios del orden de 24 horas. Estos comienzan con una configuración del Tipo A, seguida de una de tipo BPF.

1.7. Procedimientos en casos extremos en Santiago de Chile.

De acuerdo a lo estipulado en la legislación chilena vigente, corresponde a las Secretarías Regionales Ministeriales (SEREMI) del Medio Ambiente, la coordinación de los distintos servicios públicos en la gestión de los episodios críticos de contaminación. En caso de que el evento crítico afecte a más de una región la coordinación quedará a cargo del Ministerio de Medio Ambiente.

Existen medidas permanentes y paliativas durante el período de Gestión de Episodios Críticos (GEC), entre el 01 de Abril y 31 de Agosto de cada año.

Las medidas permanentes, se aplican durante todo el período GEC, las que se asocian a la implementación de restricción vehicular, aplicación de medidas de gestión de tránsito (prohibición de estacionamientos, vías de tipo exclusivas para locomoción colectiva, reversibles y permanentes), prohibición de quemas agrícolas y uso de calefactores no certificados por la SEC (Superintendencia de Electricidad y Combustibles), entre otras medidas. Las medidas paliativas en cambio se aplican cuando el intendente de la región declara un episodio crítico. La tabla 1.5 muestra las medidas frente a episodios críticos.

Estas medidas sin embargo, requieren de una estrategia comunicacional para entregar de forma oportuna la información a la comunidad. Para ello se usan los siguientes medios de comunicación:

- **Redes Sociales:** Difusión del estado de la calidad del aire a través de la cuenta de Twitter y Facebook de la Intendencia regional, SEREMI del M.A-R.M y Ministerio del Medio Ambiente.

1.7. PROCEDIMIENTOS EN CASOS EXTREMOS EN SANTIAGO DE CHILE.

- Sitios Web de los servicios regionales: Se implementa el banner de calidad del aire en diferentes sitios web de entidades públicas para mantener informada a la población del estado de calidad del aire.
- Servicios de Noticias: Radio, Televisión y Prensa escrita.
- Red de establecimientos educacionales: El SEREMI del medio ambiente mantiene un contacto permanente con más de 500 establecimientos que incluyen jardines infantiles, escuelas y liceos.
- Aplicación Móvil “Aire de Santiago”: Dicha aplicación tiene por objetivo divulgar y masificar la información de la calidad del aire de la región Metropolitana de Santiago en tiempo real, rescatando datos del sitio web, conteniendo además un menú desplegable en el cual se informa las medidas a aplicar en cada tipo de episodio, entre ellas restricción vehicular, recomendación de no realizar actividad física, prohibición de uso de calefactores, prohibición de quemas agrícolas, prohibición de circulación de vehículos de carga al interior del anillo de Américo Vespucio, entre otras. Además entrega un glosario con los términos referidos a la calidad del aire.

Nivel 1 Alerta	Nivel 2 Pre-emergencia	Nivel 3 Emergencia
Suspensión de clases de educación física y actividades deportivas: SEREMI EDUCACIÓN R.M		
Programa Integrado de la Fiscalización: SEREMI SALUD R.M / SEREMITT R.M/ CONAF R.M -SAG / CARABINEROS		
Reforzamiento del programa de aspirado y lavado de calles: GOBIERNO REGIONAL.		
Prohibición de funcionamiento de todo tipo de artefactos de calefacción residencial que utilicen leña u otros dendroenergéticos: SEREMI SALUD R.M		
Restricción vehicular Sin Sello Verde: SEREMITT R.M	Restricción vehicular Sin Sello Verde + Con Sello: SEREMITT R.M	Restricción vehicular Sin Sello Verde + Con Sello: SEREMITT R.M
NO APLICA	Paralización de fuentes estacionarias: SEREMI SALUD R.M.	Paralización de fuentes estacionarias: SEREMI SALUD R.M.

Tabla 1.5: **Medidas frente a episodios críticos de contaminación:** Se muestran los niveles de alerta existentes y los servicios públicos responsables ante estas situaciones. Fuente: Elaboración propia.

Capítulo 2

Series de tiempo

Definimos una serie temporal x_i , con $i = 1, \dots, N$ como un conjunto de observaciones las cuales son obtenidas por lo general en un tiempo de muestreo δt uniforme. Dichas observaciones pueden corresponder a datos meteorológicos como por ejemplo a la velocidad del viento, la precipitación caída en una ciudad. Otro ejemplo son las series de tiempo obtenidas de los sismógrafos que miden la intensidad y duración de un terremoto. En sistemas biológicos tenemos los electrocardiogramas, encefalogramas. En economía la evolución de la venta de cobre o combustibles, o como evoluciona el intercambio de dinero entre distintos agentes.

2.1. Estadística aplicada al análisis de Series de Tiempo

Una serie de tiempo debe ser estudiada y caracterizada para un mejor entendimiento. Para ello podemos partir estudiando la media, desviación estándar y varianzas asociadas a esta.

Definición 2.1.1 *Sea una serie temporal x_i , con $i = 1, \dots, N$ discreta, definimos la media aritmética \bar{x} de la serie temporal como:*

$$\bar{x} = \frac{1}{N} \sum_i^N x_i. \quad (2.1.1)$$

En la literatura se suele también representar la media $\langle x \rangle$ o también usando la letra griega μ cuando el número de datos N tiende al infinito. Sin embargo ya que en la práctica esto no es posible se suele usar indistintamente μ o $\bar{x}(\langle x \rangle)$. También en algunos

libros se habla del valor de expectación de una variable x , dado por $E(x)$, sin embargo debemos notar que la definición del valor de expectación está asociada a la distribución de probabilidad que siguen los datos, por lo cual hay un abuso del lenguaje. El valor de expectación será igual a la media sólo si los datos siguen una distribución simétrica como la de Gauss.

En el análisis de datos se suele estudiar cual es la fluctuación de los datos en torno a la media. Para ello usamos la desviación estándar.

Definición 2.1.2 Definimos la desviación d_i de cualquier medida x_i con respecto a la media de la serie de tiempo como:

$$d_i = x_i - \mu, \quad (2.1.2)$$

y la desviación promedio :

$$\alpha = \lim_{N \rightarrow \infty} \left[\frac{1}{N} \sum_i^N |x_i - \mu| \right]. \quad (2.1.3)$$

Una forma más fácil de conocer la dispersión de los valores x_i es mediante la desviación estándar σ . La varianza σ^2 es definida como el límite del promedio de las desviaciones en torno a la media μ al cuadrado.

$$\sigma^2 = \lim_{N \rightarrow \infty} \left[\frac{1}{N} \sum_i^N (x_i - \mu)^2 \right], \quad (2.1.4)$$

la desviación estándar es la raíz cuadrada de la varianza, dada por:

$$\sigma = \lim_{N \rightarrow \infty} \left(\left[\frac{1}{N} \sum_i^N (x_i - \mu)^2 \right] \right)^{1/2}. \quad (2.1.5)$$

Para un conjunto de datos discretos, tenemos que la varianza es dada por:

$$s = \left[\frac{1}{N-1} \sum_i^N (x_i - \bar{x})^2 \right], \quad (2.1.6)$$

sin embargo, al igual que en el caso de la media se suele usar la letra griega σ^2 en lugar de s , o también si x es la variable, su varianza se escribe como $var(x)$. Se suele denominar a la media y la varianza como momentos estadísticos donde la media corresponde al primer momento, y la varianza al segundo momento. En base a esto podemos hacer una primera clasificación de las series de tiempo.

2.1.1. Estacionaridad

Definición 2.1.3 Sea una serie temporal x_i , con $i = 1, \dots, N$. Diremos que cuando la media μ , la varianza σ^2 y la desviación estándar σ se conservan a lo largo de la señal, la serie es **débilmente estacionaria**. Por el contrario hablamos de una serie **no-estacionaria**.

Una serie temporal al ser un conjunto de datos, siguen una cierta distribución de probabilidad, donde la media, la varianza y la desviación estándar están asociadas a dicha distribución.

Definición 2.1.4 Diremos que una serie temporal es **estrictamente estacionaria** si la serie temporal x_i , con $i = 1, \dots, N$ al ser desfasada en un factor r que puede ser positivo o negativo, tal que la nueva serie es x_{i+r} , mantiene la misma distribución de probabilidad y los mismos momentos estadísticos.

2.1.2. Correlación

Consideremos dos series de tiempo x_i e y_i , donde $i = 1, \dots, N$. Esto quiere decir que ambas series contienen el mismo número de datos.

Definición 2.1.5 Llamamos a la correlación de Pearson ρ_{xy} a la relación dada por :

$$\rho_{xy} = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^N (y_i - \bar{y})^2}} = \frac{\text{cov}(x, y)}{\sigma_x \sigma_y}, \quad (2.1.7)$$

donde el numerador corresponde a la covarianza entre las series de tiempo, dada por:

$$\text{cov}(x, y) = \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y}), \quad (2.1.8)$$

y el denominador es el producto de las desviaciones estándar σ_x y σ_y respectivamente.

La ecuación 2.1.7 suele escribirse también como:

$$\rho_{xy} = \frac{N \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{N \sum x_i^2 - (\sum x_i)^2} \sqrt{N \sum y_i^2 - (\sum y_i)^2}}. \quad (2.1.9)$$

La correlación es una medida de que tan relacionadas se encuentran ambas series de tiempo y el valor de esta se encuentra en el intervalo $-1 < \rho < 1$. Si la correlación entre ambas variables es positiva diremos que están directamente correlacionadas, mientras

que si es negativa hablamos de correlación inversa. Cuando es nula diremos que no hay correlación. Es importante recalcar que si bien dos variables pueden estar fuertemente correlacionadas eso no implica causalidad, ya que podríamos caer en una falacia conocida como “Cum hoc ergo propter hoc” (‘Con esto, por causa de esto’). Supongamos por un momento que existe una correlación positiva entre x_i e y_i , No podemos decir de que x_i sea la causa de y_i ya que podría ser al revés, es decir y_i sea la causa de x_i . También podría darse de que exista un tercer factor que causa la relación entre ambas series, o que esto sea simple coincidencia.

Consideremos ahora dos series temporales discretas x_i, y_i y desplazemos esta última en un factor $r < N$ que puede ser positivo o negativo de modo que tenemos una nueva serie y_{i-r} , donde r es llamado el desfase o lag, que aplico a la serie y_i :

Definición 2.1.6 Definimos la correlación cruzada, $\hat{\rho}_{xy}$ a la relación dada por:

$$\hat{\rho}_{xy} = \frac{1}{(N-r)} \frac{\sum_{i=1}^{N-r} (x_i - \bar{x})(y_{i-r} - \bar{y})}{\sigma_x \sigma_y}, \quad (2.1.10)$$

donde N es el largo y σ_x y σ_y son las desviaciones estándar de cada serie.

Para el caso en que $y_{i-r} = x_{i-r}$, de modo tal que $\sigma_x = \sigma_y$:

Definición 2.1.7 Definimos la auto-correlación, R a la relación dada por:

$$R(r) = \frac{1}{(N-r)\sigma^2} \sum_{i=1}^{N-r} (x_i - \bar{x})(x_{i-r} - \bar{x}), \quad (2.1.11)$$

donde N y σ^2 son el largo de la serie y la varianza respectivamente. La cantidad dada por:

$$autocov(x_r) = \frac{1}{(N-r)} \sum_{i=1}^{N-r} (x_i - \bar{x})(x_{i-r} - \bar{x}), \quad (2.1.12)$$

la cual corresponde a la autocovarianza siendo una cantidad importante en el análisis de señales y en el análisis multifractal de estas ya que nos permite clasificarlas en diferentes tipos los que serán abordados próximamente en esta tesis. Se suele denominar a la autocovarianza de una señal con la letra griega $\gamma(r)$.

2.2. Regresión Lineal y tendencia

Dado un conjunto de datos $\{y_i, x_{i1}, \dots, x_{ip}\}_{i=1}^n$, con n unidades estadísticas. Una regresión lineal asume una relación lineal entre la variable dependiente y_i y el p -vector de regresión x_i . De esta forma se tiene:

$$y_i = \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \varepsilon_i = x_i' \beta + \varepsilon_i, \quad (2.2.1)$$

con $i = 1, \dots, N$, donde $'$ denota la traspuesta de x_i . Luego $x_i' \beta$ es el producto interno entre los vectores x_i y β , correspondiendo a los coeficientes de regresión y ε_i es un error aleatorio asociado a las mediciones de la variable temporal. Los coeficientes β_i son obtenidos mediante el método de los mínimos cuadrados, que consiste en minimizar los errores aleatorios ε_i . Consideremos la recta de regresión dada por:

$$y_i = \beta_1 x_i + \beta_0 + \varepsilon_i, \quad (2.2.2)$$

despejando ε_i , sumando todos los términos al cuadrado se tiene:

$$\sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (y_i - (\beta_1 x_i + \beta_0))^2. \quad (2.2.3)$$

Sea $f(\beta_0, \beta_1) = \sum_{i=1}^n \varepsilon_i^2$, minimizando con respecto a β_0 y β_1 , se tiene que:

$$\frac{\partial f}{\partial \beta_0} = 0, \quad \frac{\partial f}{\partial \beta_1} = 0. \quad (2.2.4)$$

Resolviendo el sistema de ecuaciones dado por (1.2.4) se obtiene que:

$$\beta_0 = \frac{S_y S_{xx} - S_x S_{xy}}{n S_{xx} - S_x S_x}, \quad \beta_1 = \frac{n S_{xy} - S_x S_y}{n S_{xx} - S_x S_x}, \quad (2.2.5)$$

donde $S_x = \sum_{i=1}^n x_i$, $S_y = \sum_{i=1}^n y_i$, $S_{xy} = \sum_{i=1}^n x_i y_i$, $S_{xx} = \sum_{i=1}^n x_i^2$, respectivamente.

El coeficiente β_1 que corresponde a la pendiente de la recta es lo que llamamos la tendencia lineal y nos entrega información de como varía linealmente la variable y_i con respecto a la variable x_i . En el caso de una serie temporal, esta muestra la variación de una variable con respecto al tiempo. Dicha tendencia puede ser positiva, negativa o nula. También global o local, como es el caso de una serie no estacionaria donde existen cambios en la tendencia.

2.2.1. Significancia Estadística de la Tendencia

Un test de inferencia estadística permite distinguir entre tendencias monótonamente crecientes o decrecientes bajo un cierto nivel de significancia α , que por lo general es de un 5%. La hipótesis nula $H = 0$ establece que la serie de datos a estudiar no posee tendencia. Una vez realizado el test este determina si se acepta la hipótesis nula ($H = 0$) o se rechaza ($H = 1$), con un cierto valor asociado p que es mayor a la significancia α en caso de aceptar la hipótesis nula, o menor en caso contrario.

Para el caso de la tendencia usamos el test de Mann-Kendall que es un método para datos no paramétricos. Esto quiere decir que la serie temporal a estudiar debe ser independiente. El concepto de independencia de una serie temporal asume que una observación para un tiempo t_i no depende del valor anterior t_{i-1} (**Estocacidad**). Cuando la serie es independiente y normalmente distribuida hablamos que los datos son paramétricos.

Definición 2.2.1 Para una serie temporal x_i , definimos el estadístico S como:

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{sgn}(x_j - x_i), \quad (2.2.6)$$

donde sgn corresponde a la función signo dada por:

$$\text{sgn}(x_j - x_i) = \begin{cases} 1 & \text{si } x_i < x_j \\ 0 & \text{si } x_i = x_j \\ -1 & \text{si } x_i > x_j \end{cases} \quad (2.2.7)$$

Kendall mostró que si S es normalmente distribuida con media $E(S)$ y varianza $\text{Var}(S)$, estas son dadas por:

$$E(S) = 0, \quad (2.2.8)$$

$$\text{Var}(S) = \frac{n(n-1)(2n+5)}{18}. \quad (2.2.9)$$

Cuando existen valores repetidos la varianza queda dada por:

$$\text{Var}(S) = \frac{n(n-1)(2n+5) - \sum_{i=1}^n t_i(i-1)(2i+5)}{18}, \quad (2.2.10)$$

donde t_i es el número de datos repetidos para un cierto i . El test estadístico parametrizado Z se calcula por:

$$Z_{MK} = \begin{cases} \frac{S-1}{\sqrt{\text{Var}(s)}} & \text{si } S > 0 \\ 0 & \text{si } S = 0 \\ \frac{S+1}{\sqrt{\text{Var}(s)}} & \text{si } S < 0 \end{cases} \quad (2.2.11)$$

Si $Z_{MK} > 0$, hablamos que la serie tiene una tendencia monótonamente creciente, mientras que si $Z_{MK} = 0$, la serie no tiene tendencia y finalmente si $Z_{MK} < 0$ la tendencia es monótonamente decreciente. Para testear el nivel de significancia p de la hipótesis se usa una tabla de distribución normal. Un completo desarrollo de la teoría de este método puede encontrarse en los estudios de Hamed y Rao (1997) [9] y de Da Silva (2013) [10]. Un algoritmo sumamente preciso, puede obtenerse desde <http://www.mathworks.com/matlabcentral/fileexchange/25533-mann-kendall-modified-test>.

2.3. Análisis de Fourier

Una serie de tiempo es una colección de variables aleatorias $\{x_t\}$, con $t = 1, \dots, n$ definida para todos los reales t en el caso continuo o todos los enteros t en el caso discreto. Dicha serie tiene las siguientes propiedades:

- I) El valor de expectación $E(x_t)$ de x_t es constante.
- II) EL valor de expectación $E(x_t, x_{t'})$ no depende de t y t' sólo de $t - t'$.

Ya que $E(x_t) = \mu$ es constante, este valor puede ser estimado y substraído de las series sin mayores efectos, e incluso asumirse nulo. Sin embargo, en el procedimiento de análisis de datos es mejor substraer esta cantidad. De esta forma hablaremos de una **Serie de Tiempo Estacionaria con Media Cero**. (En inglés **Zero-Mean Stationary time series**)

La propiedad II) implica que la $var(x_t)$ es constante y que $\gamma_r = cov(x_t, x_{t+r}) = \gamma_{-r}$ no depende de t .

La cantidad γ_r es la autocovarianza teórica con un desfase (lag) r de $\{x_t\}$, la cual desde el punto de vista discreto es dada por:

$$\gamma_r = \frac{1}{n} \sum_{t=|r|}^{n-1} (x_t - \mu)(x_{t-|r|} - \mu). \quad (2.3.1)$$

Como estamos considerando la media nula ($\mu = 0$) entonces

$$\gamma_r = \frac{1}{n} \sum_{t=|r|}^{n-1} (x_t)(x_{t-|r|}). \quad (2.3.2)$$

Si $r = 0$, tenemos entonces que la autocovarianza se reduce a $\frac{1}{n} \sum x_t^2$ que es justamente la varianza s o σ^2 .

2.3.1. Densidad Espectral

Sea x_t una serie temporal con $t = 1, \dots, n$. La densidad espectral (en inglés como **Power Spectrum**, **Power spectral density** o **Energy espectral density**), de una serie describe cómo la varianza de esta señal se distribuye en las frecuencias que la componen.

La distribución de la varianza de una serie puede ser descrita, o por la medida de la media y su variabilidad con respecto a esa media, o por una región de distribución acumulativa $S(f)$, donde $S(f)$ es la potencia contribuida por las diferentes frecuencias f_t hasta f .

Sea $[a, b]$ una banda de frecuencias, la cantidad de varianza contribuida a x_t en este intervalo corresponde a $S(b) - S(a)$. Si S es una función de distribución continua, entonces existe una función de densidad espectral S' . La densidad espectral S' en una frecuencia f da la razón de varianza contribuida por las frecuencias en la vecindad inmediata de f , para la varianza de x_t por unidad de frecuencia.

Al igual que en la consola de Linux, diferenciaremos entre minúsculas y mayúsculas. Llamaremos $S(f)$ a la función de distribución espectral y $s(f)$ la densidad espectral. En otros libros se les asigna la letra F y f respectivamente y a la frecuencia la letra griega ω .

Consideremos la autocovarianza teórica de x_t con un lag r dada por:

$$\gamma_r = \int_0^1 e^{2\pi i f r} s(f) df. \quad (2.3.3)$$

Si $s(f)$ es suave entonces, esta puede ser representada por una “**Serie de Fourier**”, de modo que tenemos:

$$s(f) = \sum_{r=-\infty}^{\infty} \gamma_r e^{-2\pi i f r}, \quad (2.3.4)$$

$$= \gamma_0 + 2 \sum_{r=1}^{\infty} \gamma_r \cos(2\pi f r). \quad (2.3.5)$$

Aquí γ_0 corresponde a la varianza total de la señal x_t , la cual puede ser calculada integrando la densidad espectral sobre todo el rango de frecuencias. Por medio de la expresión (2.3.5) vemos que la densidad espectral es simétrica, es decir $s(f) = s(-f)$ puesto que $\gamma_r = \gamma_{-r}$.

2.3.2. Relación entre densidad espectral y períodograma

Consideremos la transformada discreta de Fourier (DFT) de una serie x_t con $t = 1, \dots, n$

$$d_x(f) = \frac{1}{n} \sum_{t=0}^{n-1} x_t e^{-2\pi i f t}, \quad (2.3.6)$$

con $i = \sqrt{-1}$. Definiremos el autoperíodograma como:

$$I_{xx}(f) = n \times d_x(f) \times d_x^*(f), \quad (2.3.7)$$

donde $d_x^*(f)$ corresponde al complejo conjugado de $d_x(f)$. A continuación recurriremos a un artilugio matemático que nos ayudará a introducir en la definición de períodograma la autocovarianza muestral c_r . Podemos reescribir (2.3.7) como:

$$I_{xx}(f) = \frac{1}{n} \sum_t x_t e^{-2\pi i f t} \sum_{t'} x_{t'} e^{2\pi i f t'}, \quad (2.3.8)$$

$$= \frac{1}{n} \sum_t \sum_{t'} x_t x_{t'} e^{-2\pi i f (t-t')}, \quad (2.3.9)$$

donde el factor exponencial depende únicamente de $t - t'$, por lo que sólo se toman $2n - 1$ términos. Haciendo $t - t' = r$, podemos reescribir (2.3.9) como:

$$I_{xx}(f) = \frac{1}{n} \sum_{|r| < n} c_r e^{-2\pi i f r}, \quad (2.3.10)$$

donde definimos $c_r = \sum_{t-t'=r} x_t x_{t'}$ como la autocovarianza muestral de x_t . Comparando (2.3.10) con (2.3.4), nos vemos tentados a decir que el períodograma es lo mismo que la densidad espectral. Sin embargo, eso es un error.

“La Densidad Espectral corresponde al valor de expectación del Períodograma.”

$$E\{I_{xx}(f)\} \approx s(f). \quad (2.3.11)$$

La demostración puede ser encontrada en la sección 9.4 del Bloomfield [11] con mayor detalle.

Demostración:

Supongamos que x_t tiene un espectro continuo con densidad espectral $s(f)$. Además x_t , puede ser escrito como:

$$x_t = \sum_r g_r U_{t-r}, \quad (2.3.12)$$

donde $\{U_{t-r}\}$ es un proceso de ruido blanco con un desfase r y g_r es una media móvil.¹

Este proceso de ruido blanco, puede ser:

- I) Gaussiano y por tanto independiente.
- II) No Gaussiano pero independiente.
- III) No Gaussiano, dependiente, pero lo suficientemente cercano para que el limite de distribución del períodograma sea exponencial.

Entonces la función de Transferencia $G(f)$ de los filtros $\{g_r\}$ satisface $|G(f)|^2 = s(f)$ y la transformada de Fourier de un conjunto de datos $\{x_0, x_1, \dots, x_{n-1}\}$ corresponde a la función de transferencia por la transformada de Fourier del ruido blanco dado por $d_U(f)$.

$$d_x(f) = G(f)d_U(f), \quad (2.3.13)$$

por lo cual el períodograma satisface:

$$I_{xx}(f) \approx |G(f)|^2 I_U(f) \quad (2.3.14)$$

$$= s(f)I_U(f) \quad (2.3.15)$$

Esto quiere decir que el períodograma de x_t , es la densidad espectral por el períodograma de la señal de ruido blanco dada por $I_U(f)$. Es fácil ver que $E\{I_U(f)\} = 1$ y por tanto $E\{I_{xx}(f)\} \approx s(f)$ que es lo que buscábamos probar. Una demostración alternativa se da en el Shumway [12] sección 4.4, pero hay que tener cuidado con la notación.

La definición de espectro cruzado será introducida a continuación al estudiar el concepto de coherencia, pero antes hablaremos sobre el intervalo de confianza o la significancia estadística de la densidad espectral.

2.3.3. Intervalo de confianza de la densidad espectral

Un espectro suavizado está dado por:

$$\hat{s}(f) = \sum_u g_u I(f - f_u), \quad (2.3.16)$$

donde g_u , es una media móvil, y f_u es la u -ésima media de la frecuencia, en el u -ésimo intervalo.

¹Notemos que en este caso r actúa como un subíndice.

Sabemos que $E\{I(f)\} \approx s(f)$, si el promedio cubre un pequeño intervalo de frecuencias y $s(f)$ es continuo en el intervalo. Entonces $s(f)$ es aproximadamente constante en el mismo intervalo.

$$E\{I(f - f_u)\} \approx s(f), \quad (2.3.17)$$

y también,

$$E\{\hat{s}(f)\} \approx s(f) \sum_u g_u. \quad (2.3.18)$$

La varianza de una distribución aleatoria exponencialmente distribuida está dada por:

$$\begin{aligned} \text{var}\{I(f)\} &\approx s(f), \\ \text{var}\{\hat{s}(f)\} &\approx \sum_u g_u^2 s(f - f_u)^2. \end{aligned}$$

Para construir intervalos de confianza para $\hat{s}(f)$ o $\ln(\hat{s}(f))$ necesitamos aproximar $\hat{s}(f)$. La forma más simple es notando que para una suma de variables independientes, su distribución es aproximadamente la **distribución normal**. Podemos hacer una expansión de $\ln(\hat{s}(f))$ en torno a $s(f)$ dando como resultado:

$$\ln(\hat{s}(f)) \approx \ln(s(f)) + \frac{\hat{s}(f) - s(f)}{s(f)}. \quad (2.3.19)$$

Entonces:

$$\begin{aligned} E\{\ln(\hat{s}(f))\} &\approx \ln(s(f)) \\ \text{var}\{\ln(\hat{s}(f))\} &\approx \frac{\text{var}\{\hat{s}(f)\}}{s(f)^2} \\ &\approx \sigma^2, \end{aligned}$$

donde σ^2 corresponde a la varianza de los datos.

Para el $\ln(\hat{s}(f))$ y un intervalo de confianza del 95 % se tiene que $\ln(\hat{s}(f)) \pm 1.96\sigma$. El factor 1.96, viene de una tabla de probabilidades para la distribución normal. Con esto tenemos nuestras cotas para el intervalo de confianza para $\ln(s(f))$. Podemos reescribir $\ln(\hat{s}(f)) \pm 1.96\sigma = \ln(\hat{s}(f)e^{\pm 1.96\sigma})$.

Entonces tenemos que:

$$\ln(\hat{s}(f)e^{-1.96\sigma}) \leq \ln(s(f)) \leq \ln(\hat{s}(f)e^{1.96\sigma}). \quad (2.3.20)$$

El intervalo correspondiente para $s(f)$ es entonces:

$$e^{-1,96\sigma} \hat{s}(f) \leq s(f) \leq e^{1,96\sigma} \hat{s}(f). \quad (2.3.21)$$

Otra posibilidad para obtener intervalos de confianza es utilizar el test de χ^2 como una aproximación para $\hat{s}(f)$.

La media μ , la varianza σ^2 y los grados de libertad ν de χ^2 se relacionan por $\nu\sigma^2 = 2\mu^2$ o también $\nu = 2\mu^2/\sigma^2$.

De esta forma aproximamos la distribución de $\hat{s}(f)$ por la distribución χ^2 con ν grados de libertad. Para un intervalo de confianza del 95 % es dado por:

$$\frac{\nu\hat{s}(f)}{\chi_\nu^2(0.975)} \leq s(f) \leq \frac{\nu\hat{s}(f)}{\chi_\nu^2(0.025)}. \quad (2.3.22)$$

Los factores 0.975 y 0.025 son obtenidos de las tablas de probabilidad para el test χ^2 .

2.3.4. Coherencia

Esta cantidad permite conocer en qué frecuencia un par de señales x_t e y_t comparten energía. También se suele decir en la literatura que con qué fuerza se relacionan estas series. Es el equivalente de la correlación cruzada pero en el espacio de frecuencias (Espacio de Fourier). Si bien la sintaxis de un libro a otro varía, la herramienta más usada en el estudio de la coherencia es la transformada discreta de Fourier (DFT) dada por la relación (2.3.6).

Habíamos definido el autoperíodograma mediante la ecuación (2.3.7). Si consideramos otra serie y_t con el mismo número de elementos de modo que las ecuaciones (2.3.6) y (2.3.7) son validas, podemos definir el períodograma cruzado como:

$$I_{xy}(f) = d_x(f)d_y^*(f). \quad (2.3.23)$$

Definiremos el espectro cruzado dado por:

$$\hat{S}_{xy}(f) = \sum_u g_u I_{xy}(f - f_u), \quad (2.3.24)$$

donde g_u corresponde a un filtro llamado también una media móvil, que se aplica en una cierta ventana de frecuencias f_u donde u es la u -ésima media que se obtiene al dividir los datos en u -intervalos. Es preferible escribir el espectro cruzado mediante su magnitud y fase relativa.

$$\hat{S}_{xy}(f) = |\hat{S}_{xy}(f)|e^{2\pi i\hat{\phi}_{xy}(f)}, \quad (2.3.25)$$

donde la fase relativa $\hat{\phi}$, corresponde a:

$$\hat{\phi}_{xy}(f) = \frac{Im(\hat{S}_{xy}(f))}{Re(\hat{S}_{xy}(f))} \quad (2.3.26)$$

Entonces, la coherencia cuadrática estará dada por:

$$coh^2 = \frac{|\hat{S}_{xy}(f)|^2}{\hat{S}_{xx}(f)\hat{S}_{yy}(f)}. \quad (2.3.27)$$

La coherencia se puede encontrar en algunos textos como C_{xy}^2 o también k_{xy}^2 , y tiene la propiedad de que siempre se encuentra entre un valor de 0 y 1. Es importante notar que en la definición de coherencia los espectros son mediados, de no ser así ésta sería siempre exactamente uno.

Demostración: Consideremos que nuestros espectros no son suavizados de modo que las densidades espectrales son los periodogramas. Entonces:

$$\begin{aligned} C_{xy}^2 &= \frac{|I_{xy}|^2}{I_{xx}I_{yy}} \\ &= \frac{|d_x(f)d_y^*(f)|^2}{|d_x(f)d_x^*(f)||d_y(f)d_y^*(f)|} \\ &= \frac{|d_x(f)d_y^*(f)|^2}{|d_x(f)||d_x^*(f)||d_y(f)||d_y^*(f)|} \\ &= \frac{|d_x(f)d_y^*(f)|^2}{|d_x(f)d_y^*(f)||d_x^*(f)d_y(f)|} \\ &= \frac{|d_x(f)d_y^*(f)|^2}{|d_x(f)d_y^*(f)|^2} \\ &= 1 \end{aligned}$$

2.3.5. Significancia estadística de la coherencia

Lo que hemos definido hasta el momento, corresponde a la coherencia teórica k_{xy} , pero en la práctica siempre trabajamos con un estimado dado por \hat{k}_{xy} , por lo cual tenemos que considerar un intervalo de confianza estadística.

$$\hat{k}_{xy}(f) = \frac{|\hat{S}_{xy}(f)|^2}{\hat{S}_{xx}(f)\hat{S}_{yy}(f)}.$$

Si queremos probar la hipótesis nula ($k_{xy}^2(f) = 0$), en este caso tenemos que la estadística es dada por:

$$F = \frac{\hat{k}_{xy}(f)}{(1 - \hat{k}_{xy}(f))} (L - 1), \quad (2.3.28)$$

donde F es una distribución con 2 y $2L - 2$ grados de libertad. Al resolver la ecuación (2.3.28) llegamos a que el intervalo de confianza al estimar la coherencia cuadrática es dada por:

$$C_\alpha = \frac{F_{2,2L-2}(\alpha)}{L - 1 + F_{2,2L-2}(\alpha)}, \quad (2.3.29)$$

donde α corresponde a un cierto nivel de significancia.

La significancia estadística del espectro cruzado está relacionada con el estimado de la coherencia \hat{k}_{xy} . Para un intervalo de confianza del 95 %, tenemos:

$$\ln \frac{s_{xx}(f)}{s_{yy}(f)} \approx \ln \frac{\hat{s}_{xx}(f)}{\hat{s}_{yy}(f)} \pm 1.96 \sqrt{2\sigma^2(1 - \hat{k}_{xy}^2(f))}. \quad (2.3.30)$$

El análisis de Fourier ha sido ampliamente utilizado con éxito en series temporales que presentan características estacionarias. Sin embargo, este método pierde eficacia con series de tiempo no estacionarias, que presentan una alta varianza y discontinuidades. Además dicho análisis se centra sólo en el espacio de frecuencias o espacio de Fourier. Por ello requerimos una técnica más avanzada que nos permita realizar un análisis simultaneo en tiempo y frecuencia, de modo de tener una cantidad de información mucho más rica de la señal. Dicha técnica está basada en funciones especiales llamadas **Ondeletas** (**Wavelets** en inglés), con las cuales se desarrollan herramientas análogas a Fourier.

2.4. Transformada de Ondeletas

El análisis de señales utilizando ondeletas consiste en descomponer una serie temporal en un espacio de tiempo-frecuencia, determinando los modos dominantes de variabilidad y cómo estos varían en el tiempo. Una buena introducción al tema puede encontrarse en Torrence y Compo [13], en el cual se basará la descripción de esta herramienta, complementada con otras fuentes de información.

Definición 2.4.1 Sea $x(t)$ una serie temporal y $\Psi_{a,b}$ una ondeleta -real o compleja- definida por:

$$\Psi_{a,b} = \Psi_0 \left(\frac{t-b}{a} \right), \quad (2.4.1)$$

donde b es una variable de desplazamiento temporal y a un control de dilatación si $a > 1$ o contracción si $a < 1$ de la función Ψ_0 . (El ancho de la ondeleta que depende de a es llamado soporte temporal de la ondeleta.)

Se define la transformada continua de ondeletas como la convolución entre $x(t)$ y $\Psi_0 \left(\frac{t-b}{a} \right)$ dada por:

$$T(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t) \Psi_0^* \left(\frac{t-b}{a} \right) dt, \quad (2.4.2)$$

donde Ψ_0^* es el complejo conjugado de Ψ_0 y $1/\sqrt{a}$ es el factor de normalización de Ψ_0 .

Se denomina a Ψ_0 como la ondeleta madre y cumple con las siguientes propiedades:

- Es una función numerable y cuadrado integrable:

$$\int_{-\infty}^{\infty} |\Psi_0(t)| dt < \infty \quad y \quad \int_{-\infty}^{\infty} |\Psi_0(t)|^2 dt < \infty. \quad (2.4.3)$$

- Es una función de media nula y norma 1:

$$\int_{-\infty}^{\infty} \Psi_0(t) dt = 0 \quad y \quad \int_{-\infty}^{\infty} |\Psi_0(t)|^2 dt = 1. \quad (2.4.4)$$

Una de las ondeletas más conocida corresponde a la de Morlet, que es de uso común en el estudio de procesos geofísicos.

La transformada de ondeletas admite inversa definida como:

$$x(t) = \frac{1}{a^2} \int_{-\infty}^{\infty} \int_a^{\infty} T(a,b) \Psi_{a,b} da db. \quad (2.4.5)$$

Denotamos ahora las siguientes relaciones:

$$W_n(s) = T(a, b), \quad s = a, \quad n'\delta t = t, \quad n\delta t = b,$$

donde n' y n son enteros.

Definición 2.4.2 Sea $x_{n'}$ una serie temporal discreta, donde $n' = 0, \dots, N-1$ con N el largo de la serie, la cual es muestreada por lo general en un tiempo uniforme δt . Llamamos ahora a la transformada de ondeleta discreta $W_n(s)$ con $n = 0, \dots, N-1$ como la convolución entre $x_{n'}$ y Ψ , dada por:

$$W_n(s) = \sum_{n'=0}^{N-1} x_{n'} \Psi^* \left(\frac{(n' - n)\delta t}{s} \right), \quad s \neq 0 \quad (2.4.6)$$

donde Ψ^* es el complejo conjugado de Ψ la cual es una versión normalizada de la ondeleta madre Ψ_0 , vale decir.

$$\Psi \left(\frac{(n' - n)\delta t}{s} \right) = \sqrt{\frac{\delta t}{s}} \Psi_0 \left(\frac{(n' - n)\delta t}{s} \right), \quad (2.4.7)$$

donde $\sqrt{\frac{\delta t}{s}}$ corresponde al factor de normalización y es independiente de la ondeleta. Esto se demuestra fácilmente a continuación.

Demostración:

Sea A un factor de normalización tal que

$$\Psi \left(\frac{(n' - n)\delta t}{s} \right) = A \Psi_0 \left(\frac{(n' - n)\delta t}{s} \right). \quad (2.4.8)$$

Aplicando la definición de producto interno entre dos funciones f y g dado por la ecuación 2.4.9:

$$\left\langle \Psi \left(\frac{(n' - n)\delta t}{s} \right) \middle| \Psi \left(\frac{(n' - n)\delta t}{s} \right) \right\rangle = 1. \quad (2.4.9)$$

Desarrollando al lado izquierdo de (2.4.9) se tiene:

$$\begin{aligned} \left\langle \Psi \left(\frac{(n' - n)\delta t}{s} \right) \middle| \Psi \left(\frac{(n' - n)\delta t}{s} \right) \right\rangle &= \int_{-\infty}^{\infty} A^2 \Psi_0 \left(\frac{(n' - n)\delta t}{s} \right) \Psi_0^* \left(\frac{(n' - n)\delta t}{s} \right) dn', \\ &= \int_{-\infty}^{\infty} A^2 \left| \Psi_0 \left(\frac{(n' - n)\delta t}{s} \right) \right|^2 dn', \\ &= A^2 \int_{-\infty}^{\infty} \left| \Psi_0 \left(\frac{(n' - n)\delta t}{s} \right) \right|^2 dn'. \end{aligned}$$

Sea $u = \frac{(n'-n)\delta t}{s}$, de modo que $du = dn' \frac{\delta t}{s}$ y despejando $\frac{sdu}{\delta t} = dn'$. Reemplazando y considerando que toda ondeleta madre es ortonormal ($\langle \Psi_0(u) | \Psi_0(u) \rangle = 1$), se tiene:

$$\begin{aligned} A^2 \int_{-\infty}^{\infty} \frac{s}{\delta t} |\Psi_0(u)|^2 du &= 1, \\ A^2 \underbrace{\int_{-\infty}^{\infty} |\Psi_0(u)|^2 du}_1 &= \frac{\delta t}{s}, \\ A^2 &= \frac{\delta t}{s}, \\ A &= \sqrt{\frac{\delta t}{s}}. \end{aligned}$$

Hasta el momento hemos visto la definición en el espacio temporal, sin embargo es mucho mejor, desde el punto de vista computacional, trabajar en el espacio de Fourier. Sin pérdida de generalidad cambiamos el índice $n' \rightarrow n$. Tomemos ahora la transformada discreta de Fourier para la serie x_n , de modo tal que:

$$\hat{x}_k = \frac{1}{N} \sum_{n=0}^{N-1} x_n e^{-2\pi i k n / N}, \quad (2.4.10)$$

donde $k = 0, \dots, N-1$ es el índice de la frecuencia k -ésima. La transformada de Fourier de la función de ondeletas $\Psi(t/s)$ es dada por $\hat{\Psi}(s\omega_k)$. Por lo tanto, del teorema de convolución la transformada de ondeleta para una escala s es la transformada inversa de Fourier del producto:

$$W_n(s) = \sum_{k=0}^{N-1} \hat{x}_k \hat{\Psi}^*(s\omega_k) e^{i\omega_k n \delta t}, \quad (2.4.11)$$

donde la frecuencia angular ω_k se define como:

$$\omega_k = \begin{cases} \frac{2\pi k}{N\delta t} : k \leq \frac{N}{2} \\ -\frac{2\pi k}{N\delta t} : k > \frac{N}{2} \end{cases}$$

Por otra parte $\hat{\Psi}^*(s\omega_k)$ es el complejo conjugado normalizado de la función de ondeleta en el espacio de Fourier. Este último es dado por:

$$\hat{\Psi}(s\omega_k) = \sqrt{\frac{2\pi s}{\delta t}} \hat{\Psi}_0(s\omega_k), \quad (2.4.12)$$

donde $\sqrt{\frac{2\pi s}{\delta t}}$ es la constante de normalización. El procedimiento de calculo es análogo al efectuado anteriormente solo que tomamos las definiciones en el espacio de Fourier.

Usando esta normalización para cada escala s tenemos que

$$\sum_{k=0}^{N-1} |\hat{\Psi}(s\omega_k)|^2 = N, \quad (2.4.13)$$

donde N es el número de puntos que compone la señal.

2.4.1. Elección de la escala s

Hemos visto que la ondeleta posee un factor de escala s que dependiendo de su valor, puede contraer o dilatar la ondeleta. Para poder desarrollar un análisis de multiresolución, se debe tomar un conjunto de factores de escala (descritas como potencias de 2, para un ahorro del coste computacional). La relación a usar está dada por:

$$s_j = s_0 2^{j\delta j}, \quad (2.4.14)$$

donde $j = 0, 1, \dots, J$, con el máximo número de escalas dado por:

$$J = \frac{1}{\delta j} \log_2 \left(\frac{N\delta t}{s_0} \right), \quad (2.4.15)$$

s_0 se escoge como el período equivalente de Fourier que es aproximadamente $2\delta t$. El factor δj depende del ancho en el espacio espectral de la función de ondeleta. Para una ondeleta de Morlet se usan valores no mayores a 0.5, siendo este uno de los más usados. Otros valores usados son 0.25 y 0.125.

2.4.2. Cono de Influencia

Al trabajar con series de tiempo de longitud finita, pueden existir errores en el inicio y final del espectro de ondeletas, ya que la relación (2.4.11) asume que los datos son cíclicos. Una solución es hacer un proceso de padding (acolchar), con ceros antes de realizar la transformada de ondeletas y luego removerlos. (También en el caso de series con alta varianza se puede hacer el padding con la media de la señal). Al hacer el proceso de padding se usan suficientes ceros para llevar la longitud de la señal N hasta la próxima potencia de dos, así se limitan los efectos en los bordes y acelera la transformada de Fourier. El cono de influencia (COI) es la región del espectro de ondeleta donde los efectos en los bordes se vuelven importantes y se define como el tiempo de despliegue, e , que se elige como un factor de discontinuidad que decae de la forma e^{-2} , de modo que los efectos en los bordes son despreciables. La extensión del cono aumenta linealmente con el factor de escala s , siendo proporcional al ancho de la ondeleta.

2.4.3. Reconstrucción de señales y filtrado

La relación (2.4.5), nos dice que es posible reconstruir la serie a partir de los coeficientes de la transformada de ondeletas para el caso continuo. Por otro lado, la transformada de ondeletas discreta dada por (2.4.6) funciona también como un filtro pasa-banda, donde la ondeleta es la función de respuesta. Así que también es posible reconstruir la señal de forma discreta, en un proceso un poco más complicado, dada la redundancia en tiempo y escala. La reconstrucción es posible de obtener mediante la relación dada por:

$$x_n = \frac{\delta j \delta t^{1/2}}{C_\delta \Psi_0(0)} \sum_{j=0}^J \frac{Re\{W_n(s_j)\}}{s_j^{1/2}}, \quad (2.4.16)$$

donde $\Psi_0(0)$ es la ondeleta evaluada cuando $n' = n$ y C_δ es un factor de reconstrucción que depende de la ondeleta elegida Ψ_0 .

Si deseamos utilizar la reconstrucción, como un filtro pasa banda, consideramos la relación (2.4.16), pero en vez de sumar sobre todas las escalas sumamos en una banda con $j_1 < j < j_2$, de modo tal que:

$$\hat{x}_n = \frac{\delta j \delta t^{1/2}}{C_\delta \Psi_0(0)} \sum_{j=j_1}^{j_2} \frac{Re\{W_n(s_j)\}}{s_j^{1/2}}, \quad (2.4.17)$$

donde \hat{x}_n , es la señal filtrada.

2.4.4. Espectro de Ondeletas o Potencia espectral de Ondeletas

Como la función Ψ es en general compleja, la transformada de ondeleta $W_n(s)$ también lo es; por lo cual podemos definir el espectro local de ondeletas (amplitud del espectro de potencia), con su respectiva fase dado por:

$$A_n = |W_n(s)|^2, \quad \phi_n = \arctg \left\{ \frac{Im(W_n(s))}{Re(W_n(s))} \right\} \quad (2.4.18)$$

Una corrección en la definición de la amplitud de la potencia es introducida por Liu et al [14]. Esto debido a que existe una sobre estimación del espectro de potencia en las altas frecuencias (bajas periodicidades). De esta forma la amplitud del espectro de potencia para una escala s_j es dada por:

$$A_n = \frac{|W_n(s_j)|^2}{s_j}. \quad (2.4.19)$$

Podemos promediar globalmente en el tiempo para cada escala s_j , mediante la relación:

$$\overline{W}^2(s_j) = \frac{1}{N} \sum_{n=0}^{N-1} \frac{|W_n(s_j)|^2}{s_j}, \quad (2.4.20)$$

de esta forma al plotear el espectro global versus el conjunto de escalas, podemos ver como se distribuye dicha potencia en las escalas.

Por otro lado, también podemos ver como se distribuye la potencia en el tiempo para una banda de escalas, mediante:

$$\overline{W}_n^2 = \frac{\delta j \delta t}{C_\delta} \sum_{j=j_1}^{j_2} \frac{|W_n(s_j)|^2}{s_j}. \quad (2.4.21)$$

Al tomar (2.4.21) promediando en el tiempo sobre todas las escalas, obtenemos la varianza total de la señal (σ^2), la cual se conserva bajo la transformada de ondeletas por el teorema de Parseval. De esta forma, la varianza está dada por:

$$\sigma^2 = \frac{\delta j \delta t}{C_\delta N} \sum_{n=0}^{N-1} \sum_{j=0}^J \frac{|W_n(s_j)|^2}{s_j}. \quad (2.4.22)$$

2.4.5. Espectro Teórico y Nivel de Significancia

Para determinar el nivel de significancia, ya sea en el análisis de Fourier o de Ondeletas, debemos escoger un espectro de fondo apropiado. Para procesos geofísicos se asume que estos son distribuidos de forma aleatoria en torno a la media y que los procesos reales tienen distribuciones comparables. Para un espectro de Fourier se asume ruido blanco, mientras que para un espectro de ondeletas se asume ruido rojo. Los procesos de ruido blanco y ruido rojo son derivados y comparados con un proceso de Montecarlo, los cuales se usan para establecer una hipótesis nula para estudiar la significancia de un peak en el espectro de ondeletas. Un modelo simple para el ruido rojo está dado por:

$$x_n = \alpha x_{n-1} + z_n, \quad (2.4.23)$$

que corresponde a un proceso de Markov, donde α es la autocorrelación para un lag -1, x_0 y z_n , son tomados de un ruido blanco Gaussiano. Al normalizar (2.4.23), su espectro discreto de Fourier está dado por:

$$P_k = \frac{1 - \alpha^2}{1 + \alpha^2 - 2\alpha \cos(2\pi k/N)}, \quad (2.4.24)$$

donde $k = 0, \dots, N/2$ es el índice de la frecuencia. Así, eligiendo una autocorrelación lag -1 apropiada, podemos usar (2.4.24) para modelar el espectro de ruido rojo que es nuestra hipótesis nula. (Note que para $\alpha = 0$, (2.4.24) es el espectro de ruido blanco).

Si un peak del espectro de ondeletas es significativo con respecto al ruido rojo, entonces se asume que es una característica verdadera con un cierto porcentaje de significancia por lo general del 5 %.

Para una señal x_n normalmente distribuida, entonces la parte real e imaginaria de su transformada de Fourier \hat{x}_k están también normalmente distribuidas. Ya que el cuadrado de una variable normalmente distribuida sigue una distribución de χ^2 con un grado de libertad. Entonces $|\hat{x}_k|^2$ sigue una distribución de χ^2 con dos grados de libertad. De esta forma la distribución para el espectro de Fourier está dada por:

$$\frac{N|\hat{x}_k|^2}{\sigma^2} \Rightarrow \frac{1}{2}P_k\chi_2^2, \quad (2.4.25)$$

donde el subíndice 2 de χ indica que hay dos grados de libertad.

Para el espectro local de ondeletas, la distribución de éste sigue la relación:

$$\frac{|W_n|^2}{\sigma^2} \Rightarrow \frac{1}{2}P_k\chi_2^2, \quad (2.4.26)$$

para cada tiempo n y escala s .

EL intervalo de confianza para esta distribución está dado por:

$$\frac{2|W_n(s)|^2}{\chi_2^2(p/2)} \leq \sigma^2 P_k \leq \frac{2|W_n(s)|^2}{\chi_2^2(1-p/2)}, \quad (2.4.27)$$

donde $p = 0.05$ es la significancia deseada para el intervalo de confianza del 95 %.

2.4.6. Espectro cruzado de ondeletas

Sean x e y dos series temporales con transformadas de ondeletas $W_n^x(s)$ y $W_n^y(s)$, definimos el espectro cruzado de ondeletas como:

$$W_n^{xy}(s) = W_n^x(s)W_n^{y*}(s). \quad (2.4.28)$$

Los intervalos de confianza pueden ser obtenidos si ambos espectros de ondeletas siguen una distribución χ^2 con ν grados de libertad. La distribución de probabilidad está dada por:

$$f_\nu(z) = \frac{2^{2-\nu}}{\Gamma^2(\frac{\nu}{2})} z^{\nu-1} K_0(z), \quad (2.4.29)$$

donde, $K_0(z)$ es la función de Bessel de orden cero y Γ es la función gamma. La función de distribución acumulativa está dada por $p = \int_0^{Z_\nu(p)} f_\nu(z) dz$, donde $Z_\nu(p)$ es el nivel de confianza asociado con probabilidad p . Si las dos series temporales tienen espectros de Fourier P_k^x y P_k^y , entonces la distribución del espectro cruzado es:

$$\frac{|W_n^X(s)W_n^{Y*}(s)|}{\sigma_X \sigma_Y} \Rightarrow \frac{Z_\nu(p)}{\nu} \sqrt{P_k^X P_k^Y}, \quad (2.4.30)$$

con σ_X y σ_Y las desviaciones estándar respectivas. Cuando $\nu = 1$, tenemos que la ondeleta es real con $Z_1(95\%) = 2.182$, mientras para $\nu = 2$, $Z_2(95\%) = 3.999$.

2.4.7. Coherencia de ondeletas

De acuerdo con Torrence y Compo [13], la coherencia-que es la medida de la intensidad de la covarianza de las dos series en el espacio de frecuencias-, siempre vale 1. Sin embargo, S. Jevrejeva et al. [15] establecen una definición que es posible calcular. Sea R^2 , la coherencia cuadrática entre los espectros de ondeletas de modo tal que:

$$R_n(s)^2 = \frac{|S(s^{-1}W_n^{xy}(s))|^2}{S(s^{-1}|W_n^x(s)|^2) \cdot S(s^{-1}|W_n^y(s)|^2)}, \quad (2.4.31)$$

donde S mayúscula es el filtro que suaviza la señal y s es el factor de escala.

La fase de la coherencia del espectro de ondeletas cruzado está dada según Torrence y Webster [16] por:

$$\phi_n(s) = \arctan \left\{ \frac{Im(s^{-1}W_n^{xy}(s))}{Re(s^{-1}W_n^{xy}(s))} \right\}. \quad (2.4.32)$$

El filtro se diseña de acuerdo a la ondeleta elegida y está dada por:

$$S(W) = S_{scale}(S_{time}(W(s, t))), \quad (2.4.33)$$

donde S_{scale} denota el filtrado a lo largo de la escala de la ondeleta y S_{time} el filtrado en el tiempo. Estos están dados por:

$$S_{time}(W)|_s = \left(W(t, s) * c_1 e^{-\frac{t^2}{2s^2}} \right) |_s, \quad (2.4.34)$$

$$S_{time}(W)|_s = (W(t, s) * c_2 \Pi(0.6s)) |_t. \quad (2.4.35)$$

donde c_1 y c_2 son constante de normalización y Π es la función rectángulo. EL factor 0.6 está determinado de manera empírica para la ondeleta de Morlet.

Capítulo 3

Multifractalidad en Series de Tiempo

3.1. Fractales.

El concepto de fractal fue introducido por Benoit Mandelbrot en 1967 [17] para caracterizar las figuras geométricas que no pueden ser descritas por la geometría euclídea. Un fractal posee algunas de las siguientes propiedades:

- I) Una estructura complicada en un amplio rango de escalas.
- II) Repetición de estructuras en diferentes escalas, lo cual es conocido como auto similitud.
- III) Una dimensión fractal que a diferencia de un objeto euclídeo no es entero.

Un fractal puede ser originado a partir de un modelo determinista dado por un proceso de iteración (mapeo o ecuación en diferencia). Ejemplos de procesos de iteración son el conjunto de Cantor, la curva de Von Koch, el triángulo de Sierpinski y el mapa de Hénon. Estos ejemplos son autosimilares desde el punto de vista geométrico. Sin embargo existen otros tipos de fractales llamados fractales aleatorios, los cuales si bien no son exactamente autosimilares desde el punto de vista geométrico, lo son desde el punto de vista estadístico. Cada parte del atractor aleatorio tiene las mismas propiedades estadísticas del todo. Otra forma de generar un fractal es a partir de un modelo determinista regido por un conjunto de ecuaciones diferenciales (flujo) como por ejemplo el atractor de Lorenz y el atractor de Rössler. Por último una serie de tiempo obtenida a partir de datos discretos, puede presentar también condiciones de auto-similaridad ya sea estadística o geométrica. Esto porque los puntos que componen la señal están confinados a un atractor [18].

3.1.1. Ejemplos de fractales.

- I) **Conjunto de Cantor:** Consideremos una línea entre 0 y 1 ($k=0$). El segmento es dividido en tres partes iguales sacando el segmento del medio ($k=1$). De los segmentos restantes sacamos el segmento medio ($k=2$) y se repite el procedimiento una gran cantidad de veces.
- II) **Conjunto de Cantor Aleatorio:** Consideramos nuevamente la línea entre 0 y 1 ($k=0$). Sin embargo aleatoriamente retiramos cualquiera de los segmentos en cada iteración.
- III) **Curva de Von Koch:** El método de construcción de la curva de Von Koch, consiste en tomar una línea recta horizontal en el intervalo entre 0 y 1 ($k=0$). El segmento unitario es dividido en tres y la línea del medio es removida siendo reemplazada con dos segmentos iguales, ambos de igual longitud un tercio que forman un triángulo equilátero ($k=1$). El siguiente paso ($k=2$) se quita el tercio medio de cada uno de los 4 segmentos y se reemplaza con dos segmentos iguales formando nuevos triángulos equiláteros.
- IV) **Curva de Von Koch Aleatoria:** En vez de remover el tercio medio en cada iteración, se escoge aleatoriamente la posición del tercio a remover y se reemplaza por los tercios que forman el triángulo equilátero.
- V) **Atractor de Lorenz.** El atractor de Lorenz nace a partir del estudio de la convección termal de una capa horizontal de fluido calentada desde abajo. El fluido caliente se eleva debido a su flotabilidad y circula en trayectorias cilíndricas. Lorenz usó las ecuaciones de continuidad y de Navier-Stokes en conjunto con la ecuación de conducción del calor para describir la dinámica de estas trayectorias cilíndricas. Las ecuaciones que describen este sistema son dadas por:

$$\dot{x} = \sigma(y - x), \quad (3.1.1)$$

$$\dot{y} = rx - y - xz, \quad (3.1.2)$$

$$\dot{z} = xy - bz. \quad (3.1.3)$$

El término x representa la rotación del cilindro, y representa la diferencia en la temperatura en los lados opuestos del cilindro y z es la desviación del gradiente vertical de temperatura. σ es el número de Prandtl del aire que involucra la viscosidad y conductividad térmica de este, b depende de la razón ancho-alto de la capa y r es un parámetro de control representando temperatura fija entre el fondo y la cima del sistema. La no-linealidad en la ecuación 3.1.2 y 3.1.3 nombradas arriba resulta de la no-linealidad de las ecuaciones de flujo. Para valores de $\sigma = 10$, $b = 8/3$ y $r = 28$, el atractor de Lorenz consiste de dos discos, cada uno hecho de trayectorias espirales. Ciertas trayectorias dejan los discos casi perpendiculares y fluyen de un disco a otro.

- VI) **Series de tiempo.** Se ha reportado comportamiento fractal en series de tiempo generadas por sistemas complejos. Algunos ejemplos son:

- Series de tiempo geofísicas (temperatura, precipitación, eventos sísmicos),
- Series de tiempo médicas y fisiológicas (presión sanguínea, niveles de glucosa, secuencias de ADN.),
- Series de tiempo Astrofísicas (fuentes de rayos X y números de manchas solares),
- Series de tiempo técnicas (trafico de internet, tráfico en carretera),
- Series de tiempo sociales (economía, modelos de conflictos sociales).

Hasta el momento hemos introducido distintos ejemplos de fractales. Sin embargo, debemos recurrir a una cantidad que nos permita cuantificar las características, de estos objetos. Para ello usamos la Dimensión fractal que pasamos describir a continuación.

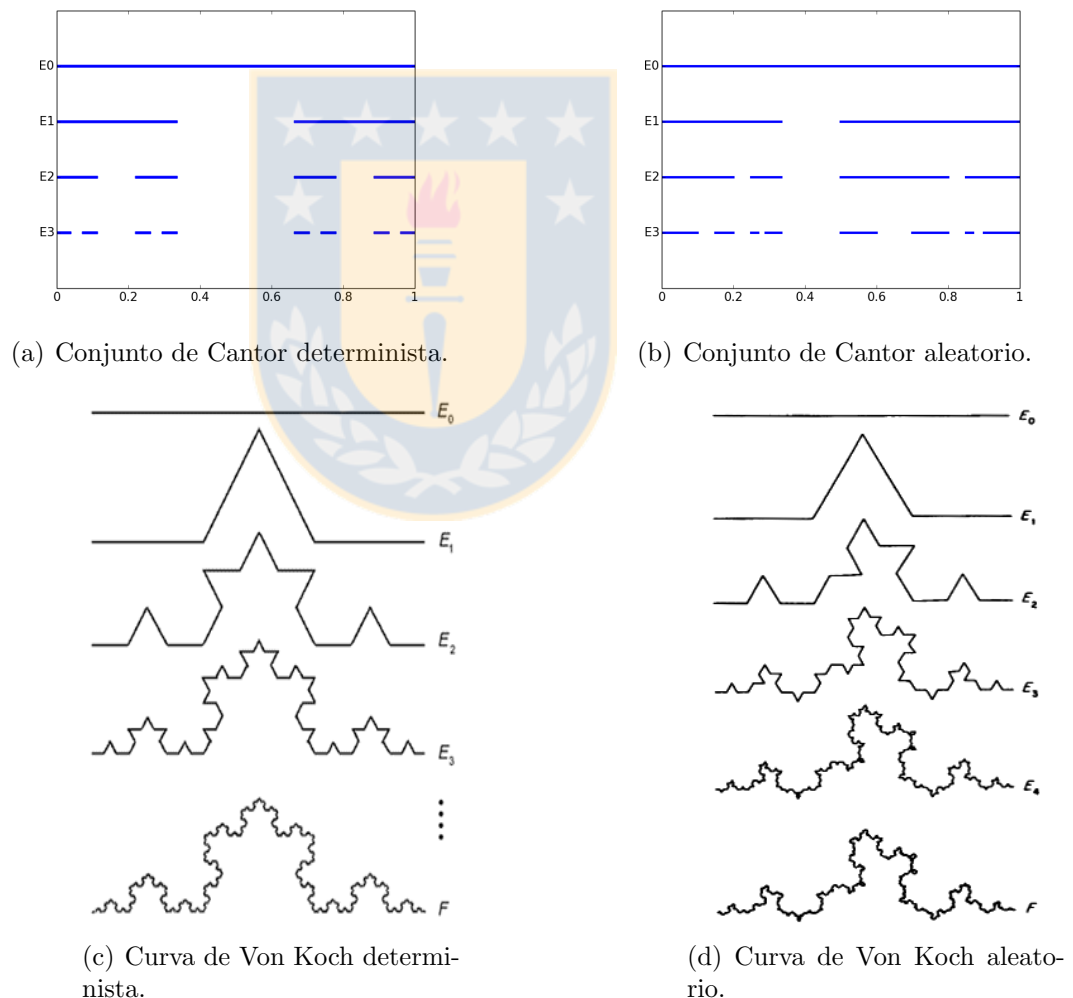
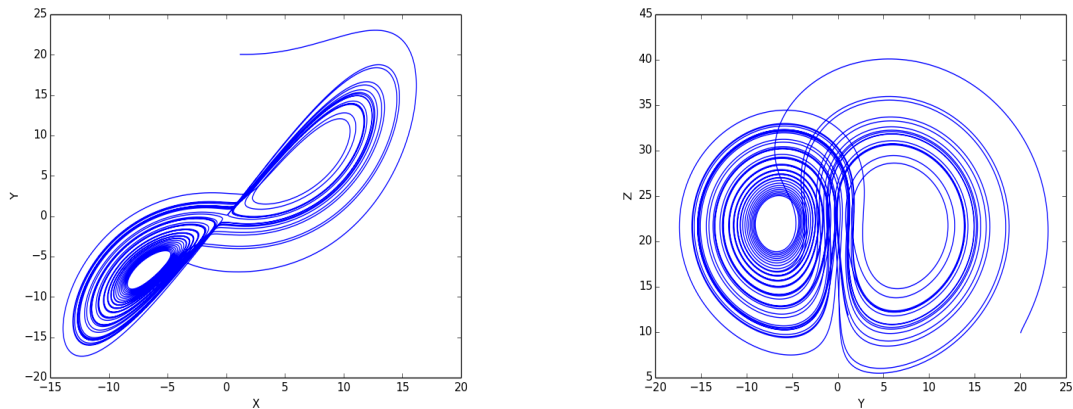


Figura 3.1: **Ejemplos de Fractales:** a) Conjunto de Cantor, c) Curva de Von Koch, los cuales corresponden a fractales a partir de mapeos deterministas. b) Conjunto de Cantor Aleatorio, d) Curva de Von Koch Aleatoria. Fuente: a) y b), Elaboración propia. c) <https://goo.gl/gWlKGL>, d) <https://goo.gl/CEdZrr>



(a) Proyección XY del atractorde Lorenz.

(b) Proyección YZ del atractorde Lorenz.

Figura 3.2: **Atractor de Lorenz:** Ejemplo de fractal determinista a partir de ecuaciones diferenciales. Fuente: Elaboración propia.

3.1.2. Dimensión Fractal

Considere una grilla de paso $1/n$ en el intervalo unitario $[0, 1]$. Es decir hay una grilla de puntos desde $0, 1/n, 2/n, \dots, (n-1)/n, 1$. Podemos ver que el número de cajas que componen la grilla dependen del tamaño del paso con él se recorro la grilla. Así, para el intervalo unitario necesitamos n grillas de tamaño $1/n$ para cubrirlo completamente. Por otro lado, para el intervalo entre 0 y 8 se necesitan $8n$ grillas de tamaño $1/n$.

La propiedad común para el intervalo unidimensional es que el número de cajas de tamaño ϵ requeridas para cubrir dicho intervalo no es más que $C(1/\epsilon)$, donde C es una constante que depende del largo del intervalo y no de ϵ .

El cuadrado $\{(x, y) : 0 \leq x, y \leq 1\}$ de lado 1 en el plano puede ser cubierto por n^2 cajas de lado $1/n$. Notemos que la diferencia con el caso anterior es el exponente 2. Cualquier rectángulo en R^2 puede ser cubierto por $C(1/\epsilon)^2$ cajas de tamaño ϵ . Por extensión una región d-dimensional requiere $C(1/\epsilon)^d$ cajas de lado ϵ .

Extendiendo esta idea a un fractal y cubriéndolo con una grilla de lado ϵ , la cantidad $C(1/\epsilon)^D$ es una relación de escala, de modo que la idea es conocer su dimensión D . Sea $N(\epsilon)$ el número de cajas de lado ϵ necesarias para cubrir un cierto conjunto S en un espacio R^m , diremos que este conjunto S es un conjunto d-dimensional que puede ser cubierto por:

$$N(\epsilon) = C \left(\frac{1}{\epsilon} \right)^D, \quad (3.1.4)$$

cajas de lado ϵ cuyo valor es pequeño. En este sentido el exponente D no necesariamente es un entero.

Aplicando el logaritmo natural a ambos lados de la ecuación (3.1.4) podemos conocer

la dimensión d mediante la relación:

$$D = \frac{\ln N(\epsilon) - \ln C}{\ln(1/\epsilon)}. \quad (3.1.5)$$

Si C es una constante para valores pequeños de ϵ , la contribución del segundo término en el numerador de la formula (3.1.5) es despreciable.

Definición 3.1.1 Sea S un conjunto acotado en R^n , este tiene una dimensión de **box-counting** (llamada también dimensión de Minkowski-Bouligand.), dada por:

$$D = \text{boxdim}(S) = \lim_{\epsilon \rightarrow 0} \frac{\ln N(\epsilon)}{\ln(1/\epsilon)}, \quad (3.1.6)$$

cuando el límite existe.

Podemos hacer simplificaciones a la definición (3.1.6) de modo que sea más fácil de usar.

- **Simplificación 1.** Es suficiente verificar que $\epsilon = b_n$, donde $\lim_{n \rightarrow \infty} b_n = 0$ y

$$\lim_{n \rightarrow \infty} \frac{\ln(b_{n+1})}{\ln(b_n)} = 1. \quad (3.1.7)$$

Teorema. Asumimos que $b_1 > b_2 > \dots$, $\lim_{n \rightarrow \infty} b_n = 0$ y que (3.1.7) es valida de modo tal que,

$$\lim_{n \rightarrow \infty} \frac{\ln N(b_n)}{\ln(1/b_n)} = D, \quad (3.1.8)$$

entonces también lo es,

$$\lim_{\epsilon \rightarrow 0} \frac{\ln N(\epsilon)}{\ln(1/\epsilon)} = D, \quad (3.1.9)$$

y por lo tanto “the box counting dimension” es D .

- **Simplificación 2.** Las cajas pueden moverse para cubrir de forma más eficiente el objeto en estudio.
- **Simplificación 3.** Otros conjuntos pueden usarse en vez de cajas, por ejemplo discos con radio ϵ muy pequeño que pueden cubrir el conjunto S .

Algunas dimensiones fractales conocidas corresponden a 0.63 para el conjunto de Cantor y 1.26 para la curva de Von Koch.

Hasta el momento hemos visto diferentes tipos de fractales. Sin embargo, esta última palabra la hemos usado indistintamente abusando de su significado, ya sea para el conjunto de Cantor, la curva de Von Koch, y el atractor de Lorenz. Los primeros corresponden a monofractales, pero el último corresponde a un multifractal.

Un multifractal posee una estructura mucho más compleja, donde diferentes regiones de un objeto poseen diferentes propiedades fractales requiriendo el uso de infinitas dimensiones para poder describirlo [19].

3.2. Formalismo Multifractal

Consideremos una serie de tiempo $\{x_i\}_{i=1}^N$, donde los x_i quedan contenidos en una trayectoria en un espacio de fase d -dimensional. Las trayectorias no llenan el espacio de fase cuando $N \rightarrow \infty$ ya que las trayectorias están sobre un atractor extraño de dimensión D , con $D < d$. Consideremos una grilla compuesta por cajas de lado ϵ muy pequeño, de modo tal que cubrimos el atractor S . Sea:

$$p_i = \lim_{N \rightarrow \infty} \frac{N_i}{N}, \quad (3.2.1)$$

donde N_i es el número de veces que la serie visita la i -ésima caja de lado ϵ . Generamos la medida sobre el atractor $d\mu(x)$

$$p_i = \int_{i\text{-caja}} d\mu(x). \quad (3.2.2)$$

Definimos el exponente de escalamiento α como:

$$p_i^q \sim \epsilon_i^{\alpha q}, \quad (3.2.3)$$

donde α es el exponente de Hölder y toma diferentes valores correspondientes a diferentes regiones de la medida y q es el q -ésimo momento.

Como dividimos el sistema en cajas de lado ϵ se sugiere que el número de veces que α toma un valor entre α' y $\alpha' + d\alpha'$ sea de la forma:

$$d\alpha' \rho(\alpha') \epsilon^{-f(\alpha')}, \quad (3.2.4)$$

donde $f(\alpha')$ es una función continua y refleja las diferentes dimensiones del conjunto sobre las singularidades de fuerza α' . Y donde $\rho(\alpha')$ es una densidad asociada al número de cajas que cubren el atractor.

En orden de determinar la función $f(\alpha)$, conocida como el espectro multifractal, la relacionamos con una propiedad observable de la medida. Para ello la vinculamos con la dimensión de Hentschel y Proccacia [18, 20].

$$D_q = \lim_{l \rightarrow 0} \left(\frac{1}{q-1} \frac{\ln \chi(q)}{\ln l} \right), \quad (3.2.5)$$

donde $\chi(q) = \sum_i p_i^q$. Con esto, D_0 es la dimensión fractal, D_1 es la dimensión de información, D_2 es la dimensión de correlación. Notamos que $\chi(q)$ depende de q , el cual se relaciona con diferentes índices de escalamiento. Reescribiendo $\chi(q)$ en espacio el continuo se tiene:

$$\chi(q) = \int d\alpha' \rho(\alpha') \epsilon^{-f(\alpha')} \epsilon^{q\alpha'} \quad (3.2.6)$$

Al tomar el límite $\epsilon \rightarrow 0$, la integral será dominada por el valor de α' que hace $q\alpha' - f(\alpha')$ más pequeño, siempre cuando $\rho(\alpha') \neq 0$. De esta forma reemplazamos $\alpha' = \alpha(q)$, el cual se define por la condición extrema:

$$\begin{aligned} \frac{d}{d\alpha'} [q\alpha' - f(\alpha')] \Big|_{\alpha'=\alpha(q)} &= 0 \\ \frac{d^2}{d\alpha'^2} [q\alpha' - f(\alpha')] \Big|_{\alpha'=\alpha(q)} &> 0 \end{aligned} \quad (3.2.7)$$

Desarrollando se encuentra:

$$\begin{aligned} f'(\alpha(q)) &= q, \\ f''(\alpha(q)) &< 0. \end{aligned} \quad (3.2.8)$$

Estas dos condiciones son importantes. La primera indica que la pendiente de $f(\alpha)$, depende de q y la segunda indica que $f(\alpha)$ posee un máximo. De esta forma, se tiene que D_q está dado por:

$$D_q = \frac{1}{q-1} [q\alpha(q) - f(\alpha(q))], \quad \alpha(q) = \frac{d}{dq} [(q-1)D_q], \quad (3.2.9)$$

donde 3.2.9 es una transformación de Legendre.

Consideremos ahora una medida que posee una estructura recursiva, por ejemplo el conjunto de Cantor. De este modo podemos conocer D_q . Dividimos el atractor S en S_i piezas, con $i = 1 \cdots N$, con una medida p_i y tamaño l_i , donde el máximo de l_i es l . Definimos la función de partición, usando la notación de Halsey et al. [18] como:

$$\Gamma(q, \tau, l) = \sum_i \frac{p_i^q}{l_i^\tau} \quad (3.2.10)$$

En un siguiente paso dividimos cada pieza, en N piezas cada una con una medida reducida por un factor p_j y tamaño l_j . Diremos que una partición de N elementos, es generadora si se cumple:

$$\Gamma(q, \tau, l^2) = [\Gamma(q, \tau, l)]^2, \quad (3.2.11)$$

donde el exponente 2 indica la segunda iteración. De esta forma podemos conocer la cantidad $\tau(q)$ conocida como el exponente de Renyi, de modo que

$$\Gamma(q, \tau(q), l^2) = 1, \quad (3.2.12)$$

y a partir de esta condición se tiene:

$$\tau(q) = (q - 1)D(q), \quad (3.2.13)$$

La relación 3.2.13 obtenida para este caso especial es también posible de mostrar a partir de la técnica MFDFA que se verá más adelante. Es posible reescribir la transformación de Legendre dada por la ecuación 3.2.9, de modo tal que:

$$\tau(q) = q\alpha(q) - f(\alpha(q)), \quad \alpha(q) = \frac{d\tau(q)}{dq}. \quad (3.2.14)$$

3.2.1. Cálculo del Espectro Multifractal

Método de Chhabra y Jensen

Ashvin Chhabra et al. [21] propusieron una metodología para la determinación directa de $f(\alpha)$ basándose en un método canónico para calcular promedios termodinámicos. Esta técnica no requiere calcular las dimensiones generalizadas de Hentschel y Procaccia, D_q , ni el uso de la transformada de Legendre.

Si se busca describir una medida singular $P(x)$ una cantidad de interés es la dimensión de Hausdorff del soporte teórico de la medida, $P(x)$. Este es simplemente la ínfima dimensión de los conjuntos en los cuales toda la medida vive. Para clases especiales de medidas que surgen de procesos multiplicativos, se tienen varios teoremas que dan información de como calcular la dimensión del soporte teórico de la medida. En particular, se conoce la entropía S , de cada uno de los procesos que está dada por:

$$S = - \sum_i P_i \log P_i, \quad (3.2.15)$$

siendo la dimensión de Hausdorff:

$$d_h = - \lim_{N \rightarrow \infty} \frac{1}{\log N} \sum_i^N P_i \log P_i. \quad (3.2.16)$$

Si se divide la medida experimental considerada, de forma tal que $P_i(l)$ corresponda a la probabilidad de un proceso multiplicativo con $N \sim l^{-1}$, la ecuación 3.2.16 provee una formula para el cálculo de la dimensión de Hausdorff del conjunto, el cual es el soporte teórico de la medida $P(x)$. Definimos una medida normalizada dada por:

$$\mu(q, l) = \frac{|P_i(l)|^q}{\sum_j |P_j(l)|^q} \quad (3.2.17)$$

El parámetro q en 3.2.17 es análogo a un microscopio que permite explorar diferentes regiones de la medida P . Para $q > 1$ se acentúan las regiones más intermitentes de P , mientras para $q < 1$ se acentúan las regiones menos intermitentes y para $q = 1$ se replica la medida original. De esta forma, se tiene:

$$\begin{aligned} f(q) &= - \lim_{N \rightarrow \infty} \frac{1}{\log N} \sum_{i=1}^N \mu_i(q, l) \log[\mu_i(q, l)] \\ &= \lim_{l \rightarrow 0} \frac{\sum_{i=1}^N \mu_i(q, l) \log[\mu_i(q, l)]}{\log(l)}. \end{aligned} \quad (3.2.18)$$

Además, es posible encontrar el valor promedio del exponente de Hölder o fuerza de singularidad, $\alpha_i = \log(P_i)/\log(l)$ con respecto a $\mu(q)$, al evaluar:

$$\begin{aligned} \alpha(q) &= - \lim_{N \rightarrow \infty} \frac{1}{\log N} \sum_{i=1}^N \mu_i(q, l) \log[P_i(l)], \\ &= \lim_{l \rightarrow 0} \frac{\sum_{i=1}^N \mu_i(q, l) \log[P_i(l)]}{\log(l)}. \end{aligned} \quad (3.2.19)$$

Las ecuaciones 3.2.18 y 3.2.19 dan un alternativa a la estimación del espectro de singularidades, calculando de forma directa y sin hacer uso de la transformada de Legendre.

Método del Análisis Multifractal de Fluctuaciones (MF DFA)

Esta técnica se divide en 5 pasos y se describe en el trabajo de Jan W. Kantelhardt et al. [22]:

- Sea x_i un conjunto de datos discretos donde $i = 1, \dots, N$. Definimos el perfil de la serie $Y(i)$ como:

$$Y(i) = \sum_{k=1}^i (x_k - \bar{x}_k). \quad (3.2.20)$$

Es necesario notar que no es obligatorio restar las media \bar{x}_k , puesto que estas son eliminadas en el proceso de eliminación de tendencias (detrend).

- Se divide el perfil en $N_s = \text{int}(N/s)$ segmentos no traslapados de igual longitud s de izquierda a derecha. Como la longitud N de los datos puede no ser un múltiplo entero de la escala s y pueden sobrar datos en los extremos se repite el procedimiento desde el extremo opuesto. Por lo tanto se tienen $2N_s$ segmentos.
- Se calcula la tendencia local de cada uno de los $2N_s$ segmentos por un ajuste de mínimos cuadrados. Entonces se determina la varianza mediante la relación:

$$F^2(s, \nu) = \frac{1}{s} \sum_{i=1}^s [Y((\nu - 1)s + i) - y_\nu(i)]^2, \quad (3.2.21)$$

para cada segmento ν , $\nu = 1, \dots, N_s$ y

$$F^2(s, \nu) = \frac{1}{s} \sum_{i=1}^s [Y(N - (\nu - N_s)s + i) - y_\nu(i)]^2, \quad (3.2.22)$$

para $\nu = N_s + 1, \dots, 2N_s$. Aquí $y_\nu(i)$ es el ajuste polinomial en el segmento ν . Dicho ajuste polinomial puede ser lineal, cuadrático, cúbico o superior. De esta forma se suele llamar esta técnica como MFDFAm, donde m es el orden del polinomio usado en el proceso de detrend.

- Promediando sobre todos los segmentos obtenemos la función de fluctuaciones de orden q , dada por:

$$F_q(s) = \left\{ \frac{1}{2N_s} \sum_{\nu=1}^{2N_s} [F^2(s, \nu)]^{q/2} \right\}^{1/q}. \quad (3.2.23)$$

- Se determina la ley de escalamiento de la función de partición analizando el gráfico log-log de $F_q(s)$ versus s para cada valor de q .

$$F_q(s) \sim s^{h(q)}, \quad (3.2.24)$$

donde $h(q)$ se conoce como el exponente generalizado de Hurst.

3.2.2. Relación entre $\tau(q)$ y $h(q)$.

Supongamos que las series x_k de longitud N es estacionaria y normalizada. Para estos casos no se requiere la técnica MFDFa ya que no hay tendencia que eliminar. De esta forma podemos usar el análisis de fluctuaciones estándar, donde usamos una versión modificada

de la varianza (ver eq 3.2.21) para cada segmento ν , $\nu = 1, \dots, N_s$ de modo tal que se tiene:

$$F_{FA}^2(s, \nu) = [Y(\nu s) - Y((\nu - 1)s)]^2. \quad (3.2.25)$$

Insertando 3.2.25 en 3.2.23 y usando 3.2.24 se tiene:

$$\left\{ \frac{1}{2N_s} \sum_{\nu=1}^{2N_s} |Y(\nu s) - Y((\nu - 1)s)|^q \right\}^{1/q} \sim s^{h(q)}. \quad (3.2.26)$$

Por simplicidad se asume que la longitud N de la serie temporal es un múltiplo entero de la escala s , obteniendo $N_s = N/s$. Por lo tanto:

$$\sum_{\nu=1}^{N/s} |Y(\nu s) - Y((\nu - 1)s)|^q \sim s^{(qh(q)-1)} \quad (3.2.27)$$

Para relacionar este formalismo con la dimensión generalizada D_q empleamos la definición del perfil dado por la ecuación 3.2.20. Es evidente que el termino $Y(\nu s) - Y((\nu - 1)s)$ en 3.2.27 es idéntico a la suma de los x_k en cada segmento ν de tamaño s . Esta suma es conocida como la caja de probabilidad $p_s(\nu)$:

$$p_s(\nu) = \sum_{k=(\nu-1)s+1}^{\nu s} x_k = Y(\nu s) - Y((\nu - 1)s) \quad (3.2.28)$$

De esta forma y asociada al exponente de Renyi $\tau(q)$, la función de partición usando la notación según Kantelhard et al. [22], está dada por:

$$Z_q(s) = \sum_{\nu=1}^{N/s} |p_s(\nu)|^q \sim s^{\tau(q)}, \quad (3.2.29)$$

A partir de 3.2.28 vemos 3.2.29 y 3.2.27 son idénticas de modo tal que comparando, se obtiene:

$$\tau(q) = qh(q) - 1. \quad (3.2.30)$$

Recordando la relación 3.2.13 se tiene:

$$D(q) = \frac{\tau(q)}{q-1} = \frac{qh(q) - 1}{q-1}. \quad (3.2.31)$$

Finalmente el espectro multifractal (o espectro de singularidades) $f(\alpha)$ relacionado con el exponente de Renyi mediante la ecuación 3.2.14, puede reescribirse en función del exponente generalizado de Hurst $h(q)$. De esta forma se tiene:

$$\alpha(q) = h(q) + q \frac{dh}{dq}, \quad f(\alpha) = q(\alpha - h(q)) + 1. \quad (3.2.32)$$

A partir de ahora continuaremos el trabajo usando la técnica MF DFA introducida por Kantelhard et al [22] y veremos como esta técnica permite distinguir y clasificar las series de tiempo.

3.3. Series de Tiempo y formalismo multifractal.

Antes de continuar introduciremos una serie de definiciones importantes que caracterizan una serie temporal, de modo tal que al aplicar el MF DFA se obtiene información de dichas características.

3.3.1. Persistencia, Correlaciones de largo y corto plazo en series de tiempo.

Se denomina **persistencia** en una serie de tiempo, a la probabilidad de que el valor de un campo no fluctúe lo suficiente desde su valor medio para cambiar de signo [23]. Consideremos el incremento $\Delta x_i = x_i - x_{i-1}$ de una serie donde $i = 1, \dots, N$ con N valores medidos en un tiempo equidistante. Uno encuentra que Δx_i puede ser persistente, independiente o anti-persistente. Para datos estacionarios la función de auto-covarianza de los incrementos,

$$C(s) = \langle \Delta x_i \Delta x_{i+s} \rangle = \frac{1}{N-s} \sum_{i=1}^{N-s} \Delta x_i \Delta x_{i+s}, \quad (3.3.1)$$

puede estudiarse para determinar el grado de persistencia. Si $C(s)$ es dividido por la varianza $\langle (\Delta x_i)^2 \rangle$, se convierte en la función de auto-correlación. Si Δx_i no están correlacionados, $C(s)$ es zero para $s > 0$.

Se dice que una serie de tiempo está **Correlacionada en el corto plazo** cuando los incrementos Δx_i se describen por $C(s)$ de modo tal que disminuye exponencialmente.

$$C(s) \sim \exp(-s/t_x), \quad (3.3.2)$$

con un tiempo de decaimiento característico t_x .

Se dice que una serie de tiempo está **Correlacionada en el largo plazo**, si $C(s)$ decae con una ley de potencia, dada por:

$$C(s) \propto s^{-\gamma}, \quad (3.3.3)$$

con el exponente γ variando entre 0 y 1 si la serie de tiempo es estacionaria.

Desde el punto vista del formalismo multifractal, el exponente de Hurst generalizado para $q = 2$, corresponde a la dimensión de correlación. Para estos casos si $h(2) > 0.5$, la serie presenta persistencia, si $h(2) < 0.5$ hablamos de anti-persistencia. Si $h(2) = 0.5$, la serie presenta comportamiento aleatorio [24].

Clasificación de Series de tiempo a partir del formalismo multifractal

- I) **Series Monofractales:** Para series de tiempo monofractales, $h(q)$ es independiente de q , ya que el escalamiento de las varianzas $F^2(s, \nu)$ es idéntico para todos los segmentos ν . De esta forma el exponente de Hurst generalizado tiene un único valor que es exactamente el exponente de Hurst, H . Así, se tiene:

$$h(q) = H, \quad (3.3.4)$$

$$\tau(q) = qH - 1. \quad (3.3.5)$$

Algunos valores conocidos de H son:

- Para series de tiempo anti-correlacionadas en el largo plazo se encuentran que H varía entre $0 < H < 0.5$.
- Para series no-correlacionadas se encuentra que $H = 0.5$
- Para series correlacionadas en el largo plazo se encuentra que $0.5 < H < 1$.

En el último caso, si la serie es estacionaria la función de autocorrelación $C(s) \sim s^{-\gamma}$, con $\gamma = 2 - 2H$. De forma alternativa, una serie de tiempo correlacionada en el largo plazo está caracterizada por una ley de potencia de modo que su densidad espectral decae de la forma:

$$S(f) \sim f^{-\beta}, \quad (3.3.6)$$

donde f es la frecuencia y $\beta = 2H - 1$.

II) **Series de tiempo multifractales correlacionadas en el largo plazo.** A continuación mostraremos un ejemplo de estas series. Consideremos un modelo binomial, donde la serie es de $N = 2^{n_{max}}$, números con $k = 1 \cdots N$ dado por:

$$x_k = a^{n(k-1)}(1-a)^{n_{max}-n(k-1)}, \quad (3.3.7)$$

donde $0.5 < a < 1$ es un parámetro y $n(k)$ es el número de dígitos iguales a 1 en representación binaria del índice k , por ejemplo $n(13) = 3$, ya que 13 corresponde al número binario 1101. Este bi-fractal puede ser visto en la figura 3.3, en diferentes acercamientos de modo que se aprecia las características de autosimilaridad. Sus parámetros como exponentes de Renyi, Hurst, Hölder y espectro multifractal pueden ser conocidos de manera exacta. De esta forma el exponente de Renyi $\tau(q)$ puede calcularse de forma directa de acuerdo a la ecuación 3.2.28. Así, se tiene:

$$p_{2s}(\nu) = p_s(2\nu - 1) + p_s(2\nu) = [(1-a)/a + 1]p_s(2\nu) = p_s(2\nu)/a, \quad (3.3.8)$$

y la función de partición está dada por:

$$Z_q(s) = \sum_{\nu=1}^{N/s} [p_s(\nu)]^q \quad (3.3.9)$$

$$= \sum_{\nu=1}^{N/2s} [p_s(2\nu - 1)]^q + [p_s(2\nu)]^q \quad (3.3.10)$$

$$= \left[\frac{(1-a)^q}{a^q} + 1 \right] \sum_{\nu=1}^{N/2s} [p_s(2\nu)]^q$$

$$= [(1-a)^q + a^q] \sum_{\nu=1}^{N/2s} [p_{2s}(\nu)]^q$$

$$= [(1-a)^q + a^q] Z_q(2s)$$

De esta forma usando 3.2.29 y 3.2.30 se tiene:

$$\tau(q) = \frac{-\ln(a^q + (1-a)^q)}{\ln 2}, \quad (3.3.11)$$

$$h(q) = \frac{1}{q} - \frac{\ln(a^q + (1-a)^q)}{q \ln 2}. \quad (3.3.12)$$

Notamos que $\tau(q)$ tiene una dependencia no lineal sobre q , indicando multifractalidad.

El exponente de Hölder y el espectro multifractal están dados respectivamente por:

$$\alpha(q) = \frac{-q(a^{q-1} + (1-a)^{q-1})}{(a^q + (1-a)^q) \ln(2)}, \quad (3.3.13)$$

$$f(\alpha) = \frac{-q^2(a^{q-1} + (1-a)^{q-1})}{(a^q + (1-a)^q) \ln(2)} + \frac{q \ln[a^q + (1-a)^q]}{\ln(2)}. \quad (3.3.14)$$

Los resultados teóricos versus los resultados al aplicar el método MF DFA pueden ser vistos en la figura 3.4. En la figura 3.4a) se observa un régimen lineal característico, donde para cada momento q , existen rectas con diferentes pendientes, las cuales corresponden a diferentes exponentes de Hurst. El polinomio de ajuste fue de orden 1. En la figura 3.4b) se observa una curva suave, donde el exponente de Hurst depende de q . En la figura 3.4c) vemos el espectro multifractal que muestra una parábola simétrica, curva que cumple con las condiciones 3.2.8 y el ancho del espectro multifractal que da una medida de cuan complejo es el bifractal estudiado.

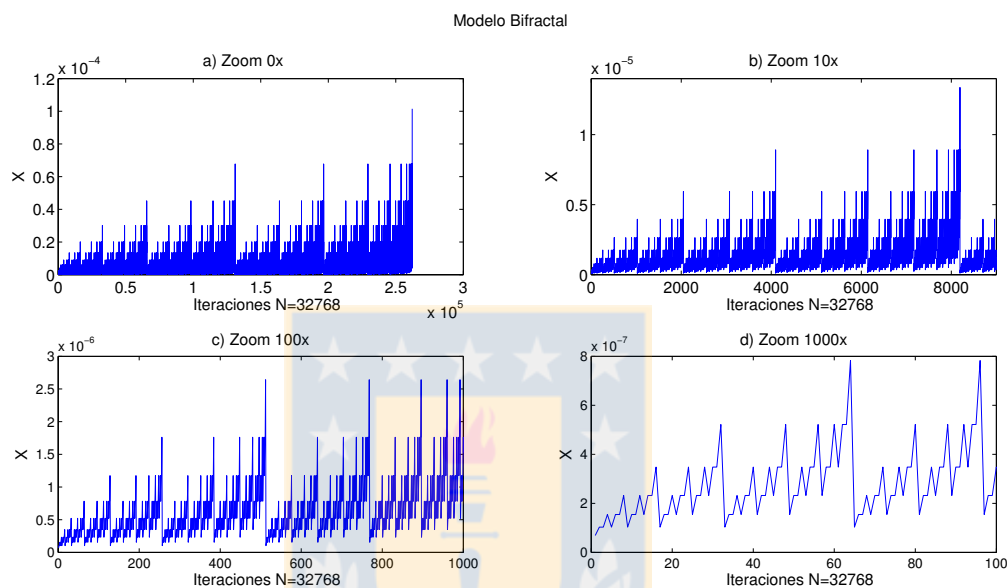


Figura 3.3: **Modelo Bifractal con N=32768 iteraciones:** Se observa la auto similitud geométrica desde un zoom normal en a), zoom 10x en b), zoom 100x en c), zoom 1000x en d). Se observa como la amplitud X cambia desde 10^{-4} a 10^{-7} . Fuente: Elaboración propia.

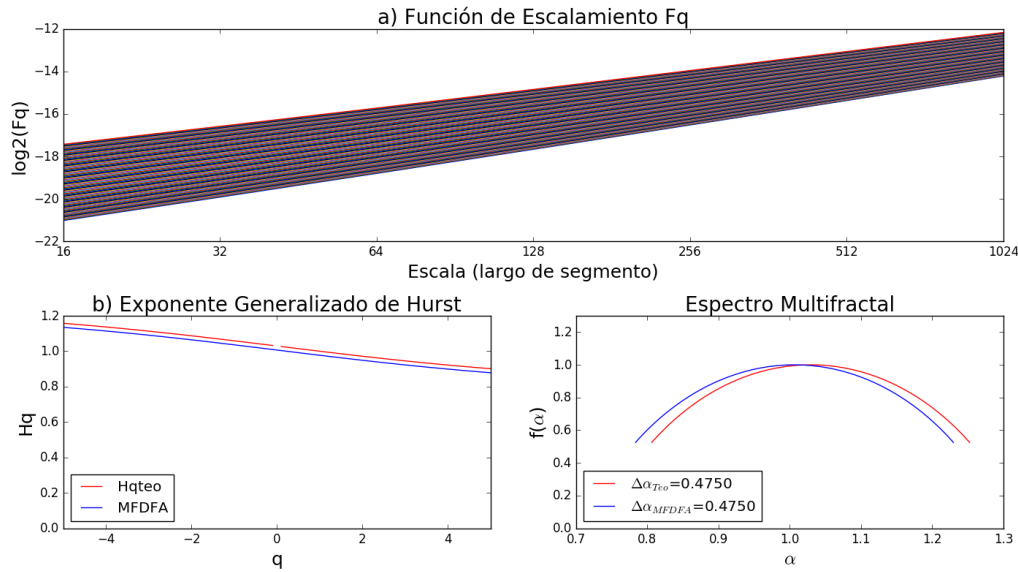


Figura 3.4: **Resultados teóricos y del método MF DFA:** a) Gráfico log-log de la función de fluctuaciones F_q versus exponente de longitud de escala s . b) Exponente de Hurst. c) Espectro multifractal. Fuente: Elaboración propia.

3.3.2. Otros ejemplos.

Se muestran a continuación ejemplos de series sintéticas que no poseen soluciones exactas a diferencia del modelo bifractal. Dichas señales sintéticas corresponden a un monofractal, ruido blanco y multifractal como lo muestra la figura 3.5 y estas señales junto con el software para el cálculo de MF DFA pueden ser obtenidas desde http://www.ntnu.edu/documents/170234/1315232/Introduction_to_MF DFA4.zip.

La columna a) en la figura 3.5 muestra una señal sintética monofractal en rojo, ruido blanco en azul y multifractal en verde. A la derecha en la columna b) se observan los regímenes lineales respectivos característicos a cada señal. En el caso de una señal monofractal observamos líneas rectas que tienden a ser paralelas. Dado que la pendiente es el exponente de Hurst, podemos deducir que éste es constante. En el caso del ruido blanco vemos un comportamiento similar. Sin embargo, la naturaleza de las señales es diferente. El caso del ruido blanco corresponde a un proceso aleatorio, mientras que el caso monofractal existe un proceso determinista subyacente. El espectro multifractal, al igual que el caso bifractal muestra diferentes pendientes para cada línea recta. Es decir, existe más de un exponente de Hurst.

La figura 3.6 a) muestra el exponente de Hurst generalizado. En rojo vemos que para la señal monofractal existe una línea paralela al eje q constante, con $H = 0.7716$, mientras que para ruido blanco (en azul) aparece un valor de $H = 0.5$ el cual es conocido en múltiples trabajos [22, 25]. En verde observamos una curva suave que depende de q .

En 3.6 b) observamos los espectros multifractales. En el caso monofractal y ruido blanco es importante notar que $f(\alpha)$ tiende a converger a un punto, mientras que para la señal multifractal sigue siendo una parábola pero, en este caso, asimétrica.

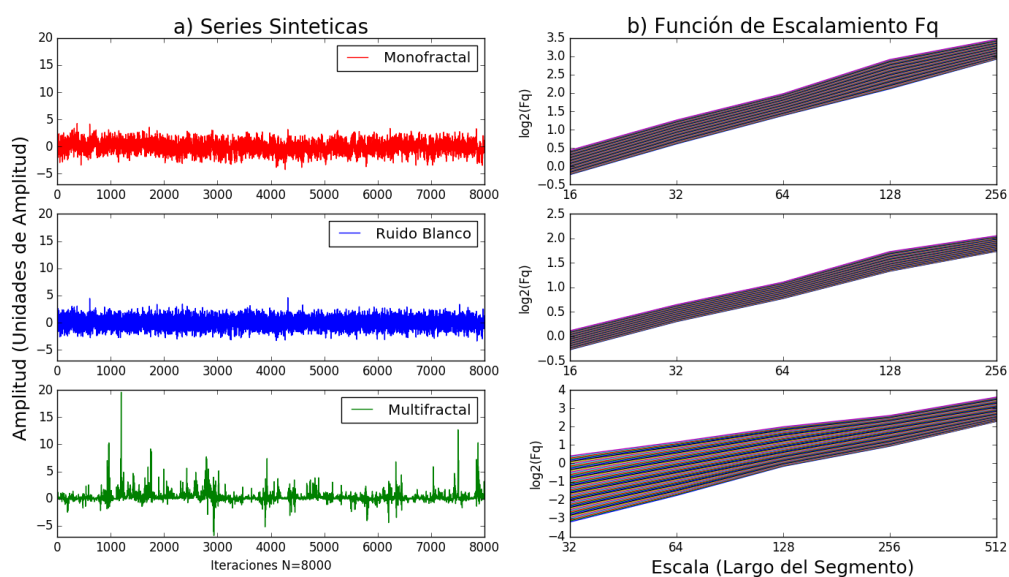


Figura 3.5: **Señales Sintéticas:** columna a) en rojo monofractal, en azul ruido blanco, en verde multifractal. columna b) regimen lineal al aplicar la técnica MF DFA, con un polinomio de orden 1, para cada señal. Fuente: Elaboración propia.

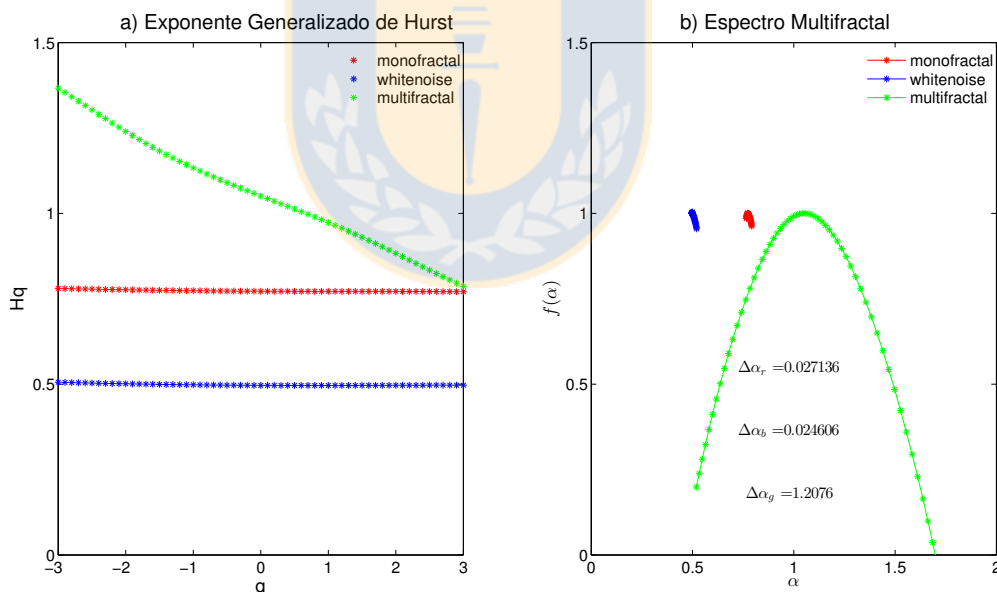


Figura 3.6: **Resultados Señales Sintéticas:** a) Exponente generalizado de Hurst. b) Espectro multifractal y sus anchos respectivos. En rojo monofractal, en azul ruido blanco, en verde multifractal. Fuente: Elaboración propia.

La figura 3.7 muestra un camino aleatorio multifractal, con $H=0.5, 0.6, 0.7, 0.8$ y 0.9 con anchos multifractales $\Delta\alpha = 0, 0.1, 0.2, 0.3, 0.4, 0.5$ y 0.6 . Se observan períodos intermitentes donde hay variaciones pequeñas y grandes de la amplitud. Las grandes va-

3.3. SERIES DE TIEMPO Y FORMALISMO MULTIFRACTAL.

riaciones en la amplitud son detectadas por los momentos positivos, mientras las pequeñas son detectadas por los momentos negativos. Dichas diferencias en las amplitudes se ven reflejadas en el ancho del espectro multifractal. A mayor ancho, existen mayores fluctuaciones, mientras a menor ancho hay menos fluctuaciones a lo largo de la señal.

La asimetría en el espectro multifractal depende de la sensibilidad del método al detectar la variación en las fluctuaciones de la señal, lo cual puede ser visto en la figura 3.8. De esta forma cuando el método no detecta grandes fluctuaciones, existe asimetría en el espectro multifractal truncándose a la izquierda. Esto se ve reflejado en el exponente generalizado de Hurst, donde para los momentos $q > 0$ el exponente de Hurst es casi constante. Por otro lado cuando el método no es capaz de detectar pequeñas fluctuaciones existe asimetría truncándose a la derecha. Esto se ve reflejado en el exponente generalizado de Hurst, donde para los momentos $q < 0$ el exponente de Hurst es casi constante.

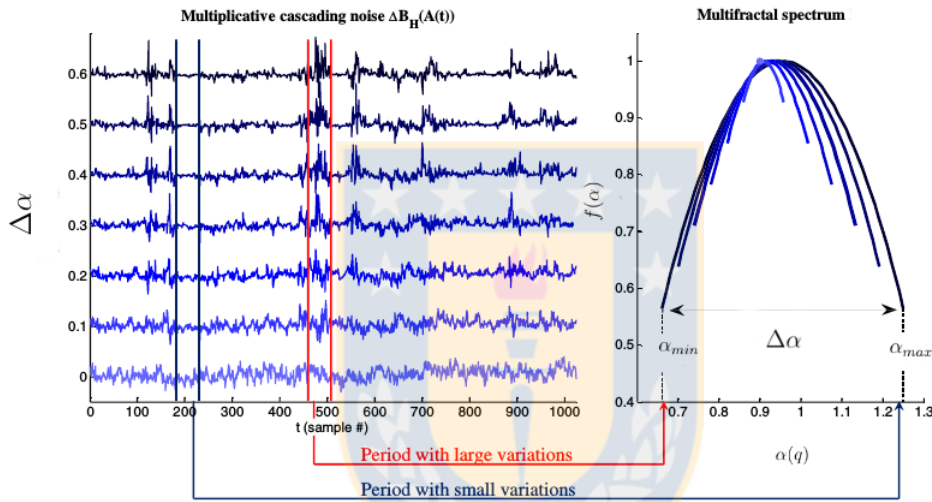


Figura 3.7: **Series de tiempo multifractal de camino aleatorio:** Se observa para ventanas en torno $t=200$ y $t=500$, como las fluctuaciones cambian en cada señal y como dichos cambios de amplitud afectan el ancho del espectro multifractal. Fuente: Ihlen (2013, p. 933, [26]) (Se cambio la notación del espectro multifractal por la usada por Halsey [18])

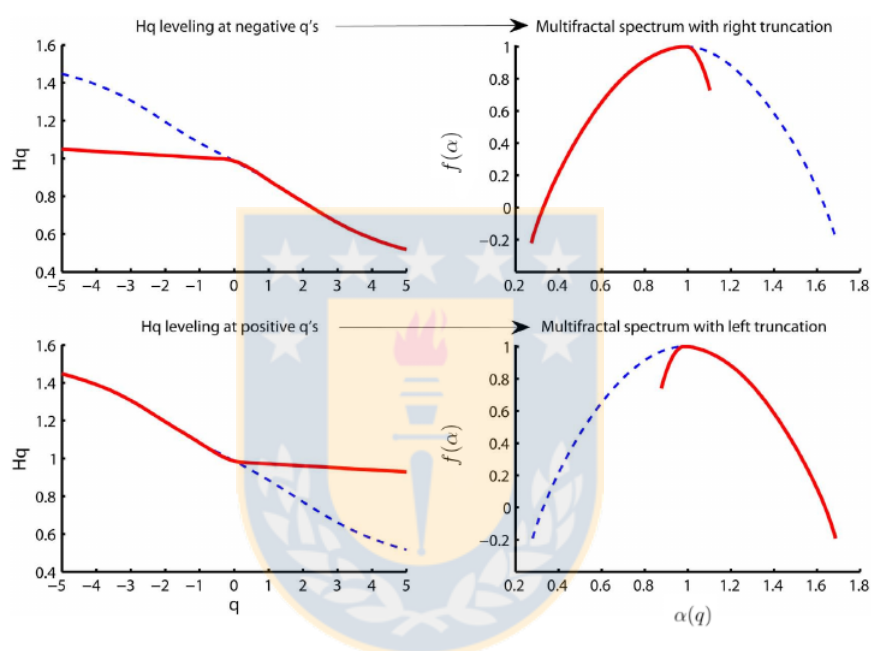


Figura 3.8: **Espectro multifractal con truncamiento:** Panel superior muestra truncamiento a la derecha. El método no es capaz de detectar pequeñas fluctuaciones. Panel inferior muestra truncamiento a la izquierda. El método no es capaz de detectar altas fluctuaciones. Fuente: Ihlen (2012, p. 12, [25]) (Se cambió la notación del espectro multifractal por la usada por Halsey [18]).

Capítulo 4

Análisis y Resultados

4.1. Base de Datos

Los datos usados para esta tesis corresponden a series de tiempo de $PM_{2.5}$ horaria de público acceso desde el sitio web del Sistema Nacional de Calidad del Aire (SINCA) [7]. Estos datos son recolectados en estaciones de monitoreo que se encuentran distribuidas en diferentes comunas del área metropolitana de Santiago, pertenecientes a la red MACAM-2 [27]. Estas cuentan con sensores capaces de medir parámetros meteorológicos como humedad relativa, temperatura ambiente, dirección y velocidad del viento. También miden parámetros contaminantes como agentes primarios (NO_x , CO_x , etc) y agentes secundarios como el ozono y dióxido de azufre (O_3 , SO_2) entre otros aerosoles. Dentro de estos aerosoles se encuentra el $PM_{2.5}$ medido por el dispositivo MET-ONE BAM1020 [28]. Las estaciones de monitoreo empleadas fueron Las Condes (LC), La Florida (LF), Parque O'Higgins (PO) y Pudahuel (PU) ya que tienen registros desde el año 2000.

	Coordinate WGS		Elevation
	Lat (S)	Long (W)	[m]
LC	33°22'26"	70°31'21"	811
LF	33°30'48"	70°35'09"	654
PO	33°27'40"	70°39'39"	562
PU	33°26'06"	70°44'52"	553

Tabla 4.1: **Coordenadas lat-long:** Se indican en grados, minutos y segundos las locaciones de las estaciones de monitoreo, junto con la elevación respecto al nivel del mar. Fuente: Elaboración propia.

En la tabla 4.1 se muestran las coordenadas latitud, longitud y elevación, mientras que en la figura 4.1 muestra la ubicación de las estaciones de monitoreo.

Para estas estaciones, se seleccionaron los datos horarios entre 01 de Enero de 2002 al 31 de Diciembre ambas del año 2015, teniendo casi 14 años de datos con un total que ascendió a $N=122712$ horas, cuyas perdidas no superan el 2.4%. Las series de tiempo

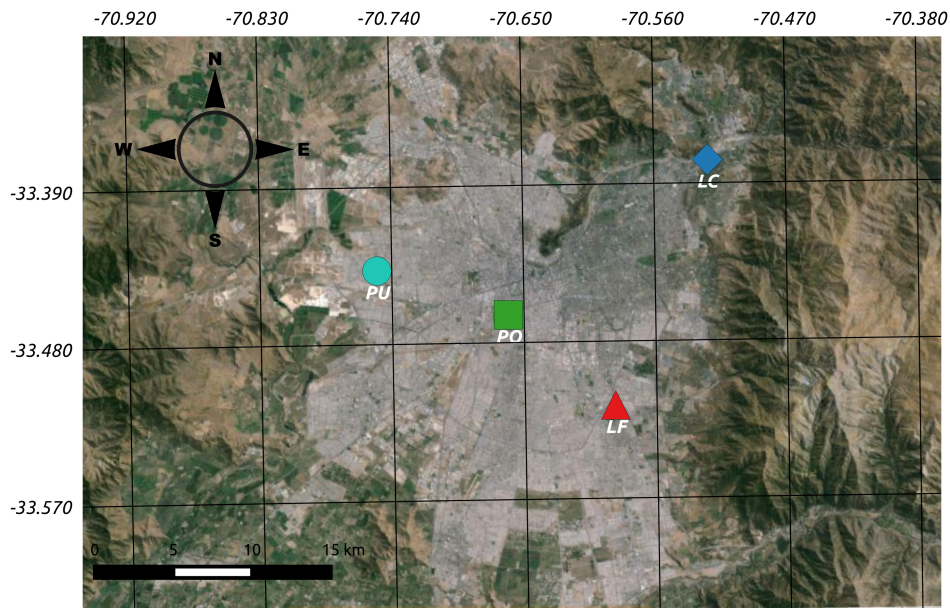


Figura 4.1: Mapa del Área metropolitana de Santiago de Chile: La grilla indica las coordenadas latitud-longitud en la cual se encuentra. Se aprecia la ubicación de las estaciones de monitoreo donde el diamante azul denota, la estación de Las Condes (LC), el triángulo rojo La Florida (LF), el círculo cian Pudahuel (PU) y el cuadrado verde Parque O'Higgins (PO). Fuente: Elaboración propia.

pueden ser vistas en la figura 4.2

La estación de monitoreo que registró más días sin datos fue Parque O'Higgins con 7 días consecutivos de perdidas desde el 13 al 20 de Abril y 9 días consecutivos entre el 10 y 19 de Julio del 2005. Mientras la estación de Pudahuel registró 9 días consecutivos sin datos durante el 01 y 10 de Marzo del 2010. El resto de los datos perdidos se distribuyen en períodos menores a 4 días consecutivos. Los datos faltantes fueron rellenados usando una interpolación lineal evitando, por un lado, afectar la distribución estadística anual y por otro lado introducir efectos no lineales.

4.2. Metodología

En una primera etapa se realizó una caracterización estadística de las series de tiempo calculando medias y desviaciones estándar, que se resumen en la tabla 4.2a).

Como criterio de selección, se calcularon medias diarias, entre inicios del 2002 a fines

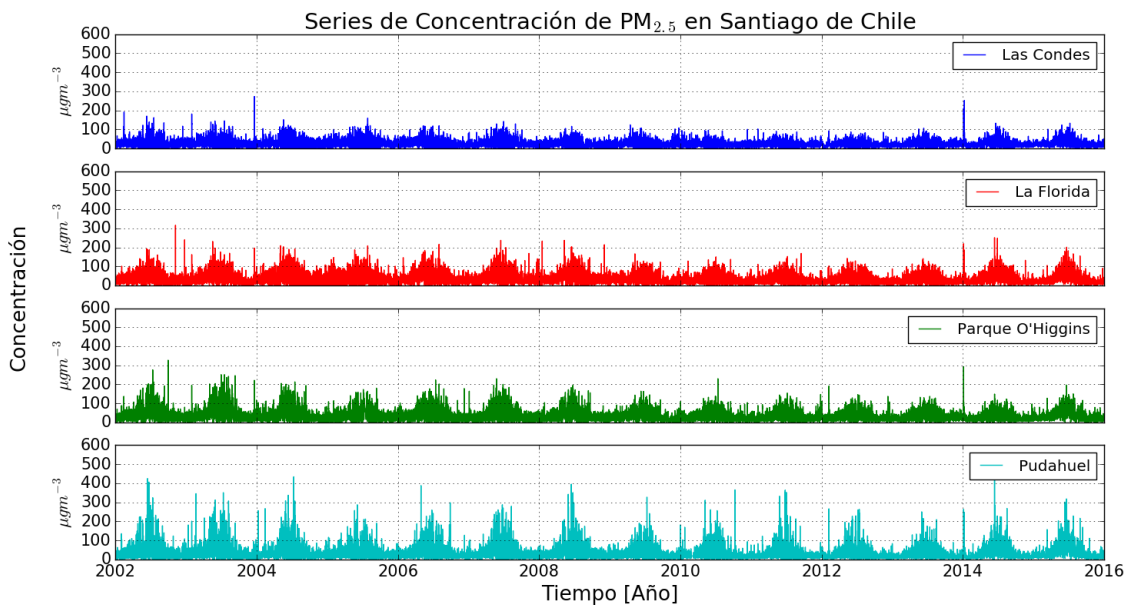


Figura 4.2: **Series de tiempo horarias de Concentración de $PM_{2.5}$ en Santiago de Chile:** Podemos observar como la comuna de Pudahuel es la que registra mayores concentraciones horarias en comparación a las demás estaciones de monitoreo. Fuente: Elaboración propia.

del 2015 (2002-2015) para conocer el papel de la varianza subsidiaria de la señal. A partir de estas señales se calcularon las tendencias y su significancia estadística usando el test de Mann-Kendall modificado según Hamed y Rao [9], el cual permite distinguir entre tendencias monótonamente crecientes o decrecientes bajo la hipótesis nula ($H=0$) de que la serie no posee tendencia. Esta hipótesis nula puede ser aceptada ($H=0$) o rechazada ($H=1$) con una significancia α por lo general del 5%. Los resultados pueden ser vistos en la tabla 4.3a).

Junto a las medias diarias, se determinaron medias mensuales y anuales las cuales pueden ser vistas en la figura 4.3.

Para estudiar el comportamiento de la tendencia de las series se dividió cada una en dos partes iguales tomando los datos desde principios de 2004 a fines de 2008 (2004-2008) y desde inicios de 2009 a fines de 2013 (2009-2013), realizando nuevamente la caracterización estadística, junto con el cálculo de tendencias y sus significancias a partir de medias diarias que se resumen en las tablas 4.2b), 4.2c), 4.3b) y 4.3c), respectivamente.

Se calcularon también medias estivales desde Octubre hasta Marzo y medias invernales desde Abril a Septiembre las cuales pueden ser vistas en las figuras 4.3a) y 4.3b), respectivamente y estos resultados pueden ser vistos en la tabla 4.4a), 4.4b), 4.4c), respectivamente, además de la razón invernal (I)/estival (E).

Se desarrolló una descomposición espectral en tiempo y frecuencia usando la ondeleta de Morlet que es de uso común en el estudio de procesos geofísicos [29] y se promedió la amplitud global de ondeletas en el tiempo entre los años 2004-2008 y entre los años 2009-2013 en cada serie de tiempo según la corrección de Liu et al [14] y la energía concentrada en torno a los peaks más importantes.

Finalmente se hizo un análisis multifractal de las series de tiempo utilizando la técnica MF DFA considerando en primera instancia los datos crudos entre los años 2002-2015, 2004-2008 y 2009-2013 para cada estación de monitoreo.

Caracterización Estadística de $PM_{2.5}$

	a) 2002-2015		b) 2004-2008		c) 2009-2013	
Estación	Media μgm^{-3}	Std μgm^{-3}	Media μgm^{-3}	Std μgm^{-3}	Media μgm^{-3}	Std μgm^{-3}
LC	23.66	16.05	24.96	16.87	21.07	12.84
LF	30.97	22.58	34.77	24.70	26.61	17.75
PO	31.23	24.39	33.90	25.86	27.22	19.15
PU	30.87	30.95	32.89	31.45	26.91	26.81

Tabla 4.2: **Medias y Desviaciones Estándar (std)**: Para series de concentración horaria de cada estación de monitoreo entre los años (a)2002-2015, (b)2004-2008 y (c)2009-2013. Fuente: Elaboración propia.

Test de Mann-Kendall (Medias diarias, $\alpha = 5\%$).

	a)Trend 2002-2015		b)Trend 2004-2008		c)Trend 2009-2013	
Estación	$\mu gm^{-3} yr^{-1}$	H	$\mu gm^{-3} yr^{-1}$	H	$\mu gm^{-3} yr^{-1}$	H
LC	-0.39	1	-1.11	1	-0.55	1
LF	-0.69	1	-0.45	0	-1.17	1
PO	-0.92	1	-0.87	1	-0.55	0
PU	-0.60	1	-0.06	0	-1.21	1

Tabla 4.3: **Tendencias y resultados del test de Mann-Kendall**: Para series de concentración de medias diarias entre los años (a)2002-2015, (b)2004-2008 y (c)2009-2013. Fuente: Elaboración propia.

Medias Invernales (I) y Estivales (E) [μgm^{-3}]

	a) 2002-2015			b) 2004-2008			c) 2009-2013		
Estación	I	E	I/E	I	E	I/E	I	E	I/E
LC	27.29	20.00	1.36	28.62	21.27	1.36	23.61	18.51	1.28
LF	38.98	22.91	1.70	42.77	26.71	1.60	32.48	20.70	1.56
PO	40.83	21.54	1.89	44.09	23.64	1.87	34.62	19.77	1.75
PU	42.52	19.15	2.22	44.15	21.57	2.04	36.85	16.89	2.18

Tabla 4.4: **Medias Invernales y Estivales**: Se muestran las medias invernales (I), estivales (E) y las razones entre estas (I/E) para las cuatro estaciones de monitoreo, entre los años (a) 2002-2015, (b) 2004-2008 y (c) 2009-2013. Fuente: Elaboración propia.

4.3. Análisis Estadístico y Tendencias

Las tablas 4.2a), 4.2b) y 4.2c) muestran las medias y desviaciones estándar entre los años 2002-2015, 2004-2008 y 2009-2013. Se observa una desviación estándar del mismo

4.3. ANÁLISIS ESTADÍSTICO Y TENDENCIAS

orden de magnitud que la media, donde destacamos que para la estación de Pudahuel (PU) los valores son casi idénticos. Así entre los años 2002-2015 la media y desviación estándar respectiva es de $30.87 \mu\text{gm}^{-3}$ y $30.95 \mu\text{gm}^{-3}$.

Al comparar 4.2b) con 4.2c) se observa que entre los años 2009-2013 para la estación de Las Condes (LC) la media disminuyó en una diferencia porcentual del 16 % y la desviación estándar un 24 %, con respecto a los años 2004-2013. Para la estación de La Florida (LF), Parque O'Higgins (PO) y Pudahuel (PU) las medias disminuyeron un 23 %, 20 % y 18 %, respectivamente. Por otro lado, las desviaciones estándar disminuyeron un 28 %, 25 % y 15 %, respectivamente.

El estudio de las tendencias muestra que la concentración disminuye a tasas diferentes en cada una de las estaciones de monitoreo. Las tablas 4.3a), 4.3b) y 4.3c) muestran dichas tendencias y la aceptación o rechazo de la hipótesis nula resultante del test de Mann-Kendall. Entre los años 2002-2015 la tabla 4.3a) muestra que las tendencias son negativas y decrecientes según el test de Mann-Kendall, donde Parque O'Higgins (PO) disminuye más rápido con un valor de $-0.92 \mu\text{gm}^{-3}\text{yr}^{-1}$. Entre los años 2004-2008 4.3b) se muestran tendencias negativas, siendo en Pudahuel (PU) casi nula. Las Condes (LC) y Parque O'Higgins (PO) tienen tendencias decrecientes según el test de Mann Kendall, disminuyendo a una tasa de $-1.11 \mu\text{gm}^{-3}\text{yr}^{-1}$ y $-0.87 \mu\text{gm}^{-3}\text{yr}^{-1}$, respectivamente. Finalmente entre los años 2009-2013 3c) se observan tendencias negativas y decrecientes según el test de Mann Kendall exceptuando Parque O'Higgins.

La figura 4.3 muestra las series de tiempo obtenidas al calcular medias diarias, mensuales y anuales, las cuales pueden ser vistas en azul, rojo y verde respectivamente. Se muestra un aumento en la concentración de $\text{PM}_{2.5}$ a partir de inicios 2014, volviendo a valores anteriores al año 2009.

La figura 4.4a) muestra las medias invernales de cada año, mientras que la figura 4.4b) las medias estivales. Se observa que la mayor concentración de $\text{PM}_{2.5}$ se produce en la temporada otoño-invierno de cada año mientras que esta disminuye considerablemente en la temporada primavera verano. En ambas figuras se muestra que en el período 2009-2013 la concentración de $\text{PM}_{2.5}$ disminuyó respecto al período 2004-2008, aumentando a partir del 2014 sólo en la temporada otoño-invierno. Cabe destacar que las medias estivales han ido disminuyendo en todo el período estudiado.

Las tablas 4a), 4b), 4c) muestra las medias invernales y estivales entre los años 2002-2015, 2004-2008 y 2009-2013 en cada estación de monitoreo. En el caso de Pudahuel (PU) el período invernal tiene valores de $42.52 \mu\text{gm}^{-3}$, $44.15 \mu\text{gm}^{-3}$ y $36.85 \mu\text{gm}^{-3}$, mientras el período estival tiene valores de $19.15 \mu\text{gm}^{-3}$, $21.57 \mu\text{gm}^{-3}$, $16.89 \mu\text{gm}^{-3}$, donde las razones entre temporadas estivales e invernales son 2.22, 2.04 y 2.18.

Al comparar las tablas 4b) y 4c) se observa que en la temporada invernal (I) entre los años 2009-2013 la concentración de $\text{PM}_{2.5}$ bajó entre un 17 % y 25 %, mientras la temporada estival (E) bajó entre un 13 % y 23 % respecto a los años 2004-2008.

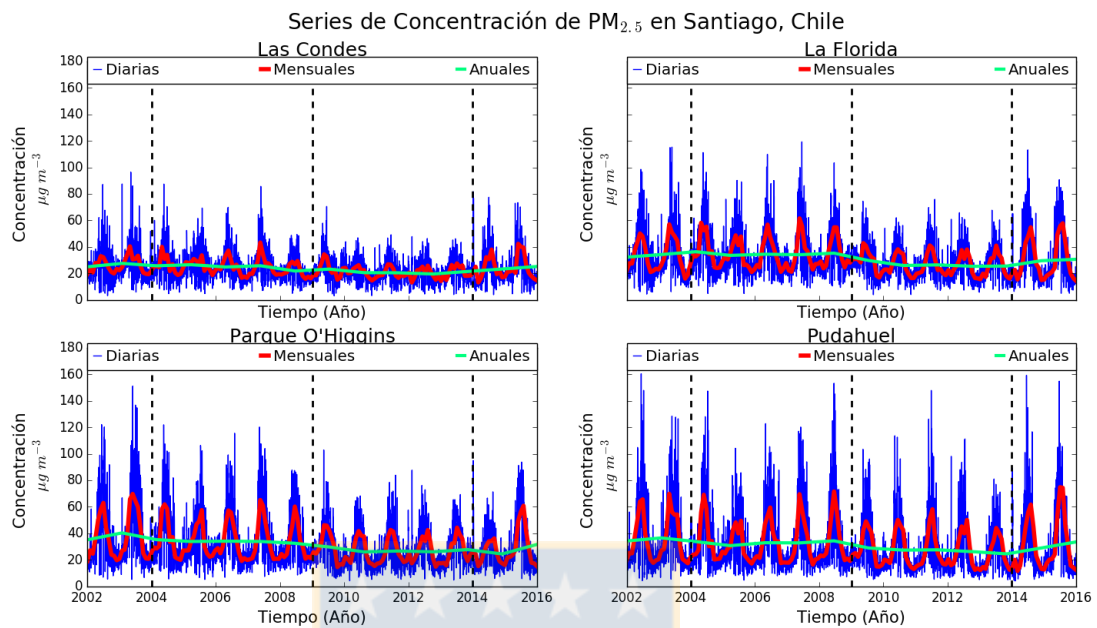


Figura 4.3: **Medias diarias, mensuales y anuales:** En azul, rojo y verde respectivamente, para cada estación de monitoreo entre el 2002-2016. Las líneas entrecortadas indican los intervalos de tiempo, desde el 1 de Enero del 2004 hasta el 31 de Diciembre del 2008 y desde el 1 de Enero del 2009 hasta el 31 de Diciembre del 2014. Fuente: Elaboración propia.

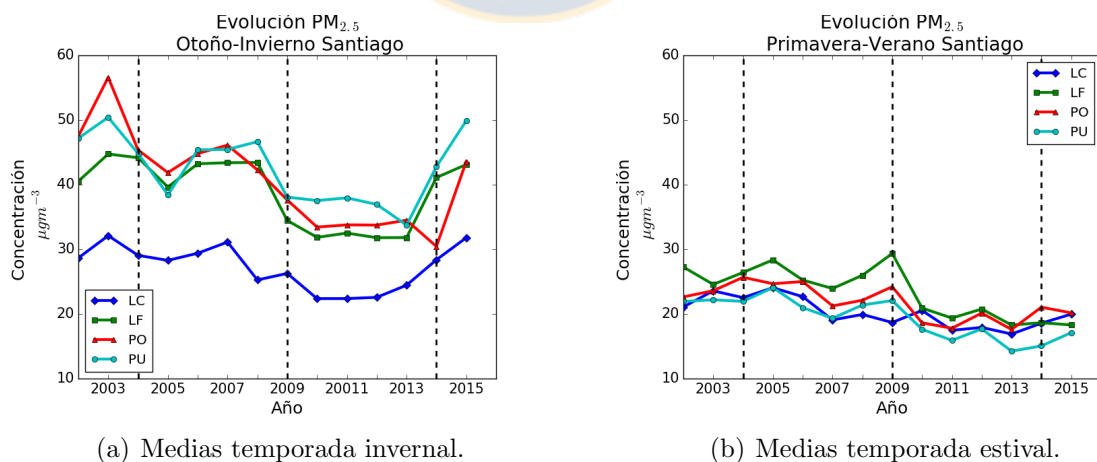


Figura 4.4: **Evolución de medias invernales y estivales:** Se aprecia la evolución de la temporada invernal a) y estival b) entre los años 2002-2015. Las líneas entrecortadas indican los intervalos de tiempo, desde el 1 de Enero del 2004 hasta el 31 de Diciembre del 2008 y desde el 1 de Enero del 2009 hasta el 31 de Diciembre del 2014. Fuente: Elaboración propia.

4.4. Ondeletas

La figura 4.5a) muestra la serie de tiempo horaria de la estación de monitoreo de Pudahuel (PU) entre los años 2002-2015, mientras en la figura 4.5b) podemos ver el mapa de contorno de la transformada de ondeleta [30], junto con la escala de colores donde el azul indica una menor amplitud que parte de $0 \mu\text{gm}^{-3}$, mientras el rojo indica una mayor concentración que no supera los $128 \mu\text{gm}^{-3}$. El eje vertical representa el período medido en días y el eje horizontal muestra como ha ido evolucionando a lo largo del tiempo entre los años 2002-2015. La línea negra muestra el cono de influencia [14, 29], donde los datos fuera de éste no pueden ser interpretados debido a los efectos de borde, por lo cual son suprimidos de la figura.

Para cualquier año podemos ver que la mayor amplitud de $\text{PM}_{2.5}$ se concentra en la temporada invernal desde Abril a Septiembre, especialmente en el ciclo diario donde el contorno es rojo, mientras que en la temporada estival de Octubre a Marzo disminuye hacia tonos más claros como por ejemplo el cyan. Este ciclo es secundado por el ciclo de medio día, donde en invierno también se aprecian tonos rojizos que tienden a ser más claros. Simultáneamente, la amplitud disminuye gradualmente hacia periodicidades más altas como el ciclo semanal (7 días, donde vemos amplitudes importantes en los años 2004 y 2008) y el ciclo mensual (30 días). No obstante, el ciclo anual es también fuerte (amarillo) y modula la señal. Las líneas verticales entrecortadas delimitan los intervalos entre los años 2004-2008 y 2009-2013 que es donde se encuentra la disminución en la concentración de $\text{PM}_{2.5}$ observada anteriormente tanto en las figuras 4.3 y 4.4, como en los s 4.2b) y 4.2c). Resultados similares al aplicar la transformada de ondeletas pueden ser vistos en las otras estaciones de monitoreo, en el Apéndice A.

La figura 4.5c) corresponde a la amplitud global de ondeletas que se obtiene promediando en el tiempo las amplitudes de los coeficientes de la transformada de ondeletas. Aquí se muestra como se distribuye la amplitud en todas las componentes de la señal, entre los años 2004-2008 (rojo) y 2009-2013 (verde) donde se aprecia que la mayor amplitud de la señal se alcanza en el ciclo anual, secundado por el ciclo diario. Se observa que el ciclo anual disminuyó en un 6% desde $17.02 \mu\text{gm}^{-3}$ a $15.84 \mu\text{gm}^{-3}$, mientras el ciclo diario disminuyó en un 10% desde $11.38 \mu\text{gm}^{-3}$ a $10.21 \mu\text{gm}^{-3}$ respecto a los años 2004-2008.

La figura 4.6 muestra el espectro global de ondeletas para cada estación de monitoreo, donde se identifican tres peaks importantes ubicados en el ciclo de medio día, diario y anual.

A partir de este espectro es posible conocer la energía total de la señal (la varianza) promediando sobre todas las periodicidades y como ésta se concentra en torno a los peaks de mayor intensidad, eligiendo una ventana de tiempo en torno a estos [11, 29]. Para este caso se seleccionaron ventanas entre 0.4 y 0.6 días (9-15 horas), 0.8 y 1.13 días (19-27 horas) y 338 a 402 días, para cada peak respectivamente.

Para Pudahuel (PU) la figura 4.6a) muestra que la varianza total entre los años 2004-2008 (rojo) y 2009-2013 (verde) fue de $974.23 [\mu\text{gm}^{-3}]^2$ y $724.15 [\mu\text{gm}^{-3}]^2$, respectivamente, disminuyendo en un 25%. Entre los años 2004-2008 la varianza fue de $91.01 [\mu\text{gm}^{-3}]^2$, $157.53 [\mu\text{gm}^{-3}]^2$ y $87.93 [\mu\text{gm}^{-3}]^2$ en torno a cada peak. Así mismo, entre los años 2009-2013 la varianza fue de $56.80 [\mu\text{gm}^{-3}]^2$, $137.50 [\mu\text{gm}^{-3}]^2$ y $76.27 [\mu\text{gm}^{-3}]^2$, respectivamente.

te. Así, se tiene que la varianza en cada peak disminuyó en un 37 %, 12 % y 13 %, relativo a los años 2004-2008.

Para Las Condes (LC) la figura 4.6b) muestra que la varianza total entre los años 2004-2008 (rojo) y 2009-2013 (verde) fue de 279.41 $[\mu\text{gm}^{-3}]^2$ y 165.77 $[\mu\text{gm}^{-3}]^2$, respectivamente, disminuyendo en un 41 %. Entre los años 2004-2008 la varianza fue de 25.21 $[\mu\text{gm}^{-3}]^2$, 66.66 $[\mu\text{gm}^{-3}]^2$ y 11.65 $[\mu\text{gm}^{-3}]^2$ en torno a cada peak. Así mismo, entre los años 2009-2013 la varianza fue de 15.06 $[\mu\text{gm}^{-3}]^2$, 34.46 $[\mu\text{gm}^{-3}]^2$ y 6.11 $[\mu\text{gm}^{-3}]^2$, respectivamente. Así, se tiene que la varianza en cada peak disminuyó en un 40 %, 48 % y 47 % relativo a los años 2004-2008.

Para La Florida (LF) la figura 4.6c) muestra que la varianza total entre los años 2004-2008 (rojo) y 2009-2013 (verde) fue de 603.75 $[\mu\text{gm}^{-3}]^2$ y 320.27 $[\mu\text{gm}^{-3}]^2$, respectivamente, disminuyendo en un 46 %. Entre los años 2004-2008 la varianza fue de 109.37 $[\mu\text{gm}^{-3}]^2$, 77.27 $[\mu\text{gm}^{-3}]^2$ y 47.04 $[\mu\text{gm}^{-3}]^2$ en torno a cada peak. Así mismo, entre los años 2009-2013 la varianza fue de 60.64 $[\mu\text{gm}^{-3}]^2$, 43.13 $[\mu\text{gm}^{-3}]^2$ y 28.84 $[\mu\text{gm}^{-3}]^2$, respectivamente. Así, se tiene que la varianza en el ciclo de medio día y diario disminuyó en un 44 %, mientras que el ciclo anual disminuyó un 38 % relativo a los años 2004-2008.

Finalmente para Parque O'Higgins (PO) la figura 4.6d) muestra que la varianza total entre los años 2004-2008 (rojo) y 2009-2013 (verde) fue de 667.68 $[\mu\text{gm}^{-3}]^2$ y 366.18 $[\mu\text{gm}^{-3}]^2$ respectivamente, disminuyendo en un 45 %. Entre los años 2004-2008 la varianza fue de 81.30 $[\mu\text{gm}^{-3}]^2$, 64.15 $[\mu\text{gm}^{-3}]^2$ y 75.75 $[\mu\text{gm}^{-3}]^2$ en torno a cada peak. Así mismo, entre los años 2009-2013 la varianza fue de 46.25 $[\mu\text{gm}^{-3}]^2$, 43.32 $[\mu\text{gm}^{-3}]^2$ y 38.14 $[\mu\text{gm}^{-3}]^2$, respectivamente. Así, se tiene que la varianza en cada peak disminuyó en un 43 %, 32 % y 50 % relativo a los años 2004-2008.

4.5. Multifractalidad

El análisis multifractal de las señales crudas de contaminación de Pudahuel, muestran en la figura 4.7 b) que para los períodos 2002-2015 y 2009-2013, estos comparten las mismas longitudes de escala, s , entre 128 y 2048, no así con el caso 2004-2008 donde las escalas s se encuentran entre 128 y 1024. Se destaca que entre dichas escalas se encuentra una diferencia de un orden de magnitud según lo establece Kantelhard [31]. Entre dichas escalas existen regímenes lineales que cumplen con la ecuación 3.2.24, encontrándose que cada línea recta posee diferente pendiente, es decir diferentes exponentes de Hurst.

La figura 4.8 a) muestra el exponente generalizado de Hurst, para cada uno de los períodos, donde se observa una curva suave con cambios de pendiente en $q = 1$, mostrando características de multifractalidad en cada período. La figura 4.8 b) muestra las curvas características del espectro multifractal, en este caso asimétricas con truncamiento en los momentos negativos. Se observa que el espectro multifractal entre los años 2004-2008 tiene un ancho multifractal de 0.4584, mientras el período 2009-2013 tiene un ancho multifractal de 0.3432, es decir este último período es menos complejo que el anterior. Para $q(2)$, se encuentra entre los períodos 2002-2015, 2004-2008, 2009-2013, los valores de 0.8033, 0.7948, 0.7793, respectivamente. Esto muestra que la serie de tiempo de Pudahuel

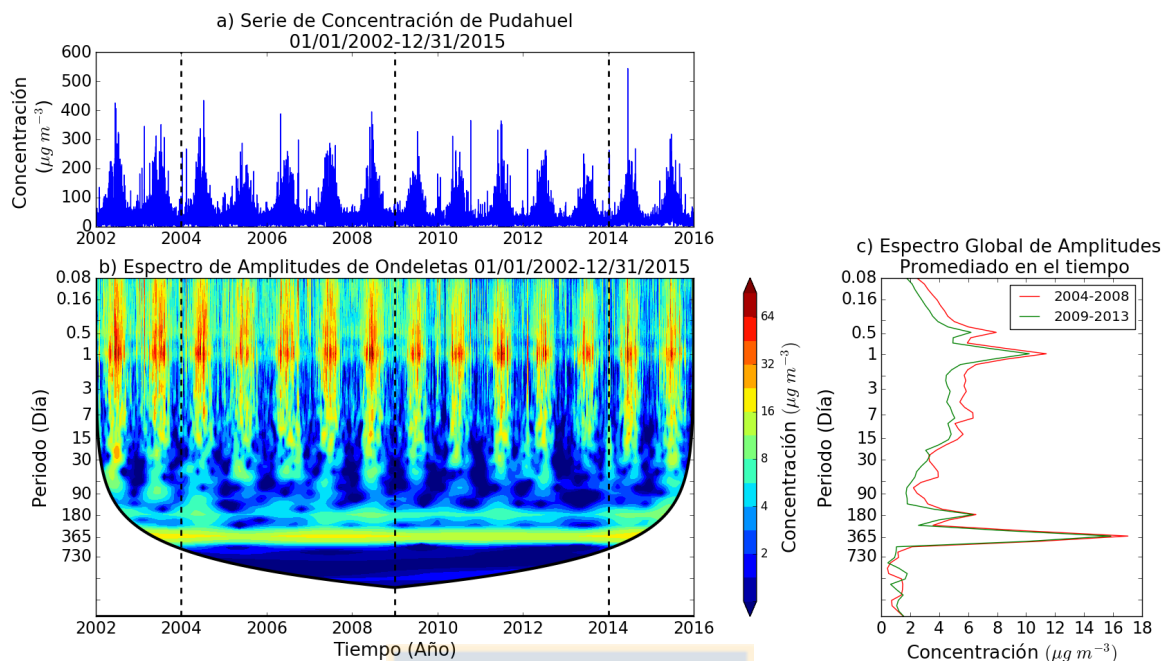


Figura 4.5: **Análisis de ondeletas en Pudahuel:** a) Serie de tiempo de Pudahuel. b) Diagrama de Contorno de la transformada de ondeletas. El eje vertical indica el período medido en días y el eje horizontal el tiempo. El área en blanco denota el exterior del cono de influencia donde no se puede hacer una interpretación de los datos debido a los efectos de borde. c) Amplitud global de ondeletas promediado en el tiempo. Fuente: Elaboración propia.

es persistente y correlacionada en el largo plazo en cada período.

Para la estación de monitoreo de Las Condes se encontraron que las escalas s varían entre 32 y 512, para los períodos 2002-2015, 2004-2008 y 2009-2013, respectivamente. El régimen lineal encontrado muestra diferentes exponentes de Hurst.

El exponente de Hurst generalizado, para Las Condes, en la figura 4.9a) muestra curvas suaves dando cuenta que la serie de contaminación en los distintos períodos corresponde a un multifractal. En la figura 4.9b) vemos que las curvas del espectro multifractal son asimétricas, donde el período 2002-2015 es mucho menos complejo que los períodos 2004-2008 y 2009-2013 con un ancho multifractal de 0.2702, 0.4120 y 0.4912, respectivamente. Para $q(2)$ los valores encontrados fueron de 0.8370, 0.8052 y 0.7789, entre los períodos 2002-2015, 2004-2008 y 2009-2013 respectivamente. Esto muestra que la serie de tiempo de Las Condes es persistente y correlacionada en el largo plazo en cada período.

Para la estación de monitoreo de La Florida se encontraron que las escalas s varían entre 128 y 2048, para los períodos 2002-2015, 2004-2008 y 2009-2013 respectivamente y el régimen lineal encontrado muestra diferentes exponentes de Hurst.

El exponente de Hurst generalizado, para La Florida, en la figura 4.10a) muestra curvas suaves dando cuenta que la serie de contaminación en los distintos períodos corresponde

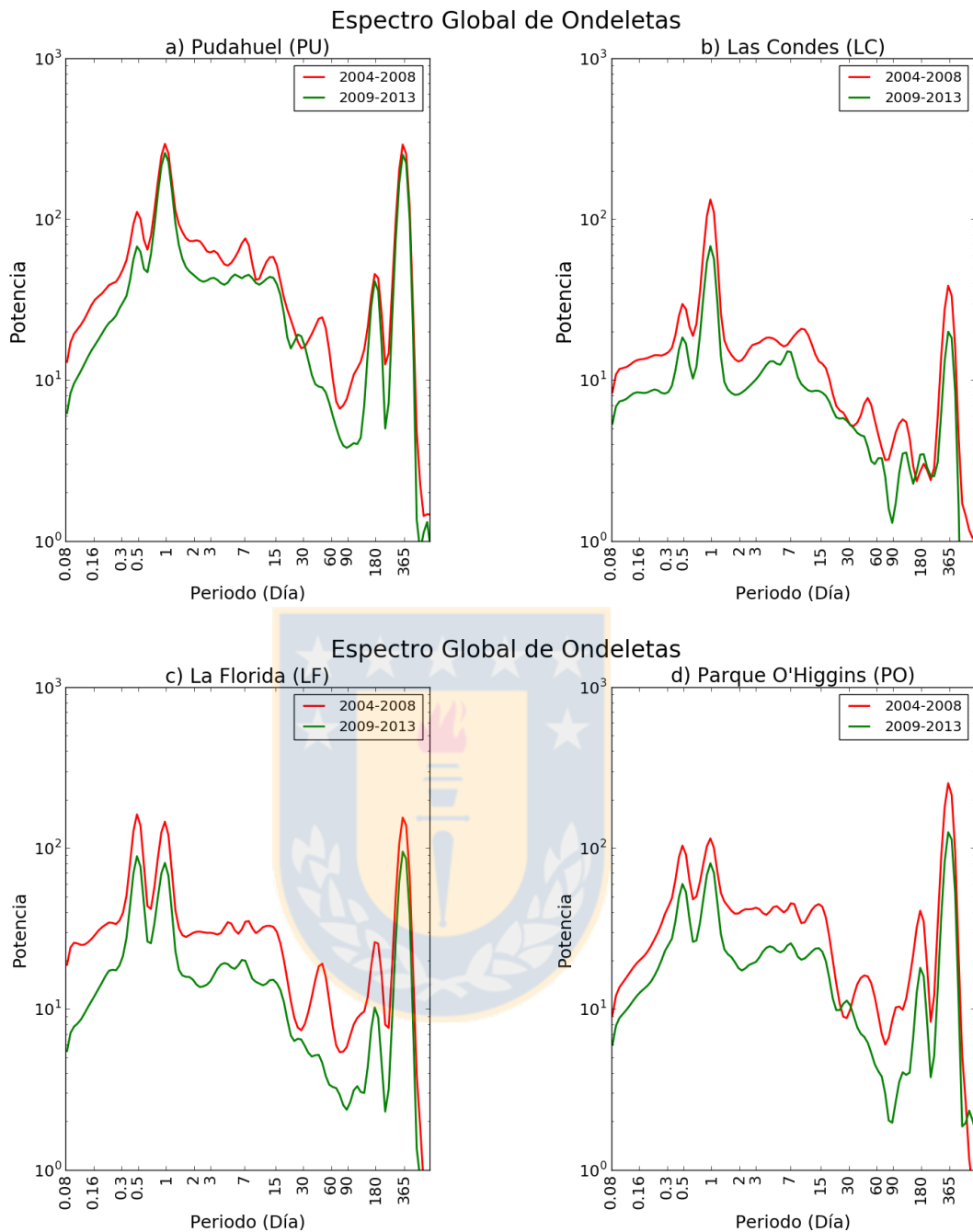


Figura 4.6: **Espectro Global de Ondeletas**: El eje horizontal corresponde al período medido en días y el eje vertical corresponde a la Potencia Espectral promediada en el tiempo. Las líneas rojo y verde indican los intervalos entre los años 2004-2008 y 2009-2013. Fuente: Elaboración propia.

a un multifractal. En la figura 4.10b) vemos que las curvas del espectro multifractal son asimétricas, donde el período 2002-2015 es mucho menos complejo que los períodos 2004-2008 y 2009-2013 con un ancho multifractal de 0.2252, 0.2958 y 0.3271, respectivamente. Para $q(2)$ los valores encontrados fueron de 0.7773, 0.7929 y 0.7499, entre los períodos 2002-2015, 2004-2008 y 2009-2013 respectivamente. Esto muestra que la serie de tiempo

4.5. MULTIFRACTALIDAD

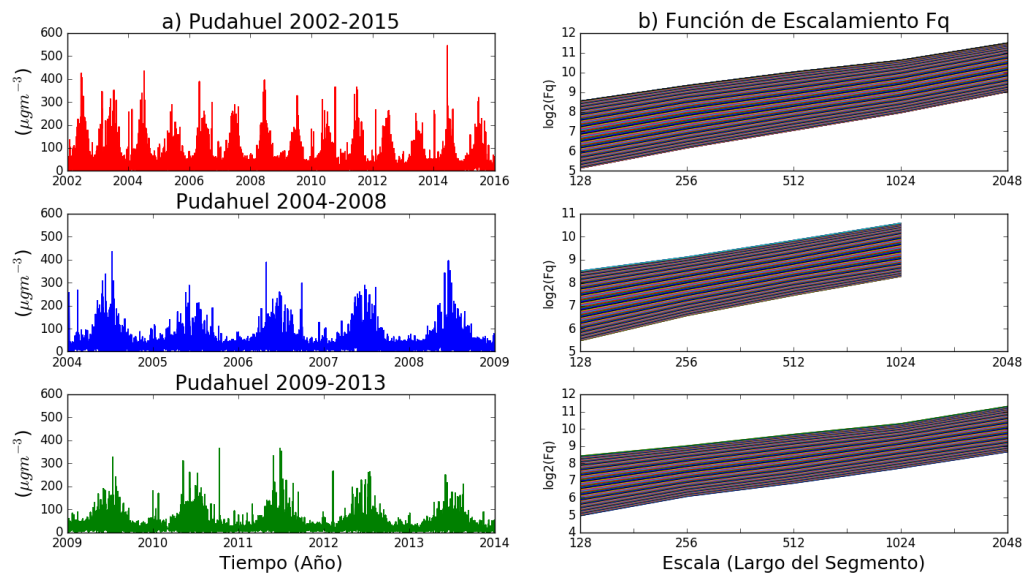


Figura 4.7: **Análisis de Pudahuel 1:** La columna a) muestra la serie de tiempo de Pudahuel entre los años 2002-2015, 2004-2008, 2009-2013. La columna b), muestra los regímenes lineales correspondientes al aplicar la técnica MFDFA sobre Pudahuel. El polinomio de ajuste para los tres casos fueron de orden 1. Fuente: Elaboración propia.

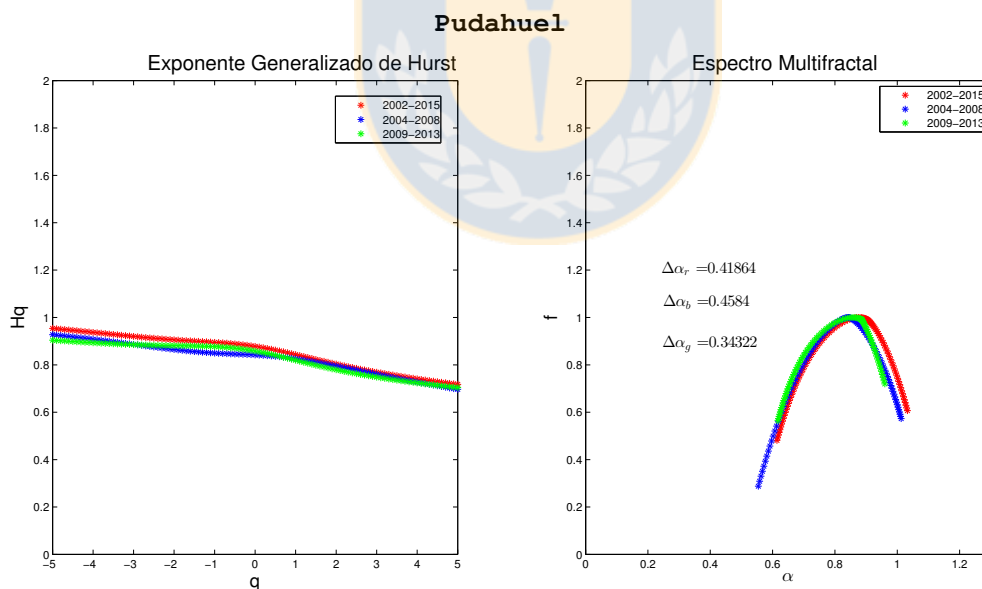


Figura 4.8: **Análisis de Pudahuel 2:** a) Exponente de Hurst. En rojo 2002-2015, en azul 2004-2008, en verde 2009-2013. b) Espectros Multifractales correspondientes. Fuente: Elaboración propia.

de La Florida es persistente y correlacionada en el largo plazo en cada período.

Para la estación de monitoreo de Parque O'Higgins se encontraron que las escalas s

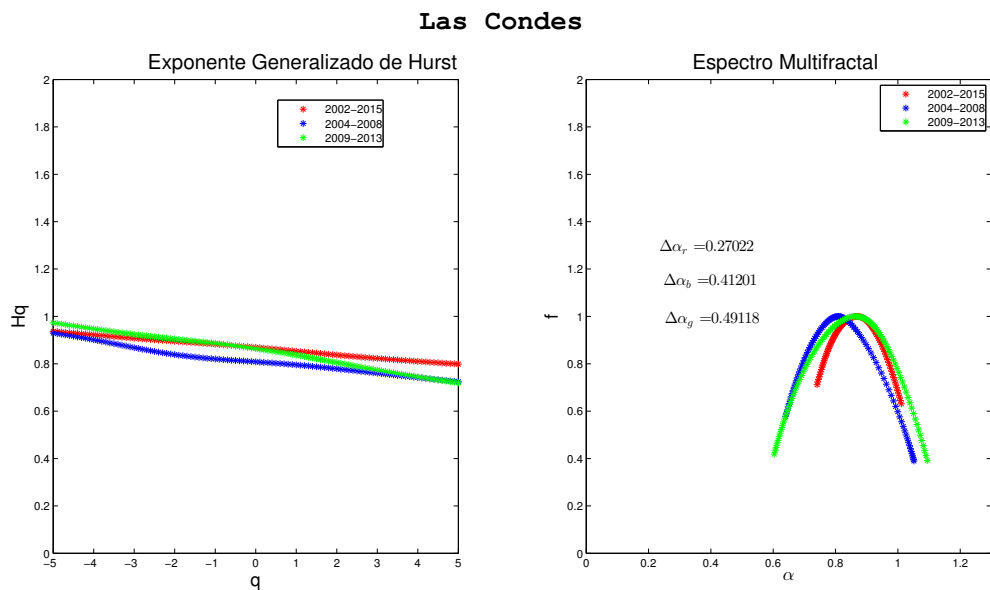


Figura 4.9: **Análisis de Las Condes:** a) Exponente de Hurst. En rojo 2002-2015, en azul 2004-2008, en verde 2009-2013. b) Espectros Multifractales correspondientes. Fuente: Elaboración propia.

varían entre 128 y 2048, para el período 2002-2015, mientras que las escalas s variaron entre 128 y 1024, para los períodos 2004-2008 y 2009-2013 respectivamente. Los regímenes lineales encontrados muestran diferentes exponentes de Hurst.

El exponente de Hurst generalizado, para Parque O'Higgins, en la figura 4.11a) muestra curvas suaves dando cuenta que la serie de contaminación en los distintos períodos corresponde a un multifractal. En la figura 4.11b) vemos que las curvas del espectro multifractal son asimétricas, donde el período 2009-2013 es mucho más complejo que los períodos 2004-2008 y 2002-2015 con un ancho multifractal de 0.6197, 0.2841 y 0.4540, respectivamente. Para $q(2)$ se encontraron valores de 0.8518, 0.8006 y 0.7432, entre los períodos 2002-2015, 2004-2008 y 2009-2013, respectivamente. Esto muestra que la serie de tiempo de Parque O'Higgins es persistente y correlacionada en el largo plazo en cada período.

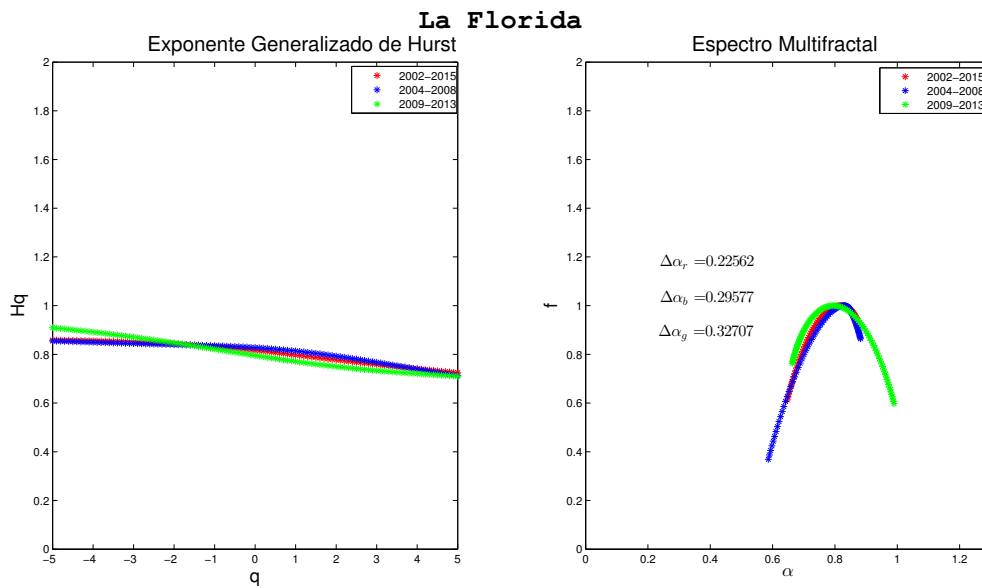


Figura 4.10: **Análisis de La Florida:** a) Exponente de Hurst. En rojo 2002-2015, en azul 2004-2008, en verde 2009-2013. b) Espectros Multifractales correspondientes. Fuente: Elaboración propia.

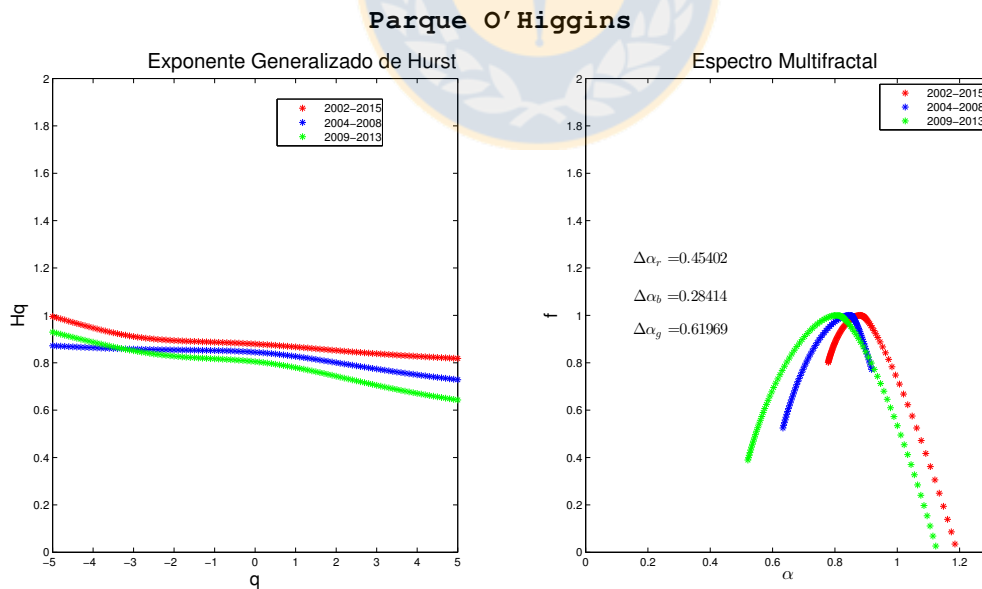


Figura 4.11: **Análisis de Parque O'Higgins:** a) Exponente de Hurst. En rojo 2002-2015, en azul 2004-2008, en verde 2009-2013. b) Espectros Multifractales correspondientes. Fuente: Elaboración propia.

Capítulo 5

Conclusión y Discusión

A partir de los datos horarios estudiados entre los años 2002-2015, se concluye que la concentración de $PM_{2.5}$ ha disminuido en el área metropolitana de Santiago con una fuerte caída a partir de inicios del año 2009 la cual se mantiene estable hasta fines del 2013 en las cuatro estaciones de monitoreo en estudio.

El análisis estadístico muestra que dicha disminución varía entre un 16 % y 23 % para las medias y entre un 15 % y 28 % para las desviaciones estándar entre los años 2009-2013 respecto a los años 2004-2008. Para las estaciones de monitoreo de Las Condes (LC) y Parque O'Higgins (PO), se muestra una disminución respecto al trabajo de Koutrakis et al [1], quienes estimaron que las razones entre la temporada invernal y estival fueron de 1.9 y 2.8 entre los años 1989-2001, las que disminuyeron a 1.36 y 1.89 entre los años 2002-2015.

Además es interesante darse cuenta que de todas las estaciones, Las Condes (LC) es la que más se acerca a la concentración media anual establecida por la legislación chilena http://www.sinia.cl/1292/articles-47699_recurso_2.pdf, y muy especialmente entre el 2009-2013 con un valor de $21.07 \mu\text{g}/\text{m}^3$, superando la norma objetivo en sólo un 1 %.

El análisis mediante ondeletas, para la estación de Pudahuel (PU), la cual registra concentraciones por sobre los $400 \mu\text{g}/\text{m}^3$ en el período invernal, indica que la mayor actividad se presenta en el ciclo diario, siendo secundado por el ciclo de medio día, en el periodo invernal (Abril-Septiembre), bajando notablemente en el periodo estival (Octubre-Marzo). Hacia el ciclo semanal, mensual y otros ciclos mayores, existe una disminución en las amplitudes de concentración de $PM_{2.5}$, tanto en el periodo estival como invernal. Sin embargo, existe un incremento en el ciclo anual, debido a que éste modula la señal. De esta manera la señal presenta tres ciclos importantes, en el ciclo de medio día, diario, y anual. La comparación entre los años 2009-2013 con respecto a los años 2004-2008, muestra que en los ciclos más importantes la amplitud decayó entre un 6 % y 10 %. Por otro lado la varianza en torno a los ciclos más importantes disminuyó entre un 12 % y 50 %.

Finalmente, el análisis multifractal (usando MF DFA) mostró que las series de tiempo originales, en las estaciones de monitoreo de Las Condes (LC), La Florida (LF), Parque O'Higgins (PO) y Pudahuel (PU) son multifractales, persistentes y correlacionadas en el

largo plazo, en el período 2002-2015, 2004-2008 y 2009-2013.

Cabe destacar que la estación de monitoreo de Parque O'Higgins (PO), se observa el mayor ancho del espectro multifractal con un valor de 0.61 entre los años 2009-2013, siendo ésta la señal que muestra mayor complejidad.



Apéndice A

Transformadas de Ondeletas otras comunas

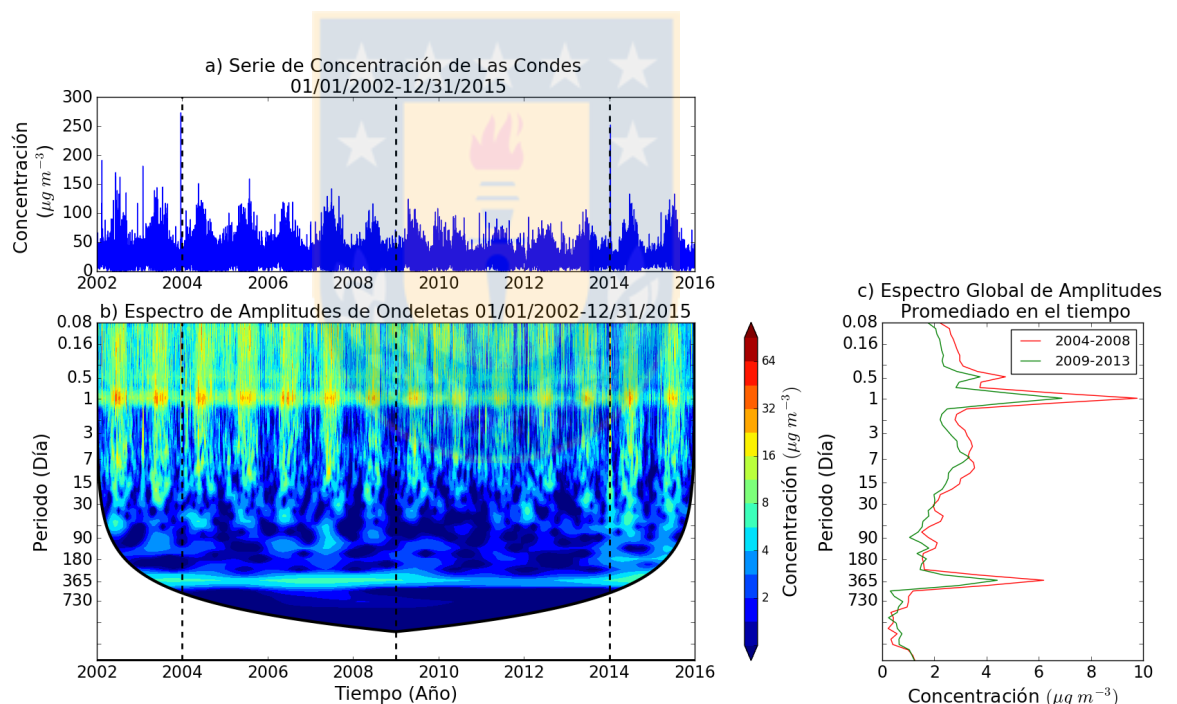


Figura A.1: **Análisis de ondeletas en Las Condes:** a) Serie de tiempo de Las Condes. b) Diagrama de Contorno de la transformada de ondeletas. El eje vertical indica el período medido en días y el eje horizontal el tiempo. El área en blanco denota el exterior del cono de influencia donde no se puede hacer una interpretación de los datos debido a los efectos de borde. c) Amplitud global de ondeletas promediado en el tiempo. Fuente: Elaboración propia.

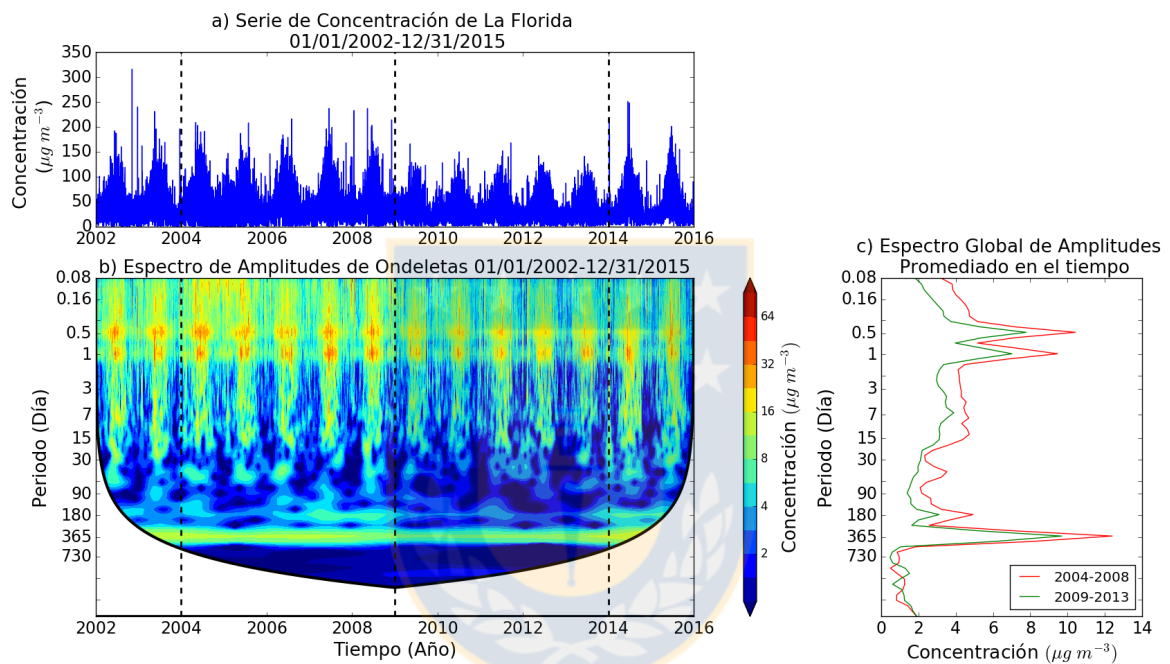


Figura A.2: **Análisis de ondeletas en La Florida:** a) Serie de tiempo de La Florida. b) Diagrama de Contorno de la transformada de ondeletas. El eje vertical indica el período medido en días y el eje horizontal el tiempo. El área en blanco denota el exterior del cono de influencia donde no se puede hacer una interpretación de los datos debido a los efectos de borde. c) Amplitud global de ondeletas promediado en el tiempo. Fuente: Elaboración propia.

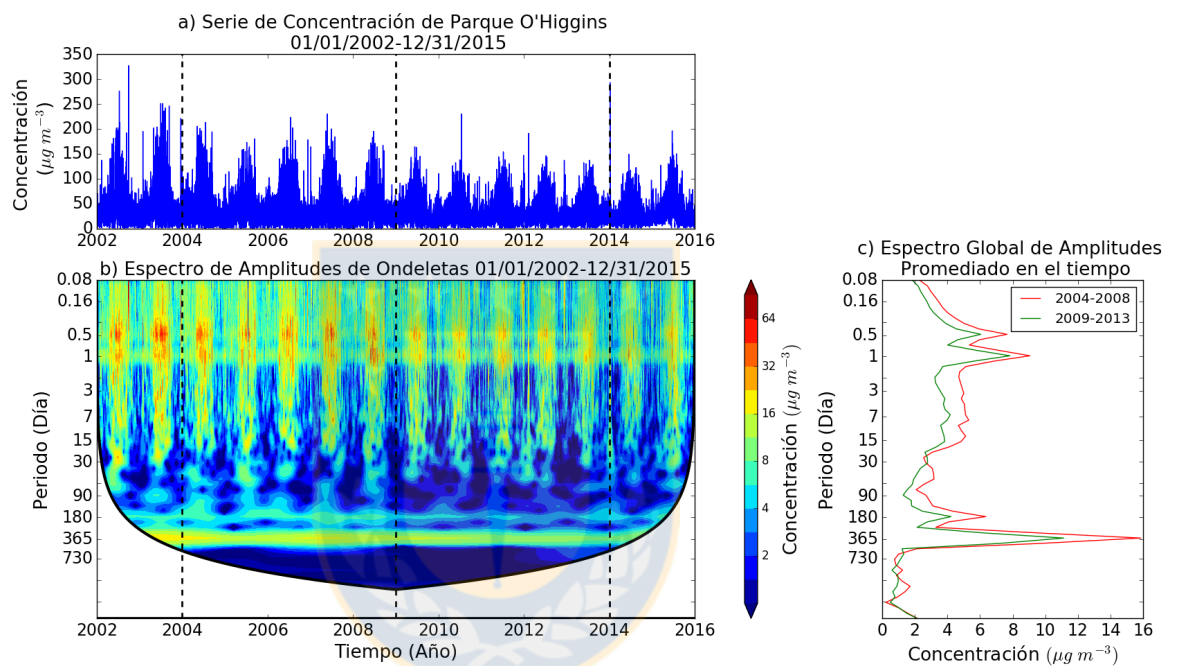


Figura A.3: **Análisis de ondeletas en Parque O'Higgins:** a) Serie de tiempo de Parque O'Higgins. b) Diagrama de Contorno de la transformada de ondeletas. El eje vertical indica el período medido en días y el eje horizontal el tiempo. El área en blanco denota el exterior del cono de influencia donde no se puede hacer una interpretación de los datos debido a los efectos de borde. c) Amplitud global de ondeletas promediado en el tiempo. Fuente: Elaboración propia.

Apéndice B

Trabajos Presentados

A partir de esta tesis se han presentado dos trabajos. Uno de ellos fue la exposición de un póster, en el XX Simposio de la Sociedad Chilena de Física (SOCHIFI), desarrollado entre los días 30 de Noviembre, 01 y 02 de Diciembre del año 2016.

También se realizo un paper, cuyo manuscrito fue enviado en Septiembre del 2016 a la revista Aerosol and Air Quality Research (AAQR).



Estudio de la concentración de $PM_{2,5}$ en Santiago de Chile entre 2009 y 2013

R. Valenzuela-Melivilu^{1,4,*}, R. Abarca-del-Rio^{2,4}, G. Salini³, F. Borotto^{1,4}, H. Astudillo^{1,4}

¹ Departamento de Física, Universidad de Concepción, Concepción, Chile

² Departamento de Geofísica, Universidad de Concepción, Concepción, Chile

³ Departamento de Matemática y Física Aplicadas, Universidad Católica de la Santísima Concepción, Concepción, Chile

⁴ Grupo de Sistemas Complejos, Universidad de Concepción, Concepción, Chile

Resumen

El estudio de las series de tiempo de concentración de $PM_{2,5}$ en 4 estaciones de monitoreo en Santiago de Chile entre los años 2002-2015, mostró una fuerte baja entre los años 2009-2013, siendo esta tan importante que en el caso de la estación de monitoreo de Las Condes, la concentración anual se aproximó por fin a la normativa chilena ($20 \mu g/m^3$ al año), superándola en solo un 1%, lo cual era previsto alcanzar solo para el año 2032 [1]. Diferentes tipos de análisis estadísticos y espectrales [2, 3] llevados a cabo sobre las 4 estaciones de monitoreo, muestran que las series de contaminación son moduladas por el ciclo anual, con una alta actividad en el ciclo de medio día y diario, especialmente durante la temporada invernal y que la energía en estas frecuencias ha cambiado en el transcurso de estos años.

Sin embargo, esta disminución está presente en todas las estaciones de monitoreo, tanto en invierno como en verano, a lo largo de todas las frecuencias. Es así, que la medias y desviaciones estándar disminuyeron entre un 15% y 28%, mientras que la energía disminuyó entre un 36% y 46% respecto a los años 2004-2008.

Esta baja hubiera sido una gran noticia, que

confirmaba la tendencia a la baja observada en trabajos anteriores [4]. Sin embargo, se observa un aumento en la concentración a partir del año 2014 y 2015, volviendo a valores previos al 2009. No obstante, dicho incremento solo se detectó en invierno.

Referencias

- [1] Biblioteca del congreso nacional de Chile. <http://bcn.cl/1v90q>.
- [2] Christopher Torrence and Gilbert P. Compo. A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society*, 79:61–78, 1998.
- [3] YonGGang Liu, X, San Liang, Robert H. Weisberg. Rectification of the Bias in the Wavelet Power Spectrum. *American Meteorological Society*, 24:2093–2102, 2007.
- [4] Jhun I, Oyola P, Moreno F, Castillo MA, Koutrakis P. $PM_{2,5}$ mass and species trends in Santiago, Chile, 1998 to 2010. The impact of fuel-related interventions and fuel sales. *Journal Air and Waste Management*, 63(2):161–169, 2013.

*Email: rodrigovalenzue@udec.cl



**A Note on the significant drop in the 2009-2013 of PM2.5
concentrations in Santiago, Chile**

Journal:	<i>Aerosol and Air Quality Research</i>
Manuscript ID	AAQR-16-09-TN-0413
Manuscript Type:	Technical Note
Date Submitted by the Author:	23-Sep-2016
Complete List of Authors:	Valenzuela-Melivilu, Rodrigo; Universidad de Concepcion, Departamento de Física; Universidad de Concepcion, Grupo de Sistemas Complejos Abarca-del-Rio, Rodrigo; Universidad de Concepcion, Departamento de Geofísica; Universidad de Concepcion, Grupo de Sistemas Complejos Salini, Giovanni; Universidad Catolica de la Santisima Concepcion, Departamento de Matemática y Física Aplicadas; Universidad de Concepcion, Grupo de Sistemas Complejos Borotto , Felix ; Universidad de Concepcion, Departamento de Física; Universidad de Concepcion, Grupo de Sistemas Complejos Astudillo , Hernán; Universidad de Concepcion, Departamento de Física; Universidad de Concepcion, Grupo de Sistemas Complejos
Keywords:	PM2.5, Air Pollution, Mann Kendall Trend Test, Wavelets

SCHOLARONE™
Manuscripts

1 **A Note on the significant drop in the 2009-2013 of PM_{2.5}**
2 **concentrations in Santiago, Chile**

3
4 **Rodrigo Valenzuela-Melivilu^{1,4*}, Rodrigo Abarca-del-Rio^{2,4}, Giovanni A.**
5 **Salini³, Félix A. Borotto^{1,4}, Hernán F. Astudillo^{1,4}.**

6
7 ^{1*} *Departamento de Física, Universidad de Concepción, Concepción, Chile.*

8 ² *Departamento de Geofísica, Universidad de Concepción, Concepción, Chile.*

9 ³ *Departamento de Matemática y Física Aplicadas, Universidad Católica de la Santísima*
10 *Concepción, Concepción, Chile.*

11 ⁴ *Grupo de Sistemas Complejos, Universidad de Concepción, Concepción, Chile.*

12
13 **Abstract**

14
15 The study of concentrations of PM_{2.5} time series in four monitoring stations at Santiago, Chile,
16 between the years 2002-2015 showed a strong reduction between the years 2009-2013. The
17 reduction is so significant that in the case of one of the monitoring stations, Las Condes, the
18 annual concentration finally approached Chilean regulation levels (20 µg/m³ per year), only
19 exceeding them by 1%, which was predicted to occur in the year 2032. Different types of
20 statistical and spectral analysis show that the series of pollution are modulated by the annual
21 cycle, with high activity in the half daily and daily cycles, especially during winter, and that the
22 energy at these frequencies has changed over the course of this period. However, this reduction is
23 present at all the monitoring stations, in winter as well as summer, and is observed at all
24 frequencies. As such, the averages and standard deviations decreased by between 15% and 28%,
25 whilst the energy was reduced by between 36% and 46% with respect to the years 2004-2008.
26 This reduction would have been excellent news, and would have confirmed a trend implied by the
27 reduction already observed in previous studies, as well as verifying the various mitigation
28 policies carried out in the region of Santiago. However, an increase in concentration is observed
29 from the years 2014 and 2015, returning to levels observed prior to 2009, although this increase
30 was only detected in winter.

31
32
33
34
35 **Keywords:** PM_{2.5}; Air Pollution; Mann-Kendall Trend Test; Wavelets.

1*Corresponding Author Email: rodrigovalenzue@udec.cl

Bibliografía

- [1] Koutrakis, P., S.N. Sax, J.A. Sarnat, B. Coull, P. Demokritou, P. Oyola, J. Garcia. Analysis of PM_{10} and $PM_{2.5}$ and $PM_{2.5-10}$ Concentrations in Santiago, Chile, From 1989 to 2001. *Journal Air and Waste Management* , 55:342–351, 2005.
- [2] Richard Toro A., Raúl G.E. Morales S., Mauricio Canales, Claudio Gonzalez-Rojas, and Manuel A. Leiva G. Inhaled and inspired particulates in Metropolitan Santiago Chile exceed air quality standards. *Building and Environment*, 79:115–123, sep 2014.
- [3] Merino M. y Llano A. Ulriksen P. Informe técnico, análisis comparativo de condiciones meteorológicas asociadas a episodios de contaminación atmosférica en Santiago durante los períodos otoño-invierno 1997, 1998, 1999, 2000 y 2001. *Laboratorio de Modelación y Contaminación Atmosférica*, 2001.
- [4] Mauricio Ilabaca, Ignacio Olaeta, Elizabeth Campos, Jeannette Villaire, Martha Maria Tellez-Rojo, and Isabelle Romieu. Association between levels of fine particulate and emergency visits for pneumonia and other respiratory illnesses among children in Santiago, Chile. *Journal of the Air & Waste Management Association*, 49(9):154–163, sep 1999.
- [5] Sabit Cakmak, Robert E. Dales, and Claudia Blanco Vidal. Air pollution and mortality in chile: Susceptibility among the elderly. *Environ Health Perspect*, 115(4):524–527, jan 2007.
- [6] WHO air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulfur dioxide. global update 2005. summary of risk assesment. http://apps.who.int/iris/bitstream/10665/69477/1/WHO_SDE_PHE_OEH_06.02_eng.pdf.
- [7] Sistema de información nacional de calidad del aire. <http://sinca.mma.gob.cl/index.php/region/index/id/M>.
- [8] Giovanni Salini Calderón. Estudio acerca del material particulado emitido en ciudades de tamaño medio al sur de Santiago de Chile. *INGE CUC*, 10(1):97–108, 2014.
- [9] Khaled H. Hamed and A. Ramachandra Rao. A modified Mann-Kendall trend test for autocorrelated data. *Journal of Hydrology*, 204(1-4):182–196, January 1998.

- [10] de Sousa A. Da Silva, D. Detecção de tendências climáticas no estado de alagoas (tend detection of climatic in alagoas state). *Revista Brasileira de Geografia Física*, 6(3), 2013.
- [11] Peter Bloomfield. *Fourier analysis of time series: an introduction; 2nd ed.* Wiley Series in Probability and Statistics. Wiley, New York, NY, 2000.
- [12] Robert H. Shumway and David S. Stoffer. *Time series analysis and its applications : with R examples.* Springer texts in statistics. Springer, New York, 2006.
- [13] Christopher Torrence and Gilbert P. Compo. A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society*, 79:61–78, 1998.
- [14] Yonggang Liu, X. San Liang, and Robert H. Weisberg. Rectification of the Bias in the Wavelet Power Spectrum. *J. Atmos. Oceanic Technol.*, 24(12):2093–2102, December 2007.
- [15] S. Jevrejeva, J. C. Moore, and A. Grinsted. Influence of the arctic oscillation and el niño-southern oscillation (ENSO) on ice conditions in the baltic sea: The wavelet approach. *Journal of Geophysical Research: Atmospheres*, 108(D21), nov 2003.
- [16] Christopher Torrence and Peter J. Webster. Interdecadal Changes in the EN-SO–Monsoon System. *J. Climate*, 12(8):2679–2690, August 1999.
- [17] B. Mandelbrot. How long is the coast of britain? statistical self-similarity and fractional dimension. *Science*, 156(3775):636–638, may 1967.
- [18] Thomas C. Halsey, Mogens H. Jensen, Leo P. Kadanoff, Itamar Procaccia, and Boris I. Shraiman. Fractal measures and their singularities: The characterization of strange sets. *Phys. Rev. A*, 33(2):1141–1151, feb 1986.
- [19] H. Eugene Stanley and Paul Meakin. Multifractal phenomena in physics and chemistry. *Nature*, 335(6189):405–409, sep 1988.
- [20] R. Dominguez-Tenreiro, L. J. Roy, and V. J. Martinez. On the multifractal character of the lorenz attractor. *Progress of Theoretical Physics*, 87(5):1107–1118, may 1992.
- [21] Ashvin B. Chhabra, Charles Meneveau, Roderick V. Jensen, and K. R. Sreenivasan. Direct determination of the $f(\alpha)$ singularity spectrum and its application to fully developed turbulence. *Physical Review A*, 40(9):5284–5294, nov 1989.
- [22] Jan W. Kantelhardt, Stephan A. Zschiegner, Eva Koscielny-Bunde, Shlomo Havlin, Armin Bunde, and H.Eugene Stanley. Multifractal detrended fluctuation analysis of nonstationary time series. *Physica A: Statistical Mechanics and its Applications*, 316(1-4):87–114, dec 2002.
- [23] Satya N. Majumdar and Deepak Dhar. Persistence in a stationary time series. *Physical Review E*, 64(4), sep 2001.

- [24] Chen Zhang, Zhiwei Ni, Liping Ni, Jingming Li, and Longfei Zhou. Asymmetric multifractal detrending moving average analysis in time series of PM2.5 concentration. *Physica A: Statistical Mechanics and its Applications*, 457:322–330, sep 2016.
- [25] Espen A. F. Ihlen. Introduction to multifractal detrended fluctuation analysis in matlab. *Frontiers in Physiology*, 3, 2012.
- [26] Espen A. F. Ihlen. Multifractal analyses of response time series: A comparative study. *Behavior Research Methods*, 45(4):928–945, mar 2013.
- [27] Rodrigo Seguel A., Raúl G.E. Morales S., and Manuel A. Leiva G. Estimations of primary and secondary organic carbon formation in PM2.5 aerosols of Santiago City, Chile. *Atmospheric Environment*, 43(13):2125–2131, apr 2009.
- [28] Met one instruments. http://www.metone.com/docs/bam1020_datasheet.pdf.
- [29] Christopher Torrence and Gilbert P. Compo. A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society*, 79:61–78, 1998.
- [30] Stephane Mallat. *A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way*. Academic Press, 3 edition, 12 2008.
- [31] Jan W. Kantelhardt. Fractal and multifractal time series. In *Mathematics of Complexity and Dynamical Systems*, pages 463–487. Springer Nature, 2012.

