



UNIVERSIDAD DE BUENOS AIRES
Facultad de Ciencias Exactas y Naturales
Departamento de Matemática

Tesis de Licenciatura

Estudio del comportamiento de representaciones de tiempo
frecuencia en presencia de ruido heterocedástico dependiente

Sofía Nieva

Directora: Dra. Ana Georgina Flesia

Fecha de Presentación:

Resumen

La estacionalidad (o periodicidad) y la tendencia son características que describen una secuencia observada, y extraer estas características es un tema importante en muchos campos científicos. Sin embargo, no es una tarea fácil para los métodos existentes analizar simultáneamente la tendencia y la dinámica de la estacionalidad, como frecuencia y amplitud variables en el tiempo, ya que la adaptabilidad a dicha dinámica y la robustez a los errores heterocedásticos dependientes en general no están garantizados. Estas tareas se vuelven aún más desafiantes cuando existen múltiples componentes estacionales. Chen et al. (2013) [3] propone un modelo no paramétrico para describir la dinámica de multicomponentes estacionales, e investiga la transformada Synchrosqueezing (SST) para extraer estas características en presencia de tendencia y errores heterocedásticos dependientes.

Se estudia la identificabilidad del modelo, se lo compara con los métodos EMD y TBATS, desarrollados para problemas similares. También se proporcionan los resultados de una serie de simulaciones y la extracción de modas de una serie temporal de movimiento de codornices japonesas.

Abstract

Seasonality (or periodicity) and trend are features describing an observed sequence, and extracting these features is an important issue in many scientific fields. However, it is not an easy task for existing methods to analyze simultaneously the trend and dynamics of the seasonality such as time-varying frequency and amplitude, since the adaptivity of the analysis to such dynamics and robustness to heteroscedastic, dependent errors is not guaranteed. These tasks become even more challenging when there exist multiple seasonal components. Chen et al. (2013) [3] propose a nonparametric model to describe the dynamics of multi-component seasonality, and investigate the Synchrosqueezing transform (SST) in extracting these features in the presence of a trend and heteroscedastic, dependent errors.

The identifiability problem of the model is studied and the model is compared with the methods EMD and TBATS, which were developed for similar problems. Also, results are provided of a series of simulations and the extraction of modes from a time series of the movement of japanese quails.

Índice general

Introducción	4
1. Preliminares	6
1.1. Análisis de series de tiempo	6
1.2. Componentes y características	7
1.3. Ruido	9
1.4. Algunos procesos	10
1.5. Otras nociones	11
2. Métodos para descomponer señales	13
2.1. BATS/TBATS	14
2.2. EMD	16
3. Teoría de Wavelets	21
3.1. Wavelets	21
3.1.1. Requisitos para una Wavelet	24
3.1.2. Ubicación y Escala	25
3.2. Transformada Wavelet Continua (CWT)	27
3.2.1. Interrogación de una señal	27
3.2.2. Gráfico de la transformada	29
3.2.3. Transformada inversa	32
3.2.4. Transformada Wavelet vs. Transformada de Fourier	34
3.3. Métodos de reasignación	34
4. Método no paramétrico	36
4.1. Transformada synchrosqueezing	36
4.2. Análisis en ausencia de ruido	38
4.3. Análisis en presencia de ruido determinístico y Gaussiano	39
4.4. Análisis en presencia de tendencia y ruido heterocedástico dependiente	40
4.4.1. Identificabilidad	40
4.4.2. Modelo	43
4.4.3. Método	44
4.4.4. Teoremas	45

5. Análisis de simulaciones y datos reales	49
5.1. Datos simulados	49
5.1.1. Configuración de la simulación	49
5.1.2. Elección de parámetros	51
5.1.3. Experimentos	53
5.2. Datos reales	60
Conclusiones	65
Bibliografía	68

Introducción

La estacionalidad (o periodicidad) y la tendencia son características que describen una secuencia observada, y extraer estas características es un tema importante en muchos campos científicos. Sin embargo, no es una tarea fácil para los métodos existentes analizar simultáneamente la tendencia y la dinámica de la estacionalidad, como frecuencia y amplitud variables en el tiempo, ya que la adaptabilidad a dicha dinámica y la robustez a los errores heterocedásticos dependientes en general no están garantizados.

Muchos de los modelos existentes, como TBATS [6], al ser paramétricos, tienen dificultades para capturar el comportamiento dinámico de la estacionalidad y los supuestos que hacen son a menudo demasiado restrictivos para la realidad. Además, si bien puede manejar errores dependientes, no hay ninguna garantía para errores heterocedásticos.

Por otra parte, el método EMD (Empirical Mode Decomposition) [10] es adaptativo y adecuado para series no estacionarias ya que puede descomponer una señal con frecuencia y estacionalidad variables. El problema que presenta es que, al utilizar varias heurísticas en el proceso, es muy difícil de analizar matemáticamente sus garantías de precisión o sus limitaciones de aplicación.

Para solucionar esto, Chen et al. [3] proponen un nuevo modelo: definen una clase de funciones para modelar la dinámica de multicomponentes estacionales y utilizan la transformada Synchrosqueezing (SST) para extraer estas características en presencia de una tendencia y errores heterocedásticos dependientes.

El objetivo de esta tesis es estudiar este modelo no paramétrico y adaptativo para entender sus garantías teóricas, sus alcances prácticos y sus ventajas y desventajas respecto a los otros métodos. Para eso, se analizarán el problema de la identificabilidad del modelo y varios resultados sobre la capacidad de la SST para reconstruir con alta precisión los componentes de la señal, tanto en ausencia de ruido ([5]), como en presencia de distintos tipos de perturbaciones ([16], [3]). Además, se realizarán una serie de simulaciones para comparar el desempeño en la práctica de SST, EMD y TBATS, y finalmente se aplicará SST a un conjunto de datos reales.

El trabajo está organizado de la siguiente manera. En el capítulo 1 definiremos algunos conceptos básicos sobre series de tiempo, sus componentes y características principales y los diferentes tipos de ruidos y procesos que utilizaremos en los capítulos siguientes.

En el capítulo 2 haremos un repaso de los dos métodos mencionados, TBATS y EMD, comúnmente usados para descomponer señales teniendo en cuenta la estacionalidad, y se discutirán algunas de sus ventajas y limitaciones.

En el capítulo 3 haremos una introducción a la teoría de wavelets, definiendo los

conceptos básicos, explicando como funcionan la transformada wavelet continua y su inversa y comparándola con la transformada de Fourier. Por último, introducimos los métodos de reasignación, entre los que se encuentra el método de synchrosqueezing.

En el capítulo 4 definiremos formalmente el método de synchrosqueezing y luego se enunciarán los teoremas presentes en [5], [16] y [3] sobre la capacidad de la SST para descomponer señales en presencia de distintas perturbaciones y la identificabilidad del modelo propuesto en [3].

Finalmente, en el capítulo 5 mostraremos los resultados de una serie de simulaciones comparando el desempeño de la transformada synchrosqueezing, TBATS y EMD en señales con distinto tipo de estacionalidad, ruido y tendencia. Mostraremos y analizaremos también los resultados de aplicar SST a series de tiempo del movimiento de codornices. El código en R de estos experimentos, junto con el código que genera las figuras presentes en este trabajo, está disponible en https://github.com/sofianieva/Tesis_de_Licenciatura.

Capítulo 1

Preliminares

A continuación definiremos algunos conceptos básicos sobre series de tiempo que utilizaremos a lo largo del trabajo.

1.1. Análisis de series de tiempo

Una *serie de tiempo* o *serie temporal* es un conjunto de observaciones x_t , cada una de las cuales se registra en un momento t específico, es decir, están indexadas en orden temporal. Una *serie a tiempo discreto* es aquella en la que el conjunto T de momentos en los que se realizan las observaciones es un conjunto discreto, como es el caso, por ejemplo, cuando las observaciones se realizan cada un intervalo de tiempo fijo. En cambio, una *serie a tiempo continuo* se obtiene cuando las observaciones se registran continuamente durante algún intervalo de tiempo, por ejemplo, cuando $T = [0, 1]$.

En procesamiento de señales, una *señal* es una función que transmite información sobre un fenómeno. En particular, una señal discreta es una serie de tiempo que consta de una secuencia de cantidades. Por eso, en este trabajo, utilizaremos con frecuencia este término como sinónimo de serie temporal, como es común en la literatura.

El *análisis de series de tiempo* comprende métodos para analizar datos de series temporales con el fin de extraer estadísticas significativas y otras características de los datos. El *pronóstico de series de tiempo* es el uso de un modelo para predecir valores futuros basados en valores previamente observados.

Los métodos para el análisis de series de tiempo pueden dividirse en dos clases: métodos en el *dominio de la frecuencia* y métodos en el *dominio del tiempo*. Un gráfico en el dominio temporal muestra la evolución de una señal en el tiempo, mientras que un gráfico en el dominio de la frecuencia muestra las componentes de la señal según la frecuencia en la que oscilan dentro de un rango determinado.

El análisis en el *dominio de tiempo-frecuencia* es el estudio de una señal en los dominios de tiempo y frecuencia simultáneamente. Este se realiza a través de *representaciones de tiempo-frecuencia*. Una señal, en función del tiempo, puede considerarse como una representación con resolución temporal perfecta. Por el contrario, la magnitud de la transformada de Fourier de la señal puede considerarse como una representación

con una resolución espectral perfecta pero sin información de tiempo porque transmite el contenido de frecuencias de la señal pero no muestra en qué momento del tiempo ocurren los diferentes eventos. Así, las representaciones de tiempo-frecuencia pueden considerarse un puente ya que proporcionan cierta información temporal y cierta información espectral simultáneamente. Por lo tanto, son muy útiles para la visualización y el análisis de señales que contienen múltiples frecuencias variables en el tiempo. Una de las formas más básicas de análisis de tiempo-frecuencia es la transformada de Fourier de tiempo corto (STFT), pero se han desarrollado técnicas más sofisticadas, en particular la transformada wavelet continua (CWT) y synchrosqueezing (SST), que estudiaremos en los capítulos 3 y 4.

Otra parte importante del análisis de series de tiempo es la selección de un modelo probabilístico adecuado para los datos. Para tener en cuenta la naturaleza posiblemente impredecible de las observaciones futuras, es natural suponer que cada observación x_t es un valor realizado de una determinada variable aleatoria X_t . Un *modelo* para los datos observados $\{x_t\}$ es una especificación de las distribuciones conjuntas (o posiblemente solo las medias y covarianzas) de una secuencia de variables aleatorias $\{X_t\}$ de las cuales se postula que $\{x_t\}$ es una realización. Con frecuencia usaremos el término serie de tiempo para referirnos tanto a los datos como al *proceso estocástico* (conjunto de variables aleatorias) del cual son una realización.

Las técnicas de análisis de series de tiempo también se pueden dividir en métodos *paramétricos* y *no paramétricos*. Los enfoques paramétricos asumen que el proceso estocástico subyacente tiene una cierta estructura que puede describirse usando una pequeña cantidad de parámetros y la tarea es estimarlos. Por el contrario, los enfoques no paramétricos estiman otras características del proceso sin asumir que tiene una estructura particular.

1.2. Componentes y características

Definimos ahora algunos componentes y características importantes que pueden presentar las series de tiempo.

Sea $\{X_t\}$ una serie de tiempo con $\mathbb{E}(X_t^2) < \infty$. La *función media* de $\{X_t\}$ está dada por $\mu_X(t) = \mathbb{E}(X_t)$ y la *función de covarianza* de $\{X_t\}$ por $\gamma_X(r, s) = \text{Cov}(X_r, X_s) = \mathbb{E}[(X_r - \mu_X(r))(X_s - \mu_X(s))]$, para todos los enteros r y s .

Una serie de tiempo $\{X_t, t \in \mathbb{Z}\}$ se dice estacionaria si tiene propiedades estadísticas similares a la serie temporal “corrida” $\{X_{t+h}, t \in \mathbb{Z}\}$, para cada entero h . Restringiéndonos a las propiedades que solo dependen de los momentos de primer y segundo orden de $\{X_t\}$, se puede definir de la siguiente manera: se dice que $\{X_t\}$ es (*débilmente*) *estacionaria* si $\mu_X(t)$ es independiente de t y $\gamma_X(t+h, t)$ es independiente de t para cada h (o sea, solo depende de h). Por otro lado, $\{X_t\}$ se dice *fuertemente o estrictamente estacionaria* si (X_1, \dots, X_n) y $(X_{1+h}, \dots, X_{n+h})$ tienen la misma distribución conjunta para todos los enteros h y $n > 0$. Notemos que si $\{X_t\}$ es fuertemente estacionaria y $\mathbb{E}(X_t^2) < \infty$ para todo t , entonces $\{X_t\}$ es débilmente estacionaria.

Sea $\{X_t\}$ una serie de tiempo estacionaria, luego la función de covarianza resulta una función de una variable ya que $Cov(X_{t+h}, X_t) = \gamma_X(t+h, t) = \gamma_X(h, 0)$. Definimos la *función de autocovarianza* de $\{X_t\}$ en el *lag* h como $\gamma_X(h) = Cov(X_{t+h}, X_t)$ y la *función de autocorrelación* como $\rho_X(h) = \gamma_X(h)/\gamma_X(0) = Cor(X_{t+h}, X_t)$.

Un ejemplo que generalizaremos en el capítulo 4 es el de *proceso Gaussiano*. $\{X_t\}$ es una serie de tiempo Gaussiana si todas sus distribuciones conjuntas son normales multivariadas, es decir, para cualquier colección finita de índices i_1, \dots, i_n , el vector aleatorio $(X_{i_1}, \dots, X_{i_n})$ tiene una distribución normal multivariada. Resulta entonces que todas sus distribuciones conjuntas están completamente determinadas por la función media y la función de autocovarianza. Si además el proceso es estacionario, la distribución conjunta de (X_1, \dots, X_n) es la misma que la de $(X_{1+h}, \dots, X_{n+h})$ para todo $h \in \mathbb{Z}, n \in \mathbb{N}$. Por lo tanto para una serie de tiempo Gaussiana, estacionalidad fuerte equivale a estacionalidad débil.

En la práctica, una serie de tiempo $\{X_t\}$ de datos reales no suele ser estacionaria. Sin embargo, para poder estudiarla y usar toda la teoría desarrollada para series estacionarias, en general podemos descomponerla en varios elementos determinísticos y que lo que reste sea un proceso estacionario. Matemáticamente, este modelo de descomposición puede representarse como:

$$X_t = f(T_t, S_t, Y_t)$$

donde T_t , la *tendencia*, es una función que varía lentamente, y representa una disminución y/o incremento a largo plazo en los datos. S_t , la *estacionalidad*, es una función periódica, cuyo período puede o no estar fijo y representa los aumentos y disminuciones de los datos a corto plazo, en intervalos más o menos regulares. Y_t , el *residuo o ruido*, es un proceso aleatorio estacionario, que explica la variación en los datos que no se puede explicar con la tendencia y la estacionalidad. La forma exacta de f depende del modelo de descomposición que se utilice. Los más clásicos son:

$$\text{Modelo aditivo: } X_t = T_t + S_t + Y_t$$

$$\text{Modelo multiplicativo: } X_t = T_t * S_t * Y_t$$

En este trabajo utilizaremos principalmente el modelo aditivo para descomponer series.

La estacionalidad puede estar a su vez compuesta de varias *componentes estacionales*. En ese caso, el análisis de la serie también tiene como objetivo descomponer la estacionalidad y estudiar individualmente cada una de estas componentes. Éstas podrían ser funciones puramente armónicas, con amplitud y período constante o podrían ser más complejas, presentando cambios en la frecuencia a través del tiempo (*frecuencia modulada*) o cambios en la amplitud (*amplitud modulada*). Encontrar métodos de descomposición que sean capaces de detectar estos cambios de forma robusta es uno de los principales objetivos de este trabajo.

La *fase* y la *frecuencia instantáneas* son conceptos importantes en el procesamiento de señales que se dan en el contexto de la representación y el análisis de funciones con frecuencia modulada. La fase instantánea (también conocida como fase local o

simplemente fase) de una función a valores complejos $s(t)$, es la función real: $\phi(t) = \arg\{s(t)\}$, donde \arg es la función de argumento complejo. La frecuencia instantánea es la tasa de cambio de fase en el tiempo, es decir $\frac{d}{dt}\phi(t)$. Uno de los métodos que veremos está inspirado en la idea de que podemos extraer la frecuencia instantánea de un componente directamente del plano tiempo-frecuencia de una representación si la imagen es lo suficientemente clara.

Otro concepto que utilizaremos en el capítulo 2 es el de *envolvente* de una señal oscilatoria, que es una curva suave delineando los extremos de la onda. La *envolvente inferior* pasa por los mínimos de la función y la *envolvente superior* por los máximos. Se puede ver un ejemplo en la figura 1.1.

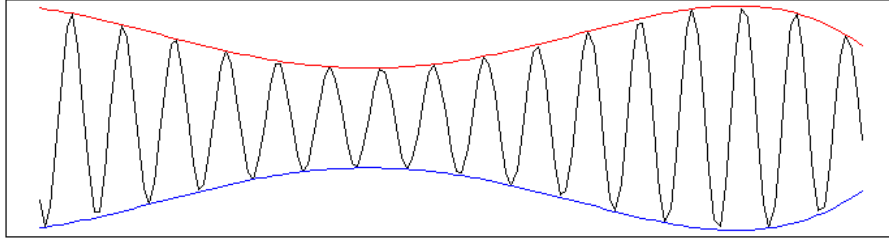


Figura 1.1: En negro, sinusoide con amplitud modulada. En rojo su envolvente superior y en azul su envolvente inferior.

1.3. Ruido

En general, las señales presentan errores aleatorios, variaciones impredecibles o modificaciones no deseadas (y, en general, desconocidas) que puede sufrir durante la captura, almacenamiento, transmisión, procesamiento o conversión. Este tipo de error es lo que en general se conoce como *ruido*. Con los modelos que definimos en la sección anterior, si tenemos una serie que no presenta tendencia o componentes estacionales, podemos considerarla una señal de ruido, generada por un proceso estocástico. Diferentes procesos estocásticos producen señales de ruido con diferentes características.

Por ejemplo, si $\{X_t\}$ es una secuencia de variables aleatorias no correlacionadas, cada una con media cero y varianza σ^2 , entonces claramente $\{X_t\}$ es estacionaria, con función de covarianza

$$\gamma_X(t+h, t) = \begin{cases} \sigma^2, & \text{si } h = 0 \\ 0, & \text{si } h \neq 0 \end{cases}$$

Este tipo de series es lo que se conoce como *ruido blanco* (con media 0 y varianza σ^2) y se indica como $\{X_t\} \sim \text{WN}(0, \sigma^2)$. Un caso particular es el *ruido blanco Gaussiano*, en el que los valores que puede tomar el ruido tienen distribución normal.

El *color del ruido* se refiere a la densidad espectral de potencia de la señal de ruido. La *densidad espectral de potencia* es la transformada de Fourier de la función de autocorrelación. Diferentes colores de ruido tienen propiedades significativamente distintas.

El ruido blanco tiene igual intensidad en diferentes frecuencias. Como el proceso no está correlacionado, su función de autocorrelación es una delta, lo que le da una densidad espectral de potencia constante e igual a σ^2 .

Otras características que pueden presentar las señales de ruido y que hace que su estudio sea más complejo son heterocedasticidad y dependencia. En estadística, un vector de variables aleatorias es *heterocedástico* si la variabilidad de la perturbación aleatoria es diferente entre los elementos del vector. Si cuantificamos la variabilidad mediante la varianza, esto implicaría que la varianza no es contante para todos los elementos de la serie. Por otro lado, decimos que dos variables aleatorias son *dependientes* si su función de distribución conjunta no puede expresarse como el producto de las funciones de distribución marginales, es decir, si no son independientes. En el caso de una señal de ruido $\{X_t\}$, decimos que es *q-dependiente* si X_s y X_t son independientes cuando $|t - s| > q$ y *q-correlacionada* si $\gamma_X(h) = 0$ cuando $|h| > q$. El ruido blanco, por ejemplo, resulta 0-correlacionado.

1.4. Algunos procesos

En el capítulo 5 realizaremos una serie de simulaciones y utilizaremos los siguientes procesos para simular el ruido.

Se dice que $\{X_t, t \in \mathbb{Z}\}$ es un *proceso autorregresivo y de medias móviles* de orden (p, q) (ARMA(p, q)) si $\{X_t\}$ es estacionario y para cada t ,

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q} \quad (1.4.1)$$

donde el proceso de innovaciones $\{Z_t\}$ es WN $(0, \sigma^2)$. Decimos que $\{X_t\}$ es un proceso ARMA (p, q) con media μ si $\{X_t - \mu\}$ es un proceso ARMA (p, q) . Los polinomios de grado p y q :

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \quad \text{y} \quad \theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$$

se denominan polinomios de autorregresión y de medias móviles, respectivamente. Si $\phi(z) \not\equiv 1$, resulta un proceso dependiente, ya que, por definición, depende de p valores pasados.

Se dice que $\{X_t, t \in \mathbb{Z}\}$ es un *proceso de heterocedasticidad condicional autorregresiva generalizada* (GARCH) cuando se asume un modelo ARMA para la varianza del error. En ese caso, en el modelo GARCH(p, q), los términos de error X_t están dados por

$$X_t = \sigma_t Z_t, \quad \text{con} \quad \sigma_t^2 = \omega + \alpha_1 X_{t-1}^2 + \cdots + \alpha_q X_{t-q}^2 + \beta_1 \sigma_{t-1}^2 + \cdots + \beta_p \sigma_{t-p}^2$$

donde p es el orden de los términos de varianza σ^2 o coeficientes GARCH, q es el orden de los coeficientes ARCH, y $\{Z_t\} \sim \text{WN}(0, 1)$.

1.5. Otras nociones

Por último, definimos conceptos que se mencionan en los teoremas del capítulo 4.

La *transformada de Fourier* es una transformación matemática empleada para transformar señales entre el dominio del tiempo y el dominio de la frecuencia. Es reversible, siendo capaz de transformar de cualquiera de los dominios al otro. En el dominio de la frecuencia muestra básicamente el espectro de frecuencias de una función. Formalmente, sea $f \in L^1(\mathbb{R})$ o $f \in L^1(\mathbb{C})$, se define la transformada de Fourier de f como la función:

$$\hat{f}(\xi) := \int_{-\infty}^{\infty} f(x) e^{-2\pi i \xi x} dx$$

El *espacio de Schwartz* o espacio de funciones de rápido decaimiento $\mathcal{S}(\mathbb{R}^n)$ es el conjunto de funciones:

$$\mathcal{S}(\mathbb{R}^n) = \{f \in C^\infty(\mathbb{R}^n) \mid \forall \alpha, \beta : \|f\|_{\alpha, \beta} < \infty\}$$

donde $\alpha \in \mathbb{N}$ y β es un multi-índice (conjunto ordenado de índices: $\beta \in \mathbb{N}^n$, $\beta = (\beta_1, \dots, \beta_n)$), $C^\infty(\mathbb{R}^n)$ es el conjunto de funciones reales suaves sobre \mathbb{R}^n y $\|\cdot\|$ es una norma definida a partir de la norma del supremo como:

$$\|f\|_{\alpha, \beta} := \|x^\alpha D^\beta f\|_\infty = \sup_{x \in \mathbb{R}^n} |x^\alpha (D^\beta f)(x)|, \text{ donde } D^\beta f := \frac{\partial^{|\beta|} f}{\partial x_1^{\beta_1} \dots \partial x_n^{\beta_n}}$$

Para funciones de \mathcal{S} vale que la integral de Fourier converge, su transformada también está en \mathcal{S} , y también funciona la fórmula de inversión.

El espacio $\mathcal{S}'(\mathbb{R}^n)$ se denomina *espacio de distribuciones temperadas* y se define como el dual topológico del espacio de Schwartz. Es decir, una distribución temperada T es un funcional lineal continuo en \mathcal{S} . Esto significa que, dados $\varphi_1, \varphi_2 \in \mathcal{S}$ y φ_n una sucesión en \mathcal{S} , T cumple:

$$T(\alpha_1 \varphi_1 + \alpha_2 \varphi_2) = \alpha_1 T(\varphi_1) + \alpha_2 T(\varphi_2), \quad \forall \alpha_1, \alpha_2 \in \mathbb{R}$$

$$\varphi_n \rightarrow \varphi \Rightarrow T(\varphi_n) \rightarrow T(\varphi)$$

Como el espacio de Schwartz tiene la propiedad de que la transformada de Fourier es un automorfismo, permite, por dualidad, extender la definición de la transformada de Fourier y su inversa a distribuciones temperadas. Entonces, la transformada de Fourier de $T \in \mathcal{S}'$ existe en el sentido de distribución y se define como $\hat{T}(\psi) = T(\hat{\psi})$ para cada función $\psi \in \mathcal{S}$. Por lo tanto, \hat{T} es nuevamente una distribución temperada.

En el capítulo 4 utilizaremos también las nociones de proceso aleatorio generalizado (GRP) y espectro de potencia de un GRP estacionario que se encuentran en [7] y definimos a continuación. Sea K el espacio de funciones C^∞ con soporte compacto de una variable. Decimos que un *funcional aleatorio* Φ se define en K , si con cada elemento $\phi(t) \in K$ hay asociada una variable aleatoria $\Phi(\phi)$. El funcional aleatorio $\Phi(\phi)$ se dice *lineal* si para cualquier elemento $\phi, \psi \in K$ y números cualquiera α y β , se tiene que $\Phi(\alpha\phi + \beta\psi) = \alpha\Phi(\phi) + \beta\Phi(\psi)$. El funcional aleatorio $\Phi(\phi)$ se dice

continuo si la convergencia en K de las funciones $\phi_{k_j}(t)$ a $\phi_j(t)$, $1 \leq j \leq n$, implica $\lim_{k \rightarrow \infty} (\Phi(\phi_{k_1}), \dots, \Phi(\phi_{k_n})) = (\Phi(\phi_1), \dots, \Phi(\phi_n))$. Entonces llamamos *proceso aleatorio generalizado* a los funcionales aleatorios lineales continuos en K . Por otro lado, sean $\phi, \psi \in \mathcal{S}$. Luego existe un funcional B_0 , que es la transformada de Fourier de alguna distribución temperada positiva, tal que:

$$\mathbb{E}[\Phi(\phi)\overline{\Phi(\psi)}] = (B_0, \phi \star \psi^*) \quad (1.5.1)$$

donde \star significa convolución. Además, se tiene

$$\mathbb{E}[\Phi(\phi)\overline{\Phi(\psi)}] = \int \hat{\phi}(\xi)\overline{\hat{\psi}(\xi)}d\eta(\xi) \quad (1.5.2)$$

donde η es la única distribución temperada positiva asociada con Φ tal que $\hat{\eta} = B_0$. En general, llamamos a $d\eta$ el *espectro de potencia* del GRP Φ .

Por último, se menciona el efecto de aliasing. En procesamiento de señales, el *aliasing*, o solapamiento, se refiere a la distorsión que ocurre cuando una señal reconstruida a partir de muestras es diferente de la señal continua original. Esto pasa cuando no se respeta el *teorema de muestreo de Nyquist-Shannon*, que establece que la frecuencia mínima de muestreo para poder reproducir la señal sin error y así evitar el aliasing debe ser al menos el doble de la del componente de frecuencia más alta presente en la señal original. Esta frecuencia mínima de muestreo se denomina *frecuencia de muestreo de Nyquist*. Viéndolo de otro modo, el teorema implica también que una frecuencia de muestreo determinada dictará la frecuencia más alta que un sistema de datos muestreados puede reproducir sin error. Esta frecuencia se denomina *frecuencia de Nyquist* y es igual a la mitad de la frecuencia de muestreo.

Capítulo 2

Métodos para descomponer señales

La detección de los distintos componentes de una señal es un área importante dentro del procesamiento de señales y el análisis de series temporales. Tener técnicas para descomponer señales puede ser muy útil ya sea para entender mejor o responder alguna pregunta sobre el proceso que generó los datos o para predecir valores futuros basados en valores previamente observados.

La estacionalidad y la tendencia son componentes de particular interés. Por ejemplo, entender la estacionalidad de una cierta enfermedad resulta útil para saber cómo y cuándo reforzar el sistema de salud y así aprovechar al máximo los recursos disponibles. Por otro lado, conocer la tendencia también es necesario para determinar si hubo un cambio a lo largo del tiempo independientemente de los altos y bajos correspondientes a la estacionalidad del proceso.

En este tipo de patrones de comportamiento oscilatorio, puede pasar que el período de oscilación varíe con el tiempo y ese cambio también contiene abundante información sobre la dinámica subyacente del proceso en estudio. Por ejemplo, se puede ver que hay una correlación entre la frecuencia variable de una señal respiratoria y las etapas del sueño [3].

Algunos problemas interesantes que surgen comúnmente al analizar señales oscilatorias son entonces los siguientes:

1. Si hay varios componentes oscilatorios dentro de la señal, ¿cómo detectarlos y estimarlos?
2. Si existe una tendencia además de los componentes oscilatorios, ¿cómo extraerla?
3. Si el patrón de los componentes oscilatorios varía en el tiempo, ¿cómo cuantificarlo/identificarlo?
4. Dado que la longitud de los datos observados se alarga con el tiempo, ¿qué tan sensible es el estimador a la longitud de la serie de tiempo observada?
5. Si los errores en diferentes puntos del tiempo son dependientes, o si la varianza del error cambia según el tiempo, ¿es el estimador robusto ante tales errores heterocedásticos dependientes?

En este capítulo haremos un breve repaso de métodos comúnmente usados para descomponer señales que tienen en cuenta la estacionalidad y discutiremos algunas de sus ventajas y limitaciones viendo cuáles de las preguntas anteriores son capaces de responder.

2.1. BATS/TBATS

Muchas series de tiempo exhiben patrones estacionales complejos. Por ejemplo, algunas tienen múltiples componentes estacionales de alta frecuencia o con períodos no enteros. También puede ocurrir que se presenten componentes estacionales no anidados, es decir, defasados y que no son unos múltiplos de otros. Los métodos BATS y TBATS se proponen en [6] como técnicas versátiles para la predicción y descomposición de series de tiempo con este tipo de estacionalidad compleja. En particular TBATS se basa en representaciones trigonométricas para manejar tanto patrones estacionales simples como complejos de una manera sencilla con menos parámetros que otros métodos.

Para formular estos modelos y que funcionen en una variedad amplia de patrones estacionales, para el método BATS, se incorpora una transformada Box-Cox (B), errores ARMA (A), tendencia (T) y K componentes estacionales (S), con lo cual queda:

$$\begin{aligned}
 y_t^{(\omega)} &= \begin{cases} \frac{y_t^\omega - 1}{\omega} & \omega \neq 0 \\ \log y_t & \omega = 0 \end{cases} \\
 y_t^{(\omega)} &= \ell_{t-1} + \phi b_{t-1} + \sum_{i=1}^K s_{t-m_i}^{(i)} + d_t \\
 \ell_t &= \ell_{t-1} + \phi b_{t-1} + \alpha d_t \\
 b_t &= (1 - \phi)b + \phi b_{t-1} + \beta d_t \\
 s_t^{(i)} &= s_{t-m_i}^{(i)} + \gamma_i d_t \\
 d_t &= \sum_{i=1}^p \varphi_i d_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \varepsilon_t
 \end{aligned} \tag{2.1.1}$$

donde y_t es la observación a tiempo t (la notación $y_t^{(\omega)}$ representa las observaciones transformadas con Box-Cox con el parámetro ω), m_1, \dots, m_K denotan los períodos constantes de las componentes estacionales, ℓ_t es el nivel local en el período t , b es la tendencia a largo plazo, b_t es la tendencia a corto plazo en el período t , $s_t^{(i)}$ representa el i -ésimo componente estacional en el momento t , d_t denota un proceso ARMA(p, q) y ε_t es un proceso de ruido blanco Gaussiano con media cero y varianza constante σ^2 . Los parámetros de suavizado están dados por α, β y γ_i para $i = 1, \dots, K$. Se adopta una tendencia amortiguada, con parámetro de amortiguación ϕ , pero se complementa con una tendencia a largo plazo b . Esto asegura que las predicciones de los valores futuros de la tendencia a corto plazo b_t converjan a la tendencia a largo plazo b en lugar de cero.

El nombre BATS, como vimos, es un acrónimo de las principales características del modelo. Sus argumentos son entonces $(\omega, \phi, p, q, m_1, m_2, \dots, m_K)$, que indican el

parámetro de Box-Cox, el parámetro de amortiguación, los parámetros de proceso AR-MA (p y q) y los períodos de las componentes estacionales (m_1, \dots, m_K).

Este modelo, si bien es útil para señales con más de un componente estacional, no puede adaptarse a estacionalidad no entera o anidada y puede requerir la estimación de una gran cantidad de parámetros iniciales, que a menudo conduce a componentes estacionales con ruido. Por eso, para tener un poco más de flexibilidad y una descomposición más suave, en [6] se introduce la siguiente representación trigonométrica de los componentes estacionales, basada en series de Fourier:

$$s_t^{(i)} = \sum_{j=1}^{k_i} s_{j,t}^{(i)}$$

$$s_{j,t}^{(i)} = s_{j,t-1}^{(i)} \cos \lambda_j^{(i)} + s_{j,t-1}^{*(i)} \sin \lambda_j^{(i)} + \gamma_1^{(i)} d_t \quad (2.1.2)$$

$$s_{j,t}^{*(i)} = -s_{j,t-1}^{(i)} \sin \lambda_j^{(i)} + s_{j,t-1}^{*(i)} \cos \lambda_j^{(i)} + \gamma_2^{(i)} d_t$$

donde $\gamma_1^{(i)}$ y $\gamma_2^{(i)}$ son parámetros de suavizado y $\lambda_j^{(i)} = 2\pi j/m_i$ (notar que $\lambda_j^{(i)}$ está fijo). Se describe el nivel estocástico del i -ésimo componente estacional con $s_{j,t}^{(i)}$, y al crecimiento estocástico del nivel del i -ésimo componente estacional, necesario para describir el cambio en el componente estacional a lo largo del tiempo, con $s_{j,t}^{*(i)}$. El número de armónicos requeridos para el i -ésimo componente estacional se indica con k_i . Se puede obtener una representación determinística de los componentes estacionales estableciendo los parámetros de suavizado en cero, ya que si no, $s_t^{(i)}$ y d_t están ambos acoplados con una fuente única de error d_t .

Luego, el método TBATS se obtiene reemplazando el componente estacional $s_t^{(i)}$ en las ecuaciones de BATS (2.1.1) por la formulación estacional trigonométrica, y la segunda ecuación por $y_t^{(\omega)} = \ell_{t-1} + \phi b_{t-1} + \sum_{i=1}^T s_{t-1}^{(i)} + d_t$. La T inicial proviene de “trigonométrica” y para proporcionar más detalles sobre su estructura, se deben dar los argumentos $(\omega, \phi, p, q, \{m_1, k_1\}, \{m_2, k_2\}, \dots, \{m_K, k_K\})$.

En general, un modelo TBATS requiere un número menor de estimaciones de valores estacionales iniciales que un modelo BATS. Además, debido a que se basa en funciones trigonométricas, se puede utilizar para modelar frecuencias estacionales no enteras y la suavidad de las estimaciones se puede controlar con el número de armónicos utilizados. Otras ventajas del TBATS son que permite la acomodación de múltiples componentes estacionales anidados y no anidados y también tiene en cuenta cualquier posible autocorrelación de los residuos. Sin embargo, existen algunas limitaciones. En particular, volviendo a las preguntas que planteamos en la sección anterior, tanto BATS como TBATS no logran responder completamente 3, 4 y 5. En primer lugar, aunque TBATS permite componentes estacionales estocásticos, los períodos estacionales en los modelos son todos fijos. Además, dado que la estimación de los parámetros depende de toda la serie de tiempo observada, la longitud de la serie juega un papel en el resultado, por lo que no es fácil responder a la pregunta 4. En cuanto a 5, aunque se pueden manejar errores dependientes, no está garantizado para errores heterocedásticos.

2.2. EMD

El método de descomposición empírica en modos (EMD) se propuso en [10] como un algoritmo que permite el análisis de tiempo-frecuencia de señales multicomponente incluso cuando no son estacionarias. Dada una señal $f(t)$, el método la descompone en varias funciones de moda intrínseca (IMF):

$$f(t) = \sum_{k=1}^K f_k(t) \quad (2.2.1)$$

donde cada función de moda intrínseca f_k es básicamente una función que oscila alrededor de 0, aunque no necesariamente con frecuencia o amplitud constante:

$$f_k(t) = A_k(t) \cos(\phi_k(t)), \text{ con } A_k(t), \phi'_k(t) > 0 \forall t \quad (2.2.2)$$

Típicamente, el cambio en el tiempo de $A_k(t)$ y $\phi'_k(t)$ es mucho más lento que el de $\phi_k(t)$, lo que significa que localmente (es decir, en un intervalo de tiempo $[t - \delta, t + \delta]$, con $\delta \approx 2\pi [\phi'_k(t)]^{-1}$) el componente $f_k(t)$ se puede considerar como una señal armónica con amplitud $A_k(t)$ y frecuencia $\phi'_k(t)$.

En [10], las condiciones que debe satisfacer una función para ser de moda intrínseca se expresan de la siguiente manera: (1) en todo el conjunto de datos, el número de extremos locales y el número de ceros de $f_k(t)$ deben ser iguales o diferir como máximo en uno; y (2) en cualquier t , el valor de la envolvente definida por los mínimos locales de $f_k(t)$ es el negativo de la envolvente correspondiente definida por los máximos locales. Es decir, son funciones que son simétricas con respecto a la media local, y pasan por cero entre dos extremos consecutivos.

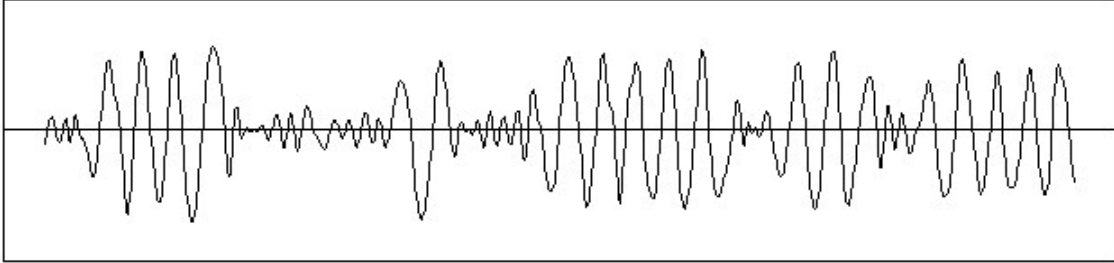


Figura 2.1: Ejemplo de una función de moda intrínseca

Después de la descomposición de $f(t)$ en sus componentes de moda intrínseca, el algoritmo EMD procede al cálculo de la “frecuencia instantánea” de cada componente. Teóricamente, como vimos en el capítulo 1, ésta viene dada por $\omega_k(t) := \phi'_k(t)$.

Si bien el algoritmo EMD puede descomponer una señal cuando tiene varios componentes estacionales, incluso si éstos varían en el tiempo, también contiene una serie de elementos heurísticos que dificultan el análisis matemático de la precisión de sus resultados. Por ejemplo, para construir una descomposición de la forma 2.2.1 el algoritmo EMD utiliza un proceso de *sifting* que consta de los siguientes pasos:

- (I) Inicializar el residuo $r_0(t) = f(t)$, y asignar $g_0(t) = r_{k-1}(t)$, $i = 1$ y $k = 1$.
- (II) Construir la envolvente inferior $Imin_{i-1}$, definida por los mínimos locales y la envolvente superior $Imax_{i-1}$, definida por los máximos locales de la señal, interpolando mediante el método de splines cúbicos.
- (III) Calcular la curva media de la señal promediando la envolvente superior y la inferior. Asignar $m_{i-1} = [Imax_{i-1} + Imin_{i-1}]/2$
- (IV) Restar la media a la señal original $g_i = g_{i-1} - m_{i-1}$, actualizar $i = i + 1$ y repetir los pasos II-IV hasta que g_i sea una función de moda intrínseca. Cuando esto ocurre, el k -ésimo componente de moda intrínseca está dado por $IMF_k = g_i$
- (v) Actualizar el residuo $r_k(t) = r_{k-1}(t) - IMF_k$. Este componente residual se trata como un dato nuevo y se somete al proceso descrito anteriormente para calcular el siguiente IMF_{k+1} .
- (vi) Repetir los pasos anteriores ya sea por un número fijo de veces o hasta que se satisfaga un cierto criterio de parada, por ejemplo, hasta que el componente residual final $r(t)$ se convierta en una función monótona.

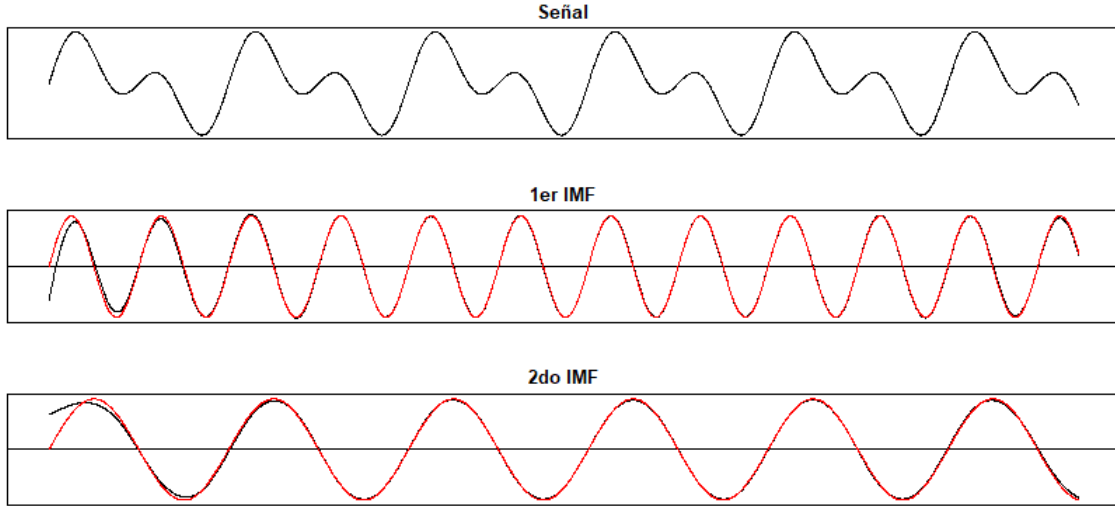


Figura 2.2: Ejemplo de descomposición de una señal sin ruido suma de dos sinusoides $f(t) = \sin(4t) + \sin(8t)$. En rojo el verdadero valor de los componentes sinusoidales y en negro la estimación obtenida con EMD para los dos primeros componentes de moda intrínseca

Debido a que el proceso de sifting se basa en gran medida en interpolaciones de máximos y mínimos, el resultado final tiene algunos problemas de estabilidad en presencia de ruido (ver figuras 2.3 y 2.4). Otros problemas que se suelen presentar son oscilaciones de amplitud muy dispar en un modo, o la presencia de oscilaciones muy similares en diferentes modos, denominado “mezcla de modos” (figura 2.4).

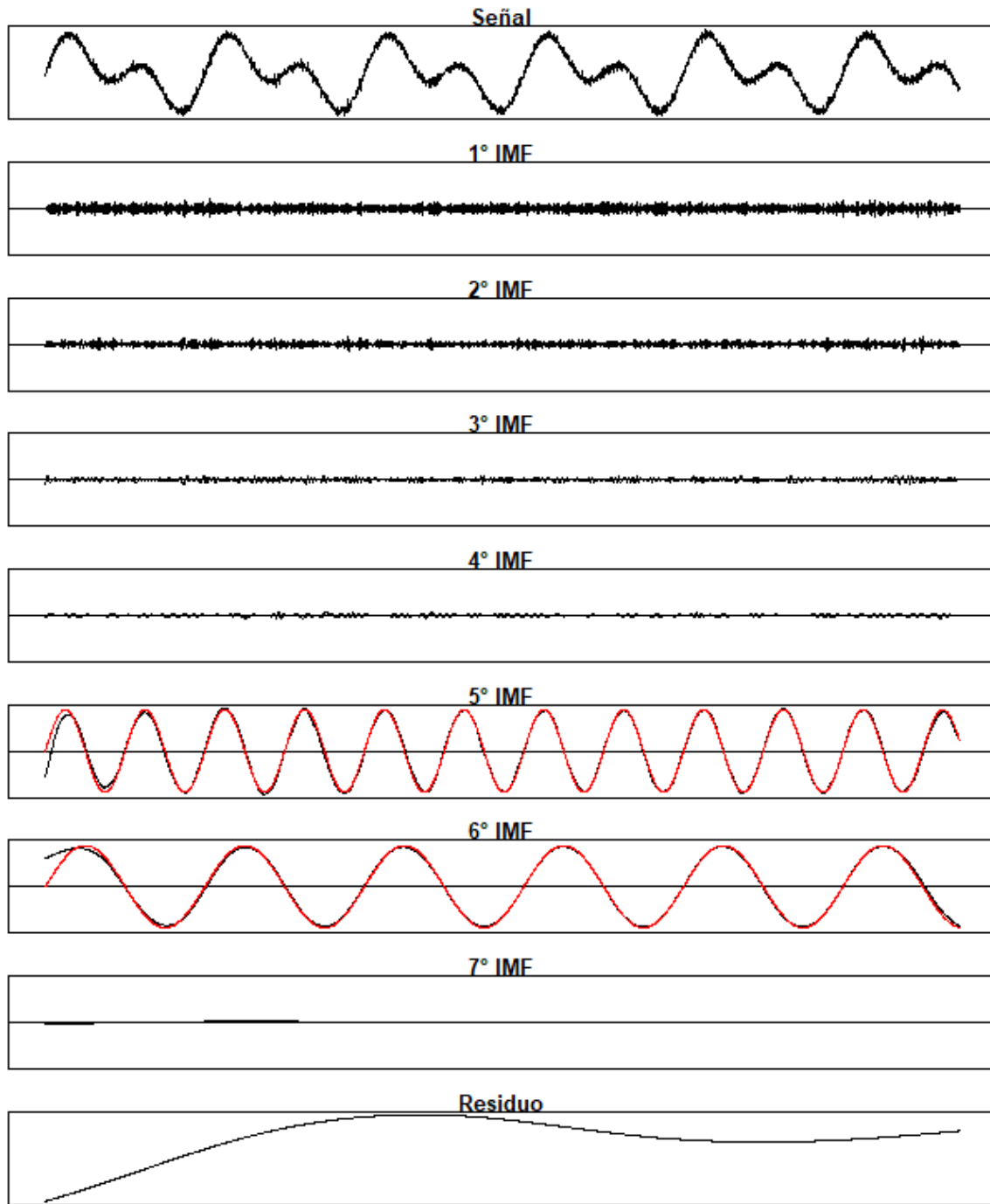


Figura 2.3: Ejemplo de descomposición de la misma señal $f(t) = \sin(4t) + \sin(8t)$ con ruido blanco, resultando la señal $g(t) = f(t) + 0,1 * WN$, con $WN = ARMA(0,0)$. Se puede ver que el algoritmo de EMD logra encontrar con bastante éxito los componentes sinusoidales como 5° y 6° IMF, mientras que los primeros 4 corresponden al ruido. El residuo está en otra escala para mostrar que si bien es muy pequeño, no es 0.

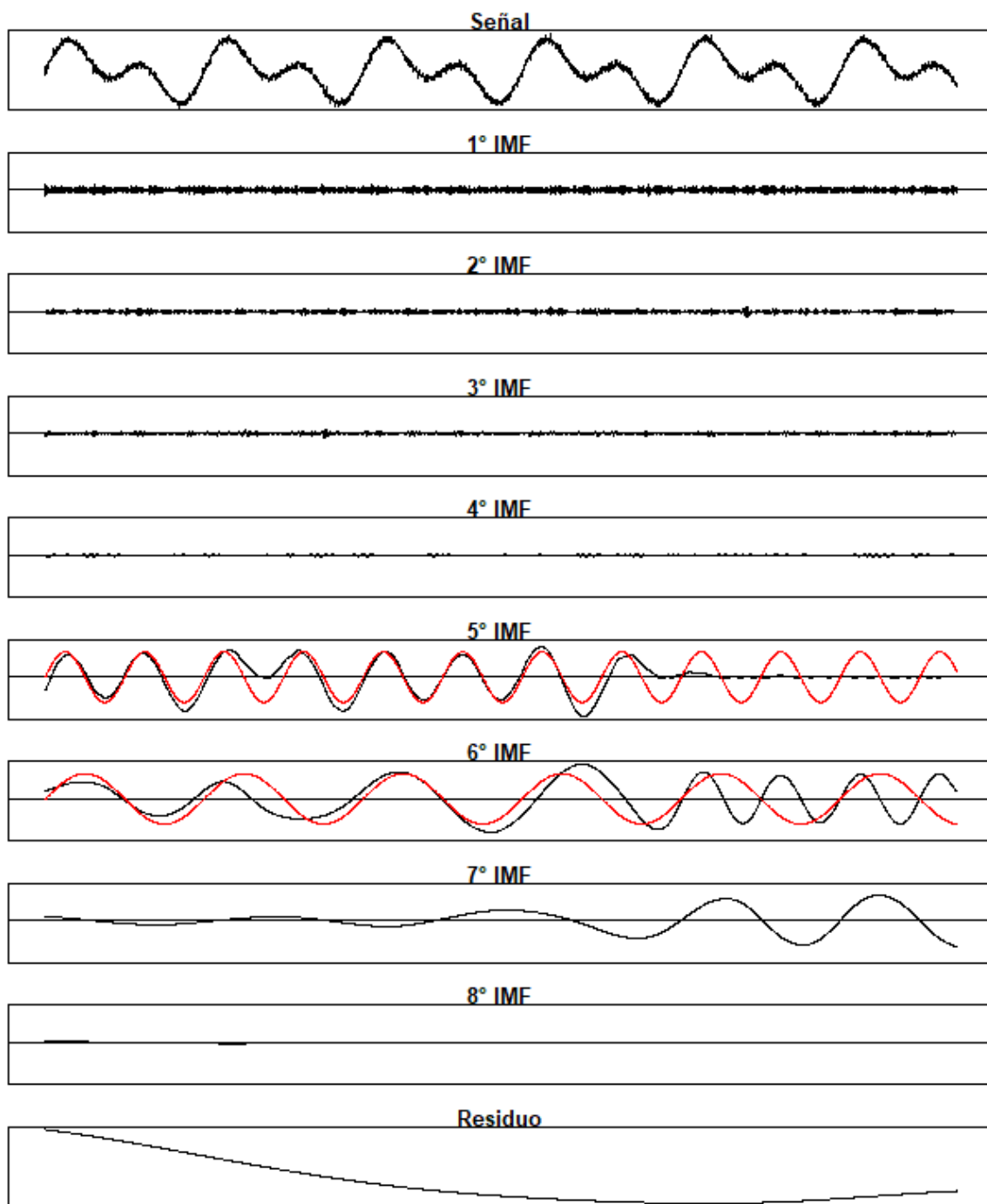


Figura 2.4: Ejemplo de descomposición de la misma señal que en 2.3 pero con otra realización del ruido blanco. Se puede observar que ahora el algoritmo de EMD encuentra 8 IMF's en lugar de 7 y además están mezclados: la última parte del 6° IMF debería ser parte del 5° para que corresponda a el componente oscilatorio $\sin(8t)$ y $\sin(4t)$ se aproxima al comienzo del 6° y el final del 7°. Esto muestra la inestabilidad del método frente al ruido.

En [17], para abordar estos problemas en la práctica se propone otro algoritmo, basado en EMD, conocido como ensamble de descomposición empírica de modos (EEMD). Este método define los componentes de moda intrínseca “verdaderos” como la media de los correspondientes obtenidos a través de EMD sobre un conjunto de señales, generadas al agregar diferentes realizaciones de ruido blanco Gaussiano de varianza finita a la señal original $f(t)$. La adición de este ruido resuelve el problema de la mezcla de modos; sin embargo, crea algunos nuevos. De hecho, la señal reconstruida incluye ruido residual y diferentes realizaciones de señal más ruido pueden producir diferentes números de modos (como se puede ver en las figuras 2.3 y 2.4). Además, plantea nuevos desafíos para la comprensión teórica matemática.

En resumen, el método de descomposición empírica en modos es útil para analizar señales no estacionarias con múltiples componentes estacionales que varían en el tiempo. Tiene varias ventajas: es adaptativo ya que la descomposición que produce sólo depende de los datos (la señal), la persona que lo utiliza no debe elegir de antemano ningún parámetro, como ocurre con otros métodos. Es local, con lo cual no se ve afectado si la longitud de los datos aumenta. Además, hace un uso eficaz de los datos porque en el proceso de sifting, la escala más larga se define por la longitud completa de los datos, con lo cual, EMD puede encontrar oscilaciones de períodos más largos que otros métodos que requieren una cierta cantidad de ciclos dentro del período de observación para identificar una frecuencia. En cuanto a la presencia de tendencia en una señal, en [13] se propone considerar el residuo final $r(t)$ como un estimador de la misma y se mencionan otros estimadores basados en EMD para distintos tipos de señales. Luego, si bien el método de EMD puede responder satisfactoriamente las primeras cuatro preguntas planteadas al comienzo de este capítulo, su punto débil está en la quinta, ya que como vimos, con ruido blanco ya hay mucha variabilidad en los resultados, con lo cual tampoco va a ser robusto ante errores más complejos, como errores heterocedásticos o dependientes. Además de ser inestable en la práctica, la interpolación en el proceso de sifting también hace muy difícil obtener resultados teóricos que garanticen robustez frente a algún tipo de error o alguna garantía de precisión en las estimaciones. Tener una técnica que permita responder adecuadamente a todas las preguntas 1-5, es lo que motivó el desarrollo de los métodos que veremos en el capítulo 4.

Capítulo 3

Teoría de Wavelets

En este capítulo hacemos una introducción a la teoría de wavelets, definiendo los conceptos básicos, explicando como funcionan la transformada wavelet continua y su inversa y comparándola con la transformada de Fourier. Por último, introducimos los métodos de reasignación, entre los que se encuentra el método de synchrosqueezing. Nos basamos principalmente en el primer capítulo de [1].

3.1. Wavelets

Las wavelets son funciones con forma de onda localizada y se utilizan para transformar una señal del dominio del tiempo a una representación de tiempo-frecuencia que presente la información de forma más útil. Esta transformación de la señal se conoce como transformada wavelet. Matemáticamente hablando, la transformada wavelet puede interpretarse como una convolución de la señal con una función wavelet.

Hay una gran variedad de wavelets para usar en el análisis de nuestros datos. La mejor para una aplicación particular depende tanto de la naturaleza de la señal como de lo que queramos obtener del análisis (es decir, qué fenómenos físicos o procesos estamos buscando interrogar, o cómo estamos tratando de manipular la señal). La siguiente figura muestra algunos ejemplos de wavelets que se usan comúnmente en la práctica.

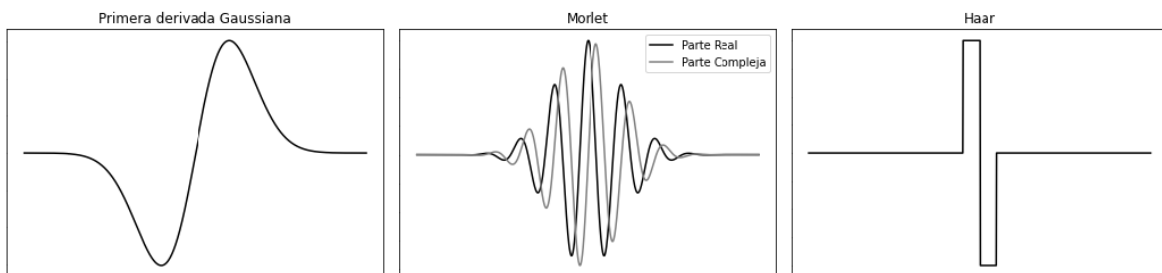


Figura 3.1: Ejemplos de wavelets

Una wavelet se puede manipular de dos maneras: se puede mover a varios lugares de la señal (figura 3.2) y se puede estirar o contraer (figura 3.3).

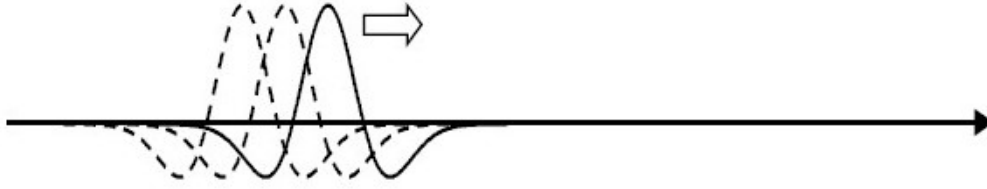


Figura 3.2: Distintas ubicaciones de una wavelet

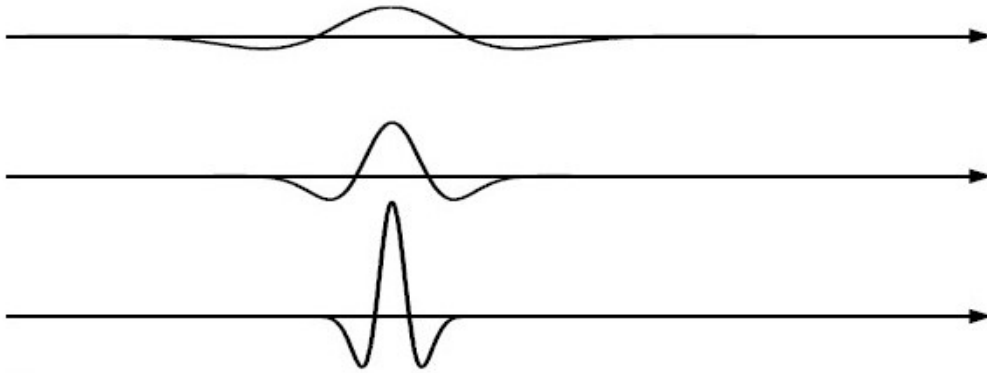


Figura 3.3: Distintas escalas de una wavelet

La figura 3.4 muestra un esquema de la transformada wavelet que básicamente cuantifica la correspondencia local de la wavelet con la señal. Si la wavelet coincide bien con la forma de la señal en una escala y ubicación específicas, como sucede en el gráfico superior de la Figura 3.4, se obtiene un valor de transformación grande. Sin embargo, si la wavelet y la señal no se correlacionan bien, se obtiene un valor de transformación bajo. El valor de la transformación se ubica entonces en el plano de transformación bidimensional que se muestra en la parte inferior de la Figura 3.4 (indicado por el punto negro). La transformada se calcula en varias ubicaciones de la señal y para varias escalas de la wavelet, llenando así el plano de transformación: esto se hace de forma suave y continua para la transformada wavelet continua (CWT) o en pasos discretos para la transformada wavelet discreta. (DWT). Así es como el gráfico de la transformada wavelet permite construir una imagen de la correlación entre la wavelet, en varias escalas y ubicaciones, y la señal.

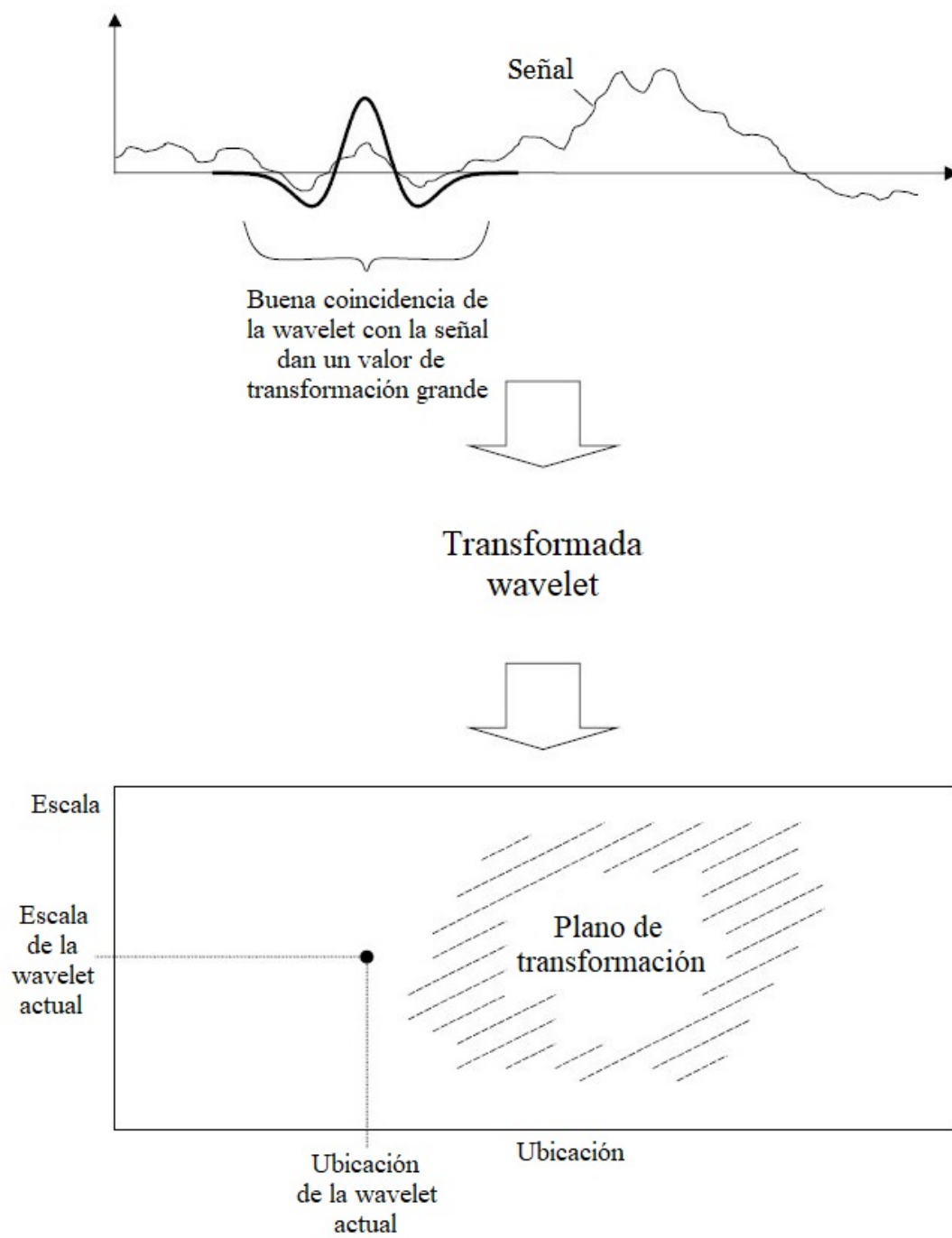


Figura 3.4: Esquema de la transformada wavelet

3.1.1. Requisitos para una Wavelet

Para ser clasificada como wavelet, una función debe satisfacer ciertas condiciones:

- Una wavelet debe tener energía finita, es decir:

$$E = \int_{-\infty}^{\infty} |\psi(t)|^2 dt < \infty \quad (3.1.1)$$

donde E es la energía de una función que equivale a la integral de su módulo al cuadrado.

- Si $\hat{\psi}(f)$ es la transformada de Fourier de $\psi(t)$, es decir:

$$\hat{\psi}(f) = \int_{-\infty}^{\infty} \psi(t) e^{-i(2\pi f)t} dt \quad (3.1.2)$$

entonces se debe cumplir la siguiente condición:

$$C_g = \int_0^{\infty} \frac{|\hat{\psi}(f)|^2}{f} df < \infty \quad (3.1.3)$$

Esto implica que la wavelet no tiene componente de frecuencia cero, $\hat{\psi}(0) = 0$, o dicho de otra forma, la wavelet $\psi(t)$ debe tener media cero. La ecuación 3.1.3 se conoce como condición de admisibilidad y C_g se denomina constante de admisibilidad. El valor de esta constante depende de qué wavelet se esté utilizando.

- Un criterio adicional que deben cumplir las wavelets complejas es que su transformada de Fourier debe ser real y decaer para las frecuencias negativas.

Ejemplo

Mexican hat o Sombrero Mejicano se le denomina a la wavelet dada por la fórmula

$$\psi(t) = (1 - t^2)e^{-t^2/2} \quad (3.1.4)$$

En realidad, la wavelet de la ecuación 3.1.4 es lo que se conoce como una *wavelet madre*, es decir, es la forma básica de la wavelet de la que luego se derivan las versiones dilatadas y trasladadas que se utilizan en la transformación.

Para el Sombrero Mejicano se tiene entonces que:

$$E = \int_{-\infty}^{\infty} |\psi(t)|^2 dt = \frac{3}{4}\pi \quad \text{y} \quad C_g = \pi \quad (3.1.5)$$

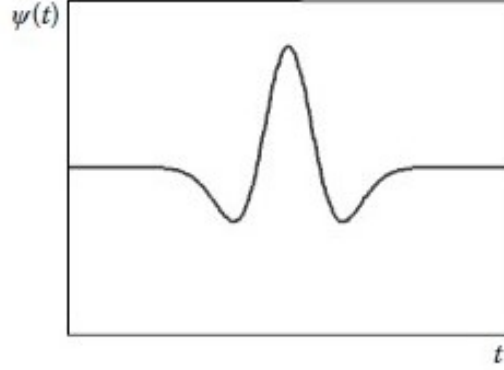


Figura 3.5: Wavelet Sombrero Mejicano

3.1.2. Ubicación y Escala

Una vez que elegimos la wavelet madre con la que vamos a trabajar, para darle la flexibilidad necesaria para ser una herramienta de análisis útil, como vimos, se le pueden aplicar dos operaciones básicas: dilatación/contracción y traslación. La dilatación de la wavelet está controlada por el parámetro de dilatación o escala, a , mientras que el movimiento de la wavelet a través del eje x es controlado por el parámetro de traslación o ubicación, b .

Se pueden entonces incluir los parámetros a y b en la definición de una wavelet. Así, las versiones dilatadas y trasladadas de una wavelet madre $\psi(t)$ se denotan $\psi((t-b)/a)$. Por ejemplo, para el Sombrero Mejicano tendríamos:

$$\psi\left(\frac{t-b}{a}\right) = \left(1 - \left(\frac{t-b}{a}\right)^2\right) e^{-\frac{1}{2}[(t-b)/a]^2} \quad (3.1.6)$$

La wavelet madre original $\psi(t)$, dada por la ecuación 3.1.4, simplemente tenía $a = 1$ y $b = 0$.

La figura 3.6 muestra la wavelet Sombrero Mexicano estirada y comprimida, respectivamente, al doble y a la mitad de su ancho original y su movimiento en el eje x de b_1 , pasando por b_2 , a b_3 .

Si se quisiera, sobre todo para aumentar la interpretabilidad de los resultados y medirlos en hz, se puede hacer una “traducción” de escalas a frecuencias. En esencia, la escala de una wavelet es inversamente proporcional a todas sus frecuencias características. Sin embargo, la relación depende de cada tipo de wavelet y solo se puede conocer de forma aproximada. La manera de calcular, dada una wavelet, a qué pseudo-frecuencia f_a corresponde una escala a es determinando la frecuencia central de la wavelet madre f_c y usar que $f_a = f_c/a$. La idea para calcular f_c es asociar la wavelet con una señal puramente periódica de frecuencia f_c , por lo tanto, se toma la frecuencia que maximiza la transformada de Fourier del módulo de la wavelet como f_c . De la relación anterior, se puede ver que la escala es inversamente proporcional a la pseudo-frecuencia.

Por ejemplo, si la escala aumenta, la wavelet se extiende más, lo que resulta en una pseudo-frecuencia más baja.

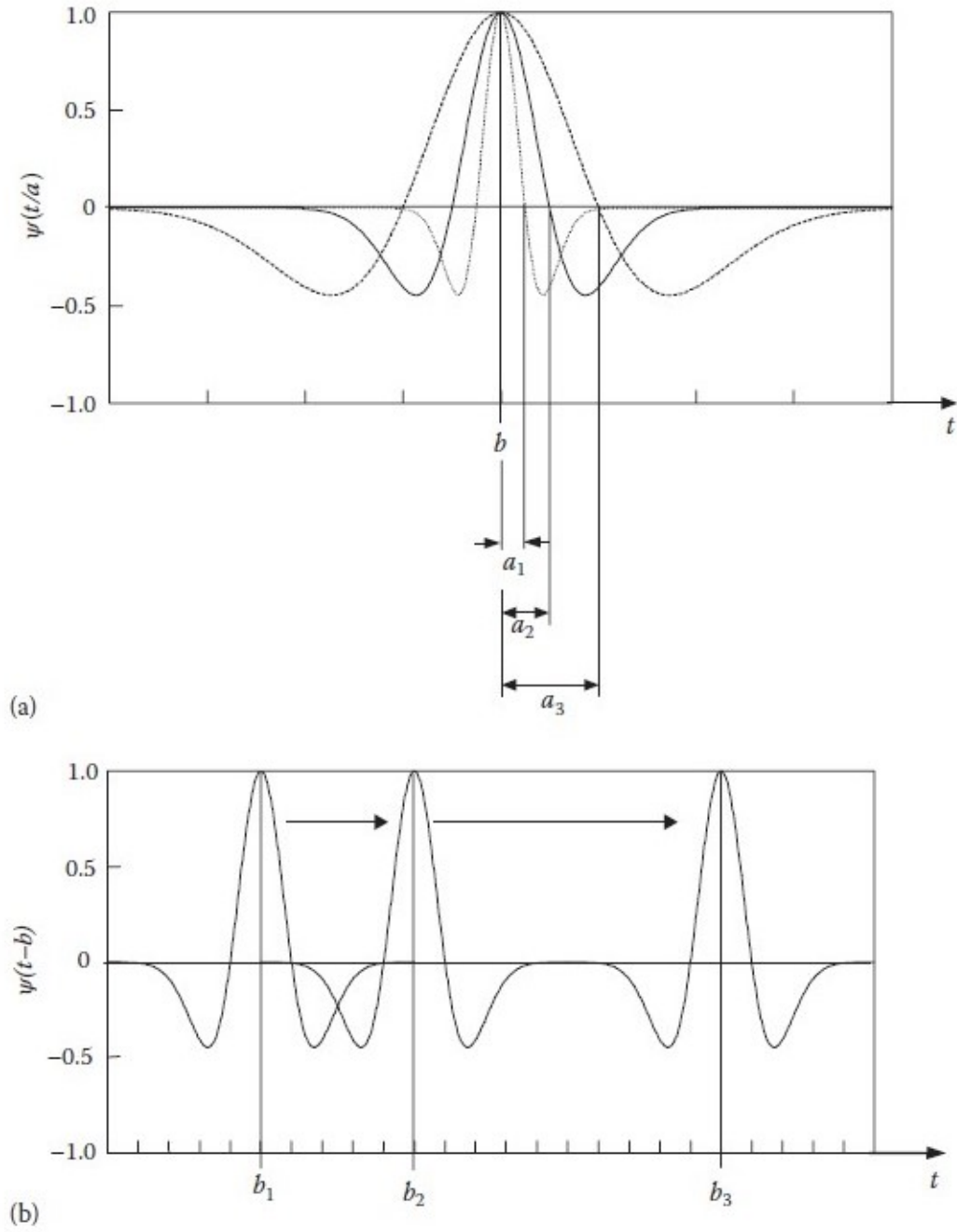


Figura 3.6: Dilatación y traslación de una wavelet: **a)** Dilatación ($a_1 = a_2/2, a_3 = 2a_2$). **b)** Traslación de b_1 , pasando por b_2 , a b_3 .

3.2. Transformada Wavelet Continua (CWT)

Ahora podemos transformar una señal, $x(t)$, usando un rango de a 's y b 's. La transformada wavelet de una señal continua con respecto a una wavelet se define como:

$$T(a, b) = w(a) \int_{-\infty}^{\infty} x(t) \psi^* \left(\frac{t - b}{a} \right) dt \quad (3.2.1)$$

donde $w(a)$ es una función de peso. El $*$ indica que el conjugado de la función wavelet se utiliza en la transformación cuando la wavelet es compleja.

Normalmente, se utiliza $w(a) = 1/\sqrt{a}$ por razones de conservación de energía (es decir, asegura que las wavelets en cada escala tengan todas la misma energía). Podemos escribir entonces la versión normalizada de la wavelet de forma más compacta como:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi \left(\frac{t - b}{a} \right) \quad (3.2.2)$$

Con esta notación, la integral de la transformada wavelet continua puede escribirse como:

$$T(a, b) = \int_{-\infty}^{\infty} x(t) \psi_{a,b}^*(t) dt \quad (3.2.3)$$

La señal $x(t)$ podría ser un corazón que late, una señal de audio, una vibración de la caja de cambios, un índice financiero o incluso una señal espacial como el perfil de una grieta o la altura de la superficie de un terreno. La transformada wavelet es particularmente útil para analizar señales aperiódicas, ruidosas, intermitentes o transitorias gracias a su capacidad para examinar simultáneamente la señal tanto en tiempo como en frecuencia. Veamos ahora mas en detalle como funciona.

3.2.1. Interrogación de una señal

La figura 3.7, muestra la mecánica de la transformada wavelet dada por la ecuación 3.2.3. En la figura, una wavelet de escala a centrada en la ubicación b se muestra superpuesta sobre una señal arbitraria. Los segmentos donde la wavelet y la señal son ambas positivas resultan en una contribución positiva a la integral de la ecuación 3.2.3, por ejemplo, la región A en la figura. Del mismo modo, los segmentos donde la wavelet y la señal son ambas negativas también resultan es una contribución positiva a la integral (región B). En cambio, las regiones donde la señal y la wavelet son de signo opuesto dan como resultado contribuciones negativas a la integral, por ejemplo, las regiones C, D y E en la figura.

La figura 3.8 muestra una wavelet a una escala fija en cuatro ubicaciones distintas de una señal. En la primera ubicación (b_1), la wavelet se encuentra en un segmento de la señal en el que las partes positivas y negativas de la señal coinciden razonablemente con las de la wavelet. Esto da como resultado un valor positivo relativamente grande de $T(a, b)$. En la ubicación b_2 , las contribuciones positivas y negativas a la integral actúan para anularse entre sí, resultando en un valor cercano a cero para $T(a, b)$. En la ubicación b_3 , la señal y la wavelet están esencialmente fuera de fase, lo que da como

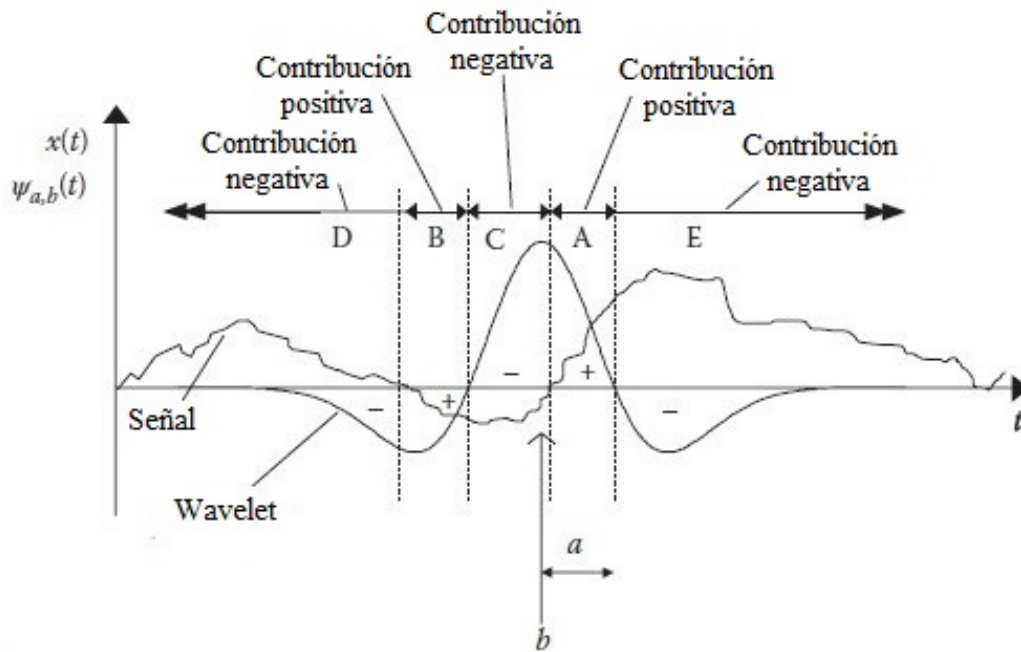


Figura 3.7: Wavelet con una escala y ubicación específica sobre una señal. Las regiones que tienen una contribución positiva en la integral están delineadas y marcadas con un signo $+$ y las que tienen contribución negativa, con un signo $-$.

resultado un gran valor negativo. En la ubicación b_4 , también están fuera de fase, sin embargo, esta vez la parte de la señal en las proximidades de la wavelet tiene una media bastante mayor a cero. Es fácil ver que la media contribuye valores positivos y negativos iguales a $T(a, b)$. Por lo tanto, solo las características locales de la señal son resaltadas por la wavelet en esta ubicación y la media es ignorada.

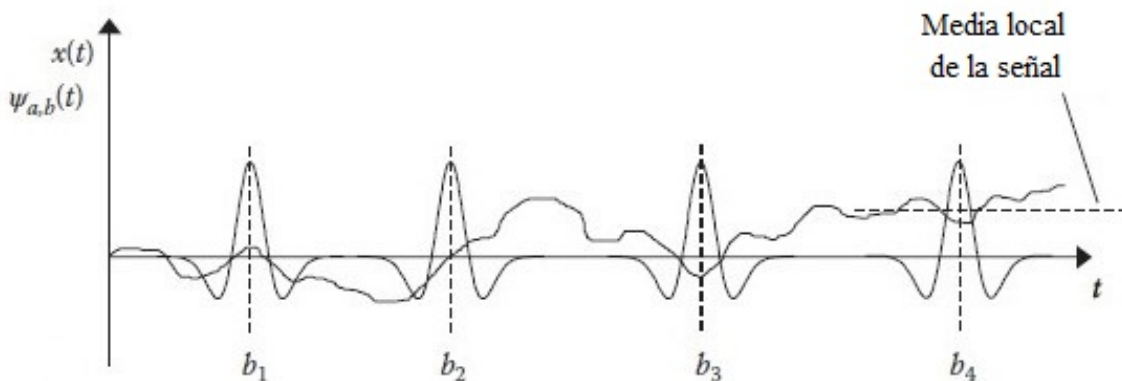


Figura 3.8: Wavelet con una escala fija en cuatro ubicaciones distintas

Es mediante este proceso que la transformada wavelet reconoce estructuras coherentes en una señal en varias escalas. Al mover la wavelet a lo largo de la señal (au-

mentando b), las estructuras coherentes relacionadas con una escala a específica en la onda se identifican y se repite este proceso en un rango de escalas hasta que todas las estructuras coherentes dentro de la señal, desde la más grande a la más pequeña, puedan ser distinguidas.

En la figura 3.9 se muestra el efecto que tiene el uso de una escala más pequeña en la transformación. En el gráfico, vemos que las partes positivas y negativas de la wavelet coinciden con aproximadamente la misma parte de la señal, produciendo un valor de $T(a, b)$ cercano a cero. Por tanto, $T(a, b)$ tiende a cero cuando la dilatación a tiende a cero. $T(a, b)$ también tiende a cero como a se vuelve muy grande (figura 3.10) ya que ahora la wavelet cubre muchas partes positivas y negativas que se repiten en la señal, de nuevo produciendo un valor cercano a cero de $T(a, b)$ en la integral de la ecuación 3.2.3. Por lo tanto, cuando la función wavelet es muy pequeña o muy grande en comparación con las características de la señal, la transformada da valores cercanos a cero.

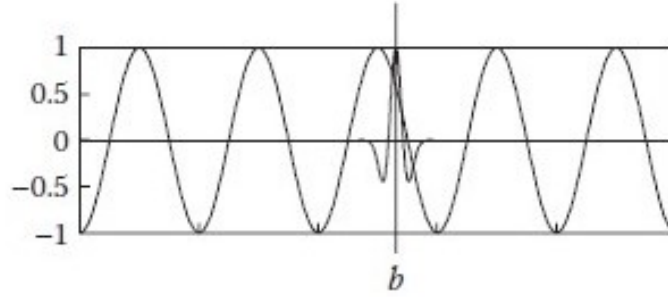


Figura 3.9: Wavelet con una escala pequeña respecto a la forma de la señal

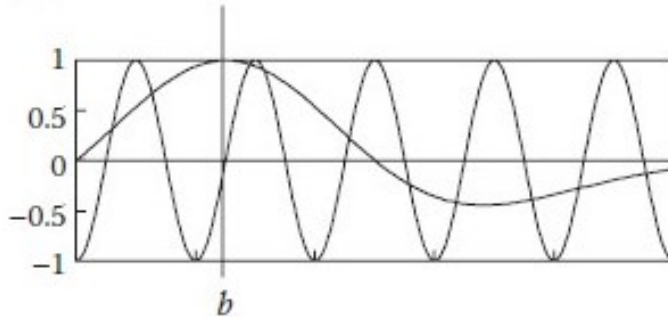


Figura 3.10: Wavelet con una escala grande respecto a la forma de la señal

3.2.2. Gráfico de la transformada

La transformada wavelet continua no suele calcularse en dilataciones arbitrarias y ubicaciones aisladas, sino más bien sobre un rango continuo de a y b . El gráfico de $T(a, b)$ contra a y b para una señal sinusoidal utilizando la wavelet Sombrero Mejicano se muestra en la figura 3.11. El gráfico de contorno es la forma más utilizada en la

práctica para representar la señal transformada resultante. Los valores cercanos a cero de $T(a, b)$ son evidentes en la gráfica tanto para valores grandes como chicos de a . Sin embargo, en valores intermedios de a , podemos ver grandes ondulaciones en $T(a, b)$ correspondiente a la forma sinusoidal de la señal. Para la wavelet Sombrero Mexicano, se requiere que la escala a sea aproximadamente una cuarta parte del período, p , de la onda sinusoidal para que haya una buena coincidencia entre la forma de la wavelet y la de la señal. En la figura 3.11b, podemos ver que los máximos y los mínimos en el gráfico de la transformación realmente ocurren en una escala de aproximadamente $0.25p$, lo que indica correlación máxima entre la wavelet y la señal a esta escala.

En el gráfico, el eje a tiene escala logarítmica. Esta es la forma más común que se ve en la práctica y también en general se muestran las escalas más pequeñas en la parte superior del gráfico. Si se traduce a frecuencias, en general se muestran al revés, es decir, las más pequeñas en la parte inferior del gráfico

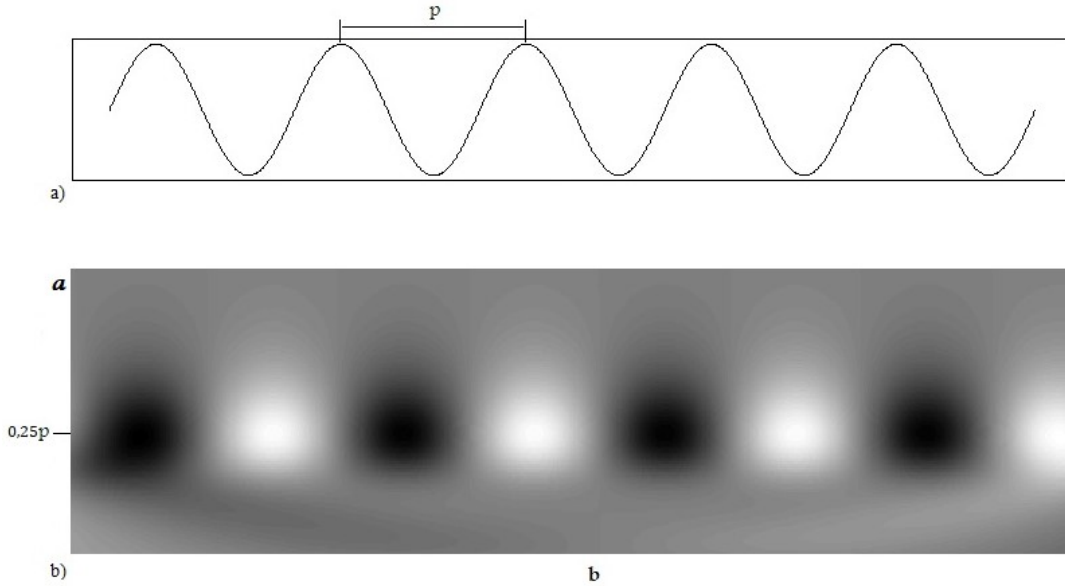


Figura 3.11: Gráfico de la transformada wavelet de una señal sinusoidal. **a)** Sinusoide de período p . **b)** Gráfico de contorno de $T(a, b)$ para la señal en a). Se utilizó una escala de grises donde el blanco corresponde a máximos de la transformada y el negro a mínimos.

La señal que se muestra en la figura 3.12 se compone de dos ondas sinusoidales, una con período p_1 y la otra con período $p_2 = p_1/5$. El gráfico de transformación muestra muy bien las dos componentes periódicas de la señal a escalas a de un cuarto de cada uno de los períodos. Este gráfico muestra claramente la capacidad de la transformada para descomponer la señal en sus diferentes componentes.

Si se utiliza una wavelet compleja, $T(a, b)$ resulta un número complejo. Por lo tanto, se pueden graficar por separado su parte real y su parte imaginaria o se la puede representar en términos de su fase $\phi(a, b) = \arctan([Im(T(a, b))]/[Re(T(a, b))])$ (figura 3.13b) y su módulo $|T(a, b)| = \sqrt{[Re(T(a, b))]^2 + [Im(T(a, b))]^2}$ (figura 3.13c). Las wavelets complejas en general producen una transformada cuyo módulo es menos oscilatorio que en el caso de una real. Esta propiedad es una ventaja para detectar y

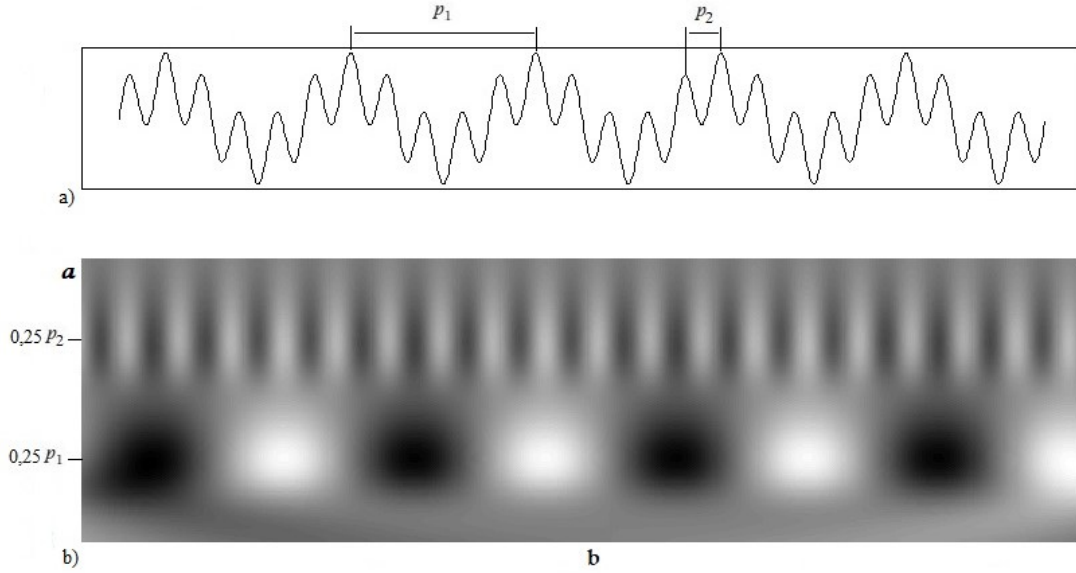


Figura 3.12: **a)** Señal compuesta por dos sinusoides. **b)** Gráfico de contorno de $T(a, b)$ para la señal en a).

rastrear frecuencias instantáneas contenidas en la señal. Una de la wavelets complejas más utilizada es la Morlet y es la que se utilizó generar las figuras 3.13b y 3.13c.

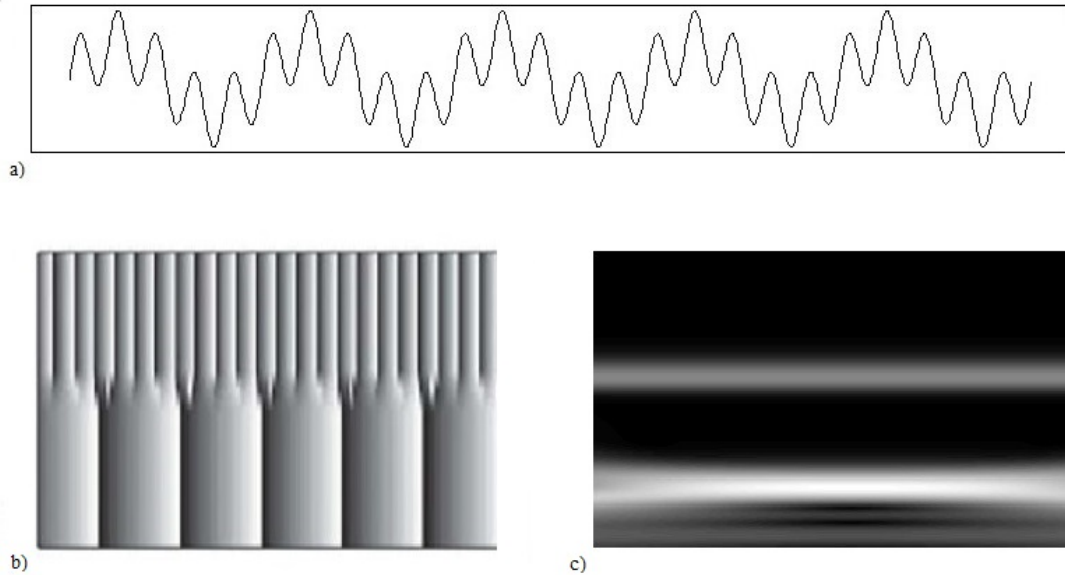


Figura 3.13: **a)** Señal compuesta por dos sinusoides. **b)** Gráfico de contorno de la fase $\phi(a, b)$ de $T(a, b)$ para la señal en a) utilizando la wavelet compleja Morlet. **c)** Gráfico de contorno del módulo de $T(a, b)$ para la señal en a) utilizando la wavelet Morlet.

3.2.3. Transformada inversa

Al igual que para la transformada de Fourier, existe una transformada wavelet inversa y se define como:

$$x(t) = \frac{1}{C_g} \int_{-\infty}^{\infty} \int_0^{\infty} T(a, b) \psi_{a,b}(t) \frac{da db}{a^2} \quad (3.2.4)$$

Esto permite recuperar la señal original a partir de su transformada wavelet al integrar en todas las escalas y ubicaciones, a y b . Observar que para la transformada inversa, se utiliza la wavelet original y no su conjugada como en la otra transformación. Si limitamos la integración sobre un rango de escalas en lugar de todas las a 's, podemos realizar un filtrado básico de la señal original.

Esto es muy útil por ejemplo para eliminar el ruido (figura 3.14), ya que este suele tener una frecuencia mucho más alta que el resto de los componentes de la señal y por lo tanto en el gráfico de la transformación va a aparecer en la zona de escalas más chicas. Se puede entonces fijar una escala a^* como umbral y asignar $T(a, b) = 0$, para todo $a < a^*$. De esta forma se estaría reconstruyendo la señal utilizando en realidad

$$x(t) = \frac{1}{C_g} \int_{-\infty}^{\infty} \int_{a^*}^{\infty} T(a, b) \psi_{a,b}(t) \frac{da db}{a^2} \quad (3.2.5)$$

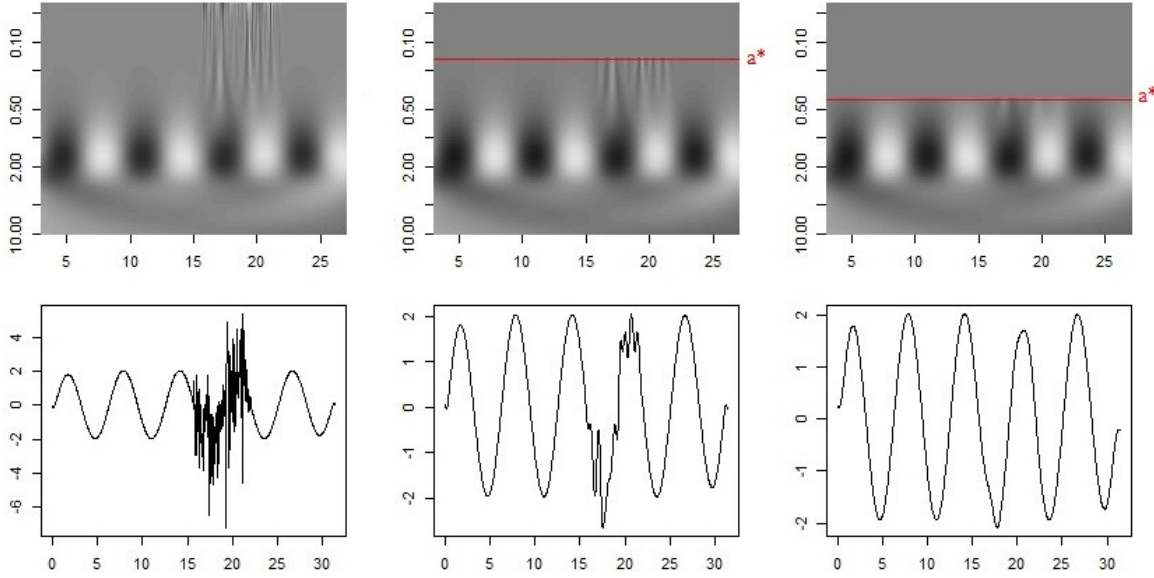


Figura 3.14: **Fila superior:** Gráficos de la transformada wavelet de una señal con ruido, filtrados con diferentes umbrales a^* (en el primero $a^*=0$, es decir, no se aplicó ningún filtro). **Fila inferior:** Reconstrucciones de la señal basadas respectivamente en los gráficos de la fila superior. Se puede observar que a mayor a^* , mayor es la reducción del ruido de alta frecuencia.

Otra aplicación es reconstruir los diferentes componentes de una señal por separado (figura 3.15). Por ejemplo, para la señal resultante de la suma de dos sinusoides con períodos p_1 y $p_2 = p_1/5$ podemos tomar una a^* tal que aproximadamente separe en el gráfico de la transformada las componentes asociadas a la onda de mayor frecuencia de la de menor frecuencia. Luego, integrando en la fórmula 3.2.4 a entre a^* y ∞ obtenemos una reconstrucción del componente con periodo p_1 e integrando entre 0 y a^* obtenemos una reconstrucción del componente con periodo p_2 .

Como se puede observar, se obtiene una coincidencia razonablemente buena entre la onda original y la reconstrucción. Sin embargo, como la información espectral en los componentes de la transformación está esparcida a través de escalas, una reconstrucción perfecta de los componentes sinusoidales individuales no es alcanzable, a diferencia de lo que se puede obtener con la transformada de Fourier en el caso específico de una señal compuesta por sinusoides.

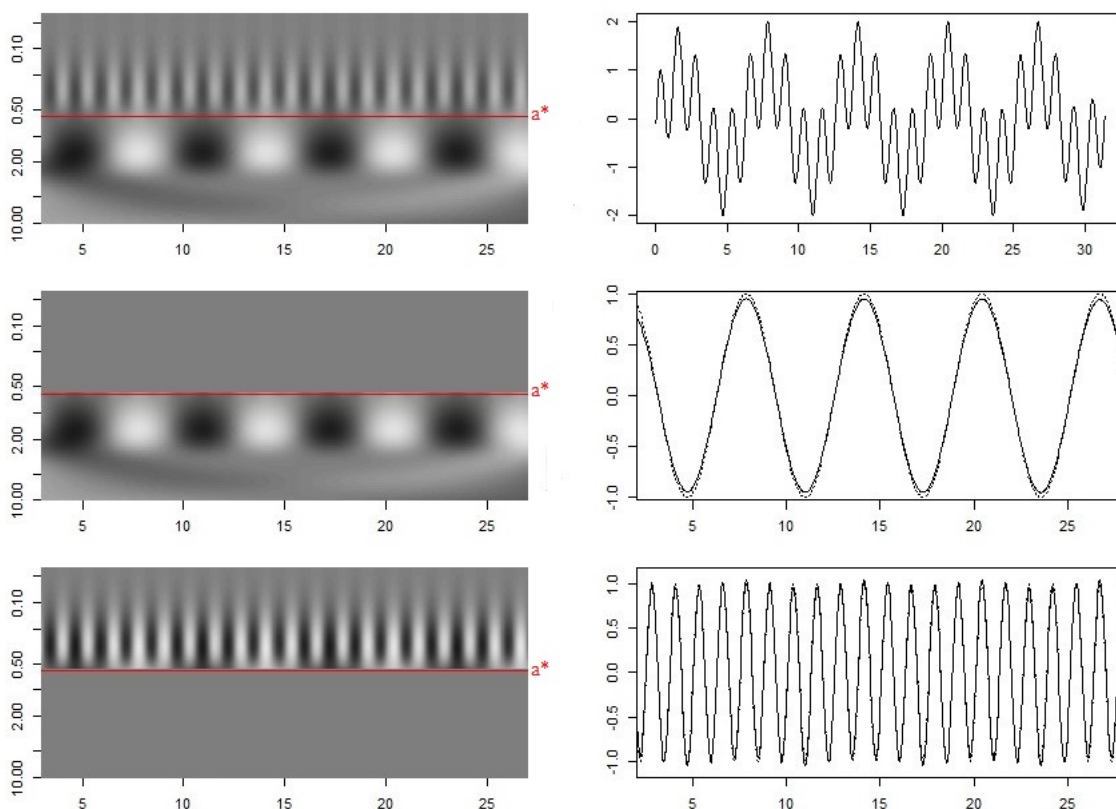


Figura 3.15: **Primera fila:** Gráfico de la transformada de una señal suma de dos sinusoides, sin filtrar, con el umbral a^* marcado en rojo y gráfico de la señal. **Segunda fila:** Gráfico de la transformada filtrando las escalas menores a a^* y reconstrucción de uno de los componentes haciendo la transformada inversa del gráfico filtrado. En línea punteada se muestra también el componente original superpuesto en el mismo gráfico que la reconstrucción. **Tercera fila:** Gráfico de la transformada filtrando las escalas mayores a a^* y reconstrucción del componentes haciendo la transformada inversa del gráfico filtrado. También, en línea punteada se muestra el componente original.

3.2.4. Transformada Wavelet vs. Transformada de Fourier

La transformada wavelet a menudo es comparada con la transformada de Fourier. La principal diferencia es que las wavelets se localizan tanto en tiempo como en frecuencia, mientras que la transformada de Fourier estándar solo se localiza en frecuencia. Es decir, dada una señal, nos dice cuanto de una frecuencia la función posee, pero no nos dice en que tiempo dicha frecuencia esta localizada.

Una forma de lidiar con este problema es utilizar la Transformada Corta de Fourier (STFT), que es aplicar una función ventana de corta duración a la señal, calcular la transformada de Fourier de los datos resultantes y repetir esto en diferentes lugares de la señal. Así podemos saber aproximadamente que frecuencias prevalecen en la señal cerca de un instante determinado. Es similar a la transformada wavelet, en el sentido de que también está localizada en tiempo y frecuencia, pero hay problemas con la compensación entre resolución de frecuencia y de tiempo.

Una buena resolución de frecuencia significa que componentes con frecuencias muy próximas pueden separarse y buena resolución de tiempo significa poder detectar el momento en el que cambian las frecuencias. El principio de incertidumbre de Heisenberg establece que la resolución temporal de la ventana es inversamente proporcional a la resolución en frecuencia: una ventana amplia proporciona una buena resolución de frecuencia pero una mala resolución de tiempo mientras que una ventana más estrecha proporciona una mejor resolución de tiempo pero una resolución de frecuencia deficiente.

Como en la STFT la longitud de la ventana es constante, la resolución de tiempo es la misma para frecuencias altas y para frecuencias bajas. Dado que una señal de alta frecuencia cambia mucho más rápido que una señal de baja frecuencia, sería ideal tener una transformación con mejor resolución de tiempo para las porciones de la señal de alta frecuencia y mejor resolución en frecuencia para las porciones de baja frecuencia.

Esta es una de las razones para la creación de la transformada wavelet que, al escalar la ventana tanto en tiempo como en frecuencia, puede dar una buena resolución de tiempo para eventos de alta frecuencia y una buena resolución de frecuencia para eventos de baja frecuencia, la combinación más adecuada para muchas señales reales.

3.3. Métodos de reasignación

Como se puede ver en los ejemplos anteriores de señales sinusoidales, la transformada wavelet produce una banda con componentes dispersos a través de escalas, distinto a la transformada de Fourier, que para una senoide produce un componente localizado presente solo en la frecuencia de la señal. Esto implica que se puede recuperar una amplitud distinta de cero incluso si la señal verdadera no tiene componente en esa coordenada de tiempo-frecuencia. Existen varios métodos de reasignación para “afinar” la representación de la transformada de una señal, uno de los más populares es *Synchrosqueezing*, que permite luego una reconstrucción más precisa.

Empecemos con un ejemplo para ver cómo funciona la idea de reasignación. Supongamos que tenemos la señal de los ejemplos anteriores, una suma de dos sinusoides de

períodos p_1 y $p_2 = p_1/5$, y por lo tanto frecuencias $f_1 = 1/p_1$ y $f_2 = 1/p_2$. El módulo de la transformada wavelet de la señal se muestra en la figura 3.16b y se puede ver la dispersión en las escalas. Sin embargo, si miramos el gráfico de la fase de la transformada (figura 3.16c), vemos claramente dos frecuencias distintas. Podemos determinar la frecuencia del ciclo de la fase en cada punto, la frecuencia instantánea, tomando la derivada de la fase local con respecto al tiempo. Luego, movemos el componente de la transformación correspondiente “hacia arriba” o “hacia abajo” en el plano de transformación a esta frecuencia. En efecto, todos los componentes de la transformación en la región marcada con “A” en el gráfico de la figura 3.16c se mueven a una posición en f_1 y todos los componentes en la región “B” se mueven a una posición en f_2 . Al hacerlo para esta señal de dos frecuencias, se obtienen entonces dos bandas concentradas como se puede ver en la figura 3.16d. La clave está entonces en que, aunque $T(a, b)$ se dispersa en a , su comportamiento oscilatorio en b apunta a la frecuencia original, independientemente del valor de a .

Existen algoritmos de synchrosqueezing basados en diferentes transformaciones de tiempo-frecuencia tales como la STFT y la CWT. En este trabajo utilizaremos el método basado en la CWT que como vimos tiene básicamente tres pasos: Primero, calcular la transformada wavelet, luego calcular las frecuencias instantáneas de la señal y por último realizar la reasignación de los componentes que corresponda. Veremos todo esto con más detalle en el próximo capítulo.

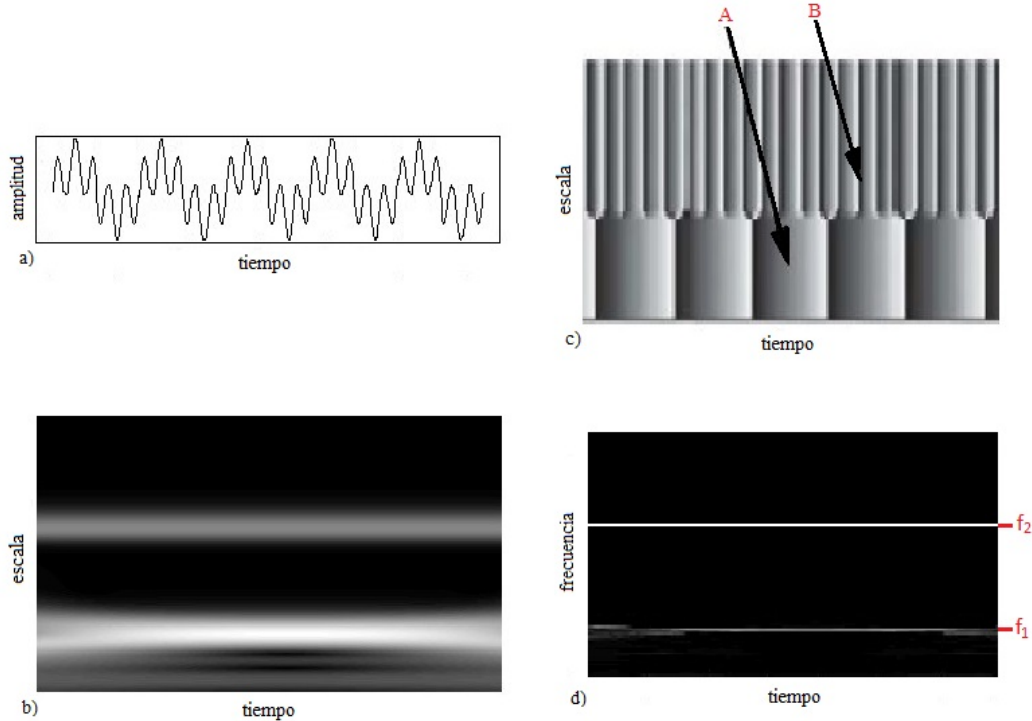


Figura 3.16: **a)** Señal compuesta por dos sinusoides. **b)** Gráfico del módulo de $T(a, b)$ para la señal en a). **c)** Gráfico de la fase de $T(a, b)$ para la señal en a). **d)** Gráfico de la transformada synchrosqueezing para la señal en a). Se puede observar que las bandas están mucho más concentradas que en b).

Capítulo 4

Método no paramétrico

En muchas aplicaciones prácticas, en una amplia gama de campos (incluidos, por ejemplo, la medicina, la biología o la ingeniería), uno se enfrenta a señales que tienen varias componentes estacionales, todas razonablemente bien localizadas en el espacio de tiempo-frecuencia, en diferentes ubicaciones. Las componentes a menudo son no estacionarias, en el sentido de que pueden presentar saltos o cambios de comportamiento, que puede ser importante capturar con la mayor precisión posible.

Como vimos la CWT proporciona una imagen un poco “borrosa” ya que suele haber una dispersión de los componentes en el plano de tiempo-frecuencia de la transformada de una señal. La transformada synchrosqueezing (SST) se desarrolló originalmente para resolver este problema y reconstruir los componentes estacionales de una señal sin contaminación de ruido. Pero como veremos en este capítulo, también se obtienen buenos resultados cuando se tienen observaciones ruidosas. Comenzaremos describiendo formalmente el método de synchrosqueezing y luego se enunciarán los teoremas presentes en [5], [16] y [3] sobre la capacidad de la SST para descomponer señales en presencia de distintas perturbaciones.

4.1. Transformada synchrosqueezing

En [5] se propone un método que captura la filosofía del enfoque de la descomposición empírica en modos (EMD) permitiendo el análisis de tiempo-frecuencia de señales multicomponente con amplitudes y frecuencias variables en el tiempo, pero construyendo los componentes de otra manera y evitando así las heurísticas que dificultan el análisis matemático. Para eso, primero se define una clase de funciones, sobre la que se establece el análisis teórico de la SST.

Definición 4.1.1. Una función $f : \mathbb{R} \rightarrow \mathbb{C}$, se dice que es de tipo moda intrínseca con precisión $\epsilon > 0$ si f y $A := |f|$ tienen las siguientes propiedades:

$$f(t) = A(t)e^{2\pi i\phi(t)}, \text{ donde} \tag{4.1.1}$$

$$A \in C^1(\mathbb{R}), \phi \in C^2(\mathbb{R}), \inf_{t \in \mathbb{R}} \phi'(t) > 0$$

$$|A'(t)|, |\phi''(t)| \leq \epsilon |\phi'(t)|, \forall t \in \mathbb{R}, \text{ y } M'' := \sup_{t \in \mathbb{R}} |\phi''(t)| < \infty$$

Llamamos función de amplitud a $A(t)$, función de fase a $\phi(t)$, frecuencia instantánea a $\phi'(t)$ y \mathcal{A}_ϵ a la clase de todas las funciones de tipo moda intrínseca con precisión $\epsilon > 0$. Intuitivamente, una señal en \mathcal{A}_ϵ es una función periódica de un solo componente que tiene funciones de amplitud y frecuencia instantánea que varían lentamente. Es esta naturaleza variable en el tiempo de $\phi'(t)$ y $A(t)$ lo que nos permite capturar el comportamiento momentáneo del sistema.

Definición 4.1.2. Una función $f : \mathbb{R} \rightarrow \mathbb{C}$, se dice que es una superposición de componentes de moda intrínseca bien separados con precisión $\epsilon > 0$ y separación $d > 0$ si puede escribirse como:

$$f(t) = \sum_{k=1}^K f_k(t) \quad (4.1.2)$$

donde las f_k son funciones de tipo moda intrínseca y sus respectivas funciones de fase ϕ_k satisfacen:

$$\phi'_k(t) > \phi'_{k-1}(t) \text{ y } |\phi'_k(t) - \phi'_{k-1}(t)| \geq d [\phi'_k(t) + \phi'_{k-1}(t)], \quad \forall t \in \mathbb{R} \quad (4.1.3)$$

Se nota como $\mathcal{A}_{\epsilon,d}$ al conjunto de todas las superposiciones de componentes de moda intrínseca bien separados con precisión $\epsilon > 0$ y separación $d > 0$. Las funciones de la clase $\mathcal{A}_{\epsilon,d}$ se componen de varios componentes oscilatorios con amplitudes y frecuencias suaves, que juntas caracterizan la dinámica de cada componente. Las frecuencias instantáneas ϕ'_k están fuertemente separadas en el sentido de que los componentes de alta frecuencia están exponencialmente más espaciados que los de baja frecuencia.

Ahora, sea $R_f : (b, \xi) \in \mathbb{R}^2 \rightarrow \mathbb{C}$ una representación de tiempo-frecuencia, por ejemplo, la obtenida aplicando la CWT a $f \in \mathcal{A}_{\epsilon,d}$. Como vimos, los métodos de reasignación concentran $R_f(b, \xi)$ al mover el valor de (b, ξ) a otro punto (b', ξ') de acuerdo a alguna regla de reasignación. El algoritmo SST, un caso especial de método de reasignación, se compone de tres pasos. Primero, se debe elegir una wavelet madre $\psi \in \mathcal{S}$ tal que $\text{supp}(\hat{\psi}) \subset [1 - \Delta, 1 + \Delta]$, donde $\Delta \ll 1$ y calcular $T_f(a, b)$, la transformada wavelet continua de $f(t)$ dada por 3.2.3. Luego, se calcula la función $\omega_f(a, b)$ definida en $\mathbb{R}^+ \times \mathbb{R}$, que juega el papel de regla de reasignación:

$$\omega_f(a, b) := \begin{cases} \frac{-i\partial_b T_f(a, b)}{2\pi T_f(a, b)} & \text{si } |T_f(a, b)| \neq 0 \\ \infty & \text{si } |T_f(a, b)| = 0 \end{cases} \quad (4.1.4)$$

Por definición, $\omega_f(a, b)$ contiene abundante información sobre las funciones de frecuencia instantánea en f . De hecho, cuando f es una función puramente armónica, $\omega_f(a, b)$ toma el valor de la frecuencia de f si es finita (los detalles de esto se pueden encontrar en [5]). Por último, la SST de $f(t)$ con umbral ϵ se define reasignando valores en la representación de $T_f(a, b)$ de acuerdo con la regla de reasignación $\omega_f(a, b)$:

$$S_{f,\epsilon}(b, \xi) := \lim_{\alpha \rightarrow 0} \int_{\{(a,b): |T_f(a,b)| \geq \epsilon\}} h_\alpha(|\omega_f(a, b) - \xi|) T_f(a, b) a^{-3/2} da \quad (4.1.5)$$

donde $(b, \xi) \in \mathbb{R} \times \mathbb{R}^+$, $\alpha, \epsilon > 0$, $h_\alpha(b) := \frac{1}{\alpha} h\left(\frac{b}{\alpha}\right)$, $h \in L^1(\mathbb{R})$, y $h_\alpha \rightarrow \delta$ débilmente cuando $\alpha \rightarrow 0$ con δ denotando la función delta de Dirac. La idea es que como la

mayoría de la dispersión ocurre en el eje de la escala, la reasignación se realiza solo a lo largo este eje, pasando del plano tiempo-escala (b, a) al plano tiempo-frecuencia $(b, \omega_f(a, b))$. Por lo tanto, en cada punto b , $S_f^T(b, \xi)$ recolecta todos los coeficientes de la CWT con escalas a en los que la CWT detecta un componente estacional con frecuencia cercana a ξ . Este operador se puede considerar luego como una inversión parcial de la CWT de f pero tomando solo pequeñas bandas alrededor de las curvas de nivel en el plano de tiempo-escala (donde $\omega_f(a, b) \approx \xi$) e ignorando el resto del plano. Al hacer $\alpha \rightarrow 0$, el dominio de inversión se concentra cada vez más en las curvas de nivel $\{(a, b) : \omega_f(a, b) = \xi\}$. Intuitivamente, este proceso de localización nos permitirá recuperar los componentes f_k más con precisión que invertir la CWT en todo el espacio. Veremos la fórmula de reconstrucción para el k -ésimo componente de f en el teorema 4.2.1.

4.2. Análisis en ausencia de ruido

Como vimos, la clase de funciones $\mathcal{A}_{\epsilon, d}$, consiste en una superposición de componentes aproximadamente armónicos localmente, bien separados y con amplitudes y frecuencia que varían lentamente. En [5] se demuestra que el método de synchrosqueezing tiene éxito en descomponer funciones arbitrarias en esta clase, identificando la frecuencia instantánea y la amplitud de cada componente. Ese es el contenido del siguiente teorema:

Teorema 4.2.1. *Sea f una función en $\mathcal{A}_{\epsilon, d}$ y sea $\tilde{\epsilon} := \epsilon^{1/3}$. Se elige una wavelet ψ tal que su transformada de Fourier $\hat{\psi}$ está soportada en $[1 - \Delta, 1 + \Delta]$, con $\Delta < d/(1 + d)$, y sea $\mathcal{R}_\psi = \int \hat{\psi}(\zeta) \zeta^{-1} d\zeta$. Consideramos la transformada wavelet continua $T_f(a, b)$ de f respecto a la wavelet ψ , y la función $S_{f, \tilde{\epsilon}}(b, \omega)$, obtenida aplicando synchrosqueezing (fórmula 4.1.5) a T_f , con umbral $\tilde{\epsilon}$. Entonces, dado ϵ (y por lo tanto $\tilde{\epsilon}$) suficientemente chicos, vale lo siguiente:*

(I) $|T_f(a, b)| > \tilde{\epsilon}$ solo si, para algún k , $(a, b) \in Z_k := \{(a, b); |a\phi'_k(b) - 1| < \Delta\}$.

(II) Para cada $k \in \{1, \dots, K\}$, y para cada par $(a, b) \in Z_k$ para el cual $|T_f(a, b)| > \tilde{\epsilon}$, se tiene que

$$|\omega_f(a, b) - \phi'_k(b)| \leq \tilde{\epsilon}$$

(III) Además, para cada $k \in \{1, \dots, K\}$, existe una constante C_1 tal que, para cualquier $b \in \mathbb{R}$,

$$\left| \left(\mathcal{R}_\psi^{-1} \int_{|\omega - \phi'_k(b)| < \tilde{\epsilon}} S_{f, \tilde{\epsilon}}(b, \omega) d\omega \right) - A_k(b) e^{2\pi i \phi_k(b)} \right| \leq C_1 \tilde{\epsilon}$$

Este teorema muestra como para $f \in \mathcal{A}_{\epsilon, d}$ la transformada synchrosqueezing puede identificar y extraer los componentes f_k de f . En (I) y (II) se ve que la SST está completamente concentrada, en el plano tiempo-frecuencia (b, ω) , en bandas estrechas alrededor de las curvas de frecuencia instantánea $\omega = \phi'_k(b)$. La última parte nos dice que invirtiendo $S_{f, \tilde{\epsilon}}$ restringiendo el dominio a la k -ésima banda concentrada basta para

reconstruir, con alta precisión, el componente $f_k = A_k(b)e^{2\pi i\phi_k(b)}$. Podemos entonces definir el siguiente estimador:

$$\tilde{f}_{k,\tilde{\epsilon}}(b) := \mathcal{R}_\psi^{-1} \int_{|\omega - \phi'_k(b)| < \tilde{\epsilon}} S_{f,\tilde{\epsilon}}(b, \omega) d\omega, \quad (4.2.1)$$

En particular, el teorema también implica que podemos recuperar las amplitudes A_k tomando el valor absoluto de la fórmula de reconstrucción del k -ésimo componente:

$$\tilde{A}_k(t) := |\tilde{f}_{k,\tilde{\epsilon}}(b)|$$

El teorema también sugiere que componentes f_k de pequeña magnitud pueden ser difíciles de detectar si su transformada wavelet tiene módulo menor que $\tilde{\epsilon}$. La demostración de 4.2.1, se basa en una serie de estimaciones y se puede encontrar en [5].

Aplicar la SST proporciona así una descomposición adaptativa, que era el objetivo de la EMD, pero a diferencia de ésta, ofrece también garantías teóricas sobre la precisión de la estimación de los componentes. Veremos en las próximas secciones que garantías similares se mantienen incluso en presencia de perturbaciones como ruido o tendencia.

4.3. Análisis en presencia de ruido determinístico y Gaussiano

En [16], los autores analizan la estabilidad de la transformada synchrosqueezing y muestran que es robusta a perturbaciones determinísticas acotadas y a ruido blanco Gaussiano. Con estos resultados se justifica entonces la aplicación de este método de descomposición a señales ruidosas o con intervalos de muestreo no uniforme, como las que muchas veces se encuentran en las ciencias naturales o la ingeniería.

El primer teorema establece que los resultados del teorema 4.2.1 esencialmente se mantienen si se perturba f con un pequeño término de error e .

Teorema 4.3.1. *Sea $f \in \mathcal{A}_{\epsilon,d}$ y supongamos que tenemos ϵ, ψ, Δ y Z_k como en el teorema 4.2.1. Sea $g = f + e$, donde $e \in L^\infty$ es un término de error que satisface $\|e\|_{L^\infty} \leq \epsilon / \max(\|\psi\|_{L^1}, \|\psi'\|_{L^1})$. Para cada k , sea $M_k \geq 1$ el “rango de frecuencia maximal” dado por $M_k = \max\left(\frac{1}{1-\Delta} \|\phi'_k\|_{L^\infty}, (1+\Delta) \left\|\frac{1}{\phi'_k}\right\|_{L^\infty}\right)$. Entonces, para cada k vale lo siguiente:*

- (I) *Si asumimos $a \in [M_k^{-1}, M_k]$, para cada punto $(a, b) \in Z_k$ con $|T_g(a, b)| > M_k^{1/2}\epsilon + \tilde{\epsilon}$, se tiene que*

$$|\omega_g(a, b) - \phi'_k(b)| \leq C_2 \tilde{\epsilon}$$

para alguna constante $C_2 = O(M_k)$. Si $(a, b) \notin Z_k$ para ningún k , entonces $|T_g(a, b)| \leq M_k^{1/2}\epsilon + \tilde{\epsilon}$

- (II) *Existe una constante $C_3 = O(M_k)$ tal que para todo $b \in \mathbb{R}$,*

$$\left| \left(\mathcal{R}_\psi^{-1} \int_{|\omega - \phi'_k(b)| \leq C_2 \tilde{\epsilon}} S_{f, M_k^{1/2}\epsilon + \tilde{\epsilon}}(b, \omega) d\omega \right) - A_k(b) e^{2\pi i \phi_k(b)} \right| \leq C_3 \tilde{\epsilon}$$

Este teorema muestra que cada componente f_k puede recuperarse con una precisión proporcional a la perturbación ϵ y a su rango de frecuencia maximal M_k , con frecuencias instantáneas intermedias (M_k cercano a 1) resultando en las mejores estimaciones.

El segundo resultado presente en [16] es que synchrosqueezing también es resistente al ruido blanco Gaussiano. Formalmente, un proceso Gaussiano generalizado estacionario (real) G es un funcional lineal aleatorio en \mathcal{S} , la clase de Schwartz, tal que todas las colecciones finitas $\{G(f_i)\}$ con $f_i \in \mathcal{S}$ tienen distribución conjunta normal o gaussiana y tienen la misma distribución para todas las posibles elecciones de f_i . Estos procesos se caracterizan por una función media $\mathbb{E}(G(f_1)) = T(f_1)$ y una función covarianza $\mathbb{E}((G(f_1) - T(f_1)) \overline{(G(f_2) - T(f_2))}) = \langle f_1, Rf_2 \rangle$ para algunos operadores $T : \mathcal{S} \rightarrow \mathcal{S}$ y $R : \mathcal{S} \rightarrow \mathcal{S}$, donde $\langle f_1, f_2 \rangle = \int_{-\infty}^{\infty} f_1(t) \overline{f_2(t)} dt$ es el producto interno de L^2 . Los procesos Gaussianos pueden verse como una generalización en dimensión infinita de distribuciones normales multivariadas. El ruido blanco Gaussiano N con potencia σ^2 es un proceso con $T = 0$ y $R = \sigma^2 I$, donde I es el operador de identidad.

Teorema 4.3.2. *Sea $f \in \mathcal{A}_{\epsilon,d}$ y supongamos que tenemos $\epsilon, \psi, \Delta, Z_k$ y M_k como en los teoremas 4.2.1 y 4.3.1, con las condiciones adicionales de que $\psi \in \mathcal{S}$ y $|\langle \psi, \psi' \rangle| < \|\psi\|_{L^2} \|\psi'\|_{L^2}$. Sea $g = f + N$, donde N es ruido blanco Gaussiano con densidad espectral ϵ^{2+p} para algún $p > 0$. Entonces, para cada k vale lo siguiente:*

- (I) *Si asumimos $a \in [M_k^{-1}, M_k]$, para cada punto $(a, b) \in Z_k$ con $|T_f(a, b)| > \tilde{\epsilon}$, existen constantes E_1 y C'_2 tales que con probabilidad $1 - e^{-E_1 \epsilon^{-p}}$,*

$$|\omega_g(a, b) - \phi'_k(b)| \leq C'_2 \tilde{\epsilon}.$$

Si $(a, b) \notin Z_k$ para ningún k , entonces para alguna constante E_2 , con probabilidad $1 - e^{-E_2 \epsilon^{-p}}$, $|T_g(a, b)| \leq \tilde{\epsilon} + \frac{1}{2} \epsilon$.

- (II) *Existe una constante C'_3 tal que con probabilidad $1 - e^{-E_1 \epsilon^{-p}}$, para todo $b \in \mathbb{R}$ se tiene:*

$$\left| \left(\mathcal{R}_{\psi}^{-1} \int_{|\omega - \phi'_k(b)| \leq C'_2 \tilde{\epsilon}} S_{f, M_k^{1/2} \epsilon + \tilde{\epsilon}}(b, \omega) d\omega \right) - A_k(b) e^{2\pi i \phi_k(b)} \right| \leq C'_3 \tilde{\epsilon}$$

La primer parte del teorema 4.3.2 dice que la potencia del error se dispersa a través del plano tiempo-frecuencia de la synchrosqueezing en lugar de acumularse en la frecuencia instantánea de un componente particular. La segunda parte implica lo mismo para todo el componente $A_k(b) e^{2\pi i \phi_k(b)}$, incluyendo la amplitud.

4.4. Análisis en presencia de tendencia y ruido heterocedástico dependiente

4.4.1. Identificabilidad

En estadística, la identificabilidad es una propiedad que un modelo debe satisfacer para que sea posible una inferencia precisa. Un modelo es identificable si es teóricamente

posible aprender los valores verdaderos de los parámetros subyacentes de este modelo después de obtener un número infinito de observaciones de él. Matemáticamente, esto equivale a decir que diferentes valores de los parámetros deben generar diferentes distribuciones de probabilidad de las variables observables. Por lo general, el modelo es identificable solo bajo ciertas restricciones técnicas, en cuyo caso el conjunto de estos requisitos se denomina condiciones de identificación.

Un modelo es no identificable si dos o más parametrizaciones son observacionalmente equivalentes. En algunos casos, aunque un modelo no sea identificable, aún es posible conocer los valores verdaderos de un cierto subconjunto de los parámetros. En este caso decimos que el modelo es parcialmente identificable. En otros casos, puede ser posible aprender la ubicación del parámetro verdadero hasta una cierta región finita del espacio de parámetros, en cuyo caso el modelo se establece como identificable.

En [12], se comenta que más de dos docenas de términos diferentes para la identificación aparecen en la literatura econométrica, los cuales incluyen identificación de conjunto, identificación causal, identificación local, identificación genérica, identificación débil, identificación en el infinito y muchas más, así como otros conceptos estrechamente relacionados con la identificación, como las normalizaciones y las diferencias en la identificación entre modelos estructurales y modelos causales de forma reducida.

En el caso particular de series de tiempo observadas sobre fenómenos con estacionalidad y tendencia, contaminados con ruidos dependientes el problema de la identificabilidad de las modas estacionales es aún más complicado de definir y estudiar. Nosotros lo vamos a hacer siguiendo el enfoque presente en [3].

Hasta ahora vimos, entre otras cosas, que utilizando la técnica de synchrosqueezing se pueden aproximar bien los componentes estacionales de una f representada como $A_k(b)e^{2\pi i\phi_k(b)}$ utilizando la fórmula de reconstrucción 4.2.1. Pero, ¿es esa la única representación de f dentro de la clase que estamos considerando?. Veamos un ejemplo: consideremos la señal

$$s(t) = \frac{1}{4} \cos([\Omega - \gamma]t) + \frac{5}{2} \cos(\Omega t) + \frac{1}{4} \cos([\Omega + \gamma]t) = \left(2 + \cos^2 \left[\frac{\gamma}{2}t\right]\right) \cos(\Omega t)$$

donde $\Omega \gg \gamma$. Esta señal puede tener más de una interpretación: se puede considerar como una suma de tres cosenos con frecuencias $\Omega - \gamma$, Ω y $\Omega + \gamma$, respectivamente, o como un solo componente con frecuencia Ω que tiene una amplitud $A(t) = 2 + \cos^2 \left[\frac{\gamma}{2}t\right]$ que varía lentamente. Cuál de las dos representaciones es “mejor” depende del problema, y diferentes representaciones conducen a diferentes interpretaciones. Quizás, si el objetivo es solo predecir algunos valores futuros, el resultado no cambia con una representación o con otra, pero sí puede ser un problema cuando uno quiere utilizar una técnica de descomposición para responder una pregunta concreta y por lo tanto no quiere ambigüedad en la interpretación. Por ejemplo, supongamos que estamos estudiando el comportamiento de ratones. Se mide el movimiento de los mismos durante un periodo prolongado de tiempo y queremos encontrar patrones o ciclos. Para eso aplicamos el método de synchrosqueezing y reconstruimos los diferentes elementos de la señal. Pero una vez que encontramos las componentes, queremos estar seguros que esa es la única posible descomposición de la señal y que no hay otra que explique los datos, porque sino, ¿cómo elegimos cuál es la que representa el verdadero comportamiento de

los ratones?.

En general, la expresión $f(t) = \sum_{k=1}^K A_k(t) \cos(2\pi\phi_k(t))$ no es única y por lo tanto existe el problema de la identificabilidad de f . Una forma de evitar este problema podría ser modelar $A(t)$ y $\phi(t)$ de alguna forma paramétrica (en [8] se hace esto para una función de un solo componente estacional y sin tendencia). Sin embargo, los supuestos paramétricos pueden ser demasiado restrictivos. Por eso en [3] se aborda este problema considerando clases de funciones, $\mathcal{A}_\epsilon^{c_1, c_2}$ y $\mathcal{A}_{\epsilon, d}^{c_1, c_2}$, en las que sólo se imponen supuestos no paramétricos sobre las funciones de amplitud y frecuencia instantánea. Estas clases son en realidad subconjuntos de la clases \mathcal{A}_ϵ y $\mathcal{A}_{\epsilon, d}$ que definimos antes, ya que solo se agregan algunas condiciones más. Luego, surge la pregunta, ¿dada una función $f(t) = A_k(t) \cos(2\pi\phi_k(t)) \in \mathcal{A}_\epsilon^{c_1, c_2}$, cuánto pueden diferir entre sí las diferentes representaciones en $\mathcal{A}_\epsilon^{c_1, c_2}$ para f ?, ¿y para $f \in \mathcal{A}_{\epsilon, d}^{c_1, c_2}$?. Este es el problema de identificabilidad que enfrentamos cuando introducimos estas clases funcionales.

El teorema 4.2.1 afirmaba que las reconstrucciones de los componentes mediante synchrosqueezing diferían poco de los componentes de una representación de f en $\mathcal{A}_{\epsilon, d}$. Como la transformada synchrosqueezing es única para cada f , ya que en particular la transformada wavelet de f es única, intuitivamente, tiene sentido pensar que si existiesen dos representaciones de f en $\mathcal{A}_{\epsilon, d}^{c_1, c_2}$, éstas no pueden diferir en mucho. Ese es el contenido de los teoremas 4.4.1 y 4.4.2. Así, si la diferencia entre distintas representaciones de f es despreciable, se resuelve el problema de la ambigüedad de la representación y podemos decir que f es “identificable”. Enunciamos las definiciones y los teoremas:

Definición 4.4.1. Para elecciones fijas de $0 < \epsilon \ll 1$ y $\epsilon \ll c_1 < c_2 < \infty$, el espacio $\mathcal{A}_\epsilon^{c_1, c_2}$ de funciones de moda intrínseca consiste de funciones $f : \mathbb{R} \rightarrow \mathbb{R}$, $f \in C^1(\mathbb{R}) \cap L^\infty(\mathbb{R})$ de la forma

$$f(t) = A(t) \cos(2\pi\phi(t)) \quad (4.4.1)$$

donde $A : \mathbb{R} \rightarrow \mathbb{R}$ y $\phi : \mathbb{R} \rightarrow \mathbb{R}$ satisfacen las siguientes condiciones para todo $t \in \mathbb{R}$:

$$A \in C^1(\mathbb{R}) \cap L^\infty(\mathbb{R}), \inf_{t \in \mathbb{R}} A(t) > c_1, \sup_{t \in \mathbb{R}} A(t) < c_2 \quad (4.4.2)$$

$$\phi \in C^2(\mathbb{R}), \inf_{t \in \mathbb{R}} \phi'(t) > c_1, \sup_{t \in \mathbb{R}} \phi'(t) < c_2 \quad (4.4.3)$$

$$|A'(t)| \leq \epsilon \phi'(t), \quad |\phi''(t)| \leq \epsilon \phi'(t) \quad (4.4.4)$$

Definición 4.4.2. Fijado $0 < d < 1$, el espacio $\mathcal{A}_{\epsilon, d}^{c_1, c_2}$ de superposiciones de funciones de moda intrínseca consiste en funciones f de la forma

$$f(t) = \sum_{k=1}^K f_k(t) \quad (4.4.5)$$

para algún $K > 0$ finito y para cada $k = 1, \dots, K$, $f_k(t) = A_k(t) \cos(2\pi\phi_k(t)) \in \mathcal{A}_\epsilon^{c_1, c_2}$ tal que ϕ_k satisface

$$\phi'_k(t) > \phi'_{k-1}(t) \quad \text{y} \quad \phi'_k(t) - \phi'_{k-1}(t) \geq d [\phi'_k(t) + \phi'_{k-1}(t)] \quad (4.4.6)$$

Teorema 4.4.1. *Supongamos que $a(t) \cos \phi(t) \in \mathcal{A}_\epsilon^{c_1, c_2}$ puede representarse de una forma diferente que también pertenece a $\mathcal{A}_\epsilon^{c_1, c_2}$, es decir,*

$$a(t) \cos \phi(t) = A(t) \cos \varphi(t) \in \mathcal{A}_\epsilon^{c_1, c_2} \quad (4.4.7)$$

Definimos $t_m \in \mathbb{R}, m \in \mathbb{Z}$ tal que $\phi(t_m) = (m + 1/2)\pi$, $s_m \in \mathbb{R}, m \in \mathbb{Z}$ tal que $\phi(s_m) = m\pi$, $\alpha(t) = A(t) - a(t)$ y $\beta(t) = \varphi(t) - \phi(t)$. Entonces $\alpha \in C^2(\mathbb{R})$, $\beta \in C^1(\mathbb{R})$, $\alpha(t_m) = 0 \forall m$, $\beta(s_m) \geq 0 \forall m$ y $\beta(s_m) = 0$ si y solo si $\alpha(s_m) = 0$. Además, tenemos que $|\alpha'(t)| \leq 3\pi\epsilon$, $|\alpha(t)| \leq \frac{4\pi^2\epsilon}{c_1}$ y $|\beta(t)| < 3\pi\epsilon$ para todo $t \in \mathbb{R}$

El teorema 4.4.1 consta de dos conclusiones. La primera es que α y β deben tener algunos “puntos bisagra” y por lo tanto, están restringidas. Esta propiedad viene de la condición de positividad de la frecuencia instantánea y de la función de amplitud. La segunda conclusión es que los valores absolutos de α , α' y β no pueden ser grandes. Esta propiedad proviene de las condiciones de “variación lenta” de la clase funcional $\mathcal{A}_\epsilon^{c_1, c_2}$ y la existencia de los puntos bisagra. Con estas propiedades, la definición de las funciones de amplitud y frecuencia instantánea es rigurosa en el sentido de que son únicas salvo un error insignificante cuando ϵ es lo suficientemente pequeño.

Teorema 4.4.2. *Supongamos que $f(t) = \sum_{l=1}^N a_l(t) \cos \phi_l(t) \in \mathcal{A}_{\epsilon, d}^{c_1, c_2}$ puede representarse de una forma diferente que también pertenece a $\mathcal{A}_{\epsilon, d}^{c_1, c_2}$, es decir,*

$$f(t) = \sum_{l=1}^N a_l(t) \cos \phi_l(t) = \sum_{l=1}^M A_l(t) \cos \varphi_l(t) \in \mathcal{A}_{\epsilon, d}^{c_1, c_2} \quad (4.4.8)$$

entonces $M = N$, $|\phi_l(t) - \varphi_l(t)| \leq E_I \epsilon$, $|\phi'_l(b) - \varphi'_l(b)| \leq E_I \epsilon$ y $|a_l(t) - A_l(t)| \leq E_I \epsilon$ para todo $l = 1, \dots, N$, donde $E_I > 0$ es una constante universal finita que depende de c_1 , c_2 y d .

En el teorema 4.4.2 se establece la identificabilidad para $\mathcal{A}_{\epsilon, d}^{c_1, c_2}$ concluyendo nuevamente que cualquier función multicomponente periódica en $\mathcal{A}_{\epsilon, d}^{c_1, c_2}$ es identificable salvo un pequeño error o sesgo de modelo de orden ϵ . Las demostraciones de estos teoremas se pueden encontrar en la sección suplementaria de [3].

4.4.2. Modelo

Una vez resuelto el problema de la identificabilidad, en [3] se modela a un proceso aleatorio $Y(t)$ con múltiples componentes estacionales y tendencia contaminadas por errores heterocedásticos dependientes de la siguiente manera:

$$Y(t) = f(t) + T(t) + \sigma(t)\Phi(t) \quad (4.4.9)$$

donde la estacionalidad $f(t) = \sum_{k=1}^K A_k(t) \cos(2\pi\phi_k(t))$ está en $\mathcal{A}_c^{c_1, c_2}$ cuando $K = 1$ y en $\mathcal{A}_{\epsilon, d}^{c_1, c_2}$ cuando $K > 1$, la tendencia $T(t)$ se modela como una función real C^1 , $\Phi(t)$ es algún proceso aleatorio generalizado (GRP) estacionario, y $\sigma \in C^\infty \cap L^\infty$ es una función suave a valores reales que se utiliza para modelar la heterocedasticidad del término de error, con $\sigma(t) > 0$. Por ejemplo, para simular $\Phi(t)$ podría tomarse un

proceso CARMA (proceso autorregresivo y de medias móviles continuo) de orden (p, q) , con $p, q \geq 0$. Notar que en este modelo, la tendencia $T(t)$ también es no paramétrica. Se especificarán algunas condiciones que debe cumplir en el teorema 4.4.3, pero intuitivamente debería ser una función de “muy baja frecuencia”. Combinando esto con los modelos no paramétricos $\mathcal{A}_\epsilon^{c_1, c_2}$ y $\mathcal{A}_{\epsilon, d}^{c_1, c_2}$ para la estacionalidad dinámica, el modelo 4.4.9 provee una descomposición general de la tendencia, la estacionalidad y el error que nos permitirá extraer información de todos los $A_k(t), \phi_k(t), \phi'_k(t)$, con $k = 1, \dots, K$, y $T(t)$ basado en observaciones de $Y(t)$, como veremos en el teorema 4.4.3

En la práctica, solo podemos acceder al proceso de tiempo continuo $Y(t)$ dado en el modelo 4.4.9 en puntos de muestreo discretos $n\tau$, donde $n \in \mathbb{Z}$ y $\tau > 0$ es el intervalo de muestreo. Entonces, consideramos el siguiente modelo de tiempo discreto:

$$Y_n = f(n\tau) + T(n\tau) + \sigma(n\tau)\Phi_n, n \in \mathbb{Z} \quad (4.4.10)$$

donde f, T y σ son como en el modelo 4.4.9 y $\Phi_n, n \in \mathbb{Z}$ es una serie de tiempo estacionaria con media cero que puede tomarse, por ejemplo, como un proceso ARMA. Podemos interpretar 4.4.10 como un modelo para un proceso a tiempo discreto $\{Y_n\}_{n \in \mathbb{Z}}$ en el que la estacionalidad y tendencia determinísticas están contaminadas por los errores heteroscedástico $\sigma(n\tau)\Phi_n, n \in \mathbb{Z}$.

4.4.3. Método

Necesitamos un método para analizar las observaciones generadas por los modelos 4.4.9 y 4.4.10 a fin de extraer información sobre la tendencia $T(t)$ y la dinámica de la estacionalidad, $A_k(t), \phi_k(t)$ y $\phi'_k(t), k = 1, \dots, K$. Si bien la transformada synchrosqueezing (SST), se introdujo originalmente para estudiar la estacionalidad dinámica sin acoplamiento con tendencia o errores aleatorios, en [3] se muestra que, teóricamente, la SST puede usarse para estimar con precisión las funciones $\phi_k(t), \phi'_k(t)$ y $A_k(t)$ cuando la tendencia está presente, y es robusta a errores aleatorios heteroscedásticos dependientes, es decir, cuando los datos se modelan mediante 4.4.9 o 4.4.10. Estos son modelos incluso más complejo que los estudiados en [16] que vimos en la sección anterior (4.3), que contenían solo ruido determinístico o Gaussiano blanco y no contenían tendencia, con lo cual amplía aún más las posibilidades de aplicación del método de synchrosqueezing manteniendo garantías teóricas sobre estabilidad.

Como en las secciones anteriores, el objetivo de los teoremas enunciados en [3] es mostrar que, de acuerdo a la regla de reasignación 4.1.4, $S_{f, \epsilon}(b, \xi)$ solo tendrá valores dominantes alrededor de $\phi'_k(t)$ lo que nos permite una estimación precisa de $\phi'_k(t)$. La diferencia es que para reconstruir el k -ésimo componente $f_k(t) = A_k(t)\cos(2\pi\phi_k(t))$ de f , se utiliza una fórmula ligeramente diferente a la fórmula de reconstrucción 4.2.1, que si bien resulta equivalente, en la práctica funciona mejor numéricamente. Los estimadores que se utilizan son entonces:

$$\tilde{f}_{k, \epsilon}^{\mathbb{C}}(t) := \mathcal{R}_\psi^{-1} \int_{\frac{1-\Delta}{\phi'_k(t)}}^{\frac{1+\Delta}{\phi'_k(t)}} T_f(a, t) \chi_{|T_f(a, t)| > \epsilon}(a) a^{-3/2} da, \quad \tilde{f}_{k, \epsilon}(t) := \Re(\tilde{f}_{k, \epsilon}^{\mathbb{C}}(t)) \quad (4.4.11)$$

donde $\mathcal{R}_\psi = \int \hat{\psi}(\zeta) \zeta^{-1} d\zeta$ como en 4.2.1, χ es la función indicadora y \Re significa tomar la parte real. El estimador para la amplitud es, igual que antes, tomar el módulo de la

fórmula de reconstrucción

$$\tilde{A}_k(t) := |\tilde{f}_{k,\epsilon}^{\mathbb{C}}(t)| \quad (4.4.12)$$

y se puede obtener un estimador de $\phi_k(t)$ a partir de la fase de la señal compleja $\frac{\tilde{f}_{k,\epsilon}^{\mathbb{C}}(t)}{\tilde{A}_k(t)}$.

Ahora, consideremos que tenemos observaciones a tiempo discreto de $f \in \mathcal{A}_{\epsilon,d}^{c_1,c_2}$, es decir, $f := \{f(n\tau)\}_{n \in \mathbb{Z}}$, donde $\tau > 0$ es el intervalo de muestreo. Si $\psi \in \mathcal{S}$, la CWT de f , $T_f(a, b)$, está bien definida y es igual a $\tau \sum_{m \in \mathbb{Z}} f(m\tau) \frac{1}{a^{1/2}} \psi\left(\frac{m\tau - b}{a}\right)$ y $\partial_b T_f(a, n\tau) = \tau \sum_{m \in \mathbb{Z}} f(m\tau) \frac{1}{a^{3/2}} \psi'\left(\frac{m\tau - n\tau}{a}\right)$. Ésta es simplemente la discretización de 3.2.3. Del mismo modo, tenemos la discretización de 4.1.4 y 4.1.5, dadas por:

$$\omega_f(a, n\tau) := \begin{cases} \frac{-i\partial_b T_f(a, n\tau)}{2\pi T_f(a, n\tau)} & \text{si } |T_f(a, n\tau)| \neq 0 \\ \infty & \text{si } |T_f(a, n\tau)| = 0 \end{cases} \quad (4.4.13)$$

$$S_{f,\epsilon}(n\tau, \xi) := \lim_{\alpha \rightarrow 0} \int_{\{(a, n\tau) : |T_f(a, n\tau)| \geq \epsilon\}} h_\alpha(|\omega_f(a, n\tau) - \xi|) T_f(a, n\tau) a^{-3/2} da \quad (4.4.14)$$

donde $\epsilon > 0$ y $n \in \mathbb{Z}$. Entonces la estimación de f_k, A_k y $\phi_k, k = 1, \dots, K$, sigue inmediatamente, por ejemplo, para $n \in \mathbb{Z}$ tenemos:

$$\tilde{f}_{k,\epsilon}^{n,\mathbb{C}} := \mathcal{R}_\psi^{-1} \int_{\frac{1-\Delta}{\phi_k'(n\tau)}}^{\frac{1+\Delta}{\phi_k'(n\tau)}} T_f(a, n\tau) \chi_{|T_f(a, n\tau)| > \epsilon} a^{-3/2} da, \quad \tilde{f}_{k,\epsilon}^n := \Re(\tilde{f}_{k,\epsilon}^{n,\mathbb{C}}) \quad (4.4.15)$$

Las discusiones anteriores se refieren a los casos en que las observaciones no están contaminadas con ruido y no contienen tendencias. Si observamos Y satisfaciendo el modelo 4.4.9, simplemente reemplazamos f en 4.1.4, 4.1.5 y 4.4.11 por Y , y consideramos el siguiente estimador de tendencia, que es un GRP en general:

$$\tilde{T} := Y - \Re\left(\int_{\frac{1-\Delta}{c_2}}^{\frac{1+\Delta}{c_1}} T_Y(a, b) a^{-3/2} da\right) \quad (4.4.16)$$

Intuitivamente, $\Re\left(\int_{\frac{1-\Delta}{c_2}}^{\frac{1+\Delta}{c_1}} W_Y(a, b) a^{-3/2} da\right)$ aproxima los componentes estacionales, que entonces se cancelarían con los de Y , dejando solo la tendencia. En la sección suplementaria de [3] se ve que efectivamente \tilde{T} aproxima a T en el sentido general de distribución.

En el caso discreto tenemos observaciones $\mathbf{Y} = \{Y_n\}_{n \in \mathbb{Z}}$ del modelo 4.4.10 así que reemplazamos f en 4.4.13, 4.4.14 y 4.4.15 por \mathbf{Y} , y luego reconstruimos la tendencia en el momento $n\tau, n \in \mathbb{Z}$, con el siguiente estimador:

$$\tilde{T}_n := \mathbf{Y}_n - \Re\left(\int_{\frac{1-\Delta}{c_2}}^{\frac{1+\Delta}{c_1}} T_{\mathbf{Y}}(a, n\tau) a^{-3/2} da\right) \quad (4.4.17)$$

4.4.4. Teoremas

Finalmente enunciamos los teoremas presentes en [3] sobre las propiedades teóricas del método de synchrosqueezing descrito en la sección anterior, tanto para el caso

continuo como discreto. Pero antes establecemos algunas hipótesis adicionales sobre el modelo y definimos un poco de notación:

Hipótesis (H1): Supongamos que la wavelet madre $\psi \in \mathcal{S}$ se elije de manera que $\hat{\psi} \subset [1 - \Delta, 1 + \Delta]$, donde $\Delta < d/(1 + d)$, y $\mathcal{R}_\psi = 1$. También supongamos que $T : \mathbb{R} \rightarrow \mathbb{R}$ está en $C^1(\mathbb{R})$ de manera que su transformada de Fourier existe en el sentido de distribución, y $|T(\psi_{a,b})|, |T'(\psi_{a,b})| \leq C_T \epsilon$ para todo $b \in \mathbb{R}$ y $a \in (0, \frac{1+\Delta}{c_1}]$, para algún $C_T \geq 0$.

Hipótesis (H2): Supongamos que el espectro de potencia $d\eta$ del GRP Φ dado satisface $\int (1 + |\xi|)^{-2l} d\eta < \infty$ para algún $l > 0$. También supongamos $\sigma \in C^\infty$ tal que $\|\sigma\|_{L^\infty} \ll 1$ y $\epsilon_\sigma := \max_{\ell=1, \dots, \max\{1, l\}} \left\{ \left\| \sigma^{(\ell)} \right\|_{L^\infty} \right\} \ll 1$, y $\text{var } \Phi(\psi) = 1$

Notación (N1): Denotamos E_0 a la constante universal que depende de los momentos de ψ y ψ' , c_1, c_2 y d . Denotamos E_T a la constante universal que depende de los momentos de ψ y ψ' , C_T, c_1, c_2 y d . Estas constantes están relacionadas con el sesgo del modelo introducido por $\mathcal{A}_{\epsilon, d}^{c_1, c_2}$ y con la tendencia. Denotamos $E_i, i = 1, \dots, 6$, a las constantes que dependen del espectro de potencia de Φ, c_1, c_2, d y los ceros y primeros momentos de $\psi^{(k)}, k = 1, \dots, l + 1$. Estas últimas están relacionadas con el proceso de error y todas se encuentran especificadas en la sección suplementaria de [3].

Ahora sí enunciamos el teorema de robustez de SST cuando la estacionalidad mas la tendencia están contaminadas por un GRP “casi” estacionario. La demostración también puede encontrarse en la sección suplementaria de [3].

Teorema 4.4.3. *Supongamos que $Y(t)$ sigue el modelo 4.4.9 y que se cumplen las hipótesis (H1) y (H2). Entonces, si ϵ es suficientemente chico, para cada $b \in \mathbb{R}$ y $\gamma > 1$ tenemos los siguientes resultados:*

(I) Para cada $a \in [\frac{1-\Delta}{c_2}, \frac{1+\Delta}{c_1}]$, con probabilidad mayor que $1 - \gamma^{-2}$, se tiene

$$\left| T_Y(a, b) - \sum_{l=1}^K A_l(b) e^{i2\pi\phi_l(b)} \sqrt{\hat{a}\hat{\psi}(a\phi'_l(b))} \right| \leq \gamma (E_1\sigma(b) + E_2\epsilon_\sigma) + E_0\epsilon \quad (4.4.18)$$

(II) Para cada $a \in Z_k(b) := [\frac{1-\Delta}{\phi'_k(t)}, \frac{1+\Delta}{\phi'_k(t)}]$ con $|T_f(a, b)| > \gamma (E_3\sigma(b) + E_4\epsilon_\sigma) + (C_T + 1)\epsilon^{1/3}$, donde $k = 1, \dots, K$, con probabilidad mayor que $1 - \gamma^{-2}$, se tiene

$$|\omega_Y(a, b) - \phi'_k(b)| \leq \frac{\gamma (E_3\sigma(b) + E_4\epsilon_\sigma) + E_0\epsilon}{|T_f(a, b)|} \quad (4.4.19)$$

(III) Con probabilidad mayor que $1 - \gamma^{-2}$, para $k = 1, \dots, K$ se tiene

$$\left| \tilde{f}_{k, E_1\sigma(b) + E_2\epsilon_\sigma}^{\mathbb{C}}(b) - A_k(b) e^{2\pi i\phi_k(b)} \right| \leq [\gamma (E_5\sigma(b) + E_6\epsilon_\sigma) + E_0\epsilon] \Delta \quad (4.4.20)$$

(IV) El estimador de tendencia satisface

$$\left| \mathbb{E} \tilde{T}(\varphi_{h,b}) - T(b) \right| \leq 2E_T \|\varphi\|_1 \epsilon \quad (4.4.21)$$

cuando $h > 0$ es suficientemente chico, donde $\varphi \in \mathcal{S}$ y $\varphi_{h,b}(t) := \frac{1}{h}\varphi\left(\frac{b-t}{h}\right) \rightarrow \delta_b$ en distribución cuando $h \rightarrow 0$.

Veamos ahora algunas observaciones. Notemos que cada cota de error en el teorema 4.4.3 consiste de dos términos. El primer término está relacionado al proceso de error $\sigma\Phi$ y su heterocedasticidad modelada por $\sigma(b)$, y el segundo término está relacionado con ϵ , el sesgo del modelo introducido cuando usamos $\mathcal{A}_c^{c_1,c_2}/\mathcal{A}_{c,d}^{c_1,c_2}$ para modelar la estacionalidad. En el primer término, $\sigma(b)$ y σ_ϵ caracterizan respectivamente el nivel de ruido y la “no estacionariedad” del proceso de error. Notemos que cuando $\sigma(b)$ domina a ϵ y ϵ_σ , el error de estimación es del mismo orden que $\sigma(b)$, que es la desviación estándar del proceso de error. En este caso, si tenemos interés en recobrar el proceso de error, se puede elegir un Δ más chico para que el error de estimación sea más chico que $\sigma(b)$ y por lo tanto con alta probabilidad la realización del proceso de error pueda ser aproximada con precisión.

Por último enunciamos el teorema que establece la robustez del método SST en el caso discreto, es decir, cuando los datos provienen del modelo 4.4.10. Para eso, presentamos las siguientes hipótesis adicionales y notación adicional:

Hipótesis (H3): Para la serie de tiempo $\mathbf{Y} = \{Y_n\}_{n \in \mathbb{Z}}$ en el modelo 4.4.10, asumimos $A_k(t) \in C^2(\mathbb{R})$ y $\sup_{t \in \mathbb{R}} |A_k''(t)| \leq \epsilon c_2$ para todo $k = 1, \dots, K$. Estas condiciones de regularidad se agregan para mostrar la interacción entre el sesgo del modelo ϵ y el efecto de la discretización. Suponemos además de (H1) que $T \in C^2$ tal que $|T''(\psi_{a,b})| \leq C_T \epsilon$ para todo $b \in \mathbb{R}$ y $a \in \left(0, \frac{1+\Delta}{c_1}\right]$.

Por otro lado, cuando la frecuencia de muestreo es lo suficientemente alta, el componente de tendencia no es significativo y podemos centrarnos en el análisis de los componentes estacionales. Pero si la frecuencia de muestreo es baja, se empezará a ver el efecto de aliasing y aparecerá un término de error adicional. Por eso, suponemos también que el intervalo de muestreo τ satisface $0 < \tau \leq \frac{1-\Delta}{(1+\Delta)c_2}$ para evitar el efecto de aliasing introducido por la discretización.

Hipótesis (H4): Suponemos $\text{var}(\Phi_n) = 1$ y $\sigma \in C^2$ para que $\|\sigma\|_{L^\infty} \ll 1$, $\epsilon_\sigma := \max\left\{\|\sigma^{(1)}\|_{L^\infty}, \|\sigma^{(2)}\|_{L^\infty}\right\} \ll 1$

Notación (N2): Llamamos $E_{\tau,0}$ a la constante universal dependiente de τ , los momentos de ψ y ψ' , c_1, c_2 y d . Llamamos $E_{T,0}$ a la constante universal dependiente de τ , los momentos de ψ y ψ' , C_T, c_1, c_2 y d . Estas constantes están relacionadas con la tendencia y el sesgo del modelo introducido por la clase $\mathcal{A}_{\epsilon,d}^{c_1,c_2}$ y están influenciadas por el intervalo de sampleo τ . Sean $E_{\tau,i}, i = 1, \dots, 6, E_{T,1}$ y $E_{T,2}$ constantes que dependen de τ , el proceso de error dado Φ, c_1, c_2, d y los zeros y los primeros momentos de ψ y ψ' .

Teorema 4.4.4. *Supongamos que $\mathbf{Y} = \{Y_n\}_{n \in \mathbb{Z}}$ sigue el modelo 4.4.10 y que se cumplen las hipótesis (H1), (H3) y (H4). Entonces, si ϵ es suficientemente chico, para cada $n \in \mathbb{Z}$ y $\gamma > 1$ tenemos los siguientes resultados:*

(I) Para cada $a \in \left[\frac{1-\Delta}{c_2}, \frac{1+\Delta}{c_1}\right]$, con probabilidad mayor que $1 - \gamma^{-2}$, se tiene

$$\left|T_Y(a, n\tau) - \sum_{l=1}^K A_l(n\tau) e^{i2\pi\phi_l(n\tau)} \sqrt{a} \hat{\psi}(a\phi'_l(b))\right| \leq \gamma \left(E_{\tau,1}\sigma(n\tau) + \tau^2 E_{\tau,2}\epsilon_\sigma\right) + E_{\tau,0}\epsilon$$

(II) Para cada $a \in Z_k(n\tau) := \left[\frac{1-\Delta}{\phi'_k(n\tau)}, \frac{1+\Delta}{\phi_k(n\tau)} \right]$ con $|T_f(a, n\tau)| > \gamma (E_{\tau,3}\sigma(n\tau) + \tau^2 E_{\tau,4}\epsilon_\sigma) + (C_T + 1)\epsilon^{1/3}$, donde $k = 1, \dots, K$, con probabilidad mayor que $1 - \gamma^{-2}$, se tiene

$$|\omega_Y(a, n\tau) - \phi'_k(n\tau)| \leq \frac{\gamma (E_{\tau,3}\sigma(n\tau) + E_{\tau,4}\epsilon_\sigma) + E_{\tau,0}\epsilon}{|T_f(a, n\tau)|}$$

(III) Con probabilidad mayor que $1 - \gamma^{-2}$, para $k = 1, \dots, K$ se tiene

$$\left| \tilde{f}_{k, E_{\tau,1}\sigma(n\tau) + \tau^2 E_{\tau,2}\epsilon_\sigma}^{n, \mathbb{C}} - A_k(n\tau) e^{2\pi i \phi_k(n\tau)} \right| \leq [\gamma (E_{\tau,5}\sigma(n\tau) + E_{\tau,6}\epsilon_\sigma) + E_{\tau,0}\epsilon] \Delta$$

(IV) Con probabilidad mayor que $1 - \gamma^{-2}$, se tiene

$$\left| \tilde{T}_n - T(n\tau) \right| \leq \gamma (E_{T,1}\sigma(n\tau) + E_{T,2}\epsilon_\sigma) + E_{T,0}\epsilon$$

En este caso, siguen valiendo las mismas observaciones que para el teorema 4.4.3. La demostración también se encuentra en la sección suplementaria de [3] y es esencialmente la misma que en el caso continuo, con algunas estimaciones adicionales específicas a la discretización para analizar el error introducido por el hecho de tener muestreo finito.

De esta forma, con el modelo no paramétrico introducido en [3], junto con la técnica adaptativa de análisis de tiempo-frecuencia SST podemos analizar series de tiempo con comportamiento estacional dinámico y tendencia suave, contaminadas por series de ruido heterocedásticas y dependientes. Además de los resultados de identificabilidad para el modelo de estacionalidad no paramétrico, tenemos resultados teóricos para justificar y cuantificar la capacidad de la SST para extraer el componente estacional y la tendencia de las observaciones ruidosas. En el próximo capítulo haremos un estudio numérico, probando la SST en datos simulados y reales y comparándola con los métodos estudiados en el capítulo 2.

Capítulo 5

Análisis de simulaciones y datos reales

5.1. Datos simulados

Para demostrar la capacidad de la transformada synchrosqueezing para detectar la estacionalidad dinámica y otras propiedades discutidas en el capítulo anterior, vamos a realizar una serie de simulaciones y comparar los resultados con los del modelo TBATS y EMD. Comparamos SST con EMD solo en el primer ejemplo de señal sin ruido, ya que como vimos, EMD es bastante inestable cuando la señal está contaminada. Por otro lado, como TBATS no está diseñado para estacionalidad dinámica, en el segundo ejemplo, la estacionalidad se compone de funciones trigonométricas puras (no varían en el tiempo). En el tercer escenario de simulación, sí consideramos a la estacionalidad como modelada por $\mathcal{A}_{\epsilon,d}^{c_1,c_2}$ y ahí se puede observar la principal diferencia entre SST y TBATS. Utilizamos los paquetes de R “EMD”, “forecast” y “SynchWave” donde están implementados los métodos de EMD, TBATS y SST, respectivamente. Las simulaciones se corrieron en una notebook Lenovo con 8GB de RAM, procesador Intel Core i5-7200U CPU @ 2.50GHz y sistema operativo Windows 10 de 64 bits. Los algoritmos fueron implementados en R versión 4.1.1 y el código está disponible en [GitHub](#).

5.1.1. Configuración de la simulación

Vamos a implementar las simulaciones presentes en [3]. Para eso, definimos dos funciones para modelar la estacionalidad:

$$\begin{aligned} s_{1,1}(t) &:= 2,5 \cos(2\pi t), \quad s_{1,2} := 3 \cos(2\pi^2 t) \\ s_1(t) &:= s_{1,1}(t) + s_{1,2}(t) \\ A_1(t) &:= (1+0,1 \cos(t)) \arctan\left(\frac{1132-200t}{87}\right) / 2+2, \quad A_2(t) := 3,5\chi_{[0, 7,5]}(t)+2\chi_{(7,5, 10]}(t) \\ \phi_1(t) &:= t + 0,1 \sin(t), \quad \phi_2(t) := 3,4t - 0,02|t|^{2,3} \\ s_{2,1}(t) &:= A_1(t) \cos(2\pi\phi_1(t)), \quad s_{2,2} := A_2(t) \cos(2\pi\phi_2(t)) \end{aligned} \tag{5.1.1}$$

$$s_2(t) := s_{2,1}(t) + s_{2,2}(t) \quad (5.1.2)$$

donde χ es la función indicadora. Notar que s_1 se compone de dos funciones armónicas y sus frecuencias no son múltiplos enteros entre sí, y s_2 modela una estacionalidad con comportamiento variable en el tiempo. Por definición, s_2 se compone de dos IMFs con frecuencias instantáneas $\phi'_1(t) = 1 + 0,1 \cos(t)$ y $\phi'_2(t) = 3,4 - 0,046t^{1,3}$. Consideramos además, dos posibles tendencias:

$$T_1(t) := 8 \left(\frac{1}{1 + (t/5)^2} + \exp(-t/10) \right), \quad T_2(t) := 2t + 10 \exp \left(-(t - 4)^2/6 \right)$$

Podemos ver el gráfico de estas funciones en la figura 5.1

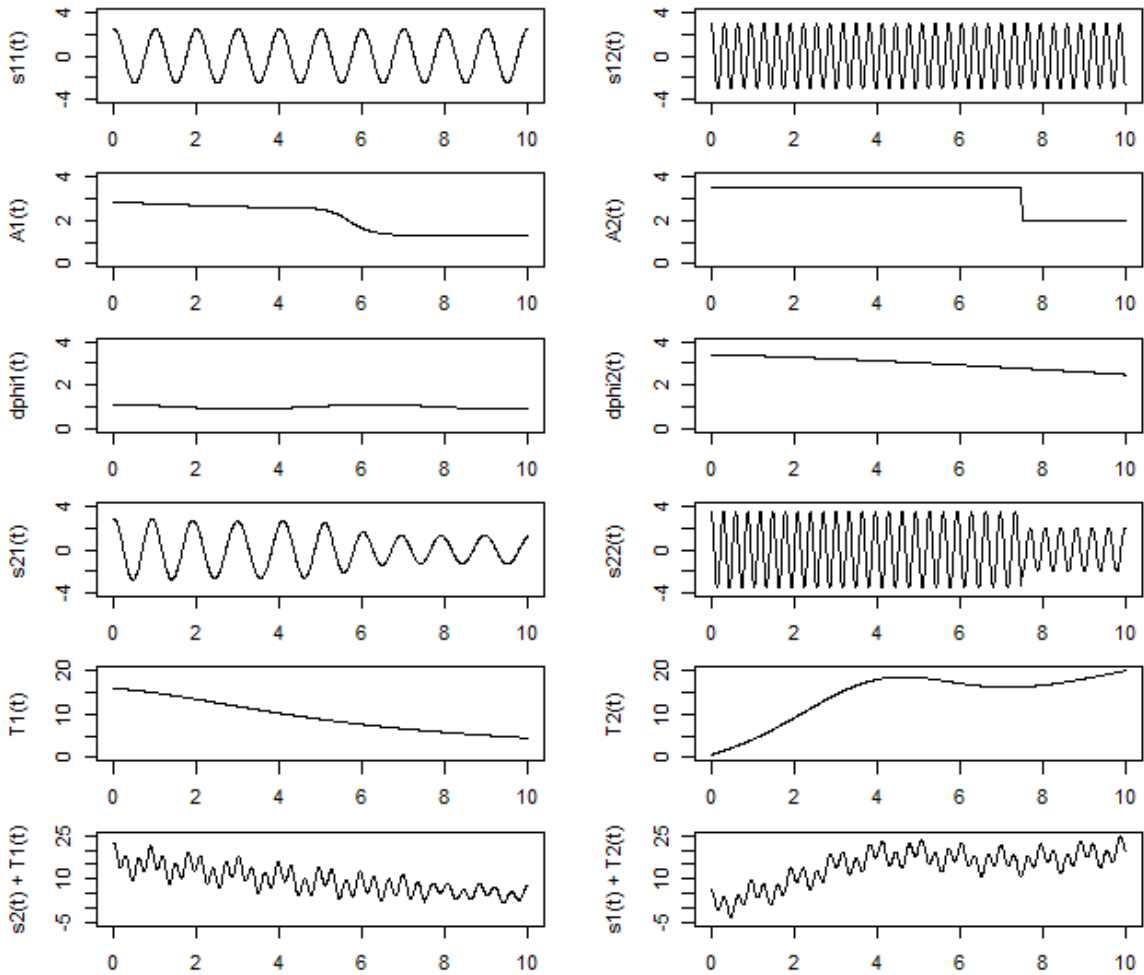


Figura 5.1: **Columna izquierda:** De arriba hacia abajo, gráficos de las funciones $s_{1,1}(t)$, $A_1(t)$, $\phi'_1(t)$, $s_{2,1}(t)$, $T_1(t)$, $s_2(t) + T_1(t)$. **Columna derecha:** De arriba hacia abajo, gráficos de las funciones $s_{1,2}(t)$, $A_2(t)$, $\phi'_2(t)$, $s_{2,2}(t)$, $T_2(t)$, $s_1(t) + T_2(t)$. Observar que las funciones de amplitud y frecuencia instantánea no son constantes.

Discretizamos estas funciones en el período de tiempo $[0, 10]$ con intervalo de muestreo de longitud $\tau = 1/100$, con lo que resultan $N = 1001$ puntos de muestreo. La serie de tiempo muestreada de una función f definida en \mathbb{R} la denotamos como $\mathbf{f} \in \mathbb{R}^N$ de manera que la l -ésima coordenada de \mathbf{f} es $f(l\tau)$ para todo $l = 1, \dots, N$. Así, las series de tiempo muestreadas de s_1, s_2, T_1 y T_2 se indican como $\mathbf{s}_1, \mathbf{s}_2, \mathbf{T}_1$ y \mathbf{T}_2 , respectivamente.

Consideramos también tres procesos aleatorios para modelar el ruido:

$$X_1(n) := 2\sigma(n\tau)X_{\text{ARMA1}}(n)$$

donde $\sigma(t) = 1 + 0,1 \cos(\pi t)$, y X_{ARMA1} es un proceso ARMA(1, 1) determinado por el polinomio de autorregresión $a(z) = 0,5z + 1$ y el polinomio de medias móviles $b(z) = 0,4z + 1$, con el proceso de innovaciones tomado como variables aleatorias i.i.d. student t_4 .

$$X_2(n) := \sigma(n\tau) \left(4X_{\text{ARMA1}}(n)\chi_{n \in [1, N/2]}(n) + X_{\text{ARMA2}}(n)\chi_{n \in [N/2+1, N]}(n) \right)$$

donde X_{ARMA2} es un proceso ARMA (1, 1) determinado por el polinomio de autorregresión $a(z) = -0,2z + 1$ y el polinomio de medias móviles $b(z) = 0,51z + 1$, con el proceso de innovaciones tomado como variables aleatorias i.i.d. student t_4 .

$$X_3(n) := 2X_{\text{GARCH}}(n)$$

donde X_{GARCH} es un proceso GARCH(2, 1) con coeficiente ARCH (0,2), coeficientes GARCH (0,2, 0,3), coeficiente constante (1) y perturbaciones $N(0, 1)$.

La serie muestreada de X_i se indica como $\mathbf{X}_i \in \mathbb{R}^N$. Notar que X_1 y X_2 son heterocedásticos y no estacionarios. La idea es probar nuestros algoritmos en algunas de las siguientes series de tiempo:

$$\mathbf{Y}_{i,j,k,\sigma_0} := \mathbf{s}_i + \mathbf{T}_j + \sigma_0 \mathbf{X}_k$$

donde $i = 1, 2$, $j = 1, 2$, $k = 1, 2, 3$ y $\sigma_0 \geq 0$ y medir el RMSE (raíz del error cuadrático medio) entre el verdadero valor y la reconstrucción de cada componente.

Recordemos que para calcular el RMSE entre la variable y y el estimador \hat{y} , observados en los tiempos $t = 1, \dots, T$, se utiliza la fórmula:

$$\text{RMSE}(y, \hat{y}) = \sqrt{\frac{\sum_{t=1}^T (\hat{y}_t - y_t)^2}{T}}$$

5.1.2. Elección de parámetros

El método de synchrosqueezing requiere previamente la elección de ciertos parámetros. Por un lado, debemos elegir una wavelet madre. Las que están implementadas en el paquete “SynchWave” de R y son aptas para synchrosqueezing son: “bump”, “gauss”, “cmhat” (sombrero mejicano complejo), “morlet” y “hhhat” (función analítica de Hilbert del sombrero hermitiano). Se probó aplicar SST a la serie $\mathbf{Y}_{1,1,1,1}$ con cada una de estas wavelets y con la que se obtuvo el menor RMSE en la reconstrucción de $s_{1,1}, s_{1,2}, T_1$ y X_1 fue con “hhhat”, por lo tanto, ésta es la que se utilizará en el resto de la simulaciones, con su parámetro μ por defecto ($\mu = 5$).

También debemos fijar K , el número de componentes estacionales. En nuestro caso, como es una simulación y por lo tanto conocemos cuales son las funciones que generaron los datos, sabemos que tanto s_1 como s_2 tienen dos componentes estacionales ($s_{1,1}, s_{1,2}$ y $s_{2,1}, s_{2,2}$, respectivamente). Pero en general, sobre todo si trabajamos con datos reales, uno no sabe exactamente el valor de K . Por eso, en [3] se propone el siguiente procedimiento para encontrarlo: Primero, se testea la hipótesis nula de que el proceso observado es estacionario, modulado por la varianza heterocedástica y la tendencia. Si no se rechaza la hipótesis nula, se puede concluir que no hay estacionariedad en la señal. De lo contrario, comenzando con $k = 1$, determinamos visualmente k componentes en el plano de tiempo-frecuencia obtenido con SST, reconstruimos los componentes estacionales y la tendencia, y luego testeamos si los residuos (obtenidos restando de la señal los k componentes estacionales reconstruidos y la tendencia) son estacionarios, modulados por la varianza heterocedástica. Si no se rechaza la hipótesis nula de estacionariedad, nos detenemos y decidimos que K es igual al valor actual de k ; de lo contrario, aumentamos el valor de k en 1, y luego buscamos otro componente estacional en el plano de tiempo-frecuencia y repetimos el procedimiento de estimación y testeo anterior. Con este enfoque iterativo, se puede determinar K .

Otros parámetros que recibe la función que calcula la transformada synchrosqueezing en R son Γ (ϵ en la fórmula 4.1.5), que representa el valor de umbral para la transformada wavelet que determina el nivel de filtro, y n_v , el número de voces, parámetro que afecta el número de escalas con el que trabajaremos. En nuestro caso utilizaremos el parámetro por defecto para Γ y $n_v = 32$. Se pueden encontrar más detalles sobre la implementación de synchrosqueezing y la elección de sus parámetros en [16].

Por último, está el problema de los efectos de frontera. En la práctica, los conjuntos de datos experimentales son finitos y la consecuencia obvia del análisis con wavelets de un conjunto de datos finitos es que, a medida que la wavelet se acerca al borde de los datos, parte de la misma comienza a “salirse” del borde. Esto crea un efecto de frontera, donde los valores de la transformación cercanos al límite de la señal están contaminados por la naturaleza discontinua del borde de la señal. Muchas veces uno desea analizar la totalidad de los datos disponibles, por lo que se han desarrollado varios métodos para lidiar con los límites de las señales de extensión finita, por ejemplo: añadir una línea de valores cero o constantes en cada extremo de la señal (relleno); continuar la señal desde el último punto hasta el primer punto (periodización); reflejar la señal en los bordes (reflexión); utilizar una extrapolación polinomial de la señal en cada extremo (ajuste polinómico), entre otros. Sin embargo, si nos enfocamos en un segmento pequeño de una señal mucho más grande disponible para nosotros, simplemente podemos usar los datos conocidos fuera del segmento en consideración. En nuestro caso, haremos esto último, ya que, si bien queremos realizar el análisis en $[0, 10]$, conocemos las funciones que generan los datos y por lo tanto podemos generar la señal más allá de ese intervalo. Entonces, vamos a generar datos en $[-4, 14]$, realizar la transformada synchrosqueezing en ese intervalo y luego “recortar” las reconstrucciones de los componentes, quedándonos solo con los valores correspondientes al $[0, 10]$ para calcular el RMSE. De esta forma logramos evitar en gran parte el error introducido por los efectos de borde, sobre todo para la estimación de la tendencia, como se puede observar en un ejemplo en la figura 5.2.

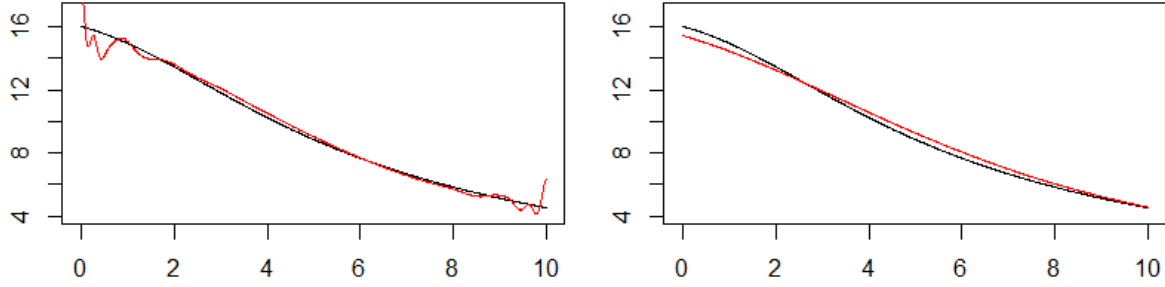


Figura 5.2: En negro el verdadero valor de la tendencia T_1 . A la izquierda, en rojo, la tendencia reconstruida de la señal $s_2 + T_1$ sin extenderla y a la derecha, en rojo, la tendencia reconstruida de la misma señal calculando la SST entre -4 y 14 y quedándonos con la reconstrucción entre 0 y 10 .

5.1.3. Experimentos

En R, para calcular la transformada synchrosqueezing de una señal utilizamos la función “synsq_cwt_fw” y para extraer y reconstruir los componentes estacionales a partir de la misma, utilizamos las funciones “curve_ext_multi” y “curve_ext_recon”. La primera extrae consecutivamente las curvas de máxima energía/mínima curvatura de la representación de SST. A medida que se extraen las curvas, sus energías asociadas se reemplazan por cero en el plano y a continuación, se vuelve a realizar la extracción de curvas en la representación restante. La segunda función reconstruye estas curvas obtenidas. Así se obtienen los estimadores $\widetilde{s}_{1,1}$, $\widetilde{s}_{1,2}$, $\widetilde{s}_{2,1}$, $\widetilde{s}_{2,2}$. Para estimar la tendencia, usamos que la transformada wavelet ignora la misma, entonces, dada una señal \mathbf{Y} , llamamos \mathbf{Y}' al resultado de aplicar la CWT a \mathbf{Y} e invertirla (con las funciones “cwt_fw” y “cwt_iw”), luego usamos el estimador $\widetilde{T} = \mathbf{Y} - \mathbf{Y}'$ para \widetilde{T}_1 o \widetilde{T}_2 , según corresponda. Por último, para reconstruir el ruido, simplemente utilizamos los residuos, es decir, restamos a la señal original los componentes estacionales y la tendencia estimados $\widetilde{r} = \mathbf{Y} - \widetilde{s} - \widetilde{T}$, donde \widetilde{s} es \widetilde{s}_1 o \widetilde{s}_2 y \widetilde{T} es \widetilde{T}_1 o \widetilde{T}_2 , según corresponda.

En el caso de EMD, utilizamos la función “emd”, que calcula los IMFs y el residuo. Determinamos visualmente cuál IMF se corresponde mejor con cada uno de los componentes estacionales que sabemos que tenemos, si es que se corresponde con alguno, y usamos el residuo como estimador para la tendencia.

Para TBATS, usamos la función “tbats”, que calcula los componentes estacionales, el componente “level”, que utilizamos como estimador de la tendencia, y los residuos, que utilizamos como estimador del ruido.

Señal sin ruido

Primero vamos a considerar la señal sin ruido $s_2 + T_1$. Vamos a comparar la precisión de los métodos STT y EMD para reconstruir los dos componentes estacionales de s_2 y la tendencia T_1 . En la figura 5.3 se pueden observar la señal y sus componentes.

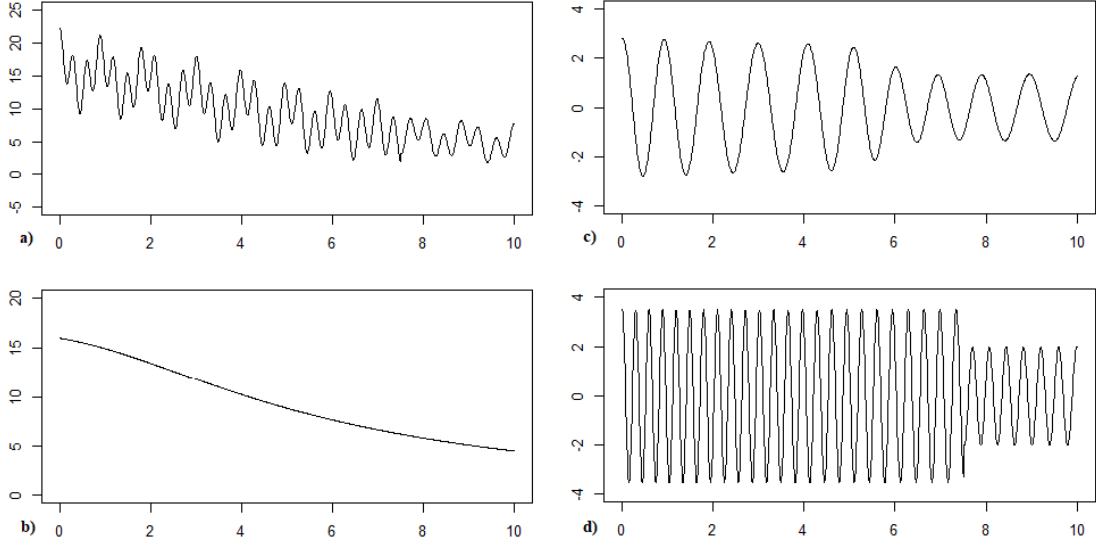


Figura 5.3: Gráficos de las funciones **a)** $s_2 + T_1$ **b)** T_1 **c)** $s_{2,1}$ **d)** $s_{2,2}$.

En la figura 5.4 se encuentran los resultados de aplicar la transformada synchrosqueezing a la señal $s_2 + T_1$ y la reconstrucción de sus componentes con este método. En la figura 5.5 se puede ver el resultado de aplicar EMD a la misma señal.

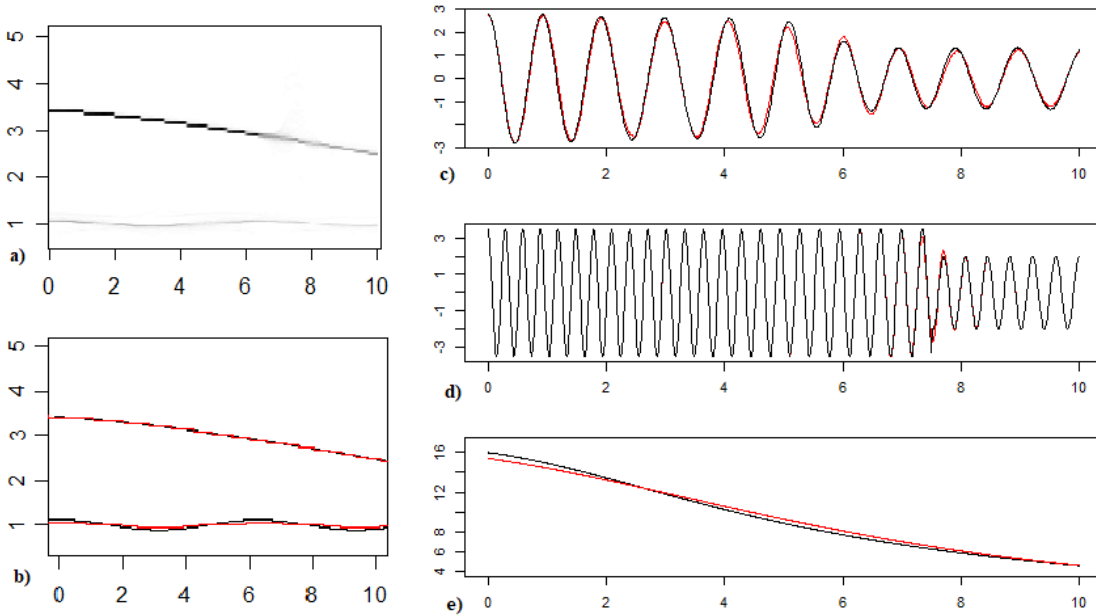


Figura 5.4: **a)** Gráfico de la transformada synchrosqueezing aplicada a la señal $s_2 + T_1$. El eje y es la frecuencia y está en escala lineal. Se utilizó una escala de grises, en blanco los puntos del plano donde la transformada es 0 y en negro los de mayor módulo. **b)** En negro el verdadero valor de las frecuencias instantáneas ϕ'_1 y ϕ'_2 y en rojo el extraído de la transformada en a). **c)** En negro $s_{2,1}$ y en rojo su reconstrucción $\widetilde{s}_{2,1}$. **d)** En negro $s_{2,2}$ y en rojo su reconstrucción $\widetilde{s}_{2,2}$. **e)** En negro T_1 y en rojo su reconstrucción \widetilde{T}_1 .

Como podemos observar en la figura 5.4a las funciones de frecuencia instantánea de $s_{2,1}$ y $s_{2,2}$ se ven claramente en las curvas dominantes de la representación de SST y a menor amplitud, menor es la intensidad de la curva. En cuanto a la reconstrucción de las componentes estacionales y la tendencia, se puede ver que, excepto para $s_{2,2}$ cerca del punto de cambio de amplitud 7,5, la reconstrucción es bastante buena.

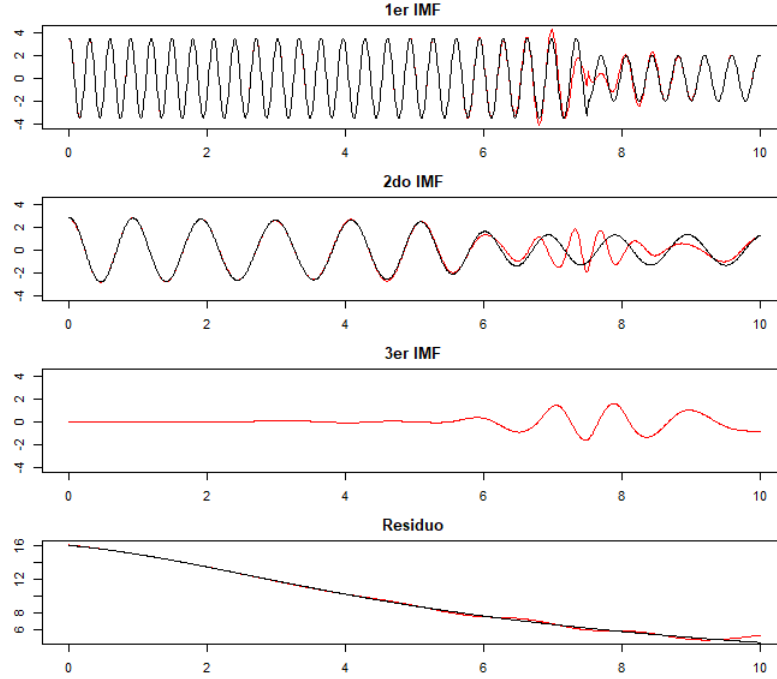


Figura 5.5: Descomposición de $s_2 + T_1$ por medio de EMD. En negro los verdaderos valores de $s_{2,2}$, $s_{2,1}$ y T_1 (de arriba hacia abajo) y superpuestos en rojo los valores de los modos y el residuo que encuentra el algoritmo. Vemos que si bien había dos componentes estacionales, el algoritmo encuentra tres IMFs.

En el caso de EMD, la reconstrucción es muy buena en las partes de la señal donde la frecuencia instantánea y la amplitud varían lentamente, pero en las zonas cercanas a los puntos donde se producen cambios grandes de amplitud, EMD no lo resuelve tan bien como SST, produciéndose una mezcla de modas. Entonces, a pesar de no requerir parámetros previos, sí requiere un mayor trabajo posterior de interpretación de los resultados, debido a que la cantidad de IMF's puede variar y pueden estar mezclados. SST requiere un estudio previo de los argumentos a utilizar, pero el resultado es más directo.

Para medir la precisión de la estimación de ambos métodos se calculó el RMSE entre cada componente y su reconstrucción. Los resultados se ven en el cuadro 5.1. En concordancia con los gráficos, SST obtuvo un menor RMSE para $\widetilde{s_{2,1}}$ y $\widetilde{s_{2,2}}$.

Método	$\widetilde{s_{2,1}}$	$\widetilde{s_{2,2}}$	$\widetilde{T_1}$
SST	0,170	0,084	0,325
EMD	0,662	0,418	0,155

Cuadro 5.1: Resultados de SST y EMD para la señal $s_2 + T_1$.

Señal con estacionalidad no dinámica

Para este segundo experimento, consideramos la señal $\mathbf{Y}_{1,1,1,1} := \mathbf{s}_1 + \mathbf{T}_1 + \mathbf{X}_1$. Vamos a generar 101 realizaciones de esta señal (lo que va a cambiar en cada realización es el proceso aleatorio con el que se simula el ruido \mathbf{X}_1) y analizarlas con SST y TBATS. Recordemos que TBATS necesita como parámetro los períodos de los componentes estacionales, no puede inferirlos de la data, con lo cual, cuando lo corremos, le pasamos los verdaderos valores de estos períodos, 100 y $100/\pi$. Como en el experimento anterior, se calculó la raíz del error cuadrático medio (RMSE) entre cada componente y su reconstrucción, para cada realización, para ambos métodos. En el cuadro 5.2 se pueden ver el promedio de los resultados para cada estimador y su desviación estándar, junto con el tiempo promedio de cada método, en segundos, y su desviación estándar. En la figura 5.6 se muestran los resultados de la descomposición hecha por cada método para la realización que tuvo como RMSE la mediana entre todas las realizaciones.

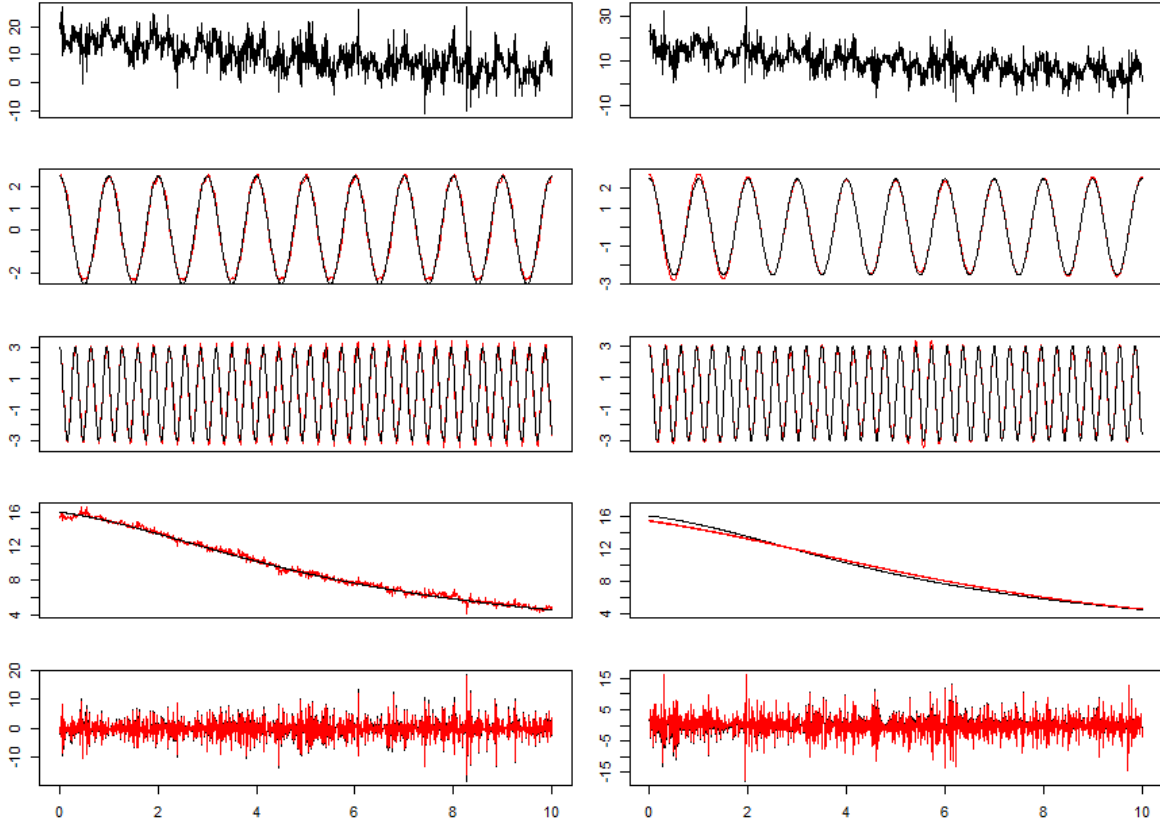


Figura 5.6: Resultado de aplicar TBATS (izquierda) y SST (derecha) a la señal $\mathbf{s}_1 + \mathbf{T}_1 + \mathbf{X}_1$. Se muestran las descomposiciones obtenidas para la realización de la señal que tuvo como error la mediana de todos los errores en cada método. La primera fila es la realización de la señal y las otras filas son, de arriba hacia abajo, los verdaderos valores de $\mathbf{s}_{1,1}$, $\mathbf{s}_{1,2}$, \mathbf{T}_1 y \mathbf{X}_1 en negro y los estimadores encontrados por cada método en rojo.

Método	$\widetilde{s}_{1,1}$	$\widetilde{s}_{1,2}$	\widetilde{T}_1	\widetilde{r}	Tiempo
SST	$0,103 \pm 0,023$	$0,181 \pm 0,024$	$0,327 \pm 0,020$	$0,391 \pm 0,024$	$2,761 \pm 0,223$
TBATS	$0,140 \pm 0,036$	$0,204 \pm 0,068$	$0,431 \pm 0,655$	$0,518 \pm 0,649$	$7,714 \pm 1,819$

Cuadro 5.2: Resultados de SST y TBATS para la señal $\mathbf{s}_1 + \mathbf{T}_1 + \mathbf{X}_1$.

Se puede ver que ambos métodos reconstruyen razonablemente bien los distintos componentes de la señal. Sin embargo, remarcamos que los componentes estacionales son puramente armónicos y además se usaron los verdaderos valores de los períodos cuando se corrió TBATS. Estas son condiciones que no resultan fáciles de cumplir cuando se trabaja con datos reales y además el estimador que se obtuvo para la tendencia contiene parte del ruido, con lo cual probablemente convendría suavizarlo antes de usarlo en alguna aplicación. Por otro lado, SST no solo no requiere de esta información si no que además el RMSE y sobre todo su desviación estándar son menores que las obtenidas con TBATS. Además, con las implementaciones que estamos utilizando, synchrosqueezing resulta una tres veces más rápido.

Señales con estacionalidad dinámica

En esta simulación también vamos a generar 101 realizaciones de algunas de las señales $\mathbf{Y}_{2,j,k,\sigma_0} := \mathbf{s}_2 + \mathbf{T}_j + \sigma_0 \mathbf{X}_k$ y las analizaremos con SST y TBATS. Notemos que estas señales tienen dos componentes estacionales dinámicos y tendencia, contaminados por diferentes procesos de error con diferentes intensidades de ruido. En la figura 5.7 se puede ver el resultado de aplicar SST a una realización de $\mathbf{Y}_{2,1,2,1}$.

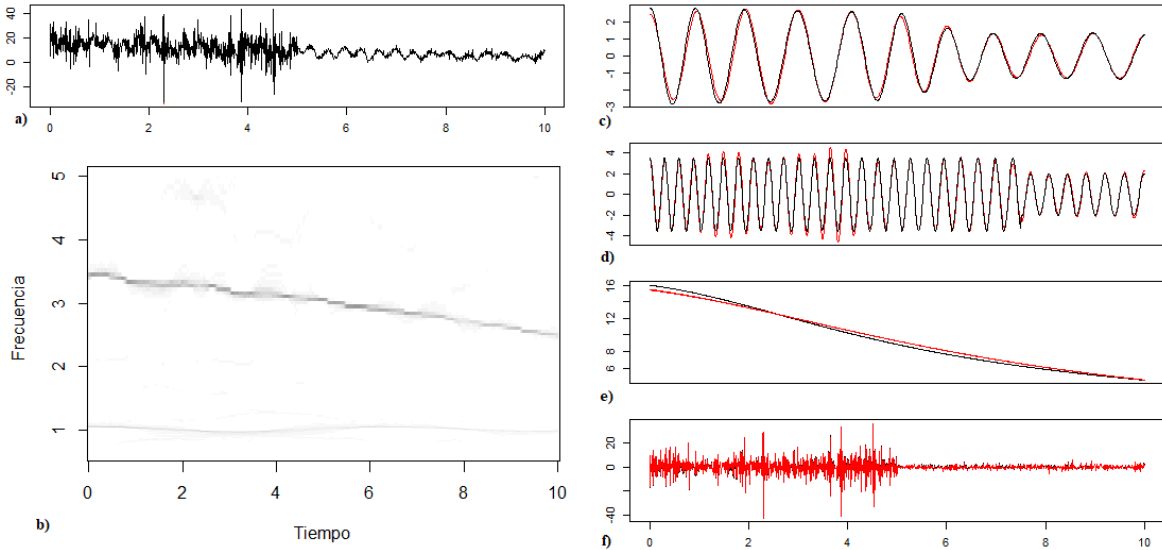


Figura 5.7: **a)** Realización de la señal $\mathbf{Y}_{2,1,2,1} := \mathbf{s}_2 + \mathbf{T}_1 + \mathbf{X}_2$ que tuvo como error de reconstrucción la mediana de todos los errores. **b)** Gráfico de la transformada synchrosqueezing aplicada a la señal en a). **c)** En negro el verdadero valor de $\mathbf{s}_{2,1}$ y en rojo su reconstrucción $\widetilde{\mathbf{s}}_{2,1}$. **d)** En negro $\mathbf{s}_{2,2}$ y en rojo su reconstrucción $\widetilde{\mathbf{s}}_{2,2}$. **e)** En negro \mathbf{T}_1 y en rojo su reconstrucción $\widetilde{\mathbf{T}}_1$. **f)** En negro \mathbf{X}_2 y en rojo su reconstrucción $\widetilde{\mathbf{X}}_2$.

En el cuadro 5.3 se puede ver el rendimiento de SST. Se muestra la media del RMSE para cada componente y el tiempo de cómputo promedio, en segundos, junto con sus desviaciones estándar para distintos valores de j , k y σ_0

(j, k, σ_0)	$\widetilde{s}_{1,1}$	$\widetilde{s}_{1,2}$	\widetilde{T}_1	\widetilde{r}	Tiempo
(1, 2, 0.5)	$0,206 \pm 0,018$	$0,164 \pm 0,020$	$0,327 \pm 0,015$	$2,960 \pm 0,199$	$2,378 \pm 0,270$
(2, 2, 0.5)	$0,205 \pm 0,018$	$0,164 \pm 0,020$	$2,124 \pm 0,016$	$3,629 \pm 0,166$	$2,945 \pm 0,494$
(1, 2, 1)	$0,241 \pm 0,035$	$0,279 \pm 0,043$	$0,333 \pm 0,029$	$0,499 \pm 0,037$	$2,765 \pm 0,262$
(2, 2, 1)	$0,241 \pm 0,035$	$0,279 \pm 0,043$	$2,124 \pm 0,031$	$2,156 \pm 0,030$	$2,661 \pm 0,169$
(1, 3, 0.5)	$0,253 \pm 0,042$	$0,279 \pm 0,032$	$0,332 \pm 0,029$	$1,967 \pm 0,078$	$2,482 \pm 0,085$
(2, 3, 0.5)	$0,253 \pm 0,042$	$0,279 \pm 0,032$	$2,125 \pm 0,029$	$2,877 \pm 0,083$	$2,758 \pm 0,146$
(1, 3, 1)	$0,368 \pm 0,076$	$0,504 \pm 0,064$	$0,348 \pm 0,058$	$0,719 \pm 0,069$	$2,824 \pm 0,128$
(2, 3, 1)	$0,368 \pm 0,076$	$0,504 \pm 0,064$	$2,128 \pm 0,058$	$2,220 \pm 0,059$	$2,411 \pm 0,120$

Cuadro 5.3: Resultados de aplicar SST a señales de la forma $\mathbf{Y}_{2,j,k,\sigma_0} := \mathbf{s}_2 + \mathbf{T}_j + \sigma_0 \mathbf{X}_k$.

Del cuadro 5.3 podemos concluir que se obtienen mejores resultados cuando el nivel de ruido σ_0 es menor y que si bien las estimaciones de las componentes estacionales en general son muy buenas, la forma en que implementamos el estimador de la tendencia no funcionaba tan bien para \mathbf{T}_2 . Notemos también que cuando σ_0 disminuye, el RMSE promedio y la desviación estándar del residuo $\widetilde{r} = \mathbf{Y}_{2,j,k,\sigma_0} - \widetilde{\mathbf{s}}_{2,1} - \widetilde{\mathbf{s}}_{2,2} - \mathbf{T}_j$, que aproxima $\sigma_0 \mathbf{X}_k$, aumentan, debido a la existencia del sesgo del modelo ϵ , que se mantiene fijo, como se indica en el teorema 4.4.4 y debido al ruido, el rendimiento de SST para estimar la amplitud de las componentes estacionales no es tan bueno como al estimar la frecuencia instantánea. De todos modos, la oscilación no está distorsionada, como sí ocurre en los otros métodos y se ve en el cuadro 5.3 que la estimación de la estacionalidad no cambia cuando cambia la tendencia y tampoco varía mucho con el ruido. Podemos decir entonces que los estimadores son robustos a diferentes tipos de errores, aunque las métricas empeoren levemente para las componentes estacionales si comparamos con las obtenidas en la señal sin ruido $\mathbf{Y}_{2,1,k,0}$ (cuadro 5.1).

Por otro lado, en el cuadro 5.4 se puede ver el rendimiento de TBATS. De nuevo, se muestra la media del RMSE para cada componente y el tiempo de cómputo promedio junto con sus desviaciones estándar para distintos valores de j , k y σ_0 . Se agrega además una columna donde se muestra la cantidad de veces, entre las 101 iteraciones, que el algoritmo falló, es decir, que no pudo encontrar una solución. En la figura 5.8 se puede ver el resultado de aplicar TBATS a una realización de $\mathbf{Y}_{2,1,2,1}$ y de $\mathbf{Y}_{2,2,3,0,5}$.

(j, k, σ_0)	$\widetilde{s}_{1,1}$	$\widetilde{s}_{1,2}$	\widetilde{T}_1	\widetilde{r}	Tiempo	Fallas
(1, 2, 0.5)	$1,393 \pm 0,072$	$2,098 \pm 0,016$	$1,187 \pm 0,851$	$3,726 \pm 0,796$	$8,926 \pm 2,690$	1
(2, 2, 0.5)	$1,412 \pm 0,048$	$2,096 \pm 0,023$	$1,363 \pm 0,615$	$3,738 \pm 0,598$	$8,876 \pm 2,862$	1
(1, 2, 1)	$1,411 \pm 0,046$	$2,116 \pm 0,022$	$0,865 \pm 0,199$	$1,897 \pm 0,147$	$10,447 \pm 3,394$	3
(2, 2, 1)	$1,411 \pm 0,045$	$2,112 \pm 0,019$	$1,293 \pm 1,379$	$1,433 \pm 1,303$	$9,618 \pm 2,615$	3
(1, 3, 0.5)	$1,443 \pm 0,078$	$2,047 \pm 0,187$	$1,361 \pm 0,550$	$2,671 \pm 0,188$	$7,477 \pm 2,994$	4
(2, 3, 0.5)	$1,443 \pm 0,061$	$2,061 \pm 0,161$	$1,575 \pm 0,449$	$2,691 \pm 0,177$	$8,441 \pm 2,989$	2
(1, 3, 1)	$1,509 \pm 0,119$	$2,103 \pm 0,124$	$1,608 \pm 0,556$	$1,775 \pm 0,210$	$5,767 \pm 1,756$	1
(2, 3, 1)	$1,517 \pm 0,131$	$2,103 \pm 0,141$	$1,713 \pm 0,439$	$1,782 \pm 0,214$	$5,716 \pm 2,098$	0

Cuadro 5.4: Resultados de aplicar TBATS a señales $\mathbf{Y}_{2,j,k,\sigma_0} := \mathbf{s}_2 + \mathbf{T}_j + \sigma_0 \mathbf{X}_k$.

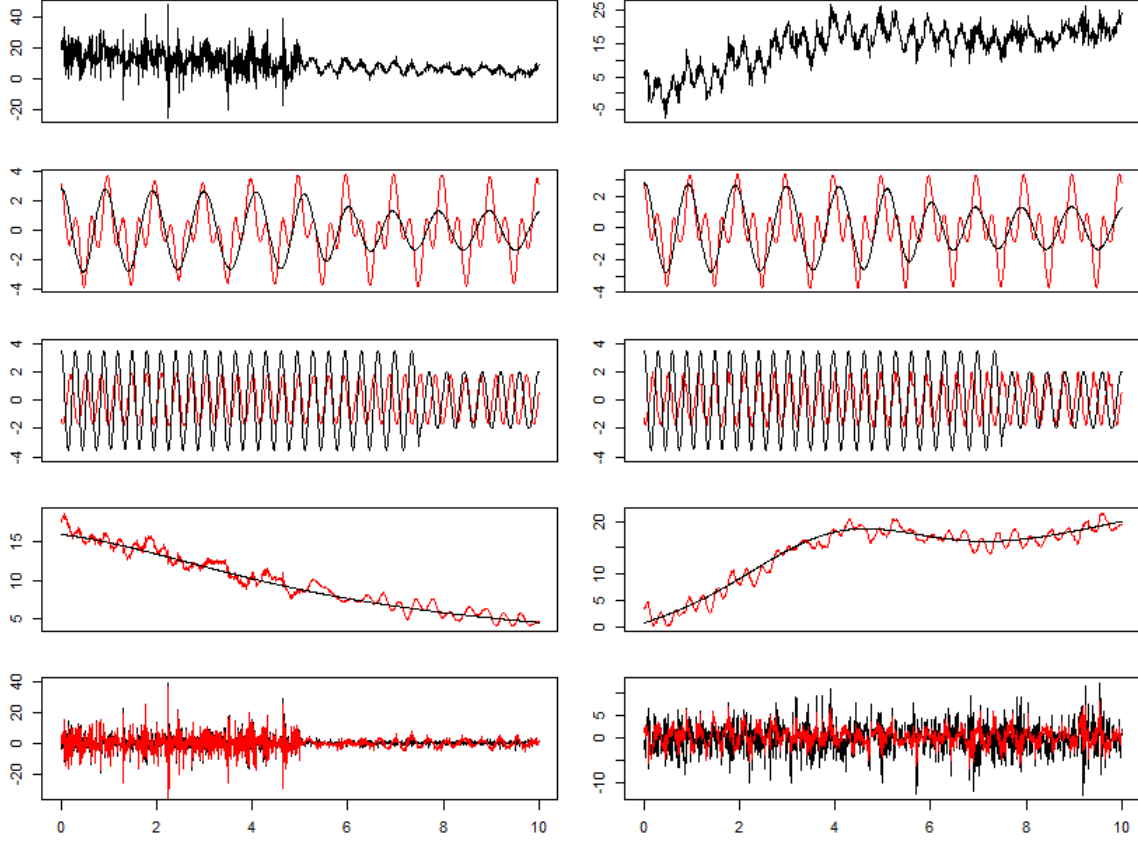


Figura 5.8: Resultado de aplicar TBATS a las señales $\mathbf{Y}_{2,1,2,1} := \mathbf{s}_2 + \mathbf{T}_1 + \mathbf{X}_2$ (izquierda) y $\mathbf{Y}_{2,2,3,0,5} := \mathbf{s}_2 + \mathbf{T}_2 + 0,5\mathbf{X}_3$ (derecha). Se muestran las descomposiciones obtenidas para la realización de la señal que tuvo como error la mediana de todos los errores. La primera fila es la realización de la señal y las otra filas son, de arriba hacia abajo, los verdaderos valores de $\mathbf{s}_{2,1}$, $\mathbf{s}_{2,2}$, \mathbf{T}_j y \mathbf{X}_k en negro y los estimadores encontrados por el método en rojo.

En este caso, con estacionalidad modelada por la clase $\mathcal{A}_{\epsilon,d}^{c_1,c_2}$ y ruido heterocedástico dependiente, TBATS no funciona bien, dado que la señal no satisface los supuestos paramétricos del modelo. En particular, para algunas realizaciones de estas señales el algoritmo no encuentra ninguna componente estacional, y cuando las encuentra, tiende a suavizar los cambios repentinos y encontrar oscilaciones incluso donde no existen. Se puede ver en la figura 5.8 que la amplitud de los estimadores $\widetilde{\mathbf{s}}_{2,1}$ y $\widetilde{\mathbf{s}}_{2,2}$ varía muy poco y que el componente estacional de baja frecuencia determinado por TBATS contiene extremos locales artificiales dentro de cada oscilación. Estos extremos locales artificiales pueden dar lugar a una mala interpretación de la dinámica del sistema y por lo tanto deben tenerse en cuenta al utilizar TBATS para el análisis de la misma. A su vez, esto se traslada también en más error en la estimación de la tendencia y el ruido para compensar las oscilaciones extras.

5.2. Datos reales

En esta sección analizaremos los datos del movimiento de codornices japonesas correspondientes al trabajo [9] y disponibles públicamente en [FigShare](#).

La locomoción o movimiento se estudia en diversas áreas de la biología como la cronobiología y fisiología. Los patrones de locomoción temporales y espaciales reflejan tanto períodos de descanso como la necesidad de moverse de un lugar a otro para satisfacer necesidades fisiológicas y de comportamiento. Los ritmos circadianos de 24 hs característicos de los ciclos de sueño/vigilia se encuentran en muchas especies animales y también se observan en la locomoción ritmos ultradianos con periodos inferiores a 24 hs (por ejemplo de 12, 8 o 6 hs), muchas veces relacionados con la conducta alimentaria.

Aunque el principio básico de la locomoción parece simple, el comportamiento motor surge de las interacciones entre los sistemas muscular, esquelético, nervioso, respiratorio y circulatorio del animal y el medio ambiente. De estas interacciones surgen dinámicas complejas: además del comportamiento rítmico, la actividad locomotora exhibe patrones fractales robustos e invariantes de escala (es decir, patrones compuestos de partes que se asemejan al todo en diferentes escalas) y correlaciones temporales de largo alcance (es decir, el proceso se caracteriza por una memoria infinita, donde la función de autocorrelación decae según una ley de potencias). Matemáticamente hablando, los fractales pueden considerarse opuestos a los ritmos oscilatorios suaves. Nosotros nos concentraremos en el análisis de los ritmos, con períodos de varias horas, utilizando SST. Si bien no se estudió teóricamente la robustez de la transformada synchrosqueezing a este tipo de patrones fractales, veremos si en la práctica es capaz de extraer componentes oscilatorios significativos, que concuerden con lo que empíricamente se sabe de los ciclos circadianos y ultradianos de los animales en estudio.

Como se describe en [9], se grabó la actividad de locomoción de 24 codornices japonesas hembras aisladas visualmente en su entorno de casa-caja (cajas de madera de 40x40x40). Las codornices están separadas en dos grupos de 12 y utilizaremos los datos correspondientes al grupo 2. Los datos de comportamiento se registraron de forma continua durante 6,5 días a una frecuencia de muestreo de 0,5 s ($> 10^6$ datos) en forma de una serie temporal de estados mutuamente excluyentes: móvil / inmóvil. En un momento dado, si el pájaro se estaba moviendo se registró un 1 o un 0 si estaba inmóvil. Cada día, durante un período de aproximadamente 30 minutos se realizaban tareas de mantenimiento que interferían con el experimento y por lo tanto los datos de esos períodos no se incluyeron en las series.

Como no tiene mucho sentido aplicar SST a una serie de ceros y unos, para reducir el ruido se agrupan y promedian los datos de a 6 minutos. Como el intervalo de muestreo era de 0,5 segundos, esto significa que cada punto de la nueva serie se obtiene tomando 720 puntos de la serie original y promediándolos. Así, para cada codorniz, en lugar de una serie de > 500000 puntos, tenemos una serie de largo 1479 con valores entre 0 y 1 indicando la proporción de tiempo en movimiento del animal cada 6 minutos, con lo cual al multiplicarlo por 100, lo tenemos en términos de porcentaje. Se puede ver en la figura 5.9 un ejemplo de esta serie procesada. A estas series le aplicaremos las transformadas wavelet y synchrosqueezing con parámetros “bump” para la wavelet madre y $n_v = 32$, el resto por defecto.

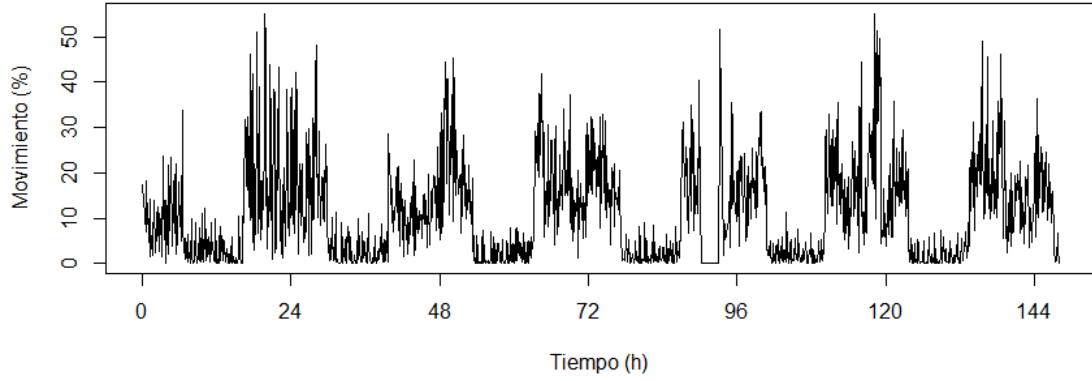


Figura 5.9: Serie promediada a 6 minutos correspondiente a la codorniz 5 del grupo 2.

Por otro lado, para que los gráficos de la transformada wavelet y la transformada synchrosqueezing resulten más útiles para el análisis de los ritmos circadianos y ultradianos, se realizó una traducción de las escalas y frecuencias de los mismos a períodos en horas utilizando las siguientes relaciones basadas en el procesamiento de los datos: $periodo = \frac{escala}{3000}$ para CWT y $periodo = \frac{1}{3600 * frecuencia}$ para SST. En las figuras 5.10 y 5.11 se pueden ver los gráficos de las transformadas de las series de algunas de las codornices.

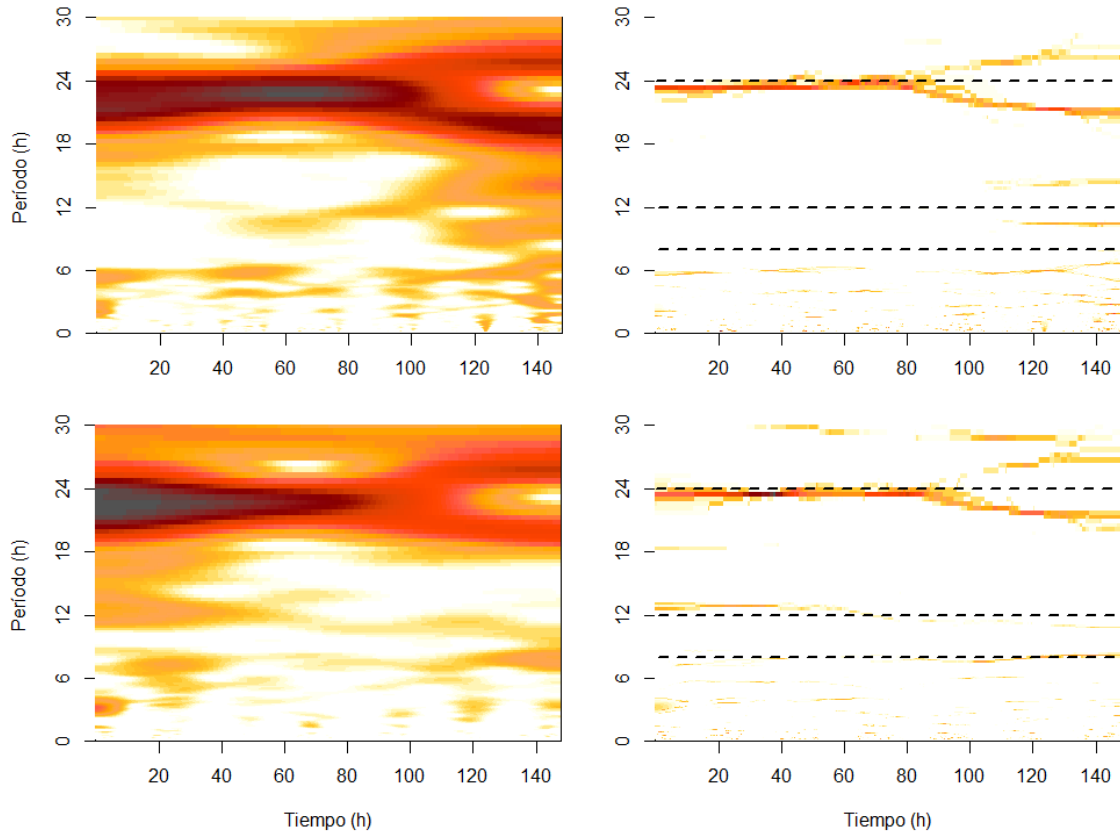


Figura 5.10: Transformada wavelet (izquierda) y synchrosqueezing (derecha) de las codornices 2 (arriba) y 4 (abajo). Se marcaron los períodos de 24, 12 y 8 hs en negro punteado.

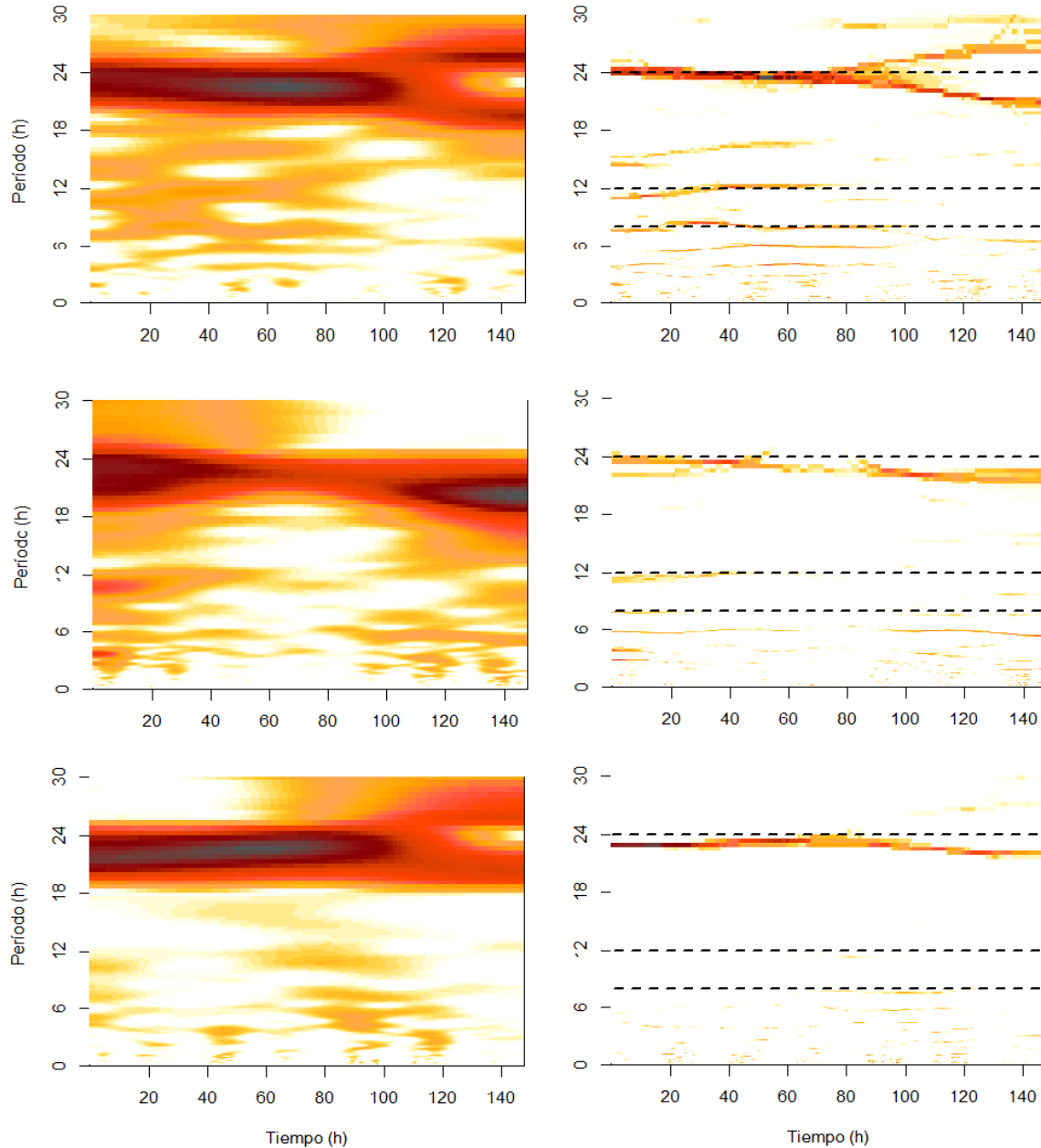


Figura 5.11: Transformada wavelet (izquierda) y synchrosqueezing (derecha) de las codornices 7, 9 y 11 (de arriba hacia abajo). Se marcaron los períodos de 24, 12 y 8 hs en negro punteado.

Como se puede observar en las figuras 5.10 y 5.11, en los gráficos de la transformada synchrosqueezing las curvas se encuentran muchos más concentradas que en los de la transformada wavelet, lo cual hace posible que se vea a simple vista la presencia de ciclos con períodos de 12, 8 o 6 hs que no son tan notorios en los gráficos de la transformada wavelet. Además, si bien en todos los gráficos se advierte la presencia del ciclo circadiano de aproximadamente 24 hs, con la transformada wavelet se ve mucho más disperso. Cabe remarcar que, debido a la eliminación de aproximadamente 30 minutos de datos cada día, el ciclo circadiano se ve en varios casos levemente por debajo de la línea de 24 hs,

porque los datos solo contemplan 23.5 hs por día. En los otros ciclos ultradianos puede explicar un poco la variabilidad, pero no es tan significativo.

Al estar trabajando con seres vivos surgen algunas complicaciones en la interpretación de los resultados que no teníamos con los datos simulados. Por un lado, podemos considerar las series como réplicas independientes de un proceso con mucha variabilidad o podemos considerar que cada una proviene de una distribución diferente, ya que cada codorniz tiene su propio reloj biológico y en ese caso podemos estudiar la fuerza de cada componente en la especie. Por ejemplo, se puede ver en los gráficos de SST que todas presentan un ciclo circadiano de alrededor de 24 hs. Las codornices 4, 7 y 9 también presentan en algún momento un ciclo ultradiano de 12 hs, luego, si bien este componente no es tan fuerte como el de 24 hs, sí se manifiesta en varios animales. El de 8 hs se puede observar en las codornices 4 y 7 y en otras se observa más marcado un ciclo de 6 hs, con mayor o menor intensidad. Por otro lado, los ruidos son más complejos que los estudiados en el capítulo 4, con lo cual, si bien se sabe empíricamente que los animales suelen tener ritmos internos, para estos datos no tenemos garantías teóricas sobre la existencia y la identificabilidad de los componentes encontrados, sobre todos los de mayor frecuencia (menor período) que se encuentran más contaminados por el ruido. De todos modos, la transformada synchrosqueezing resulta una herramienta muy útil de visualización y análisis en conjunto con el conocimiento de dominio.

Como hicimos con los datos simulados, también podemos extraer las componentes periódicas de los gráficos de SST, lo cual puede ser muy útil en experimentos donde se quiere estudiar si hay un cambio en el comportamiento de los animales cuando se introduce un cambio en el ambiente, ya que aislando visualmente cada componente estos cambios son más notorios. No es el caso de estos datos que estamos estudiando, ya que las condiciones del experimento se mantuvieron constante durante los 6,5 días, pero de todos modos, los ciclos de las codornices también dependen de variables internas que no se pueden controlar, con lo cual igual pueden variar. Tomamos para descomponer la serie correspondiente a la codorniz 5 del grupo 2, cuya transformada synchrosqueezing puede verse en la figura 5.12.

Podemos observar en la figura 5.12 que el algoritmo extrae una curva con período aproximadamente de 24 hs, otra de aproximadamente 12 hs y la tercera curva es en parte un ciclo de 6 hs y en parte uno de 8 hs. Si bien el ciclo de 8 hs parece estar presente en toda la duración del experimento, su amplitud va cambiando y esto hace que el método de extracción tome en parte la curva correspondiente al período de 6 hs cuando ésta es más intensa. En la figura 5.13 se puede ver la reconstrucción de estas tres componentes y en particular, en la curva que mezcla los ciclos de 6 y 8 hs se ve el cambio en la frecuencia, alrededor de la hora 90, cuando pasa a ser más intenso el ciclo de 6 que el de 8. También podemos ver en la figura 5.13, al superponer las curvas extraídas del gráfico de synchrosqueezing con la serie original, que si bien el ciclo de 24 se corresponde bastante con la señal, visualmente los ciclos de 12, 6 y 8 hs no son tan evidentes en los datos, pero al sumarlos todos explican mejor el movimiento del animal en conjunto que el de 24 hs individualmente. Es decir, SST nos permite extraer ciclos que no son evidentes a simple vista pero que son importantes para explicar el comportamiento de los datos.

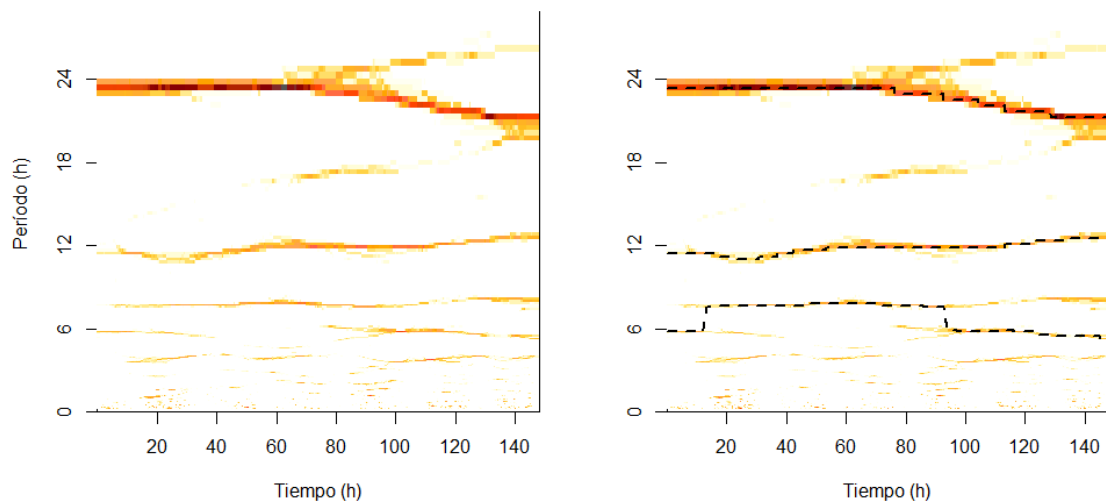


Figura 5.12: A la izquierda, gráfico de la transformada synchrosqueezing de la serie de movimiento de la codorniz 5. A la derecha, el mismo gráfico y superpuestas en negro las curvas de máxima energía/mínima curvatura encontradas automáticamente por la función “curve_ext_multi” cuando se le pide extraer 3 curvas.

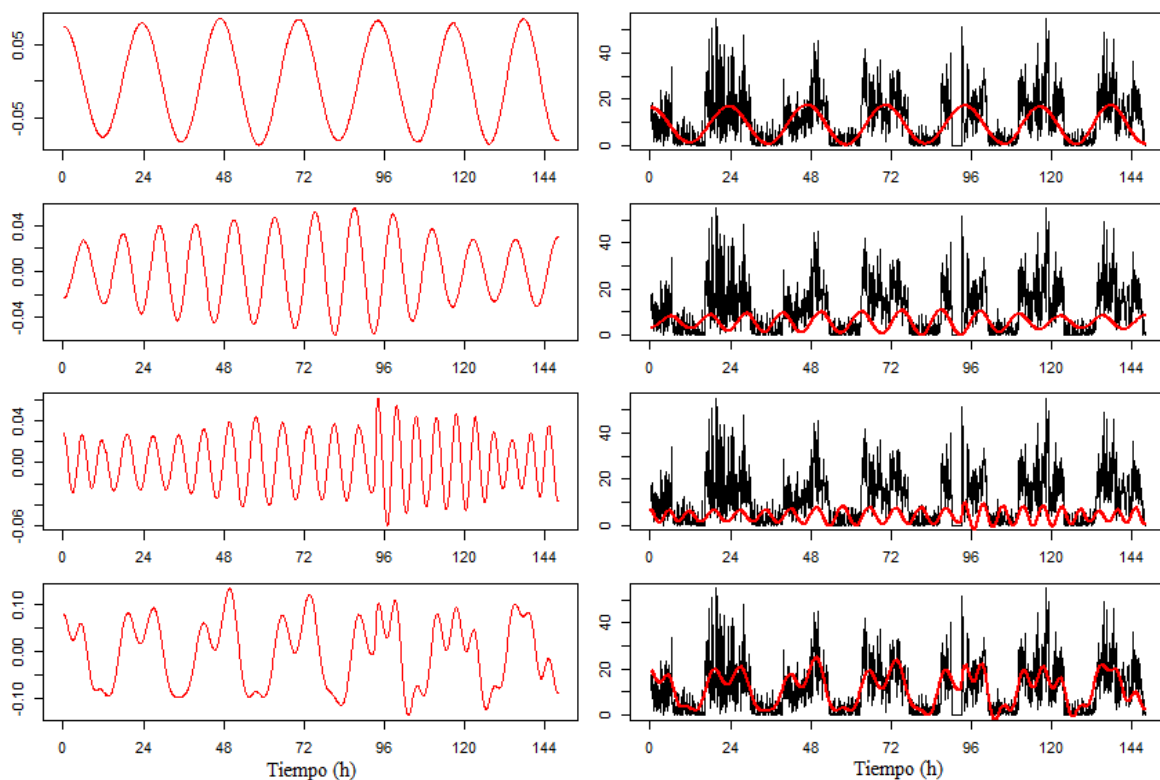


Figura 5.13: A la izquierda, de arriba hacia abajo, las tres componentes extraídas del gráfico en la figura 5.12 (aproximadamente 24, 12 y 6/8 hs) y el resultado de la suma de estas tres componentes. A la derecha, las mismas señales que a la izquierda (multiplicadas por 100) superpuestas con la serie de movimiento de la codorniz 5 (figura 5.9).

Conclusiones

El problema de la estimación de la descomposición de señales en suma de componentes ha sido tratado de diversas formas a lo largo del tiempo en diversas ramas de aplicación, desde la econometría hasta el procesamiento de señales provenientes de problemas ingenieriles. En particular, existen muchos métodos disponibles en la literatura estadística para descomponer señales que tienen en cuenta tanto la estacionalidad como la tendencia. En este trabajo estudiamos tres de ellos: SST, EMD y TBATS.

Por un lado, estudiamos el problema de la identificabilidad de los componentes encontrados con SST. En estadística, la identificabilidad es una propiedad que un modelo debe satisfacer para que sea posible una inferencia precisa. Un modelo es no identificable si dos o más parametrizaciones son observacionalmente equivalentes. Sin embargo, en otras áreas el término no siempre se utiliza con ese significado. En [12], por ejemplo, se comentan más de dos docenas de términos diferentes para la identificación de modelos solamente en la literatura econométrica.

En el caso particular de series de tiempo observadas sobre fenómenos con estacionalidad y tendencia contaminadas con ruido, el problema de la identificabilidad de las modas estacionales es aún más complicado. En el capítulo 4 vimos teoremas que nos permiten afirmar que si f está en la clase $\mathcal{A}_{\epsilon,d}^{c_1,c_2}$ la diferencia entre distintas representaciones de f es despreciable, con lo que podemos decir que f es “identificable”, y que al aplicar SST estamos encontrando la única función que explica nuestros datos y que no hay otra observacionalmente equivalente. Sin embargo, cuando trabajamos con datos reales, éstos no siempre cumplen las hipótesis de los teoremas, con lo cual, no vamos a tener las garantías teóricas sobre la existencia y la identificabilidad de los componentes encontrados. De todos modos, la transformada synchrosqueezing resulta una herramienta muy útil de visualización y análisis en conjunto con el conocimiento de dominio.

Por otro lado, analizamos la capacidad de cada modelo para detectar tendencia y múltiples componentes estacionales con comportamiento variable en el tiempo y su estabilidad frente a cambios en la longitud de la serie y la presencia de errores heterocedásticos dependientes, tanto en la teoría como en la práctica, por medio de simulaciones. A continuación resumimos las principales ventajas y desventajas de cada método.

TBATS resulta útil para pronóstico de series de tiempo pero tiene algunas limitaciones cuando se utiliza para analizar data histórica. Si bien las componentes estacionales están acopladas con un proceso de error ARMA que le da algo de variabilidad, los períodos estacionales son todos fijos, con lo cual, como se vió en las simulaciones, el método tiene dificultades cuando la frecuencia de las componentes varía en el tiempo. Al

ser paramétrico puede ser demasiado restrictivo para datos del mundo real. Las violaciones de los supuestos paramétricos, como el de períodos fijos, pueden causar no solo un gran sesgo en la estimación de la estacionalidad, sino también oscilaciones espurias en la estimación de la tendencia, ya que cualquier dinámica estacional inexplicable tendría que atribuirse a la tendencia. En cambio, vimos que métodos no paramétricos como EMD o la transformada Synchrosqueezing responden mejor a un sistema dinámico, incluso si las frecuencias y las amplitudes que varían en el tiempo son desconocidas.

Además, en los métodos paramétricos convencionales, el análisis de la estacionalidad depende de toda la serie de tiempo, lo que los hace sensibles a la longitud de la misma. Mientras que, dado que la estacionalidad se modela de forma no paramétrica y el algoritmo de synchrosqueezing es de naturaleza local, SST es insensible a la longitud de la serie de tiempo observada; lo mismo vimos que ocurría para EMD.

Observamos también que la transformada synchrosqueezing y EMD comparten muchas características que los diferencian de métodos paramétricos como TBATS, pero la principal diferencia entre ellos es su capacidad de responder la pregunta 5 que habíamos planteado al comienzo del capítulo 2. Los procesos aleatorios de errores son a menudo dependientes y pueden ser heterocedásticos, y como vimos, el método de synchrosqueezing es robusto a este tipo de errores, y hay garantías teóricas de ello, mientras que EMD puede ser muy inestable y difícil de analizar matemáticamente.

Una ventaja que tiene EMD sobre synchrosqueezing, sin embargo, es que es totalmente adaptativo, ya que no requiere como entrada nada más que los datos. Para utilizar synchrosqueezing, uno debe elegir de antemano una wavelet madre. Si bien decimos que el método es adaptativo ya que las constantes que aparecen en los errores de estimación en los teoremas 4.4.3 y 4.4.4, dependen solo de los momentos de orden superior de la wavelet madre elegida, pero no de su forma, y por lo tanto, la elección de la wavelet madre no es crucial para asegurar las buenas propiedades de la transformación, dependiendo de cómo esté implementado en la práctica, puede funcionar mejor para ciertas wavelets. También, otro argumento que necesita la técnica de synchrosqueezing es K , el número de componentes estacionales. EMD en cambio calcula la cantidad de IMF's automáticamente como parte del algoritmo de sifting, aunque tampoco es perfecto porque, como vimos, un pequeño cambio en la función de entrada puede generar un número distinto de IMF's entre otros problemas. Entonces, a pesar de no requerir parámetros previos, sí requiere un mayor trabajo posterior de interpretación de los resultados, por ejemplo, como el que se realiza en [4], que propone la identificación significativa de modas mediante el estudio de la transformada de Fourier de los residuos.

En definitiva, no hay un método universal que funcione perfecto bajo cualquier circunstancia. Depende que tipo de serie queremos estudiar y que pretendemos obtener del análisis de esa serie. Pero es importante entender las fortalezas y debilidades de cada uno para saber porque pueden fallar y cuál conviene usar en cada caso.

Por último, hay muchos temas de estudio interesantes que se desprenden de lo que vimos. Por ejemplo, en este trabajo nos centramos en encontrar descomposiciones para las series de tiempo y no estudiamos el problema del pronóstico de series temporales, es decir, la predicción de valores futuros. Una forma de predecir valores, teniendo una descomposición de la señal es extrapolar las componentes estacionales, la tendencia y los residuos (que aproximan el error). Con lo cual, si bien SST no fue originalmente

formulada con ese fin, se la puede utilizar para realizar predicciones a corto plazo y sería interesante estudiar su capacidad para esto y si se puede extender para predicciones a largo plazo.

También, como mencionamos, uno de los problemas de SST es que requiere de la elección de varios parámetros, como la wavelet madre o el umbral Γ , con lo cual sería interesante investigar sobre métodos para que este proceso de elección de parámetros sea automático, y que el método sea completamente adaptativo en la práctica.

Bibliografía

- [1] ADDISON, P. S. *The Illustrated Wavelet Transform Handbook*. CRC Press, 2da edición, capítulo 1, 2017.
- [2] BROCKWELL, P. J. AND R. A. DAVIS. *Introduction to Time Series and Forecasting*. Springer, 2da edición, capítulos 1, 2, 3 y 10, 2002.
- [3] CHEN, Y.-C., M.-Y. CHENG AND H.-T. WU. *Nonparametric and adaptive modeling of dynamic seasonality and trend with heteroscedastic and dependent errors*. J. R. Stat. Soc. B, vol. 76, pp. 651-682, 2014.
- [4] CHO, S., R. SHAHRIAR AND U. CHONG. *Identification of significant intrinsic mode functions for the diagnosis of induction motor fault*. J. Acoust. Soc. Am., vol. 136, pp. 72-77, 2014.
- [5] DAUBECHIES, I., J. LU, AND H.-T. WU. *Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool*. Appl. Comput. Harmon. Anal. 30, pp. 243–261, 2011.
- [6] DE LIVERA, A. M., R. J. HYNDMAN, AND R. D. SNYDER. *Forecasting Time Series With Complex Seasonal Patterns Using Exponential Smoothing*. J. Am. Stat. Assoc., vol. 106, No. 496, pp. 1513-1527, 2011.
- [7] GEL'FAND, I. M. AND N. Y. VILENKIN. *Generalized function theory: Vol. 4, Applications of Harmonic Analysis*. Academic Press, capítulo 3, 1964.
- [8] GENTON, M. G. AND P. HALL. *Statistical inference for evolving periodic functions*. J. R. Stat. Soc. B, vol. 69, No. 4, pp. 643-657, 2007.
- [9] GUZMÁN, D. A., S. PELLEGRINI, A. G. FLESIA, M. A. AON, R. H. MARIN AND J. M KEMBRO. *High resolution, week-long, locomotion time series from Japanese quail in a home-box environment*. Scientific Data, vol. 3, artículo 160036, 2016.
- [10] HUANG, N. E., Z. SHEN, S. R. LONG, M. C. WU, H. H. SHIH, Q. ZHENG, N.-C. YEN, C. C. TUNG AND H. H. LIU. *The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis*. Proc. R. Soc. Lond. A, vol. 454, pp. 903–995, 1998.

- [11] LEIS, J. W. *Digital Signal Processing Using MATLAB for Students and Researchers*. John Wiley & Sons, capítulo 3, sección 8, 2011.
- [12] LEWBEL, A. *The Identification Zoo - Meanings of Identification in Econometrics*. Journal of Economic Literature, vol. 57, pp. 835-903, 2019.
- [13] MHAMDI, F., M. JAÏDANE-SAÏDANE AND J.-M. POGGI. *Trend Extraction for seasonal Time Series Using Ensemble Empirical Mode Decomposition*. Advances in Adaptive Data Analysis, vol. 3, no. 3, pp. 363-383, 2011.
- [14] OSGOOD B.. *Lecture Notes for EE 261, The Fourier Transform and its Applications*. Electrical Engineering Department, Stanford University, capítulos 2 y 4, 2007.
- [15] TARY, J. B., R., H. HERRERA AND M. VAN DER BAAN. *Analysis of time-varying signals using continuous wavelet and synchrosqueezed transforms*. Phil. Trans. R. Soc., vol. 376, 20170254, 2018.
- [16] THAKUR, G., E. BREVDÖ, N. S. FOCKER AND H.-T. WU. *The synchrosqueezing algorithm for time-varying spectral analysis: robustness properties and new paleoclimate applications*. Signal Processing 93, pp. 1079-1094, 2013.
- [17] WU, Z. AND N. E. HUANG. *Ensemble empirical mode decomposition: A noise-assisted data analysis method*. Advances in Adaptive Data Analysis, vol. 1, no. 1, pp. 1-41, 2009.