



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA ELÉCTRICA

UN ENFOQUE MODERNO PARA LA ESTIMACIÓN ESPECTRAL PROBABILÍSTICA

TESIS PARA OPTAR AL GRADO DE
MAGÍSTER EN CIENCIAS DE LA INGENIERÍA, MENCIÓN ELÉCTRICA
MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL ELÉCTRICO

LERKO CALEB ARAYA HERNÁNDEZ

PROFESOR GUÍA:
FELIPE TOBAR HENRÍQUEZ

MIEMBROS DE LA COMISIÓN:
JORGE SILVA SÁNCHEZ
SUSANA EYHERAMENDY DUERR

Este trabajo ha sido parcialmente financiado por CONICYT N 22171830,
FONDECYT-Iniciación 1171165 y CONICYT-PIA AFB-170001.

SANTIAGO DE CHILE
2019

RESUMEN DE LA TESIS PARA OPTAR
AL GRADO DE MAGÍSTER EN CIENCIAS DE
LA INGENIERÍA, MENCIÓN ELÉCTRICA
Y AL TÍTULO DE INGENIERO CIVIL ELÉCTRICO
POR: LERKO CALEB ARAYA HERNÁNDEZ
FECHA: 2019
PROF. GUÍA: FELIPE TOBAR HENRÍQUEZ

UN ENFOQUE MODERNO PARA LA ESTIMACIÓN ESPECTRAL PROBABILÍSTICA

El presente trabajo propone y estudia dos métodos bayesianos para abordar el problema de estimación espectral: uno paramétrico y uno no-paramétrico. En esta ocasión se abordará una variante del problema de estimación espectral, en que las muestras u observaciones están en el dominio de las frecuencias, similar al problema de interferometría. Actualmente, Chile es una de las grandes potencias en astronomía, siendo el norte de Chile el hogar del interferómetro más grande del mundo, llamado Atacama Large Millimeter/submillimeter Array (ALMA).

El desarrollo metodológico de este trabajo consistió en modelar el dominio temporal (o espacial) y luego -informalmente- *pasar el modelo temporal a través de la transformada de fourier*. Para el modelamiento se utiliza inferencia bayesiana, puesto que esta es un marco de trabajo que promete incorporar la incertidumbre natural de los problemas en el mundo real. De esta manera, el enfoque probabilístico que nos brinda la inferencia bayesiana nos permitirá establecer intervalos de confianza y caracterizar la incertidumbre de la reconstrucción.

Los resultados experimentales están divididos en dos partes. La primera, está relacionada con la validación de los métodos, mientras que la segunda consiste en probar los modelos en imágenes reales. La etapa de validación a su vez se dividió en otros dos experimentos: El primer experimento es un caso unidimensional, en el que se probó el desempeño del método para realizar una regresión de una señal temporal la cual cumple todos las hipótesis propuestas en cada método; el segundo experimento consiste en realizar una regresión de una señal bidimensional, es decir, una imagen que cumpla con las hipótesis de cada método. El propósito de esto es realizar una validación en un escenario sintético y completamente controlado, en los cuales se pueda evaluar correctamente el desempeño de los métodos.

Los resultados para cada método son prometedores, entre los cuales se puede observar la correcta validación de ambos métodos, la robustez de los métodos para imágenes reales y la facilidad de extensión para ocupar los métodos en otros problemas o incluso con muestras en el dominio temporal. En ese sentido, el trabajo a futuro puede alcanzar una gran profundidad y el pulimento de estos métodos se puede entender tanto como la aplicación particular de estos métodos en un problema o el desarrollo de nueva teoría para la sofisticación de los métodos.

*Y Dios dijo:
"Hagamos a los seres humanos a nuestra imagen,
hagamos que reflejen nuestra naturaleza."*

Tabla de Contenido

Índice de Tablas	vii
Índice de Ilustraciones	viii
1. Introducción	1
1.1. Motivación	1
1.2. Objetivo	2
1.2.1. Objetivo General	2
1.2.2. Objetivo Específico	2
1.3. Hipótesis	2
1.4. Metodología de trabajo	3
1.5. Aporte	3
1.6. Estructura	3
2. Marco Teórico	4
2.1. Operaciones en matrices	4
2.1.1. Producto de Kronecker	4
2.1.2. Descomposición de Cholesky	4
2.2. Fundamentos de probabilidades	5
2.2.1. Variable aleatoria	5
2.2.1.1. Variable aleatoria a valores complejos	5
2.2.2. Vector Aleatorio	6
2.2.3. Distribución Gaussiana multivariada	6
2.2.3.1. Distribución Gaussiana multivariada condicional	6
2.2.4. Distribución Gaussiana multivariada a valores complejos	7
2.3. Modelos de mezcla de bases	7
2.4. Estimación	8
2.4.1. Estimador de máxima verosimilitud	8
2.4.2. Inferencia Bayesiana	8
2.4.2.1. Máximo a posteriori	9
2.5. Algoritmos de optimización	10
2.5.1. Método de Newton	10
2.5.2. Broyden–Fletcher–Goldfarb–Shanno (BFGS)	11
2.5.3. Powell	11
2.6. Radio Astronomía: Interferometría	12
2.6.1. Imagen Sucia y Haz Sucio	13

2.6.2. <i>Snapshot</i> e interferometría dinámica	13
3. Propuesta	15
3.1. Modelo paramétrico: Suma de Gaussianas	15
3.1.1. Modelo de ruido	16
3.1.2. Buscando la distribución posterior del modelo	17
3.1.3. Eligiendo el prior sobre los parámetros	17
3.2. Modelo no-paramétrico: Reconstrucción espectral Bayesiana	18
3.2.1. Un modelo de máxima verosimilitud para las observaciones ruidosas .	20
3.2.2. Inferencia utilizando la distribución posterior	20
3.2.3. Caso bidimensional (imágenes)	21
4. Análisis y Resultados	22
4.1. Modelo paramétrico: suma de Gaussianas	22
4.1.1. Caso unidimensional	22
4.1.2. Caso bidimensional	24
4.1.3. Caso real: Inteferometría	27
4.2. Modelo no-paramétrico: Reconstrucción espectral Bayesiana	32
4.2.1. Caso unidimensional	32
4.2.2. Caso dos-dimensional	34
4.2.3. Caso real: Inteferometría	36
5. Conclusiones y trabajo a futuro	40
6. Bibliografía	42
7. Apéndice	45
7.1. Vectorización de una matriz	45
7.2. Producto de Kronecker	45
7.3. Procesos Gaussianos	46

Índice de Tablas

4.1. Tabla de parámetros para suma de Gaussianas. Número de bases: 6.	23
4.2. Error cuadrático medio normalizado (NMSE, por su sigla en inglés).	23
4.3. Error cuadrático medio normalizado.	33

Índice de Ilustraciones

2.1. Arreglo de antenas y píxeles muestreados. (a) Arreglo de antenas. (b) Muestreo estático o <i>snapshot</i> . (c) Muestreo dinámico.	14
3.1. Modelo gráfico del problema. Se tiene la señal en el tiempo (o espacio), que está ligada a través de la transformada de Fourier a su espectro y viceversa. Las observaciones se realizan en el espectro. La notación θ representa los parámetros de entrada θ para el modelo paramétrico y los hiper-parámetros del kernel $\mathcal{K}_\theta(x, x')$ en el caso del modelo no-paramétrico.	15
4.1. Columna izquierda: Reconstrucción en frecuencia (arriba) y tiempo (abajo) para $\sigma_{\text{noise}} = 1$. Columna derecha: Reconstrucción en frecuencia (arriba) y tiempo (abajo) para $\sigma_{\text{noise}} = 6$	23
4.2. (a) Arreglo de antenas dispuesto de manera uniforme, cada cruz representa una antena dispuesta en el terreno. (b) Patrón de observación: En blanco frecuencias observadas, en negro frecuencias no observadas. Este arreglo de 81 antenas genera el 38,53 % de las observaciones.	24
4.3. Reconstrucción utilizando 2 bases. Columna izquierda: Imágenes originales. Columna central: Imágenes reconstruidas con modelo de suma de Gaussianas. Columna derecha: Diferencia entre imágenes. Arriba: Dominio espacial NMSE = $1,04 \cdot 10^{-31}$. Centro: Dominio espectral parte real NMSE = $1,83 \cdot 10^{-31}$. Abajo: Dominio espectral parte imaginaria NMSE = $1,57 \cdot 10^{-31}$	25
4.4. Reconstrucción utilizando 20 bases. Columna izquierda: Imágenes originales. Columna central: Imágenes reconstruidas con modelo de suma de Gaussianas. Columna derecha: Diferencia entre imágenes. Arriba: Dominio espacial NMSE = $3,10 \cdot 10^{-5}$. Centro: Dominio espectral parte real NMSE = $5,46 \cdot 10^{-6}$. Abajo: Dominio espectral parte imaginaria NMSE = $5,78 \cdot 10^{-5}$	26
4.5. Catálogo de cuerpos celestes.	28
4.6. Reconstrucción HD142527 utilizando 20 bases. Columna izquierda: Imágenes originales. Columna central: Imágenes reconstruidas con modelo de suma de Gaussianas. Columna derecha: Diferencia entre imágenes. Arriba: Dominio espacial NMSE = $8,45 \cdot 10^{-3}$. Centro: Dominio espectral parte real NMSE = $6,50 \cdot 10^{-3}$. Abajo: Dominio espectral parte imaginaria NMSE = $1,32 \cdot 10^{-2}$	29

4.7.	Reconstrucción IRAS16293 utilizando 50 bases y 27,51 % de los datos. Columna izquierda: Imágenes originales. Columna central: Imágenes reconstruidas con modelo de suma de Gaussianas. Columna derecha: Diferencia entre imágenes. Arriba: Dominio espacial NMSE = 8,62. Centro: Dominio espectral parte real NMSE = $9,63 \cdot 10^{-2}$. Abajo: Dominio espectral parte imaginaria NMSE = $7,85 \cdot 10^{-3}$	30
4.8.	Reconstrucción IRAS16293 utilizando 50 bases y 5 % de los datos. Columna izquierda: Imágenes originales. Columna central: Imágenes reconstruidas con modelo de suma de Gaussianas. Columna derecha: Diferencia entre imágenes. Arriba: Dominio espacial NMSE = $5,94 \cdot 10^{-5}$. Centro: Dominio espectral parte real NMSE = $4,10 \cdot 10^{-1}$. Abajo: Dominio espectral parte imaginaria NMSE = $3,47 \cdot 10^{-1}$	31
4.9.	Reconstrucción IRAS16293 utilizando 10 bases y 5 % de los datos. Columna izquierda: Imágenes originales. Columna central: Imágenes reconstruidas con modelo de suma de Gaussianas. Columna derecha: Diferencia entre imágenes. Arriba: Dominio espacial NMSE = 2,61. Centro: Dominio espectral parte real NMSE = $6,18 \cdot 10^{-3}$. Abajo: Dominio espectral parte imaginaria NMSE = $5,85 \cdot 10^{-3}$	32
4.10.	Columna izquierda: Reconstrucción en frecuencia (arriba) y tiempo (abajo) para $\sigma_{\text{noise}} = 1 \cdot 10^{-3}$. Columna derecha: Reconstrucción en frecuencia (arriba) y tiempo (abajo) para $\sigma_{\text{noise}} = 0,5$	33
4.11.	Reconstrucción utilizando BSR. De izquierda a derecha, primera columna: imágenes originales; segunda columna: imágenes reconstruidas con modelo de suma de Gaussianas; tercera columna: desviación estándar; cuarta columna: diferencia entre imágenes. Arriba: Dominio espacial NMSE = $8,72 \cdot 10^{-5}$. Centro: Dominio espectral parte real NMSE = $8,69 \cdot 10^{-5}$. Abajo: Dominio espectral parte imaginaria NMSE = $8,75 \cdot 10^{-5}$	35
4.12.	Representación vectorial de la imagen. Reconstrucción de espectro utilizando el 51 % de los datos. Se utilizó como observaciones la mitad derecha de las frecuencias, sin incluir el par de frecuencias (0, 0).	36
4.13.	Reconstrucción utilizando BSR. De izquierda a derecha, primera columna: imágenes originales; segunda columna: imágenes reconstruidas con BSR; tercera columna: desviación estándar; cuarta columna: diferencia entre imágenes. Arriba: Dominio espacial NMSE = $3,00 \cdot 10^{-2}$. Centro: Dominio espectral parte real NMSE = $5,90 \cdot 10^{-2}$. Abajo: Dominio espectral parte imaginaria NMSE = $4,96 \cdot 10^{-5}$	37
4.14.	Reconstrucción de HD142527 utilizando BSR. De izquierda a derecha, primera columna: imágenes originales; segunda columna: imágenes reconstruidas con BSR; tercera columna: desviación estándar; cuarta columna: diferencia entre imágenes. Arriba: Dominio espacial NMSE = $1,17 \cdot 10^{-3}$. Centro: Dominio espectral parte real NMSE = $5,92 \cdot 10^{-5}$. Abajo: Dominio espectral parte imaginaria NMSE = $1,69 \cdot 10^{-2}$	38

4.15. Reconstrucción de IRAS16293 utilizando BSR. De izquierda a derecha, primera columna: imágenes originales; segunda columna: imágenes reconstruidas con BSR; tercera columna: desviación estándar; cuarta columna: diferencia entre imágenes. Arriba: Dominio espacial $NMSE = 3,76 \cdot 10^{-2}$. Centro: Dominio espectral parte real $NMSE = 2,27 \cdot 10^{-2}$. Abajo: Dominio espectral parte imaginaria $NMSE = 5,39 \cdot 10^{-2}$ 39

Publicaciones

Esta tesis está basada en los siguientes artículos que resumen el desarrollo gradual del trabajo:

Conferencias con revisión de pares

- **L. Araya-Hernández, J. F. Silva, A. Osses and F. Tobar**, *A Bayesian Mixture-of-Gaussians Model for Astronomical Observations in Interferometry*, in 2017 CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON), 2017.

Escuelas de verano

- **L. Araya-Hernández, J. F. Silva, A. Osses and F. Tobar**, *Recovering Astronomical Observations Through Interferometry Using an RBF Model*, in XII IEEE CIS Escuela de Verano Latino-Americana en Inteligencia Computacional (EVIC), 2016.
- **L. Araya-Hernández, J. F. Silva, A. Osses and F. Tobar**, *Hierarchical vs Non-Hierarchical Priors for a Gaussian Mixture Model in Radio Astronomy*, in XIII IEEE CIS Escuela de Verano Latino-Americana en Inteligencia Computacional (EVIC), 2017.*

*Ganador del segundo lugar en la competencia de posters.

Capítulo 1

Introducción

1.1. Motivación

El espectro de una señal es ampliamente estudiado en distintas disciplinas, debido a que en él existe información de interés que puede indicar la existencia de un fenómeno o a que a través de él podemos recuperar la señal original. Ejemplos de esto son variados: Desde detección de arritmias en electrocardiogramas, detección de símbolos en protocolos de comunicaciones o recuperación de imágenes con hiper-resolución en radioastronomía. Es por este motivo, que tener buenas representaciones de este espectro es crucial a la hora de estudiarlo.

El teorema de muestreo de Nyquist [1, 2] enuncia que la frecuencia de muestreo en el espacio-temporal debe ser mayor o igual a dos veces el ancho de banda de la señal de interés para asegurar una correcta reconstrucción del espectro. Se puede notar que este teorema realiza dos suposiciones fuertes: (i) Existe una tasa de muestreo temporal regular, (ii) las mediciones (observaciones) son libres de ruido. Estas hipótesis no siempre se pueden garantizar, puesto que un caso típico es que las mediciones están sujetas a ruido asociada al instrumento de medición. Por otro lado, la existencia de una frecuencia de muestreo regular no se puede asegurar en todas las aplicaciones, como en radioastronomía, donde las antenas son compartidas, por lo cual las observaciones se tienen solo cuando la antena apunta al objeto de interés. Bajo estos escenarios, aplicar el teorema de muestreo no es viable, por lo cual se necesitan otras técnicas de reconstrucción de espectro. La forma más convencional de realizar esto es realizar una interpolación de los datos, pero nuevamente, esto no asegura obtener una buena representación del espectro.

Debido a que la transformada de Fourier es una transformación lineal, el problema inverso tiene las mismas dificultades, esto es, obtener la señal en el espacio temporal a través de observaciones de su espectro. Este problema es el que se abordará y es exactamente el caso de interferometría [3], la cual es una técnica de reconstrucción de señales temporales (o espaciales) utilizando un arreglo de antenas. La utilización de esta técnica se ha masificado, debido a que, en el caso particular de radioastronomía, permite obtener imágenes con hiper-resolución. En esta aplicación, las antenas simulan un lente óptico cuyo diámetro está en directa relación con la distancia más grande entre antenas. En ese sentido, recientemente se

ha llevado a cabo *interferometría de muy larga base*, la cual consiste en realizar interferometría con radio observatorios en distintos lugares del mundo. Es por este creciente interés en técnicas de interferometría que el estudio de las técnicas de reconstrucción de imágenes a través del muestreo del espectro es vital. Los algoritmos clásicos de reconstrucción de imágenes en interferometría son CLEAN [4] y MEM [5], los cuales tienen la problemática de no ofrecer un método para representar la incertidumbre asociada a la reconstrucción [6].

La inferencia Bayesiana [7, 8, 9] ofrece una potente herramienta para abordar este tipo de problemas, puesto que este enfoque incorpora de manera natural la incertidumbre. Dentro de los métodos Bayesianos existen los métodos paramétricos y los no-paramétricos. En esta tesis se hará una propuesta de dos modelos: el primero, será un modelo paramétrico, mientras que el segundo uno no-paramétrico.

1.2. Objetivo

1.2.1. Objetivo General

Desarrollar métodos Bayesianos capaces de reconstruir señales temporales (o imágenes) y espectrales, a través del muestreo conjunto de la parte real e imaginaria del espectro, ofreciendo una medida de la incertidumbre de la reconstrucción.

1.2.2. Objetivo Específico

- Desarrollar un método paramétrico de inferencia Bayesiana para la reconstrucción de señales.
- Desarrollar un método no-paramétrico de inferencia Bayesiana para la reconstrucción de señales.
- Probar la capacidad de regresión de ambos métodos con el estudio de casos sintéticos.
- Probar los métodos en datos reales.

1.3. Hipótesis

Los métodos Bayesianos son capaces de modelar la incertidumbre intrínseca de los problemas y de este modo ofrecer una regresión y una medida de incertidumbre.

1.4. Metodología de trabajo

La metodología de trabajo consiste principalmente en validar el método y posteriormente probar los métodos en imágenes reales. Para esto el trabajo se dividirá en:

1. Modelar matemáticamente el problema
 - Modelar problema en el dominio temporal (o espacial).
 - Calcular el modelo inducido en el dominio espectral.
2. Validar el modelo a través del muestreo espectral de una señal sintética temporal (caso unidimensional).
3. Validar el modelo a través del muestreo espectral de una imagen sintética (caso dos-dimensional).
4. Probar el modelo en imágenes interferométricas reconstruidas por algoritmos tradicionales.

1.5. Aporte

El principal aporte de este trabajo consiste en el desarrollo de algoritmos capaces de resolver el problema de estimación espectral a través del muestreo del espectro, y en particular el problema de interferometría. Estos métodos deben ser capaces de caracterizar el error por punto (o píxel) estimado. Además, algo importante de este método es que, al igual que Lomb-Scargle [10, 11], entrega una estimación completa del espectro, es decir, de su parte real y su parte imaginaria en conjunto y no estima directamente la potencia espectral.

El modelo paramétrico es un aporte en sí para la caracterización de los cuerpos celestes, esto da un modelo matemático capaz de hacer la regresión y por esta misma razón, al ser un modelo paramétrico, el modelo puede ayudar a analizar y caracterizar los cuerpos celestes. En tanto, el modelo no-paramétrico es un terreno no explorado en cuanto a transformar distribuciones normales utilizando transformaciones lineales, en particular, transformando al dominio de Fourier.

1.6. Estructura

La tesis está estructurada de la siguiente manera: El capítulo 2 presenta el marco teórico, el cual contiene los fundamentos básicos para entender la totalidad de los métodos propuestos; el capítulo 3 expone el modelamiento teórico de los métodos propuestos, esto implica la modelación de la señal, la modelación del proceso de muestro y la forma de entrenamiento en cada modelo; El capítulo 4 muestra los principales resultados de cada modelo, en ella se inicia con resultados del caso unidimensional, el caso dos dimensional y el caso con imágenes reales de cuerpos celestes; finalmente, el capítulo 4 presenta las principales conclusiones asociadas al desarrollo de la tesis, además de presentar los trabajos a futuro de la misma.

Capítulo 2

Marco Teórico

2.1. Operaciones en matrices

A continuación se introducen algunas operaciones matriciales que serán útiles a la hora de modelar el problema.

2.1.1. Producto de Kronecker

El producto de Kronecker [12] es una operación entre dos matrices que da como resultado una matriz definida por bloques. Para una definición formal de producto de Kronecker ver Apéndice 7.2.

Proposición 2.1 *Sea A, X, B matrices en $\mathbb{R}^{n_A \times m}$, $\mathbb{R}^{m \times n}$ y $\mathbb{R}^{n \times m_B}$ respectivamente. Entonces la vectorización del producto se puede escribir como:*

$$AXB = C \quad / \text{vec}(\cdot) \quad (2.1)$$

$$\text{vec}(AXB) = \text{vec}(C) \quad (2.2)$$

$$(B^T \otimes A)\text{vec}(X) = \text{vec}(C) \quad (2.3)$$

2.1.2. Descomposición de Cholesky

La descomposición (o factorización) de cholesky es una propiedad de las matrices definidas positivas útil para resolver sistemas lineales con simetría, o para definir matrices de covarianzas definidas positivas [13].

Definición 2.2 (Descomposición de Cholesky) *Sea A matriz definida positiva, entonces A admite una descomposición de Cholesky dada por una matriz triangular inferior L y su traspuesta tal que*

$$A = LL^\top \tag{2.4}$$

donde L es llamada matriz de Cholesky o factor de Cholesky.

2.2. Fundamentos de probabilidades

2.2.1. Variable aleatoria

El nombre variable aleatoria (v.a.) es sugerente, puesto que entrega la intuición de que una variable aleatoria es una variable que toma valores aleatoriamente. Esta última es una buena interpretación, pero para ser precisos, una variable aleatoria no es una variable y tampoco es aleatoria. Para ser precisos, una variable aleatoria es una función entre el espacio de probabilidad y un espacio medible \mathbb{I} .

Definición 2.3 (Variable aleatoria) *Una variable aleatoria X se define sobre (Ω, \mathcal{A}, P) un espacio de probabilidad y (E, \mathcal{E}) espacio medible, entonces se dice que X es una variable aleatoria cuando:*

$$X : \Omega \rightarrow E \tag{2.5}$$

Donde Ω es el espacio muestral, \mathcal{A} σ -álgebra sobre Ω y P función de probabilidad. Además, E es un conjunto y \mathcal{E} es una σ -álgebra sobre E .

Observación Usualmente se tiene que $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ en este caso se habla de variable aleatoria a valores reales.

2.2.1.1. Variable aleatoria a valores complejos

Una variable aleatoria compleja es una generalización de una variable aleatoria a valores reales donde el espacio de llegada son complejos.

Definición 2.4 (Variable aleatoria compleja) *Se dice que Z es una v.a. valores complejos (o simplemente una v.a. compleja) del espacio de probabilidad (Ω, \mathcal{A}, P) , si el espacio medible es $(\mathbb{C}, \mathcal{B}(\mathbb{C}))$, es decir, \mathbf{Z} es una función definida por:*

$$Z : \Omega \rightarrow \mathbb{C} \tag{2.6}$$

y además $\Re\{Z\}$, $\Im\{Z\}$ son v.a. sobre (Ω, \mathcal{A}, P) .

2.2.2. Vector Aleatorio

Definición 2.5 (Vector aleatorio) *Un vector aleatorio ($\vec{v}.a.$) o multivariable aleatoria*

$$\mathbf{X} = (X_1, \dots, X_n)^\top$$

es un vector columna, donde $\forall i$ X_i es v.a. sobre (Ω, \mathcal{A}, P) .

2.2.3. Distribución Gaussiana multivariada

Definición 2.6 (Distribución normal multivariada) *Sea \mathbf{X} un $\vec{v}.a.$, se dice que \mathbf{X} distribuye como una Gaussiana multivariada si su función de distribución es*

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2} |\Sigma_{\mathbf{X}}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu)^\top \Sigma_{\mathbf{X}}^{-1}(\mathbf{x} - \mu)\right) \quad (2.7)$$

y se denota como:

$$\mathbf{X} \sim \mathcal{MVN}(\mu, \Sigma_{\mathbf{X}}) \text{ o simplemente } \mathbf{X} \sim \mathcal{N}(\mu, \Sigma_{\mathbf{X}}) \quad (2.8)$$

Donde:

$$\mathbb{E}[\mathbf{X}] = \mu_{\mathbf{X}} \quad (2.9)$$

$$\text{Cov}[\mathbf{X}, \mathbf{X}] = \mathbb{E}[(\mathbf{X} - \mu_{\mathbf{X}})^\top (\mathbf{X} - \mu_{\mathbf{X}})] = \Sigma_{\mathbf{X}} \text{ semi-definida positiva.} \quad (2.10)$$

2.2.3.1. Distribución Gaussiana multivariada condicional

Sea \mathbf{X} e \mathbf{Y} $\vec{v}.a.$ de modo que el $\vec{v}.a.$ conjunto distribuye como una Gaussiana multivariada, es decir,

$$\begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} \sim \mathcal{MVN}\left(\begin{bmatrix} \mu_{\mathbf{X}} \\ \mu_{\mathbf{Y}} \end{bmatrix}, \mathcal{K}\right) \quad (2.11)$$

donde

$$\mathcal{K} = \begin{bmatrix} \text{Cov}[\mathbf{X}, \mathbf{X}] & \text{Cov}[\mathbf{X}, \mathbf{Y}] \\ \text{Cov}[\mathbf{Y}, \mathbf{X}] & \text{Cov}[\mathbf{Y}, \mathbf{Y}] \end{bmatrix} = \begin{bmatrix} \Sigma_{\mathbf{X}} & \Sigma_{\mathbf{X}, \mathbf{Y}} \\ \Sigma_{\mathbf{Y}, \mathbf{X}} & \Sigma_{\mathbf{Y}} \end{bmatrix}. \quad (2.12)$$

Entonces, la variable aleatoria condicional $\mathbf{X}|\mathbf{Y}$ también es una gaussiana multivariada de manera que:

$$\mathbb{E}[\mathbf{X}|\mathbf{Y}] = \mu_{\mathbf{X}} - \Sigma_{\mathbf{X}, \mathbf{Y}} \Sigma_{\mathbf{Y}}^{-1} (\mathbf{Y} - \mu_{\mathbf{Y}})^\top \quad (2.13)$$

$$\text{Cov}[\mathbf{X}|\mathbf{Y}] = \Sigma_{\mathbf{X}} - \Sigma_{\mathbf{X}, \mathbf{Y}} \Sigma_{\mathbf{Y}}^{-1} \Sigma_{\mathbf{X}, \mathbf{Y}}^\top. \quad (2.14)$$

2.2.4. Distribución Gaussiana multivariada a valores complejos

Definición 2.7 (Distribución normal multivariada compleja) *Sea \mathbf{Z} un $\vec{v}.a.$ complejo de dimensión k , se dice que \mathbf{Z} distribuye Gaussiana y se denota como*

$$\mathbf{Z} \sim \mathcal{C}\text{-MVN}(\mu_{\mathbf{Z}}, C, P) \text{ o simplemente } \mathbf{Z} \sim \mathcal{C}\text{-N}(\mu_{\mathbf{Z}}, C, P) \quad (2.15)$$

Si la función de distribución está dada por:

$$f_{\mathbf{Z}}(\mathbf{z}) = \frac{1}{\pi^k (|\mathbf{R}_{zz}|)^{1/2}} \exp \left\{ -\frac{1}{2} [(\bar{\mathbf{z}} - \bar{\mu})^\top \quad (\mathbf{z} - \mu)^\top] \underbrace{\begin{bmatrix} C & P \\ \bar{P} & \bar{C} \end{bmatrix}^{-1}}_{\mathbf{R}_{zz}^{-1}} \begin{bmatrix} \mathbf{z} - \mu \\ \bar{\mathbf{z}} - \bar{\mu} \end{bmatrix} \right\}, \quad (2.16)$$

donde $\mu_{\mathbf{Z}}$ es la esperanza de la variable aleatoria, C su covarianza y P su pseudo-covarianza, los cuales están definidos como:

$$\mu_{\mathbf{Z}} = \mathbb{E}[\mathbf{Z}] \quad (2.17)$$

$$C = \mathbb{E}[(\mathbf{Z} - \mu_{\mathbf{Z}})^\dagger (\mathbf{Z} - \mu_{\mathbf{Z}})] \quad (2.18)$$

$$P = \mathbb{E}[(\mathbf{Z} - \mu_{\mathbf{Z}})^\top (\mathbf{Z} - \mu_{\mathbf{Z}})]. \quad (2.19)$$

El operador $(\cdot)^\dagger$ representa el operador hermítico. Además, notar que la covarianza $C \in \mathbb{R}$, mientras que la pseudo-covarianza $P \in \mathbb{C}$.

2.3. Modelos de mezcla de bases

Los modelos de mezcla de bases, también llamados suma de bases, son modelos que están determinados por la suma ponderada de funciones base ϕ_i . Es decir, sea $y \in Y$, $A = \{\alpha_i\}_{i=1}^N$ donde $\forall i \quad \alpha_i \in \mathbb{R}$, entonces el modelo de mezclas de bases está dado por

$$y(x) = \sum_{i=1}^N \alpha_i \phi_i(x). \quad (2.20)$$

Donde las funciones $\forall i \quad \phi_i \in Y$ usualmente tienen parámetros θ_i asociados, de esta manera, usualmente se puede escribir

$$\phi_i(x) = f(x, \theta_i). \quad (2.21)$$

Es importante notar que $\theta_i \in \mathbb{R}^d$, con d la cantidad de parámetros de la base, de este modo, la cantidad de parámetros totales del modelo de mezcla de bases es $N \cdot (d + 1)$.

2.4. Estimación

El problema de estimación de parámetros se centra en encontrar un valor que aproxime un parámetro, usualmente de un modelo asociado, para esto, las dos técnicas más utilizadas son *estimador de máxima verosimilitud* y el *estimador de máximo a posteriori* [14, 15], este último se encuadra en el contexto de *inferencia Bayesiana*. Por este motivo, describiremos y definiremos ambos estimadores.

2.4.1. Estimador de máxima verosimilitud

Definición 2.8 (Función de verosimilitud) *Sea \mathbf{Y} \vec{v} .a. de observaciones siendo y una realización de este, además sea \mathbf{X} \vec{v} .a. del modelo latente, entonces se define la función de verosimilitud como:*

$$\mathcal{L}(\mathbf{X}, y) = \mathbb{P}(\mathbf{Y} = y | \mathbf{X}) \quad (2.22)$$

Definición 2.9 (Estimador de máxima verosimilitud) *Se define el estimador de máxima verosimilitud como el \mathbf{X} que maximiza la función de verosimilitud:*

$$\hat{\mathbf{X}}_{ML} = \arg \max_{\mathbf{X}} \mathcal{L}(\mathbf{X}, y). \quad (2.23)$$

Para efectos computacionales, y dado que usualmente se utilizan funciones de las familias de las exponenciales para \mathbf{X} e \mathbf{Y} , es común computar la función de log-verosimilitud negativa

$$\text{NLL}(\mathbf{X}, y) = -\log(\mathcal{L}(\mathbf{X}, y)), \quad (2.24)$$

donde el estimador de máxima verosimilitud está dado por:

$$\hat{\mathbf{X}}_{ML} = \arg \min_{\mathbf{X}} \text{NLL}(\mathbf{X}, y). \quad (2.25)$$

Observación Cuando la variable aleatoria \mathbf{Y} distribuye como una Gaussiana multivariable, entonces la expresión de NLL tiene la siguiente expresión cerrada:

$$\text{NLL}(\mathbf{X}, y) = \log(2\pi|\Sigma_{\mathbf{Y}}|) + \frac{1}{2}(y - \mu_{\mathbf{Y}})^\top \Sigma_{\mathbf{Y}}^{-1}(y - \mu_{\mathbf{Y}}). \quad (2.26)$$

2.4.2. Inferencia Bayesiana

La inferencia Bayesiana [14, 8] es un potente marco de trabajo que permite realizar inferencia basados en el teorema de Bayes, el cual fue propuesto por Thomas Bayes en el año 1763. El principal fuerte de este marco de trabajo es su capacidad natural para integrar la incertidumbre natural del problema a modelar.

Teorema 2.10 (Bayes) Sea X e Y $\vec{v}.a.$, el teorema de Bayes enuncia que es posible obtener la probabilidad de la condicional $\mathbb{P}(\mathbf{X}|\mathbf{Y})$ conociendo la condicional inversa $\mathbb{P}(\mathbf{Y}|\mathbf{X})$ y las probabilidades marginales $\mathbb{P}(\mathbf{X})$ y $\mathbb{P}(\mathbf{Y})$:

$$\mathbb{P}(\mathbf{X}|\mathbf{Y}) = \frac{\mathbb{P}(\mathbf{Y}|\mathbf{X})\mathbb{P}(\mathbf{X})}{\mathbb{P}(\mathbf{Y})}. \quad (2.27)$$

Este teorema tiene un importante alcance a la hora de hacer estimación, puesto que permite encontrar la distribución posterior del modelo.

2.4.2.1. Máximo a posteriori

El estimador de máximo a posteriori (MAP) es una aplicación del teorema de Bayes. Sea \mathbf{Y} $\vec{v}.a.$ de observaciones, además \mathbf{X} $\vec{v}.a.$ del modelo latente, entonces gracias al teorema de Bayes sabemos que es posible obtener la probabilidad posterior, conociendo el función de verosimilitud y las probabilidades marginales.

Definición 2.11 (Estimador de máximo a posteriori) *El estimador de Máximo a posteriori se define como se define como:*

$$\hat{\mathbf{X}}_{MAP} = \arg \max_{\mathbf{x}} \mathbb{P}(\mathbf{Y} = y|\mathbf{X}) \quad (2.28)$$

Observación Aplicando el teorema de Bayes se tiene que:

$$\begin{aligned} \hat{\mathbf{X}}_{MAP} &= \arg \max_{\mathbf{x}} \frac{\mathbb{P}(\mathbf{Y} = y|\mathbf{X})\mathbb{P}(\mathbf{X})}{\mathbb{P}(\mathbf{Y} = y)} \\ &= \arg \max_{\mathbf{x}} \frac{\mathcal{L}(X, y)\mathbb{P}(\mathbf{X})}{\mathbb{P}(\mathbf{Y} = y)}, \end{aligned} \quad (2.29)$$

Se debe notar que esta expresión es similar a maximizar la verosimilitud en conjunto a $\mathbb{P}(\mathbf{X})$ puesto que el término $\mathbb{P}(\mathbf{Y} = y)$ funciona como un factor de escala y no depende de \mathbf{X} , ya que si \mathbb{X} el espacio donde vive \mathbf{X} , entonces:

$$\mathbb{P}(\mathbf{Y} = y) = \int_{\mathbf{x} \in \mathbb{X}} P(\mathbf{Y} = y|\mathbf{X})P(\mathbf{X})d\mathbf{X}. \quad (2.30)$$

Finalmente, nuevamente por efectos computacionales y dado que usualmente las funciones a optimizar son de la familia de las exponenciales, se suele minimizar la versión neg-log del problema:

$$\begin{aligned} \hat{\mathbf{X}}_{MAP} &= \arg \min_{\mathbf{x}} -\log \mathbb{P}(\mathbf{Y} = y|\mathbf{X}) \\ &= \arg \min_{\mathbf{x}} \text{NLL}(\mathbf{X}, y) - \log \mathbb{P}(\mathbf{X}) + \log \mathbb{P}(\mathbf{Y}) \\ &= \arg \min_{\mathbf{x}} \text{NLL}(\mathbf{X}, y) - \log \mathbb{P}(\mathbf{X}). \end{aligned} \quad (2.31)$$

2.5. Algoritmos de optimización

Los algoritmos aquí presentados intentan resolver una variante particular de los problemas de optimización, en que no hay restricciones en las funciones. De esta manera, el problema a resolver es:

$$\min_x f(x) \tag{2.32}$$

Si $f(x)$ es diferenciable y convexa, entonces una condición necesaria y suficiente que debe cumplir el punto óptimo x^* está dado por:

$$\nabla f(x^*) = 0 \tag{2.33}$$

Luego encontrar el punto que minimiza la función f se reduce a resolver la ecuación (2.33). En algunos casos es posible encontrar analíticamente la solución para (2.33), pero lo normal es que la ecuación sea lo suficientemente complicada como para no poder resolverla analíticamente, es por este motivo que se utilizan algoritmos de optimización, los cuales usualmente buscan iterativamente candidatos x_n para x^* de modo que:

$$\lim_{n \rightarrow \infty} f(x_n) = f(x^*) \tag{2.34}$$

De este modo, se detiene el algoritmo cuando se llega a un n_0 tal que

$$f(x_{n_0}) - f(x_{n_0-1}) < \varepsilon, \tag{2.35}$$

con $\varepsilon > 0$ la precisión deseada [16, 17].

2.5.1. Método de Newton

El método de Newton [18] es un método iterativo para resolver el sistema en (2.32). Este se basa en actualizar la propuesta x_k con la actualización

$$x_{k+1} = x_k - \alpha \nabla^2 f(x_k)^{-1} \nabla f(x_k) \tag{2.36}$$

donde $\nabla(\cdot)$ y $\nabla^2(\cdot)$ es el operador Jacobiano y Hessiano respectivamente. Además, generalmente se utiliza α variable en cada paso, aunque algunas variaciones del algoritmo utilizan α fijo. El algoritmo se resume en el Pseudo-código 1.

Algorithm 1 Método de Newton

- 1: Inicializar λ^2
 - 2: **while** $\lambda^2/2 < \text{eps}$ **do**
 - 3: Calcular p_k resolviendo $\Delta x = -\nabla^2 f(x_k)^{-1} \nabla f(x_k)$
 - 4: Calcular $\lambda^2 = \nabla f(x_k)^\top \nabla^2 f(x_k)^{-1} \nabla f(x_k)$
 - 5: Determinar α utilizando una búsqueda lineal
 - 6: Actualizar $x_{k+1} = x_k + \alpha \Delta x$
 - 7: **end while**
-

2.5.2. Broyden–Fletcher–Goldfarb–Shanno (BFGS)

El algoritmo de optimización BFGS es un método numérico para optimizar funciones escalares sin restricciones [19]. El pseudo-código del algoritmo se muestra en 2. Este método se dice de cuasi-Newton, puesto que se basa en el método de Newton, pero no utiliza el Hessiano exacto para actualizar, más bien, utiliza una aproximación de este dada por:

$$B_{k+1} = B_k + \frac{y_k y_k^\top}{y_k^\top s_k} - \frac{B_k s_k s_k^\top B_k^\top}{s_k^\top B_k s_k} \quad (2.37)$$

donde

$$s_k = \alpha p_k \quad (2.38)$$

$$y_k = \nabla f(x_{k+1}) - \nabla f(x_k). \quad (2.39)$$

Algorithm 2 BFGS

- 1: Inicializar B_0
 - 2: Inicializar $g_0 = \nabla f(x_0)$
 - 3: **while** $\|\nabla f(x_k)\| < \text{eps}$ **do**
 - 4: Calcular p_k resolviendo $B_k p_k = -\nabla f(x_k)$
 - 5: Calcular $s_k = \alpha p_k$
 - 6: Actualizar $x_{k+1} = x_k + s_k$
 - 7: Calcular $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$
 - 8: $B_{k+1} = B_k + \frac{y_k y_k^\top}{y_k^\top y_k} - \frac{B_k y_k y_k^\top B_k^\top}{y_k^\top B_k y_k}$
 - 9: **end while**
-

2.5.3. Powell

El método de Powell no es un algoritmo basado en derivadas, y es similar al “*algoritmo del taxi*”, en el que la búsqueda del mínimo se realiza “avanzando en diferentes paradas (puntos)” [20]. El método se resume en el pseudo-código del algoritmo 3. Powell ajusta la función

$$p(\alpha) = a\alpha^2 + b\alpha + c, \quad (2.40)$$

utilizando tres puntos y encuentra su mínimo, luego actualiza uno de los puntos para avanzar en cada paso iterativo.

Algorithm 3 Powell

```
1: Inicializar  $p_0 = f(x_0)$ 
2: Evaluar  $p_1 = p(\beta) = f(x_0 + \beta)$ 
3: for cada  $i = 1$  hasta  $n$  do
4:   if  $p_1 < p_0$  then
5:      $p(2\beta) = f(x_0 + 2\beta)$ 
6:   else
7:      $p(-\beta) = f(x_0 - \beta)$ 
8:   end if
9:   Resolver ecuaciones para  $a, b, c$ 
10:  Calcular  $\alpha^* = -b/2c$ 
11:  if  $\alpha^*$  es mínimo de  $p(\alpha)$  then
12:     $p_0 = p_1$ 
13:     $p_1 = p(\alpha^*)$ 
14:  else
15:     $p_0 = \arg \min_p \{|p(\alpha^*) - p_0|, |p(\alpha^*) - p_1|\}$  # tomar el punto más cercano a  $p(\alpha^*)$ 
16:     $p_1 = p(\alpha^*)$ 
17:  end if
18: end for
```

2.6. Radio Astronomía: Interferometría

La interferometría es una técnica utilizada en radioastronomía para reconstruir señales espaciales (imágenes). Esta técnica permite obtener una resolución mucho mayor que cualquier telescopio existente, puesto que se puede hacer una analogía entre esta técnica y un telescopio óptico, en la que el diámetro del telescopio está en directa relación con la distancia más grande entre dos antenas.

La técnica de interferometría involucra antenas apuntando al mismo punto en el cielo nocturno, las cuales captan la señal proveniente del espacio. El computador donde se procesa dicha información se llama *correlador*, puesto que el procesamiento que se realiza es precisamente correlacionar las señales captadas de pares de antenas en un período determinado, usualmente del orden de milisegundos a un par de segundos. El objetivo de este computo es obtener las muestras del espectro espacial de la imagen, la cual está respaldada por el teorema de Van Cittert-Zernike [3]. Dicha frecuencia¹ está determinada por el vector entre ambas antenas correlacionadas, denominado *baseline*.

Teorema 2.12 (Van Cittert-Zernike) *Sea $\mathcal{U} \times \mathcal{V}$ el espacio bidimensional de frecuencias y $\mathcal{X} \times \mathcal{Y}$ el espacio local (o espacio de la imagen), el teorema enuncia que si un interferómetro*

¹No confundir la frecuencia de la onda electromagnética con la frecuencia espacial.

tiene un campo de visión $A(x, y)$, $(x, y) \in X \times Y$, entonces la visibilidad $V(u, v)$, $(u, v) \in \mathcal{U} \times \mathcal{V}$, de un frente de onda de intensidad $I(x, y)$, $(x, y) \in X \times Y$, está dada por:

$$V(u, v) = \int \int_{\mathcal{X} \times \mathcal{Y}} A(x, y) I(x, y) \exp\{-2\pi j(ux + vy)\} dx dy \quad (2.41)$$

Observación Si el campo de visión es suficientemente pequeño, entonces la salida del correlador (mediciones) y las visibilidades coinciden. Luego la medición tomada por un par de antenas con *baseline* proyectado $\hat{\mathbf{D}}_{\lambda_0} = (\hat{u}, \hat{v})$ permite obtener $V(\hat{u}, \hat{v})$. De esta manera, si no hay redundancia en el proceso de grillado, la cantidad de mediciones simultaneas estaría dada por $\frac{N_a(N_a-1)}{2}$, con N_a la cantidad de antenas.

El principal problema de esta técnica es que sin importar la cantidad de antenas dispuestas en terreno, no se puede medir el plano de Fourier completo. Para ejemplificar esto, si se quiere medir la frecuencia $(\hat{u}, \hat{v}) = (0, 0)$, esto implicaría tener un *baseline* proyectado $\hat{\mathbf{D}}_{\lambda_0} = (0, 0)$ lo cual es físicamente imposible, puesto que esto implicaría tener dos antenas en la misma posición, es decir, una encima de otra. Con esto en mente, se puede notar que el problema de interferometría es estimar la imagen $I(x, y)$, a través del muestreo parcial del plano de Fourier $V(u, v)$.

2.6.1. Imagen Sucia y Haz Sucio

Un primer acercamiento para abordar el problema es utilizar la *imagen sucia* I_d [4] (*dirty image*, en inglés), que consiste en tomar la transformada inversa de Fourier de las mediciones existentes y las no existentes asignarles el valor cero. Formalmente, sea $S(u, v) \in \mathbb{R}$ la función de muestreo que queda definida por:

$$S(u, v) = \begin{cases} 1 & \text{si } (u, v) \text{ es medido} \\ 0 & \sim . \end{cases} \quad (2.42)$$

Además, sea $V_d \in \mathbb{C}$ el haz sucio (*dirty beam*, en inglés), $I_d = \mathcal{F}^{-1}[V_d(u, v)] \in \mathbb{C}$ y $B(x, y) = \mathcal{F}^{-1}[S(u, v)] \in \mathbb{C}$ el mapa sucio (*dirty map*, en inglés). Entonces el proceso de interferometría está completamente definido por:

$$V_d(u, v) = S(u, v) \cdot V(u, v) \quad (2.43)$$

$$I_d(x, y) = B(x, y) * I(x, y). \quad (2.44)$$

2.6.2. Snapshot e interferometría dinámica

El problema de interferometría consiste en obtener mediciones el espectro, estas mediciones están completamente determinadas por la forma del arreglo de antenas utilizado y además las muestras observadas dependen de la posición de la Tierra respecto al cuerpo. De esta manera, el movimiento de la Tierra define trayectorias elípticas para las observaciones, por ende se pueden definir dos casos:

- **El caso estático o *snapshot***, que se refiere a tomar observaciones en un intervalo de tiempo muy pequeño, lo cual produce muestras puntuales como las de la figura 2.1b.
- **El caso dinámico** que corresponde a la horas de observación. Cuando se considera el movimiento de la tierra, las observaciones generan caminos elípticos como los mostrados en la Figura 2.1c. Estos caminos han sido estudiados en [3, 21, 22].

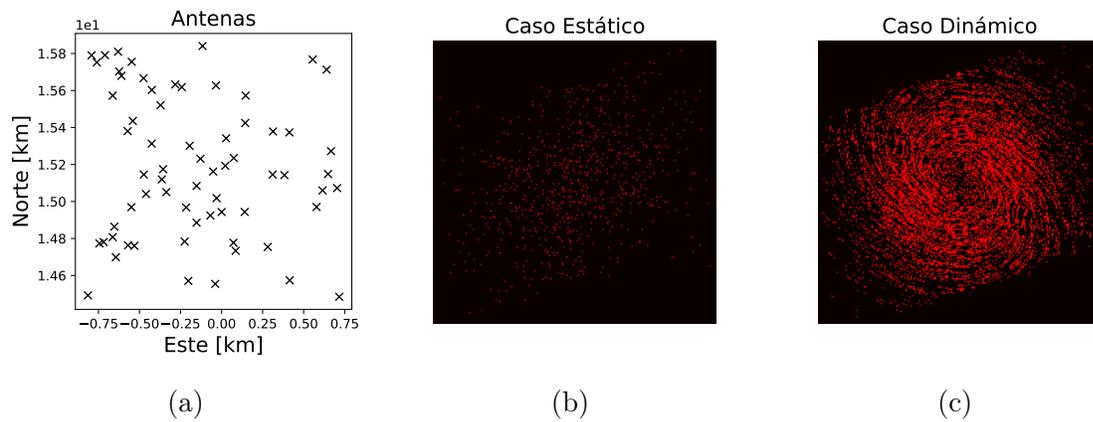


Figura 2.1: Arreglo de antenas y píxeles muestreados. (a) Arreglo de antenas. (b) Muestreo estático o *snapshot*. (c) Muestreo dinámico.

Capítulo 3

Propuesta

La propuesta para resolver este problema se basa en dos modelos Bayesianos, para lo cual se seguirá la misma estrategia tanto en el modelo paramétrico, como en el modelo no-paramétrico. Esta estrategia está resumida en el modelo gráfico de la Figura 3.1. En ella se resume el problema y se muestra la idea de modelar el espacio donde la señal, para obtener un modelo para el espacio donde vive el espectro.

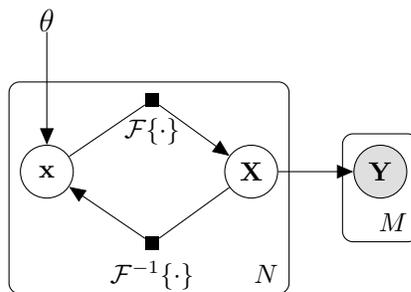


Figura 3.1: Modelo gráfico del problema. Se tiene la señal en el tiempo (o espacio), que está ligada a través de la transformada de Fourier a su espectro y viceversa. Las observaciones se realizan en el espectro. La notación θ representa los parámetros de entrada θ para el modelo paramétrico y los hiper-parámetros del kernel $\mathcal{K}_\theta(x, x')$ en el caso del modelo no-paramétrico.

3.1. Modelo paramétrico: Suma de Gaussianas

El desarrollo de una primera versión de este modelo fue presentado en [23]. Considere una imagen $I : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ y un modelo para esta basado en una mezcla de Gaussianas (SoG por su sigla en inglés *Sum of Gaussians*):

$$I(x, y) = \sum_{i=0}^{N_B-1} \beta_i \psi_i(x, y) \quad (3.1)$$

con $\psi_i(x, y)$ las funciones bases dadas por

$$\psi_i(x, y) = \exp \left\{ -\frac{1}{2} \begin{bmatrix} x - C_x^i \\ y - C_y^i \end{bmatrix}^T \Sigma_i^{-1} \begin{bmatrix} x - C_x^i \\ y - C_y^i \end{bmatrix} \right\}, \quad (3.2)$$

$$\Sigma_i = \begin{bmatrix} \sigma_{1i}^2 & \sigma_{12i} \\ \sigma_{12i} & \sigma_{2i}^2 \end{bmatrix} \quad (3.3)$$

la cual está compuesta por σ_{1i}^2 varianza de x , σ_{2i}^2 varianza de y , $\sigma_{12i} = \sigma_{21i}$ covarianza en x, y , además los centros C_x^i, C_y^i en el dominio espacial.

De acuerdo al teorema de Van Cittert-Zernike [4], asumiendo un área de visión lo suficientemente pequeña y aprovechando las propiedades lineales de la transformada de Fourier, la visibilidad $V : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{C}$ es también una mezcla de exponenciales a valores complejos dados por la ecuación [24, 23]:

$$V(u, v) = \sum_{i=1}^{N_B} \alpha_i \phi_i(u, v) \quad (3.4)$$

donde:

$$\phi_i(u, v) = \exp \left\{ 2\pi^2 \frac{a_i v^2 + b_i u^2 - c_i uv}{4a_i b_i - c_i^2} - 2\pi j(e_i v + d_i u) \right\} \quad (3.5)$$

$$\alpha_i = \frac{\beta_i}{\sqrt{4a_i b_i - c_i^2}}, \quad (3.6)$$

con

$$a_i = -\frac{1}{2} \frac{\sigma_{2i}^2}{\sigma_{1i}^2 \sigma_{2i}^2 - \sigma_{12i}^2}; b_i = -\frac{1}{2} \frac{\sigma_{1i}^2}{\sigma_{1i}^2 \sigma_{2i}^2 - \sigma_{12i}^2}; c_i = \frac{\sigma_{12i}}{\sigma_{1i}^2 \sigma_{2i}^2 - \sigma_{12i}^2} d_i = C_x^i; e_i = C_y^i. \quad (3.7)$$

Observación Notemos que si bien este modelo está escrito para el caso dos-dimensional, en el caso unidimensional es análogo y basta notar que solo existe un solo centro y la matriz de covarianza es solo un escalar que representa el *lengthscale* de la Gaussiana.

3.1.1. Modelo de ruido

Recordando que las observaciones están en el dominio de los complejos, se presenta el modelo de ruido en el caso general a valores complejos, similar a lo propuesto en [25], esto es introduciendo correlación entre los canales de la parte real e imaginaria, basado en un proceso latente propuesto en la ecuación (3.4).

$$V_{obs}(u, v) = \sum_{i=1}^{N_B} \alpha_i \phi_i(u, v) + \eta(0, C, P) \quad (3.8)$$

donde $\eta(0, C, P)$ es una Gaussiana multivariada compleja mencionada en la sección 2.2.4.

3.1.2. Buscando la distribución posterior del modelo

El objetivo de esta propuesta es encontrar una representación de la incertidumbre del espectro latente resultado de las observaciones no-medidas y del ruido del modelo. Por este motivo, se toma un enfoque completamente Bayesiano y calcularemos la distribución posterior de los parámetros del modelo propuesto. En ese sentido, considerando $U = \{u_k, v_k\}_{k=1}^{N_s}$ y las observaciones por $V = \{V_k\}_{k=1}^{N_s}$, luego la distribución posterior del modelo paramétrico (denotando $\theta = \{C, P, \Sigma_i, \alpha, C_x^i, C_y^i\}$) está dada por

$$p(\theta|U, V) = p(\theta) \prod_{k=1}^{N_s} p(V_k, u_k, v_k|\theta) \quad (3.9)$$

$$= p(\theta) \prod_{k=1}^{N_s} \frac{1}{\pi|\mathbf{R}_{zz}|} \exp \left\{ -\frac{1}{2} e_k^T \mathbf{R}_{zz}^{-1} e_k \right\} \quad (3.10)$$

donde \mathbf{R}_{zz} está definido mediante C y P en la sección 2.2.4 y e_k es el error o discrepancia entre el valor propuesto por el modelo y la k -ésima observación

$$e_k = \begin{bmatrix} V_k - V(u_k, v_k) \\ \bar{V}_k - V(u_k, v_k) \end{bmatrix}. \quad (3.11)$$

Encontrar la posterior completa permite calcular barras de error y otros estadísticos de mayor orden. En particular, también se puede calcular el máximo a posteriori de los parámetros para una estimación puntual. Este está dado por

$$\theta^* = \underset{C, P, l, \alpha_i, C_x^i, C_y^i}{\operatorname{argmax}} p(C, P, l, \alpha_i, C_x^i, C_y^i|U, V). \quad (3.12)$$

Estos parámetros encontrados, pueden ser utilizados para calcular la imagen I .

3.1.3. Eligiendo el prior sobre los parámetros

Puesto que se utilizará máximo a posteriori, se debe elegir el prior sobre los parámetros $p(\theta)$. Después de evaluar distintos priors y hacer una revisión de la literatura para distintos priors, se seleccionó los siguientes priors para los centros y los pesos:

$$\begin{aligned} C^i &\sim \mathcal{N}((S_x, S_y)/2, \sigma_C^2) \\ \alpha_i &\sim \mathcal{N}(0, 1/\sigma_\alpha^2) \\ \sigma_\alpha &\sim \mathcal{N}(0, \sigma^2) \end{aligned}$$

donde S_x y S_y son los largos de la imagen en el eje x e y respectivamente. Notar que el prior para los pesos α_i está construido de manera jerárquica por lo que la distribución de α queda completamente determinada por la ecuación (3.13).

$$p(\alpha) = \frac{1}{\sqrt{2\pi}l} \exp\left(-\frac{\alpha^2\sigma_\alpha^2}{2}\right) \int_0^\infty \exp\left(-\frac{\sigma_\alpha^2}{\sigma^2}\right) d\sigma_\alpha \quad (3.13)$$

Para asegurar la propiedad de definida positiva de la matriz de covarianza, se utilizó un distribución prior de acuerdo al método de Lewandowski-Kurowicka-Joe (LKJ) [26].

$$R_{zz} \sim \text{LKJCholesky}(\eta_0, d_0).$$

Para el caso de los coeficientes de la matriz Σ_i , se aseguró la positividad de la matriz a través de la factorización de Cholesky:

$$\Sigma_i = \Lambda \Lambda^\top \quad (3.14)$$

$$\begin{bmatrix} \sigma_{1i}^2 & \sigma_{12i} \\ \sigma_{12i} & \sigma_{2i}^2 \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 \\ \lambda_2 & \lambda_3 \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ \lambda_2 & \lambda_3 \end{bmatrix}^\top \quad (3.15)$$

$$= \begin{bmatrix} \lambda_1^2 & \lambda_1\lambda_2 \\ \lambda_1\lambda_2 & \lambda_2^2 + \lambda_3^2 \end{bmatrix}. \quad (3.16)$$

Luego sobre los parámetros λ_1 , λ_2 y λ_3 se define las distribuciones:

$$\lambda_1 \sim \text{Half} - \mathcal{N}(\sigma_{\lambda_1}) \quad \lambda_2 \sim \mathcal{N}(0, \sigma_{\lambda_2}^2) \quad \lambda_3 \sim \text{Half} - \mathcal{N}(\sigma_{\lambda_3}). \quad (3.17)$$

3.2. Modelo no-paramétrico: Reconstrucción espectral Bayesiana

Se procederá definiendo el modelo para series de tiempo discretas. La elección de este modelo será de una distribución multivariable aleatoria normal (MVN) con media y covarianza arbitraria. Por consiguiente, se modelará la serie $\mathbf{x} = [x_1, \dots, x_N]^\top \in \mathbb{R}^N$, donde el subíndice de $x_i \in \mathbb{R}$ denota el tiempo, distribuyendo como

$$\mathbf{x} \sim \text{MVN}(m, \Sigma) \quad (3.18)$$

Cabe destacar que considerar una distribución normal es muy general. A través del control de la función de covarianza se puede general estructuras lo suficientemente complejas. De hecho, muchos métodos asumen implícitamente Gaussianidad o ellos pueden ser derivados de asumir Gaussianidad. Por ejemplo, los modelos AR producen caminos Gaussianos, el método de Lomb-Scargle [10, 11] reconstruye los valores faltantes interponiendo componentes sinusoidales que pueden ser replicadas con un modelo MVN con matriz de covarianza periódica, y el periodograma de Welch [27] promedia sobre periodogramas parciales, a través del supuesto de que el ruido se cancela en promedio.

Después de considerar el modelo generativo MVN para series de tiempo en la ecuación (3.18) se puede, informalmente, *pasar el modelo a través de la transformada de Fourier* para encontrar la distribución que es inducida en el espectro. Esto puede lograr por medio de encontrar la ley del espectro de Fourier de \mathbf{x} , denotando \mathbf{X} , dado por

$$\mathbf{X} = \text{DFT}(\mathbf{x}), \quad \mathbf{x} \sim \text{MVN}(m, \Sigma) \quad (3.19)$$

donde la expresión $\text{DFT}(\cdot)$ denota el operador de la transformada de Fourier discreta. Recordando que la DFT de $\mathbf{x} \in \mathbb{R}^N$ es un vector a valores complejos, que está definido como

$$\mathbf{X}_k = \frac{1}{\sqrt{N}} \sum_{n=1}^N e^{-jkn/N} x_n \quad (3.20)$$

donde j denota la unidad imaginaria y k es el índice de la frecuencia. En forma vectorial, se puede expresar el espectro \mathbf{X} como

$$\mathbf{X} = W^\top \mathbf{x} \quad (3.21)$$

donde W es la matriz de Fourier dada por

$$[W]_{nk} = \frac{1}{\sqrt{N}} e^{-jkn/N} = \frac{1}{\sqrt{N}} (\cos(kn/N) - j \sin(kn/N)). \quad (3.22)$$

Además, de modo de evitar tratar con vectores aleatorios complejos, se define la parte real e imaginaria de la matriz de Fourier respectivamente por

$$W_r = \Re W \quad \text{and} \quad W_i = \Im W \quad (3.23)$$

para representar el espectro de \mathbf{x} a través de su parte real e imaginaria respectivamente:

$$\mathbf{X}_r = W_r^\top \mathbf{x}, \quad \mathbf{X}_i = W_i^\top \mathbf{x}. \quad (3.24)$$

Puesto que la representación de vectores a valores reales del espectro en la ecuación (3.24) es una transformación lineal de la serie de tiempo \mathbf{x} , entonces las transformaciones de los vectores \mathbf{X}_r y \mathbf{X}_i son también conjuntamente Gaussianos, es decir,

$$\begin{bmatrix} \mathbf{X}_r \\ \mathbf{X}_i \end{bmatrix} \sim \text{MVN}(m_{\mathbf{X}}, \Sigma_{\mathbf{X}}) \quad (3.25)$$

donde el vector de medias y la matriz de covarianza está dada simplemente por la aplicación de la transformación lineal en la ecuación (3.24) a la media y a la matriz de covarianza de la serie de tiempo, esto es,

$$m_{\mathbf{X}} = \begin{bmatrix} W_r^\top m \\ W_i^\top m \end{bmatrix} \quad (3.26)$$

$$\Sigma_{\mathbf{X}} = \begin{bmatrix} W_r^T \Sigma W_r & W_r^T \Sigma W_i^T \\ W_i^T \Sigma W_r & W_i^T \Sigma W_i^T \end{bmatrix}. \quad (3.27)$$

El modelo definido anteriormente constituye una distribución prior sobre el espacio de frecuencias discretas y este fue construido transportando la medida de probabilidad sobre el espacio temporal a una medida de probabilidad sobre el espacio frecuencial, a través del operador discreto de Fourier.

3.2.1. Un modelo de máxima verosimilitud para las observaciones ruidosas

El objetivo es realizar inferencia de las observaciones ausentes del espectro usando el modelo generativo definido en la ecuación (3.25). Para hacerlo, introduciremos el siguiente modelo relativo a todos los valores (latentes) del espectro $\mathbf{X}_r, \mathbf{X}_i$ y un subconjunto de observaciones ruidosas correspondientes a la parte real e imaginaria del espectro dado respectivamente por \mathbf{Y}_r y \mathbf{Y}_i y definido por:

$$\mathbf{Y}_r = H^T \mathbf{X}_r + \varepsilon_r \quad (3.28)$$

$$\mathbf{Y}_i = H^T \mathbf{X}_i + \varepsilon_i. \quad (3.29)$$

donde se asume que ε_r y ε_i son ruidos blancos Gaussianos independientes uno de otro con matriz de covarianza diagonal $\sigma_\varepsilon^2 \mathbf{I}$, la cual representa el ruido no corrompe cruzadamente las observaciones, es decir, el ruido para la parte real e imaginaria es independiente. Adicionalmente, la matriz de observación H es definida por elementos por

$$H(p, q) = \begin{cases} 1 & \text{si } \mathbf{X}_r(p) \text{ y } \mathbf{X}_i(p) \text{ es observado a trav\u00e9s de } \mathbf{Y}_r(q) \text{ y } \mathbf{Y}_i(q) \text{ respectivamente,} \\ 0 & \text{si no.} \end{cases} \quad (3.30)$$

Notemos que $H \in \mathcal{M}_{N \times M}$ con $M < N$ el n\u00famero de observaciones. Finalmente, es importante recalcar que la distribuci\u00f3n de $[\mathbf{Y}_r, \mathbf{Y}_i]$ es tambi\u00e9n Gaussiano con media y varianza

$$m_{\mathbf{Y}} = \begin{bmatrix} H^T W_r^T m \\ H^T W_i^T m \end{bmatrix} \quad (3.31)$$

$$\begin{aligned} \Sigma_{\mathbf{Y}} &= \mathbb{E}(\mathbf{Y}^T \mathbf{Y}) \\ &= \mathbb{E}([\mathbf{Y}_r^T, \mathbf{Y}_i^T] \begin{bmatrix} \mathbf{Y}_r \\ \mathbf{Y}_i \end{bmatrix}) \\ &= \begin{bmatrix} H^T W_r^T \Sigma W_r H + \sigma_\varepsilon^2 \mathbf{I} & H^T W_r^T \Sigma W_i H \\ H^T W_i^T \Sigma W_r H & H^T W_i^T \Sigma W_i H + \sigma_\varepsilon^2 \mathbf{I} \end{bmatrix}. \end{aligned} \quad (3.32)$$

3.2.2. Inferencia utilizando la distribuci\u00f3n posterior

La inferencia se realizar\u00e1 utilizando la distribuci\u00f3n posterior $\mathbb{P}(\mathbf{X}_r, \mathbf{X}_i | \mathbf{Y}_r = y_r, \mathbf{Y}_i = y_i)$ (por simplicidad de notaci\u00f3n, simplemente $\mathbb{P}(\mathbf{X} | \mathbf{Y} = y)$) el cual, al ser todos estos vectores conjuntamente Gaussianos, entonces la condicional es tambi\u00e9n Gaussiana. Por lo que esta distribuci\u00f3n queda completamente definida por su media y varianza, entonces:

$$m_{\mathbf{X}|\mathbf{Y}} = m_{\mathbf{X}} + \Sigma_{\mathbf{X},\mathbf{Y}} \Sigma_{\mathbf{Y}}^{-1} (\mathbf{Y} - m_{\mathbf{Y}}) \quad (3.33)$$

$$\Sigma_{\mathbf{X}|\mathbf{Y}} = \Sigma_{\mathbf{X}} - \Sigma_{\mathbf{X},\mathbf{Y}} \Sigma_{\mathbf{Y}}^{-1} \Sigma_{\mathbf{X},\mathbf{Y}}^T. \quad (3.34)$$

con

$$\begin{aligned}
\Sigma_{\mathbf{X}, \mathbf{Y}} &= \mathbb{E}(\mathbf{X}^\top \mathbf{Y}) \\
&= \mathbb{E}([\mathbf{X}_r^\top, \mathbf{X}_i^\top] \begin{bmatrix} \mathbf{Y}_r \\ \mathbf{Y}_i \end{bmatrix}) \\
&= \begin{bmatrix} W_r^\top \Sigma W_r H & W_r^\top \Sigma W_i H \\ W_i^\top \Sigma W_r H & W_i^\top \Sigma W_i H \end{bmatrix}.
\end{aligned} \tag{3.35}$$

3.2.3. Caso bidimensional (imágenes)

Para poder abordar el caso dos dimensional, se seguirá la estrategia de llevar el problema bidimensional al caso unidimensional [28], para esto sea $I \in \mathbb{R}^{N \times N}$ y $V \in \mathbb{C}^{N \times N}$, entonces se trabajará con las versiones vectoriales $\vec{I} \in \mathbb{R}^{N^2}$ y $\vec{V} \in \mathbb{C}^{N^2}$ respectivamente (ver Apéndice 7.1). Además, recordando que las visibilidades son la transformada de Fourier de la imagen, entonces se puede escribir las visibilidades aprovechando la bilinealidad de la transformada de Fourier discreta en dos dimensiones. Esto implica que las visibilidades se pueden escribir de la siguiente forma¹:

$$V = W^\top I W \tag{3.36}$$

donde W se definió previamente en la ecuación (3.20). A la ecuación (3.36) se le aplicará el operador de vectorización, de modo de aprovechar la propiedad (2.1) del producto de Kronecker

$$V = W^\top I W \quad / \text{vec}(\cdot) \tag{3.37}$$

$$\vec{V} = (W^\top \otimes W^\top) \vec{I} \tag{3.38}$$

$$\vec{V} = \mathcal{W}^\top \vec{I} \tag{3.39}$$

Finalmente, como el par \vec{V}, \vec{I} están ligados por transformación lineal, se utiliza la formulación presentada en la sección anterior.

¹Esta formulación está escrita a conveniencia, puesto que W es simétrica, por lo tanto $W^\top = W$.

Capítulo 4

Análisis y Resultados

Para mostrar el desempeño de ambos métodos, se estudiará el desempeño de estos en el caso de series de tiempo (unidimensional) e imágenes (dos-dimensional). En ambos casos la métrica a utilizar es el error cuadrático medio normalizado (NMSE, por su sigla en inglés)¹. Además, se abordará una versión simplificada del problema de interferometría, denominado *snapshot*, que consiste en no aprovechar el movimiento terrestre para obtener muestras.

4.1. Modelo paramétrico: suma de Gaussianas

4.1.1. Caso unidimensional

La motivación detrás de evaluar el desempeño en el caso unidimensional es medir la capacidad de regresión del método y frente a una señal completamente conocida y que sigue el modelo. De esto modo, el objetivo de este experimento es reconstruir el espectro latente utilizando observaciones parciales y corruptas por ruido. Esto permite evaluar el desempeño real del modelo ante el ruido.

Nuestro primer caso de estudio es una serie de tiempo generada por una suma de Gaussianas mostrada en azul en las Figuras 4.1c y 4.1d. Para el ajuste de los modelos ésta se muestrea uniformemente entre $[-0,3, 0,3]$ en el dominio frecuencial. Los parámetros de la señal a reconstruir se resumen en la Tabla 4.1. Este caso sintético contempla dos sub-casos en donde la señal es corrompida con bajo ruido ($\sigma_{\text{noise}} = 1$) y con algo ruido ($\sigma_{\text{noise}} = 6$), de modo de representar dos casos hipotéticos, uno en que el sensor mide con alta precisión y otro en que la precisión del sensor es baja.

El desempeño del modelo unidimensional de suma de Gaussianas se comparará contra un modelo típico en estimación de señales, llamado proceso Gaussiano (GP, ver Apéndice

¹NMSE = $\sum_{i=1}^N (y_i - \hat{y}_i)^2 / \sum_{i=1}^N y_i^2$, donde y_i y \hat{y}_i es el i -ésimo valor del espectro verdadero y la estimado respectivamente.

Parámetro	Valor Real	Valor Estimado $\sigma_{\text{noise}} = 1$	Valor Estimado $\sigma_{\text{noise}} = 6$
β	[0,3; 0,5; -2; 2; 3; 1]	[1,5; 2,4; -0,66; -1,9; 0,8; 3,3]	[-2,7; 1,6; 1,2; 0,6; 3,0; 2,43]
C	[10; 12; 30; 45; 75; 88]	[98,1; 45,9; 95,3; 29,3; 13,0; 74,6]	[28,5; 32,7; 98,68; 0,4; 75,7; 51,29]
l	[10; 12; 7; 10; 10; 30]	[16,6; 9,9; 11,4; 7,512,2; 10,3]	[6,9; 10,1; 13,3; 14,3; 7,5; 10,27]

Tabla 4.1: Tabla de parámetros para suma de Gaussianas. Número de bases: 6.

	SoG		Gaussian Process	
	Real	Imaginario	Real	Imaginario
$\sigma_{\text{noise}} = 1$	$3,69 \cdot 10^{-3}$	$3,78 \cdot 10^{-3}$	197,35	328,33
$\sigma_{\text{noise}} = 6$	$4,51 \cdot 10^{-2}$	$5,36 \cdot 10^{-2}$	292,07	379,75

Tabla 4.2: Error cuadrático medio normalizado (NMSE, por su sigla en inglés).

7.3) [13]. Los resultados de este caso se presentan en las Tablas 4.1 y 4.2, y en la Figura 4.1. La figura mencionada muestra en azul la verdadera señal (o ground truth, en inglés), en rojo se presenta la regresión generada por el modelo de suma de Gaussianas entrenada con BFGS en el framework pymc3. Finalmente, en verde se muestra la regresión realizada con un GP.

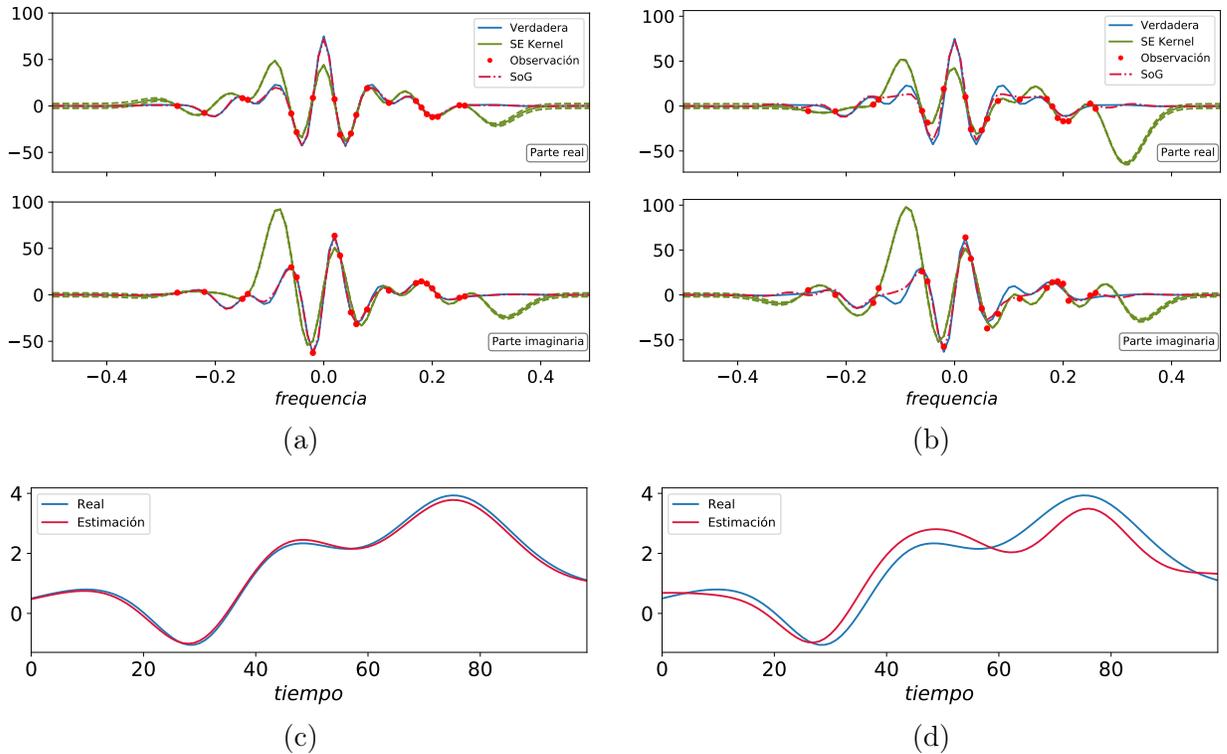


Figura 4.1: Columna izquierda: Reconstrucción en frecuencia (arriba) y tiempo (abajo) para $\sigma_{\text{noise}} = 1$. Columna derecha: Reconstrucción en frecuencia (arriba) y tiempo (abajo) para $\sigma_{\text{noise}} = 6$.

Considerando tanto el caso con bajo ruido y alto ruido, se puede notar que las reconstrucciones espectrales ofrecidas por el método son muy ajustadas y con NMSE muy bajo en ambos casos. También, como era de esperarse, el error aumenta con el ruido subiendo un or-

den de magnitud en el caso de alto ruido. Esto se puede concluir gracias a que el experimento contempla las mismas muestras, pero con mayor varianza de ruido.

En el caso de la reconstrucción de serie de tiempo, se puede notar que la reconstrucción es similar a la señal original tanto en el caso con bajo ruido como en el caso con alto ruido, esto también se aprecia en el bajo NMSE. Se puede notar que existe un error de estimación en el caso caso de alto ruido, pero a pesar de esto, la serie de tiempo estimada conserva la forma que está dada por la fundamental y los primeros modos.

4.1.2. Caso bidimensional

También se validará la capacidad del método propuesto para reconstruir imágenes desde su espectro con valores faltantes. Se generó una imagen de 100×100 píxeles, la cual se reconstruyó utilizando la mezcla de dos Gaussianas con los siguientes parámetros.

- **Pesos:** $\alpha_1 = 1$ y $\alpha_2 = 1$.
- **sigmas:** $\sigma_{1_1}^2 = 20, \sigma_{1_2} = 10, \sigma_{2_1}^2 = 20, \sigma_{2_2}^2 = 10, \sigma_{12_1} = 10$ y $\sigma_{12_2} = -5$.
- **Centros:** $[(50, 40), (40, 60)]$.
- **Desviaciones estándar del ruido:** $\sigma_x^2 = \sigma_y^2 = 0,00$ (Caso sin ruido).

Luego se computó la transformada de Fourier discreta para esta imagen y se removieron algunos puntos de la data dejando el 38,53% de los datos. Estos datos restante son las observaciones, los cuales provienen de un arreglo interferométrico de 81 antenas dispuestas de manera uniforme como se puede ver en 4.2a. Se utilizaron estas mediciones para entrenar el modelo y realizar la reconstrucción.

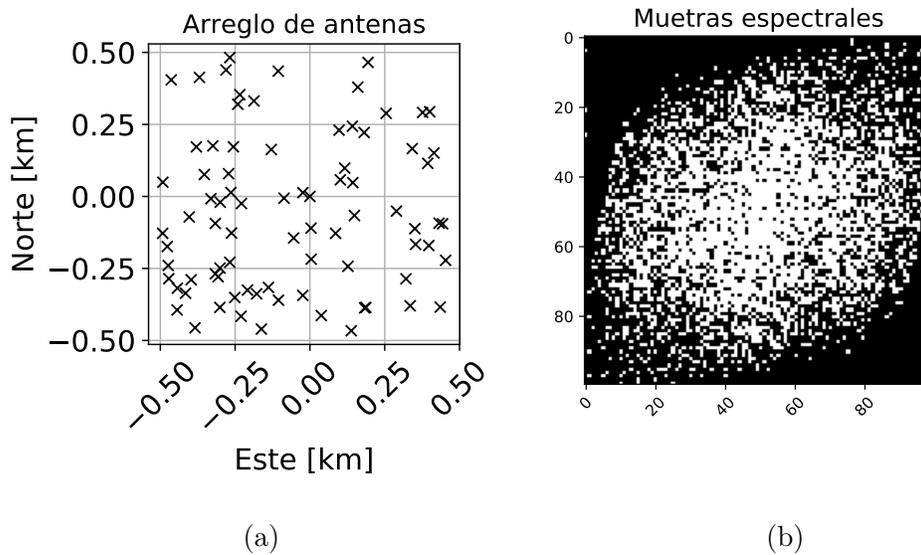


Figura 4.2: (a) Arreglo de antenas dispuesto de manera uniforme, cada cruz representa una antena dispuesta en el terreno. (b) Patrón de observación: En blanco frecuencias observadas, en negro frecuencias no observadas. Este arreglo de 81 antenas genera el 38,53% de las observaciones.

Los modelos entrenados corresponden a SoG con 2 bases y 20 bases, los cuales se muestran en las Figuras 4.3 y 4.4 respectivamente. Es importante recalcar que el objetivo de estudiar las variaciones en la cantidad de bases es estudiar la sensibilidad del método. Recordemos que aumentar la cantidad de bases implica aumenta la cantidad de parámetros a estimar, en este caso la cantidad de parámetros a optimizar es de 14 y 122 respectivamente. De esta manera, como es obvio, el costo computacional también aumenta.

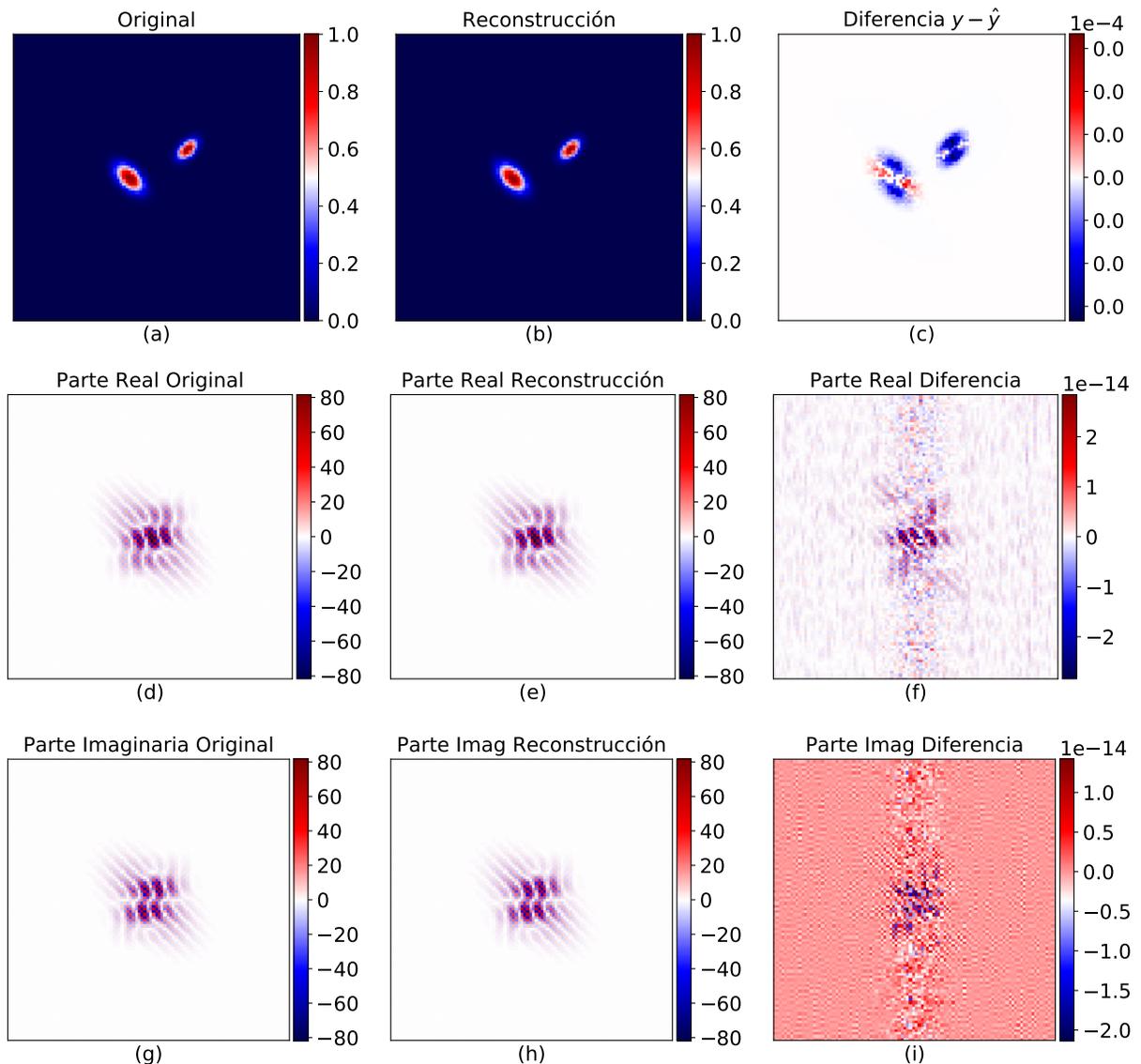


Figura 4.3: Reconstrucción utilizando 2 bases. Columna izquierda: Imágenes originales. Columna central: Imágenes reconstruidas con modelo de suma de Gaussinas. Columna derecha: Diferencia entre imágenes. Arriba: Dominio espacial NMSE = $1,04 \cdot 10^{-31}$. Centro: Dominio espectral parte real NMSE = $1,83 \cdot 10^{-31}$. Abajo: Dominio espectral parte imaginaria NMSE = $1,57 \cdot 10^{-31}$.

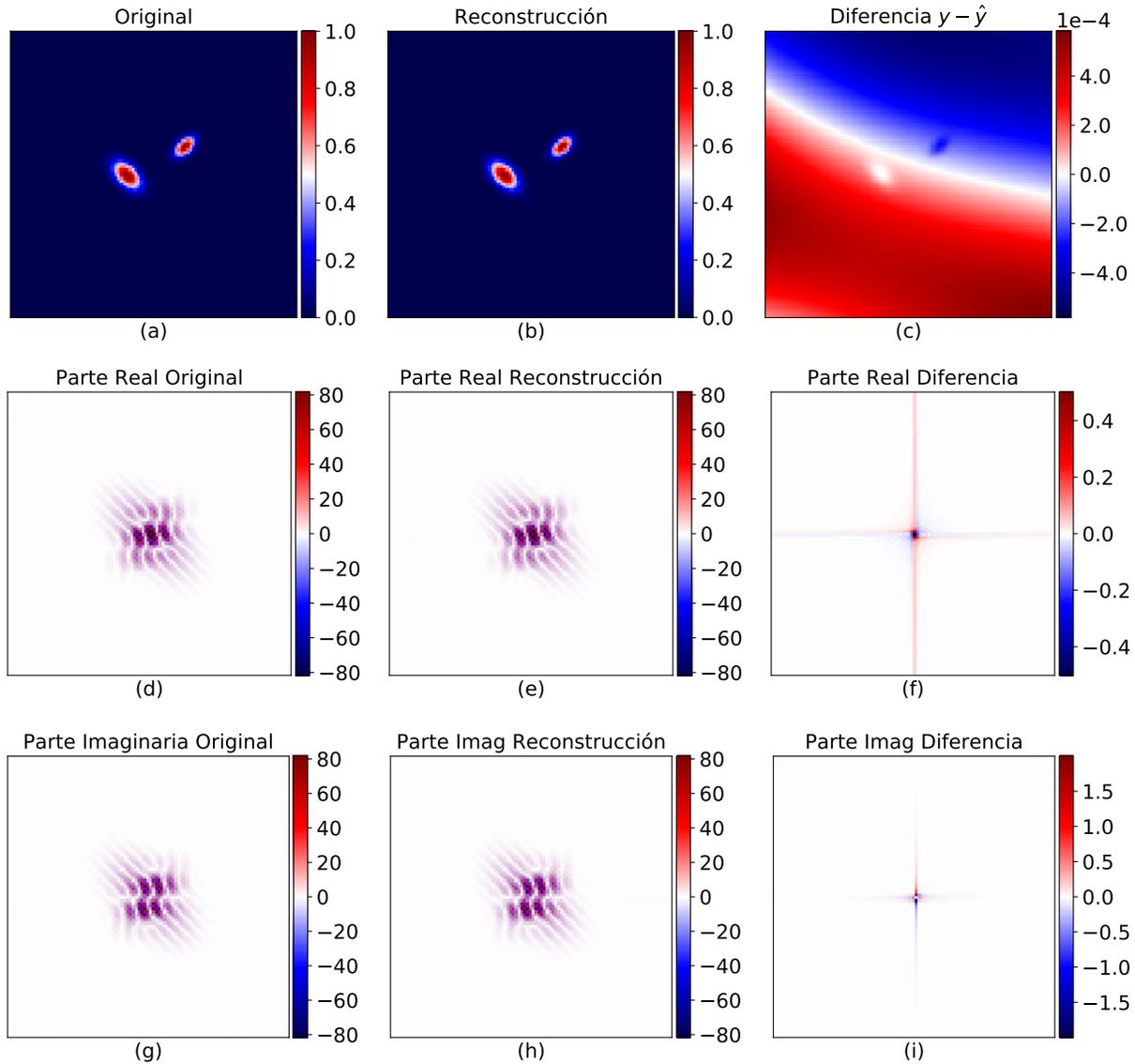


Figura 4.4: Reconstrucción utilizando 20 bases. Columna izquierda: Imágenes originales. Columna central: Imágenes reconstruidas con modelo de suma de Gaussianas. Columna derecha: Diferencia entre imágenes. Arriba: Dominio espacial $\text{NMSE} = 3,10 \cdot 10^{-5}$. Centro: Dominio espectral parte real $\text{NMSE} = 5,46 \cdot 10^{-6}$. Abajo: Dominio espectral parte imaginaria $\text{NMSE} = 5,78 \cdot 10^{-5}$.

La primera observación es a nivel visual: se puede observar que la reconstrucción es bastante similar, lo cual se correlaciona en términos del error NMSE obtenido que es del orden de 10^{-31} y 10^{-5} . De este modo, a pesar de que las reconstrucciones están alejadas en términos de error, visualmente no se nota diferencia.

Otra observación de este modelo es que al aumenta la cantidad de bases el error aumentó notablemente en varias ordenes de magnitud. A pesar de esto, el error sigue estando en rangos aceptables, es decir, orden del error mucho menor al de la estimación. Esto podría ser un problema si para cierta cantidad de bases el error comienza a aumentar y no tiene un comportamiento asintótico.

Un resultado interesante de estas reconstrucciones es que en el caso de la Figura 4.4c se puede observar que hay un cambio brusco denotada por una zona azul y otra roja. Este cambio se puede atribuir a que el sistema encontró una base con un *lengthscale* muy grande pero un peso α_i muy pequeño, lo cual genera este off-set y además repercute en el error. Esto último es atribuible a que el sistema está en un subóptimo.

4.1.3. Caso real: Inteferometría

Finalmente, el último experimento a realizar es ver como el algoritmo se comporta con imágenes reales. Se tiene un catálogo de seis imágenes mostradas en la Figura 4.5 las cuales fueron reconstruidas con MEM [29, 5] y CLEAN [4]. En este caso se reconstruirán solo las imágenes de la Figura 4.5f y 4.5d conocidas como HD 142527 y IRAS 16293 respectivamente. La elección de estas imágenes está en que HD142527 es un cuerpo clásico para validar los métodos interferométricos en la literatura [30, 31], mientras que IRAS 16293 son dos fuentes puntuales y no presentan estructuras complejas, lo cual cubre los dos tipos de imágenes que se puede encontrar en esta disciplina.

Las reconstrucciones del método SoG se muestran en las Figuras 4.6, 4.7, 4.8 y 4.9. En ellas se muestran los resultados de la misma manera que los anteriores. Si bien en este caso no hay una imagen que se pueda utilizar como validación o referencia (*ground truth*), se comparará con la imagen reconstruida por MEM.

En el caso de HD142527 se simuló un arreglo de 81 antenas² las cuales son capaces de capturar el 28.35 % de los datos en el espectro en una imagen de 128×128 píxeles. Los resultados de esta reconstrucción se muestran en la Figura 4.6 los que resumen la reconstrucción utilizando 20 bases. Se puede observar claramente que la reconstrucción espacial logra recuperar el cuerpo principal, pero texturas que están fuera del cuerpo principal no fueron recuperadas correctamente, como se observa en la Figura 4.6c. Por otro lado, las reconstrucciones espectrales son muy similares a su original, al menos a nivel visual no se observan grandes diferencias. En cuanto al error, para todas las reconstrucciones el NMSE es muy bajo y es una buena representación de lo buena que es la estimación realizada por el modelo.

Según los expertos, algo importante o deseable en este cuerpo celeste es recuperar es al punto central; sin embargo, se observa que en este caso el algoritmo no fue capaz de recuperar el punto central.

²Este arreglo de antenas es el mismo que la Figura 4.2a.

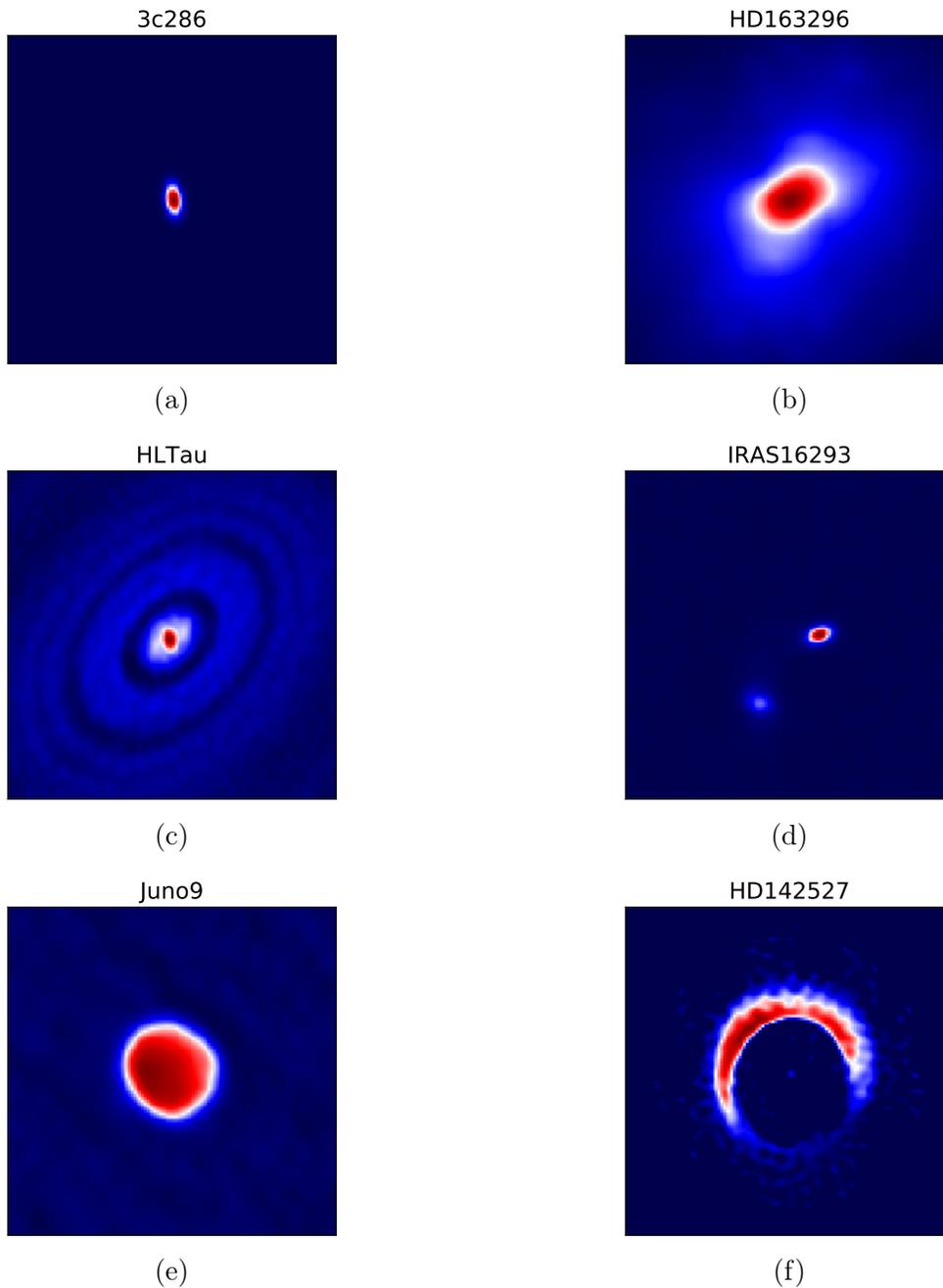


Figura 4.5: Catálogo de cuerpos celestes.

Para el caso de IRAS se simularon distintos escenarios que son interesantes observar, el primero consiste en un arreglo de antenas capaz de obtener el 27,51 % de las frecuencias, mientras que el segundo solo alcanza el 5 %. Además, en este último caso, para enriquecer el análisis se simularon 50 bases y 10 bases.

En la Figura 4.7 se puede observar que la reconstrucción es similar a la imagen original, pero añade unas bandas verticales que se atribuyen a encontrar un sub-óptimo en la optimización. Por otro lado, a nivel espectral, se puede observar que estas bandas verticales en el dominio espacial se atribuyen a bandas horizontales de baja potencia cercanas a la frecuencia

$v = 0$ en el espectro, las cuales se aprecian mucho mejor en las imágenes residuales 4.7f y 4.7i.

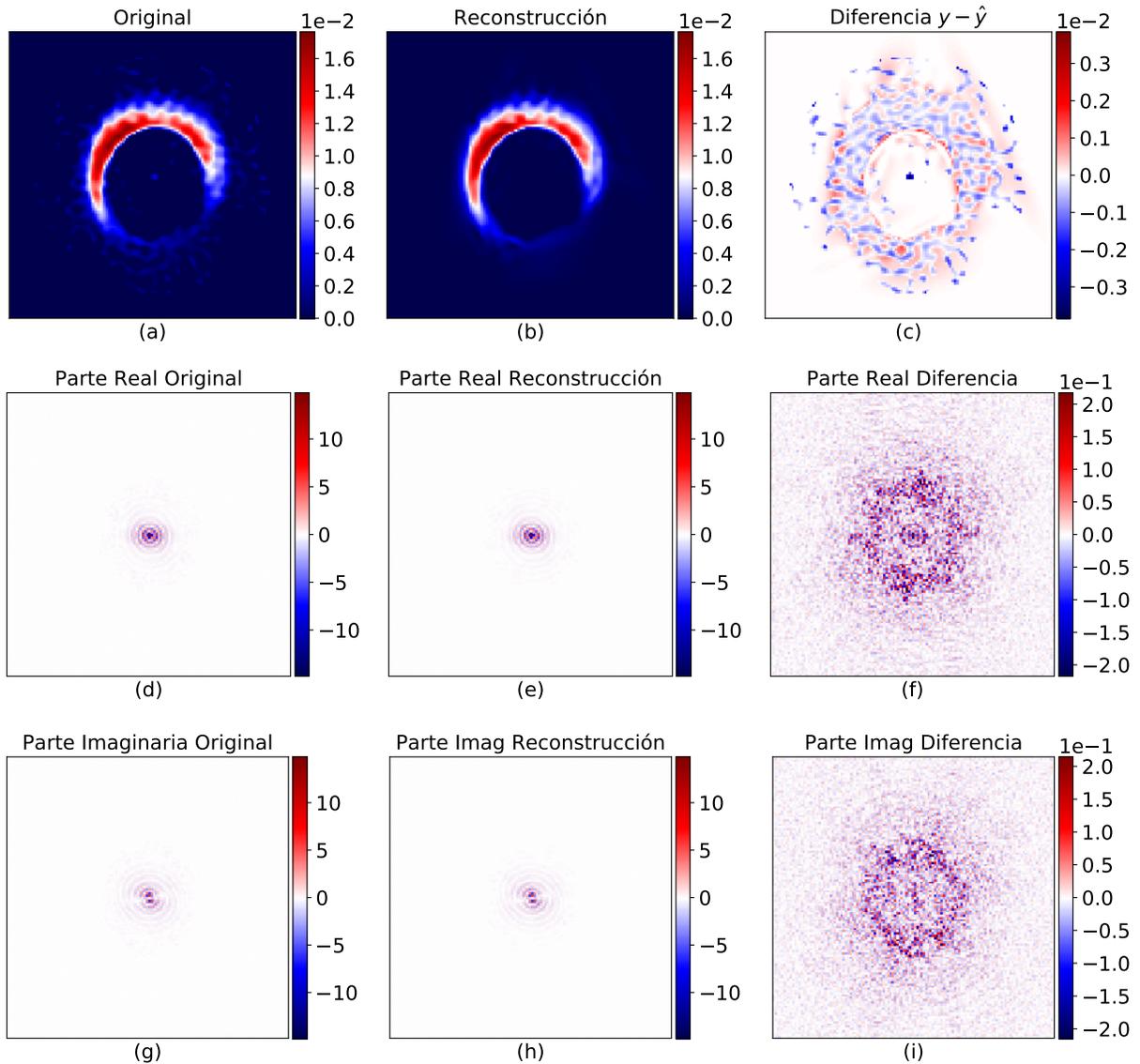


Figura 4.6: Reconstrucción HD142527 utilizando 20 bases. Columna izquierda: Imágenes originales. Columna central: Imágenes reconstruidas con modelo de suma de Gaussianas. Columna derecha: Diferencia entre imágenes. Arriba: Dominio espacial $\text{NMSE} = 8,45 \cdot 10^{-3}$. Centro: Dominio espectral parte real $\text{NMSE} = 6,50 \cdot 10^{-3}$. Abajo: Dominio espectral parte imaginaria $\text{NMSE} = 1,32 \cdot 10^{-2}$.

En cuanto a la Figura 4.8, esta muestra una reconstrucción de IRAS 16293, pero con solo el 5% de los datos. Se puede observar que el método SoG aún reconstruye las dos fuentes principales, pero esta vez agrega algunos artefactos propios de la reconstrucción. En este caso, se puede notar que estos artefactos afectan directamente a la bajas frecuencias, lugar donde está concentrado la mayor cantidad de la potencia espectral. Por otro lado, la Figura 4.9 muestra una reconstrucción con las mismas muestras, pero esta vez la cantidad de bases disminuye a 10. Este procedimiento se realizó puesto que se conjeturó que los artefactos del

caso con 50 bases corresponden a casos en que el algoritmo no puede decidir que hacer con las bases sobrantes, por lo que al disminuir el número de bases se eliminarán estos artefactos. Esta hipótesis se validó con este experimento.

Algo que es importante mencionar de estas tres ultimas reconstrucciones es que la imagen original tiene un pequeño ruido de fondo imperceptible a nivel visual, pero que este método logra quitar completamente.

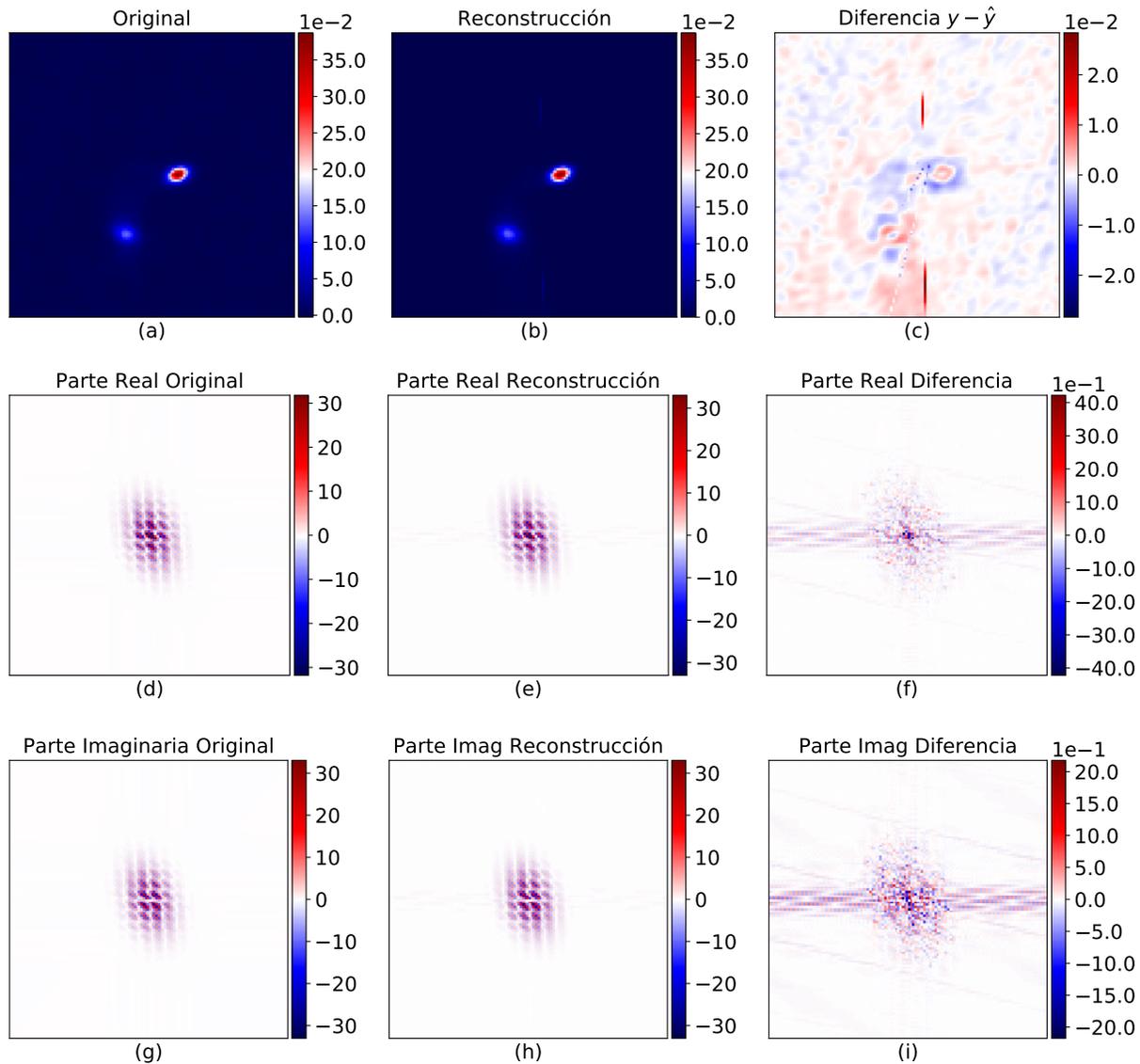


Figura 4.7: Reconstrucción IRAS16293 utilizando 50 bases y 27,51 % de los datos. Columna izquierda: Imágenes originales. Columna central: Imágenes reconstruidas con modelo de suma de Gaussianas. Columna derecha: Diferencia entre imágenes. Arriba: Dominio espacial NMSE = 8,62. Centro: Dominio espectral parte real NMSE = $9,63 \cdot 10^{-2}$. Abajo: Dominio espectral parte imaginaria NMSE = $7,85 \cdot 10^{-3}$.

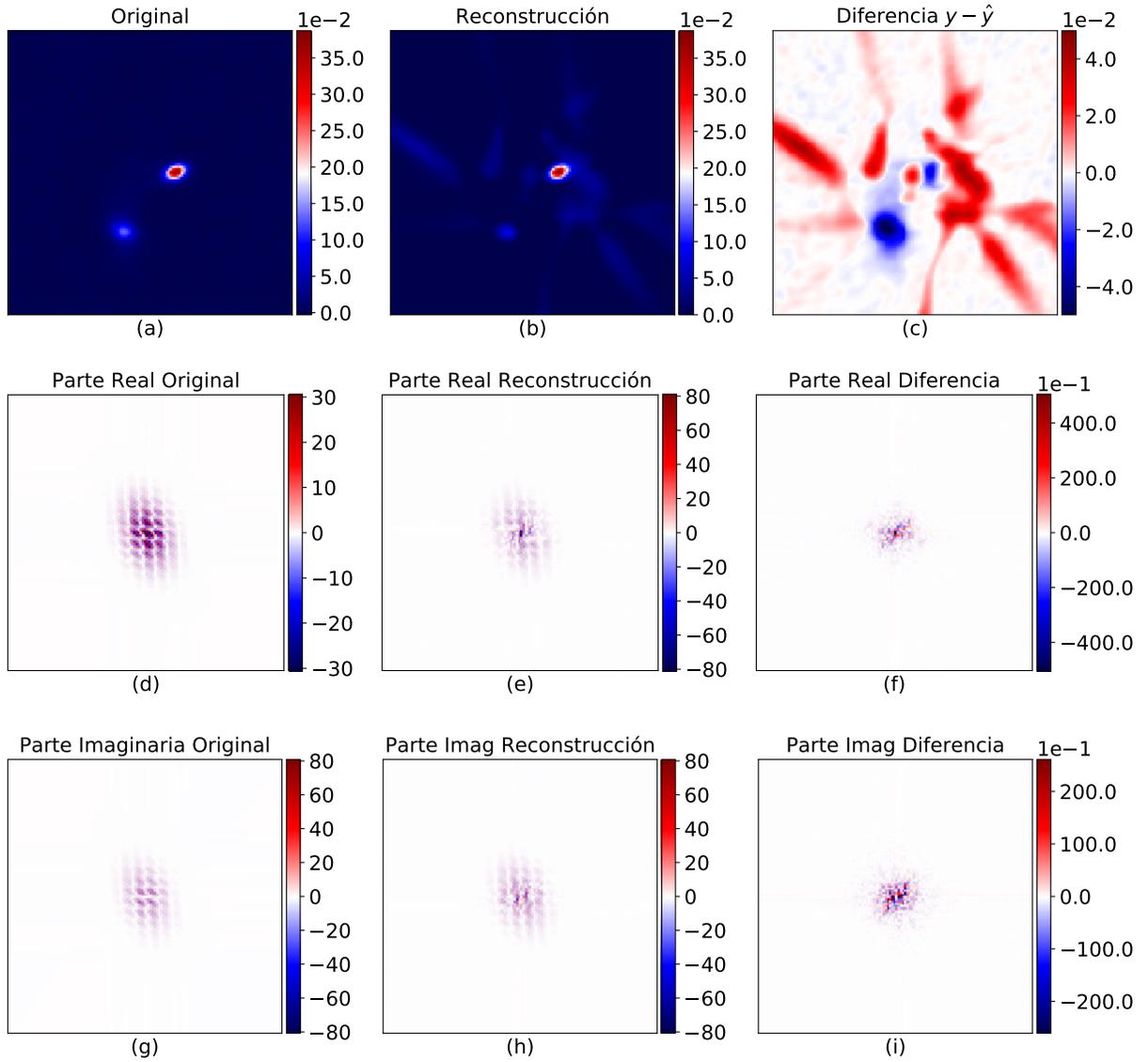


Figura 4.8: Reconstrucción IRAS16293 utilizando 50 bases y 5% de los datos. Columna izquierda: Imágenes originales. Columna central: Imágenes reconstruidas con modelo de suma de Gaussianas. Columna derecha: Diferencia entre imágenes. Arriba: Dominio espacial NMSE = $5,94 \cdot 10^{-5}$. Centro: Dominio espectral parte real NMSE = $4,10 \cdot 10^{-1}$. Abajo: Dominio espectral parte imaginaria NMSE = $3,47 \cdot 10^{-1}$.

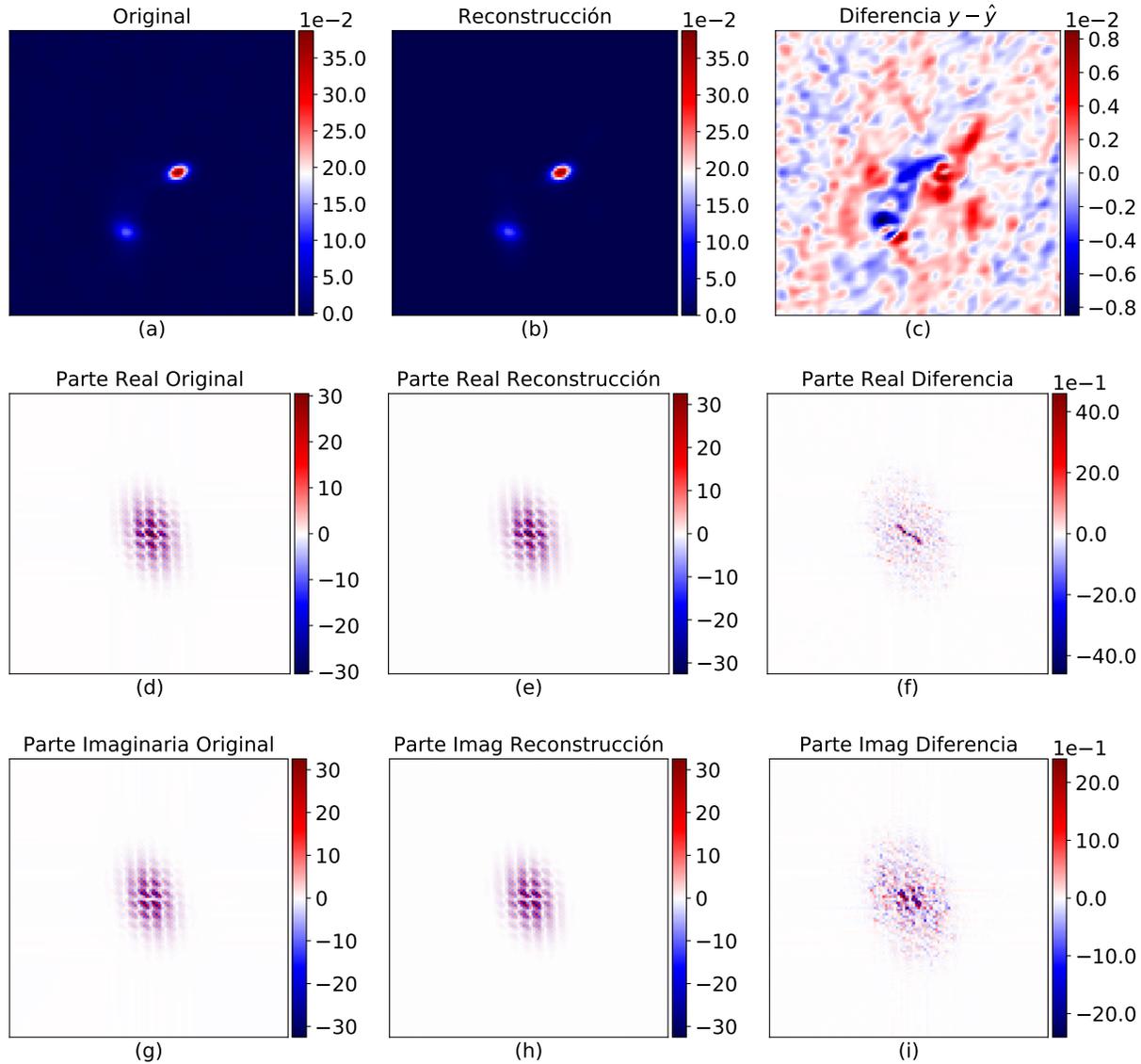


Figura 4.9: Reconstrucción IRAS16293 utilizando 10 bases y 5% de los datos. Columna izquierda: Imágenes originales. Columna central: Imágenes reconstruidas con modelo de suma de Gaussianas. Columna derecha: Diferencia entre imágenes. Arriba: Dominio espacial NMSE = 2,61. Centro: Dominio espectral parte real NMSE = $6,18 \cdot 10^{-3}$. Abajo: Dominio espectral parte imaginaria NMSE = $5,85 \cdot 10^{-3}$.

4.2. Modelo no-paramétrico: Reconstrucción espectral Bayesiana

4.2.1. Caso unidimensional

El estudio para el modelo no paramétrico considera un muestreo uniforme de un proceso Gaussiano con kernel exponencial cuadrático (SE, por su sigla en inglés) mostrado en la

Ecuación (4.1). Recordemos que un GP es una distribución prior sobre funciones continuas, luego, para este experimento se generó una trayectoria continua en el tiempo desde un GP y luego se consideró su versión muestreada uniformemente. Luego de esto, se computó el espectro de la serie calculando la DFT y se realizaron dos experimentos. El primero consiste en remover valores aleatorios del espectro dejando el 4% del espectro total, sumarle un ruido Gaussiano con una varianza pequeña $\sigma_{\text{noise}} = 1 \cdot 10^{-3}$ y luego reconstruir el espectro; el segundo es similar, pero en esta ocasión el ruido tiene una varianza mayor de $\sigma_{\text{noise}} = 0,5$.

$$K(t, t') = \sigma^2 \exp(-\alpha(t - t')^2) \quad (4.1)$$

	BSR		Gaussian Process	
	Real	Imaginary	Real	Imaginary
$\sigma_{\text{noise}} = 1 \cdot 10^{-3}$	$1,14 \cdot 10^{-3}$	$0,97 \cdot 10^{-3}$	1152,92	814,86
$\sigma_{\text{noise}} = 0,5$	$6,84 \cdot 10^{-2}$	$6,82 \cdot 10^{-2}$	1313,96	1227,73

Tabla 4.3: Error cuadrático medio normalizado.

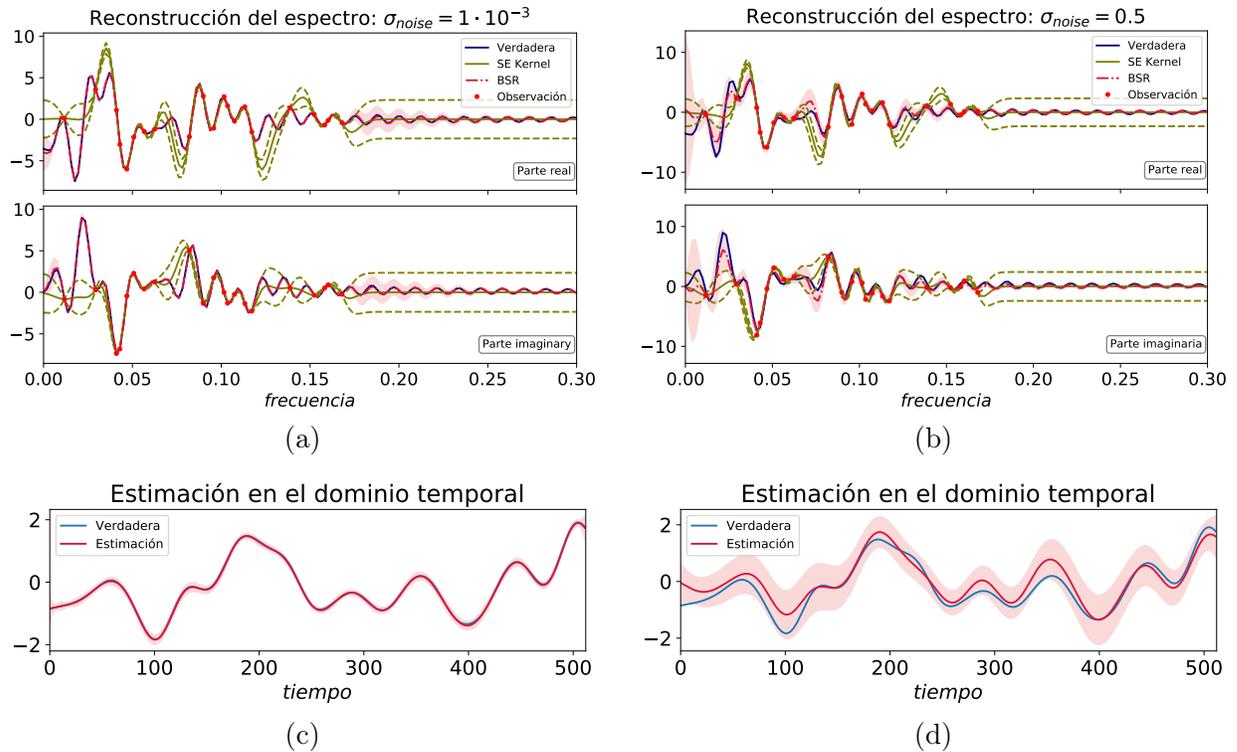


Figura 4.10: Columna izquierda: Reconstrucción en frecuencia (arriba) y tiempo (abajo) para $\sigma_{\text{noise}} = 1 \cdot 10^{-3}$. Columna derecha: Reconstrucción en frecuencia (arriba) y tiempo (abajo) para $\sigma_{\text{noise}} = 0,5$.

Análogamente al experimento realizado para la suma de Gaussianas, la motivación de este escenario es evaluar la capacidad de regresión del método y, de este modo, comparar la regresión obtenida con la señal verdadera.

Las Figuras 4.10a y 4.10b muestran las reconstrucciones espectrales de la parte real (arriba) y de la parte imaginaria (abajo) utilizando BSR y GP. La reconstrucción en rojo es la media y el intervalo del 95 % de confianza de BSR, la verdadera señal se muestra en azul y en puntos rojos las observaciones, además, la reconstrucción utilizando el GP con kernel exponencial cuadrático y su banda de 95 % de confianza se muestra en verde. La Tabla 4.3 muestra los resultados en NMSE para las reconstrucciones de la parte real e imaginaria antes mencionadas. Notar que la reconstrucción de BSR es insesgada y las barras de error son ajustadas a pesar de la baja cantidad de muestras. También observemos que las estimaciones son insesgadas para las altas frecuencias a pesar de no contar con observaciones, el prior sugiere que la energía en de altas frecuencias decae a cero. De hecho, como consecuencia de la estacionariedad de kernel exponencial cuadrático, las barras de error para el GP son constantes, ignorando que la potencia espectral es finita.

Cabe destacar que al aumentar la desviación estándar del ruido a $\sigma_{\text{noise}} = 0,5$ la reconstrucción recupera de manera muy similar las señal latente. También, en las zonas donde no hay observaciones las barras de error contienen completamente al espectro verdadero. Notemos que el NMSE es bajo y solo aumenta un orden de magnitud, a pesar de que la desviación estándar del ruido aumentó en dos órdenes de magnitud.

Además de las reconstrucciones espectrales, se consideró el caso de ingeniería inversa, en donde las observaciones están en el espectro y se desea reconstruir la serie de tiempo. Para realizar esto, el problema temporal pudo ser reconstruido gracias a que la relación entre la serie de tiempo y su espectro es que son conjuntamente Gaussianas [32], lo cual es tratable gracias a la propiedad lineal de la transformada de Fourier.

Las Figuras 4.10c y 4.10d muestran las reconstrucciones obtenidas con los mismos datos de los experimentos anteriores los cuales representan el 4 % total del espectro. Para ambos casos se puede observar como las reconstrucciones son muy ajustada en media y en barras de error. Notar además que la reconstrucción espectral es muy ajustada aún cuando no existen observaciones para $\omega > 0,2$, esto puesto que el prior fuerza a que la estimación se vaya a cero en altas frecuencias.

4.2.2. Caso dos-dimensional

En primer lugar, se realizó una reconstrucción utilizando una imagen de 50×50 píxeles proveniente de un kernel exponencial cuadrático. Las muestras proveniente del mismo arreglo de antenas utilizados en el modelo SoG presentados en la Figura 4.2a. En este escenario la Figura 4.11 muestra los resultados en cada dominio, es decir, en el dominio espacial y espectral en parte real e imaginaria.

Estos resultados tienen un alto rendimiento en términos de NMSE y se observa que el método reconstruye bien la imagen, sin embargo, el orden de magnitud de la desviación estándar es solo un orden de magnitud inferior a la media. En contra posición a esta situación, la imagen residual (diferencia) es del orden de 10^{-2} , por lo que la desviación en media es relativamente insesgada. Es interesante notar que la señal tiene una mayor incertidumbre en el área en donde se concentra la potencia espectral.

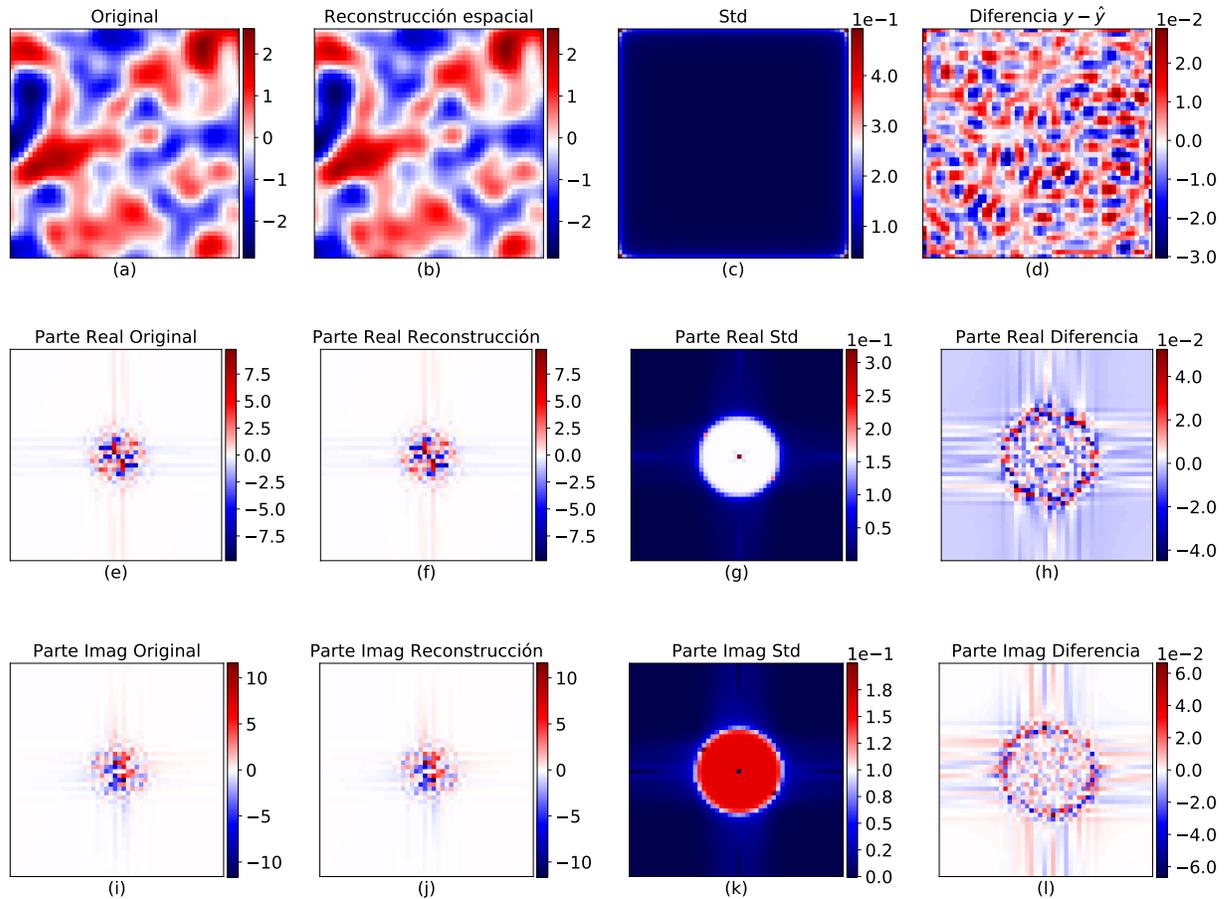


Figura 4.11: Reconstrucción utilizando BSR. De izquierda a derecha, primera columna: imágenes originales; segunda columna: imágenes reconstruidas con modelo de suma de Gaussianas; tercera columna: desviación estándar; cuarta columna: diferencia entre imágenes. Arriba: Dominio espacial $NMSE = 8,72 \cdot 10^{-5}$. Centro: Dominio espectral parte real $NMSE = 8,69 \cdot 10^{-5}$. Abajo: Dominio espectral parte imaginaria $NMSE = 8,75 \cdot 10^{-5}$.

El segundo experimento consiste en reconstruir la señal completamente utilizando la mitad de los datos. Este experimento está diseñado para observar la capacidad de regresión de BSR aprovechando a la propiedad de paridad e imparidad de la transformada de Fourier. Para realizar esto se tomaron las frecuencias correspondientes a la primera mitad. Adicionalmente, no se tomó la frecuencia correspondiente al par $(0, 0)$ puesto que este es el único punto que no se ve afectado por la paridad o imparidad. En la Figura 4.12 se muestra el escenario estudiado, se puede ver la vectorización de la imagen y los puntos tomados, además de ver la reconstrucción en su forma vectorial en rojo y la verdadera señal en azul.

Observar que en esta reconstrucción el único punto que no se interpoló correctamente es el $(u, v) = (0, 0)$, pero el método lo compensó asignándole una gran incertidumbre. Del mismo modo, notar que el resto de las frecuencias fueron completamente determinadas y con bandas de error muy bajas, puesto que el sistema es capaz de reconocer que el espectro es par en su parte real e impar en su parte imaginaria. En ese sentido, la repercusión directa de una mala estimación de la frecuencia $(0, 0)$ es que la estimación temporal es sesgada, puesto que la frecuencia $(0, 0)$ es la frecuencia que entrega el *off-set* o sesgo en el dominio temporal, lo

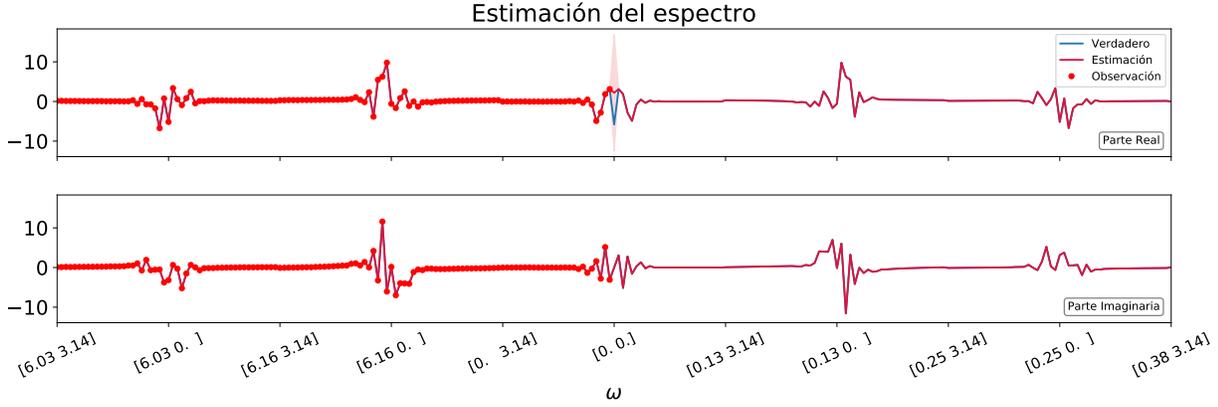


Figura 4.12: Representación vectorial de la imagen. Reconstrucción de espectro utilizando el 51% de los datos. Se utilizó como observaciones la mitad derecha de las frecuencias, sin incluir el par de frecuencias $(0, 0)$.

cual se puede corroborar en la Figura 4.13d.

4.2.3. Caso real: Inteferometría

Al igual que en SoG, se reconstruirá la imagen de HD142527, esta vez se utilizará su versión de 50×50 píxeles por razones computacionales y de tiempo. Este método es mucho más costoso en términos computacionales que el método de SoG. El método incluye nuevamente la utilización de un arreglo de 60 antenas, las cuales se disponen aleatoriamente con distribución uniforme en un terreno. Esta distribución logra captar aproximadamente el 54% de las frecuencias espaciales, concentradas en las bajas frecuencias.

El caso de HD142527 es un caso en que la reconstrucción que se obtiene es relativamente similar y mejora el desempeño frente a su par paramétrico SoG, en este caso, se puede observar que tanto el cuerpo como parte importante de la estructura exterior del cuerpo celeste. En este caso, el NMSE es menor, aunque no considerablemente menor. Además, la estrella central de HD142527 también se logró recuperar, lo que indica que el método supera a SoG recuperando estructuras más pequeñas y complejas.

Las reconstrucciones espectrales ambas tienen en común que el error es muy pequeño entorno a $(u, v) = (0, 0)$, puesto que es el lugar donde más densidad de muestras se tiene, aumentando drásticamente la desviación estándar en los lugares donde no hay muestras. También, a pesar de que no haya muestras en alta frecuencia, se puede observar que la desviación estándar se va a cero en esa zona, puesto que el sistema logra aprender que la potencia está concentrada en las bajas frecuencias.

Para el caso de IRAS 16293, la reconstrucción se muestra en 4.15. En este caso lo primero que se observa es que el sistema logra reconstruir las dos fuentes puntuales, pero el método agrega unos artefactos asociados al ruido de fondo en la imagen, esto quiere decir que el fondo de la imagen no es cero, es más, está dado por una componente sinusoidal. Este

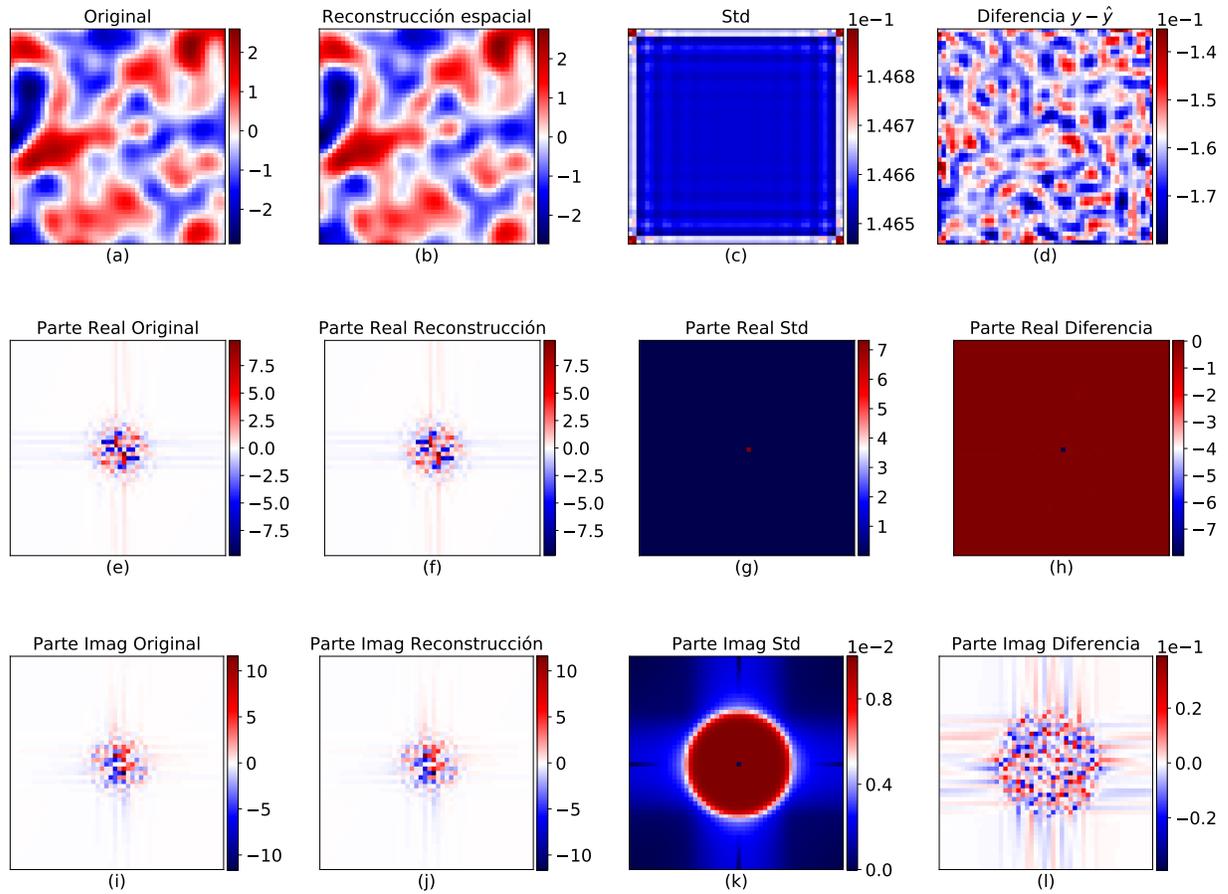


Figura 4.13: Reconstrucción utilizando BSR. De izquierda a derecha, primera columna: imágenes originales; segunda columna: imágenes reconstruidas con BSR; tercera columna: desviación estándar; cuarta columna: diferencia entre imágenes. Arriba: Dominio espacial NMSE = $3,00 \cdot 10^{-2}$. Centro: Dominio espectral parte real NMSE = $5,90 \cdot 10^{-2}$. Abajo: Dominio espectral parte imaginaria NMSE = $4,96 \cdot 10^{-5}$.

comportamiento es clásico cuando se eliminan frecuencias, como cuando se intenta reconstruir una señal cuadrada y se corta en cierta cantidad de frecuencias. En este caso, se puede observar que en el espectro pasa algo similar, en que si bien a nivel visual los espectros original y reconstruidos (Figuras 4.15e, 4.15f parte real y 4.15g, 4.15l parte imaginaria) son similares, la imagen cuenta con algunos cambios en las altas frecuencias, lo cual repercute en estos artefactos de fondo asociados. En ese mismo sentido, se puede observar en la Figura 4.15c que el error en el dominio espacial es más alto proporcionalmente en la escala (representado por un azul claro en la escala), comparado con las otras reconstrucciones, pero sigue conservando las propiedades de simetría.

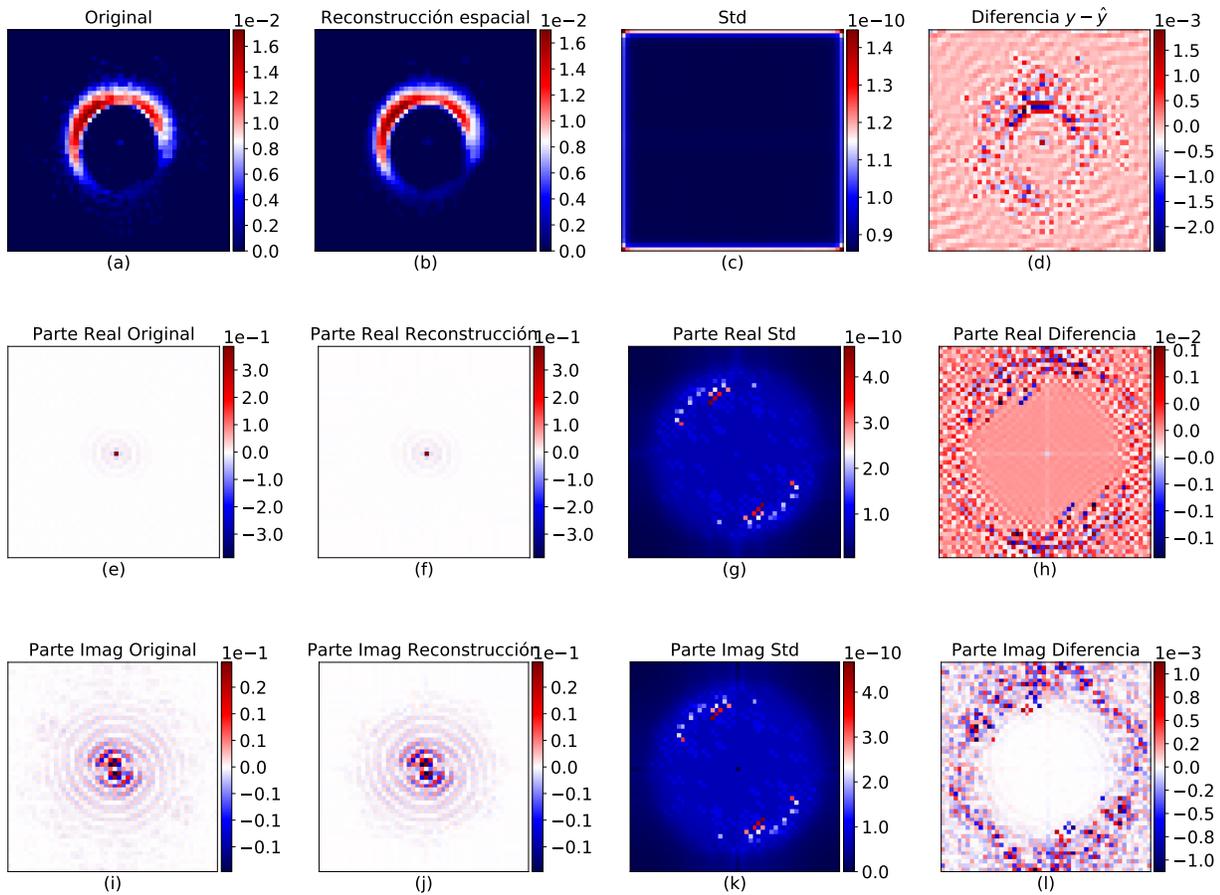


Figura 4.14: Reconstrucción de HD142527 utilizando BSR. De izquierda a derecha, primera columna: imágenes originales; segunda columna: imágenes reconstruidas con BSR; tercera columna: desviación estándar; cuarta columna: diferencia entre imágenes. Arriba: Dominio espacial $\text{NMSE} = 1,17 \cdot 10^{-3}$. Centro: Dominio espectral parte real $\text{NMSE} = 5,92 \cdot 10^{-5}$. Abajo: Dominio espectral parte imaginaria $\text{NMSE} = 1,69 \cdot 10^{-2}$.

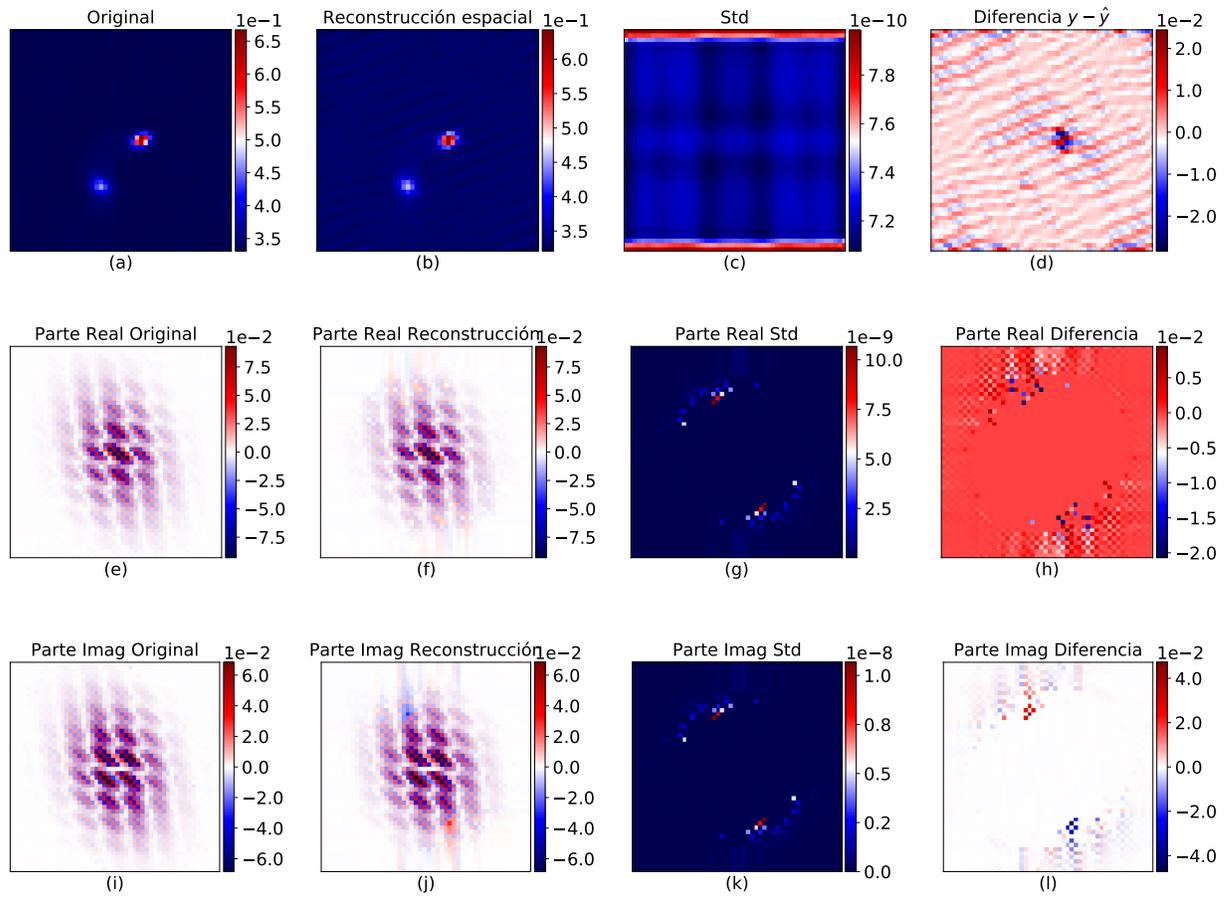


Figura 4.15: Reconstrucción de IRAS16293 utilizando BSR. De izquierda a derecha, primera columna: imágenes originales; segunda columna: imágenes reconstruidas con BSR; tercera columna: desviación estándar; cuarta columna: diferencia entre imágenes. Arriba: Dominio espacial $NMSE = 3,76 \cdot 10^{-2}$. Centro: Dominio espectral parte real $NMSE = 2,27 \cdot 10^{-2}$. Abajo: Dominio espectral parte imaginaria $NMSE = 5,39 \cdot 10^{-2}$.

Capítulo 5

Conclusiones y trabajo a futuro

En este trabajo se presentará dos herramientas para realizar estimación espectral a través de muestras en el dominio de Fourier. Adicionalmente, estos métodos son fácilmente extrapolables al caso en que el muestreo es en el dominio temporal (o espacial), caso común fuera del ambiente de la interferometría. La metodología propuesta ayuda a entender las limitantes de estos métodos y permitió confirmar la hipótesis propuesta.

El método SoG tiene una gran fortaleza al ser un método paramétrico, obteniendo resultados prometedores, en especial con fuentes puntuales, logrando realizar reconstrucciones de bajo error con tan solo el 5% de los datos. Sin embargo, al depender netamente de la similitud entre el modelo y los datos, el método se centra primero en encontrar las formas que disminuyen drásticamente el error, en desmedro de detalles fundamentales y de interés para el mundo científico. Otra fortaleza es que el método SoG está formulado completamente en continuo, lo por lo que la formulación no requiere de un grillado en la cual se pierda información.

Mejorar el método SoG implica realizar un esfuerzo en la no dependencia de la reconstrucción y la cantidad de bases a utilizar. Una opción clara sería extender el modelo de suma de gaussianas a el modelo de suma infinita de gaussianas propuesto por Rasmussen [33]. Otra solución sería crear un prior para los pesos que penalice la cantidad de pesos que no sean cero, de esta manera generar una regularización del modelo.

Por otro lado, el método BSR presenta una fortaleza al recuperar los detalles. Este método centra todo su esfuerzo en encontrar la relación de cómo covaría dos puntos en el espectro. En ese sentido, el correcto modelamiento del problema en el dominio espacial es vital y es un punto crucial a estudiar en el futuro de este método. Por ejemplo, en el caso de la IRAS16293 el espectro tiene componentes periódicas no nulas en altas frecuencias, por lo que el si se eligiera una estructura de matriz de covarianza capaz de resolver este comportamiento en frecuencia, la estimación mejoraría considerablemente. Adicionalmente, si bien el problema está formulado en discreto, la formulación no es excluyente para el caso continuo, puesto que la grilla en donde se realiza la estimación se puede hacer tan densa como se necesite.

En términos de implementación, BSR es un método que solo necesita encontrar la máxima

verosimilitud para estimar, por lo que realizar una implementación que incluya autoderivación para calcular los gradientes al momento de optimizar es un claro avance. En ese sentido se propone una implementación en `pytorch`, dado que es un marco de trabajo que tiene un amplio soporte por la comunidad.

En cuanto a las apreciaciones generales, este trabajo representa un gran avance en estimación espectral bayesiana en sí mismo. La idea de calcular modelos inducidos por transformaciones es un mundo por explorar y partir por transformaciones conocidas como la transformada de Fourier es solo el comienzo. Las aplicaciones en otros espacios se pueden extender de manera simple, siempre y cuando se pueda obtener una forma cerrada para el modelo inducido.

Finalmente, se puede concluir satisfactoriamente sobre la capacidad de incorporar la incertidumbre implícita de los problemas utilizando inferencia bayesiana. Si bien para el método SoG no se implementó un método que muestre la incertidumbre por punto reconstruido, la extensión es simple muestreando la distribución posterior. Por otro lado, el método BSR intrínsecamente ofrece este parámetro de incertidumbre puntual, lo cual es una gran ventaja a la hora de reportar descubrimientos científicos.

Capítulo 6

Bibliografía

- [1] H. Nyquist. Certain topics in telegraph transmission theory. *Transactions of the American Institute of Electrical Engineers*, 47(2):617–644, Apr 1928.
- [2] C. E. Shannon. Communication in the presence of noise. *Proceedings of the IEEE*, 86(2):447–457, Feb 1998.
- [3] G. B. Taylor, C. L. Carilli, and R. A. Perley, editors. *Synthesis Imaging in Radio Astronomy II*, volume 180 of *Astronomical Society of the Pacific Conference Series*, 1999.
- [4] J. A. Högbom. Aperture synthesis with a non-regular distribution of interferometer baselines. *Astronomy and Astrophysics Supplement*, 15:417, Jun 1974.
- [5] K. Maisinger, M. P. Hobson, and A. N. Lasenby. A maximum entropy method for reconstructing interferometer maps of fluctuations in the cosmic microwave background radiation. *Monthly Notices of the Royal Astronomical Society*, 290(2):313–326, 1997.
- [6] M. Lochner, I. Natarajan, J. T. L. Zwart, O. Smirnov, B. A. Bassett, N. Oozeer, and M. Kunz. Bayesian inference for radio observations. *Monthly Notices of the Royal Astronomical Society*, 450(2):1308–1319, 2015.
- [7] K. P. Murphy. *Machine Learning: A Probabilistic Perspective*. The MIT Press, 2012.
- [8] S. Kay. *Modern spectral estimation : Theory and application*. Englewood Cliffs, N.J. : Prentice Hall, 1988.
- [9] P. Stoica and M. Randolph. *Spectral Analysis of Signals*. Prentice Hall, Inc., 01 2005.
- [10] N. R. Lomb. Least-squares frequency analysis of unequally spaced data. *Astrophysics and Space Science*, 39(2):447–462, Feb 1976.
- [11] J. D. Scargle. Studies in astronomical time series analysis. II - statistical aspects of spectral analysis of unevenly spaced data. *The Astrophysical Journal*, 263:835, Dec

1982.

- [12] R. Horn. *Topics in Matrix Analysis*. Cambridge University Press, 1994.
- [13] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, 2005.
- [14] T. K. Moon and W. C. Stirling. *Mathematical Methods and Algorithms for Signal Processing*. Pearson, 1999.
- [15] J. V. Candy. Bayesian estimation. In *Bayesian Signal Processing*, pages 20–51. John Wiley & Sons, Inc., Aug 2016.
- [16] S Boyd and L Vandenberghe. *Convex Optimization, With Corrections 2008*. Cambridge University Press, 2004.
- [17] D. Bertsimas and J. N. Tsitsiklis. *Introduction to Linear Optimization (Athena Scientific Series in Optimization and Neural Computation, 6)*. Athena Scientific, 1997.
- [18] P. Deuffhard. *Newton Methods for Nonlinear Problems*. Springer, 2005.
- [19] M. Avriel. *Nonlinear Programming: Analysis and Methods (Dover Books on Computer Science)*. Dover Publications, 2003.
- [20] M. J. D. Powell. An efficient method for finding the minimum of a function of several variables without calculating derivatives. *The Computer Journal*, 7(2):155–162, 1964.
- [21] F. Boone. Interferometric array design: Distributions of fourier samples for imaging. *Astronomy & Astrophysics*, 386(3):1160–1171, May 2002.
- [22] R. Levanda and A. Leshem. Synthetic aperture radio telescopes. *IEEE Signal Processing Magazine*, 27(1):14–29, Jan 2010.
- [23] L. Araya-Hernández, J. F. Silva, A. Osses, and F. Tobar. A Bayesian mixture-of-Gaussians model for astronomical observations in interferometry. In *2017 CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON)*, 2017.
- [24] F. Tobar, T. Bui, and R. Turner. Learning stationary time series using Gaussian processes with nonparametric kernels. In *Proc. of Neural Information Processing Systems 28*, pages 3483–3491, 2015.
- [25] F. Tobar and R. E. Turner. Modelling of complex signals using Gaussian processes. In *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 2209–2213, 2015.
- [26] D. Lewandowski, D. Kurowicka, and H. Joe. Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, 100(9):1989 – 2001, 2009.

- [27] P. Welch. The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics*, 15(2):70–73, June 1967.
- [28] Y. Rivenson and A. Stern. Compressed imaging with a separable sensing operator. *IEEE Signal Processing Letters*, 16(6):449–452, June 2009.
- [29] R. Narayan and R. Nityananda. Maximum entropy image restoration in astronomy. *Annual Review of Astronomy and Astrophysics*, 24(1):127, Sep 1986.
- [30] M. Cárcamo, P.E. Román, S. Casassus, V. Moral, and F.R. Rannou. Multi-GPU maximum entropy image synthesis for radio astronomy. *Astronomy and Computing*, 22:16–27, Jan 2018.
- [31] ALMA Partnership et al. The 2014 alma long base campaign first results from high resolution observations toward the hl tau region - iopscience. *The Astrophysical Journal*, 808(1):L3, July 2015.
- [32] F. Tobar. Bayesian Nonparametric Spectral Estimation. In *Advances in Neural Information Processing Systems*. 2018.
- [33] C. E. Rasmussen. The infinite Gaussian mixture model. In *In Advances in Neural Information Processing Systems 12*, pages 554–560. MIT Press, 2000.

Capítulo 7

Apéndice

7.1. Vectorización de una matriz

La vectorización de una matriz es un operador que trabaja sobre matrices arbitrarias.

Definición 7.1 Sea $M \in \mathcal{M}_{n \times m}(E)$ una matriz, se define la vectorización por columnas como la función $\text{vec} : E^{n \times m} \rightarrow E^{n \cdot m}$ ta que

$$M = \begin{bmatrix} m_{11} & \dots & m_{1N} \\ m_{21} & \dots & m_{2N} \\ \vdots & & \vdots \\ m_{n1} & \dots & m_{nm} \end{bmatrix} \Rightarrow \text{vec}(M) = [m_{11}, \dots, m_{n1}, m_{12}, \dots, m_{n2}, \dots, m_{1m}, \dots, m_{nm}]^T. \quad (7.1)$$

y denotaremos $\vec{M} = \text{vec}(M)$.

7.2. Producto de Kronecker

Definición 7.2 (Producto de Kronecker) Sea una $A \in \mathcal{M}_{n_A \times m_A}$ matriz

$$A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{n_A m_A} \end{bmatrix} \quad (7.2)$$

y una matriz $B \in \mathcal{M}_{n_B \times m_B}$ dada. El producto de Kronecker, denotado como $A \otimes B$, está definido por bloques como:

$$A \otimes B = \begin{bmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & & \vdots \\ a_{n_A 1}B & \cdots & a_{n_A m_A}B \end{bmatrix}. \quad (7.3)$$

7.3. Procesos Gaussianos

Definición 7.3 (Proceso Gaussiano) *Un proceso gaussiano es una colección variables aleatorias, que cumple con que al tomar cualquier conjunto finito de puntos distribuyen conjuntamente gaussianos.*

Observación Un proceso gaussiano queda completamente definido por una función de medias $m(\mathbf{x})$ y una función de covarianza $k(\mathbf{x}, \mathbf{x}')$ definidas por

$$m(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})] \quad (7.4)$$

$$k(\mathbf{x}, \mathbf{x}') = \mathbb{E}[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))], \quad (7.5)$$

y denotamos un proceso gaussiano como

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')). \quad (7.6)$$