

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-27

# **SEPARACIÓN CIEGA DE FUENTES MEDIANTE PCA E ICA, ORIENTADA A LA SELECCIÓN DE LONGITUDES DE ONDA DISCRIMINANTES EN EL MODELO TEÓRICO DE INTERACCIÓN LUZ-PIEL POR REFLECTANCIA DIFUSA**

Autor

Camilo Alejandro Bermúdez Mejía

Programa Académico

Ingeniería Electrónica

Director(es) del trabajo de grado

July Andrea Galeano Zea

Maria Constanza Torres Madroñero

**INSTITUTO TECNOLÓGICO METROPOLITANO**

**2018**

	<b>INFORME FINAL DE TRABAJO DE GRADO</b>	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

## RESUMEN

---

Hasta la actualidad, la diagnosis mediante biopsia ha sido muy utilizada como prueba definitiva para determinar diferentes patologías. Sin embargo, esta técnica presenta alta sensibilidad, causa dolor y compromete el tejido del paciente. Como alternativa, técnicas no invasivas basadas en espectrometría de tejido biológico son investigadas para ofrecer una alternativa de diagnóstico. Para el caso de la piel, una alternativa de realizarla consiste en la adquisición de imágenes hiper/multiespectrales sobre una región de interés del tejido afectado, en la cual la posición de cada pixel determina una firma espectral de reflectancia. Sin embargo, estas firmas medidas consisten de una mezcla de señales fuente (firmas espectrales) que se relacionan con la absorción por los diferentes componentes biológicos de la piel.

De acuerdo con lo anterior, el siguiente producto de laboratorio de investigación pretende estimar las firmas espectrales fuente (o subyacentes) correspondientes a la reflectancia difusa de piel humana, a partir de un modelo de interacción luz piel y un método de separación ciega de fuentes. Este producto se desarrolla como etapa previa para la diagnosis de Leishmaniasis cutánea a partir de la cuantificación de los componentes biológicos en la piel, mediante el procesamiento de imágenes multispectrales con técnicas de separación ciega de fuentes.

Para el cumplimiento del objetivo central se determinaron 5 etapas metodológicas: 1 – simulación del modelo de interacción luz–piel por reflectancia difusa, 2 – implementación del análisis de componentes principales (PCA) para eliminar correlación entre las fuentes de las firmas de reflectancia simuladas, 3 – estimar las firmas espectrales fuente utilizando el análisis de componentes independientes (ICA), 4 – Implementación de búsqueda exploratoria de proyecciones, 5 – comparación de las fuentes estimadas mediante ICA y la búsqueda exploratoria de proyecciones, con el propósito de validar el método de separación utilizado, y 6 – comparar visualmente los espectros de absorción simulados con

	<b>INFORME FINAL DE TRABAJO DE GRADO</b>	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

los reconstruidos a partir de ICA, para identificar los rangos de longitudes de onda donde la forma de la firma espectral sea coherente con los espectros de absorción simulados.

Aunque tanto el análisis de componentes independientes como la búsqueda exploratoria de proyecciones estimaron las mismas señales fuente, la última técnica resulta exhaustiva debido a que las fuentes son estimadas secuencialmente (el cual utiliza además la ortogonalización de Gram–Schmidt), mientras que el análisis de componentes independientes las estima al mismo tiempo, siendo este óptimo para realizar la separación BSS en problemas generales. Sin embargo, debido a que ICA estima las formas de onda de las señales fuente, al compararlas con los espectros de absorción en la dermis, epidermis y sangre (por oxi/deoxihemoglobina), se evidencia una alteración en la forma de las firmas en la región del espectro ultravioleta ( $\lambda < 380nm$ ) y del infrarrojo cercano ( $\lambda > 800nm$ ). Esto puede deberse a la poca densidad de datos utilizados, pues debería obtenerse una mejor estimación de las fuentes para una mayor cantidad de mezclas.

De los resultados obtenidos, se concluye que las formas de las firmas reconstruidas en el rango del espectro visible comprendido entre 500 y 600nm son coherentes con los espectros simulados. Esta banda espectral coincide con la utilizada por (Yudovsky & Pilon, 2010) en su trabajo de inversión del modelo mediante simulación Monte–Carlo.

*Palabras clave:* modelo de interacción luz–piel, análisis de componentes principales e independientes, separación ciega de fuentes, reflectancia difusa.

	<b>INFORME FINAL DE TRABAJO DE GRADO</b>	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

## RECONOCIMIENTOS

---

Expreso profundo agradecimiento a la Dra. July Andrea Galeano Zea y a la Dra. Maria Constanza Torres Madroñero por su interés, dedicación y total apoyo durante las asesorías y acompañamiento en la realización de este producto de laboratorio.

Agradezco al docente investigador Sebastián Roldán Vasco por su interés en mi proceso formativo en investigación y por ser mi mentor de tiempo extra.

Adicionalmente agradezco a mis amigos Johnatan, Deivid y Oscar por brindarme su ayuda en las situaciones complicadas y a Andrea por sus palabras llenas de esperanza en los momentos más difíciles.

Finalmente, doy las gracias al Instituto Tecnológico Metropolitano y al grupo de investigación MATyER, el cual está adscrito a la facultad de ingenierías, por brindar los espacios, equipos y recursos que se utilizaron durante el desarrollo de este trabajo.

El presente trabajo de grado hace parte del proyecto de investigación “Desarrollo y evaluación de un sistema portátil no invasivo basado en imágenes multiespectrales para el diagnóstico y seguimiento de tratamientos de úlceras cutáneas causadas por Leishmaniasis” financiado por COLCIENCIAS bajo el código 57186.

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

## ACRÓNIMOS

---

*BSS* Separación Ciega de Fuentes (*Blind Source Separation*, del inglés).

*EEG* Electroencefalografía (*Electroencephalography*, del inglés).

*EKG* Electrocardiograma (*Elektrokardiogramm*, del alemán).

*GSO* Ortogonalización de Gram-Schmidt (*Gram-Schmidt Orthogonalization*, del inglés).

*ICA* Análisis de Componentes Independientes (*Independent Component Analysis*, del inglés).

*MATyER* Materiales Avanzados y Energía (Grupo de investigación adscrito a la facultad de ingenierías del Instituto Tecnológico Metropolitano).

*PCA* Análisis de Componentes Principales (*Principal Component Analysis*, del inglés).

*pdf* función de densidad de probabilidad (*probability density function*, del inglés).

*SVD* Descomposición de Valor Singular (*Singular Value Decomposition*, del inglés).

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

## TABLA DE CONTENIDO

1. INTRODUCCIÓN .....	7
1.1. GENERALIDADES .....	7
1.1.1. Pertinencia .....	7
1.1.2. Justificación.....	7
1.1.3. Planteamiento del problema .....	8
1.2. OBJETIVOS.....	9
1.2.1. General.....	9
1.2.2. Específicos.....	9
1.3. ORGANIZACIÓN DE LA TESIS .....	9
2. MARCO TEÓRICO .....	11
2.1. MODELO DE INTERACCIÓN LUZ-PIEL POR REFLECTANCIA DIFUSA.....	11
2.2. ANÁLISIS DE COMPONENTES PRINCIPALES .....	14
2.2.1. Pasos para implementar PCA.....	14
2.2.2. Observaciones.....	16
2.3. ANÁLISIS DE COMPONENTES INDEPENDIENTES.....	17
2.3.1. El problema de separación entre señales.....	18
2.3.2. Principios básicos para la separación de señales.....	19
2.3.3. Pasos para implementar ICA.....	20
2.3.4. Observaciones.....	25
3. METODOLOGÍA.....	26
4. RESULTADOS Y DISCUSIÓN.....	29
5. CONCLUSIONES, RECOMENDACIONES Y TRABAJO FUTURO .....	34
REFERENCIAS .....	38
APÉNDICE .....	40

	<b>INFORME FINAL DE TRABAJO DE GRADO</b>	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

# 1. INTRODUCCIÓN

---

## 1.1. GENERALIDADES

### 1.1.1. Pertinencia

Aunque la biopsia ha sido una técnica muy utilizada como prueba definitiva para diagnosticar diversas patologías, esta presenta alta sensibilidad, inclusive llegando a errar el diagnóstico en algunos casos. Además, por ser una forma invasiva de diagnosis, la biopsia puede comprometer el tejido bajo estudio y a la vez causar dolor al paciente.

Como alternativa, técnicas no invasivas basadas en espectrometría de tejido biológico son investigadas para dar una diagnosis temprana y más precisa.

Entre tales técnicas, aquellas que se basan en el análisis de imágenes por reflectancia difusa permiten estimar la concentración de oxi/deoxihemoglobina y agua, los cuales presentan un espectro de absorción definido en el rango visible y en el cercano infrarrojo.

### 1.1.2. Justificación

La reflectancia difusa ha sido modelada matemáticamente con el propósito de describir la interacción de la luz con los componentes estructurales y bioquímicos que conforman el tejido (Yudovsky & Pilon, 2010; Zonios & Dimou, 2006). Igualmente, las propiedades ópticas provenientes de diversos tejidos *in vitro* han sido reportadas en la literatura (Cheong, Prah, & Welch, 1990).

Actualmente se desarrollan métodos de diagnóstico basados en el análisis de imágenes hiper/multiespectrales ya que permiten obtener información sobre la composición del tejido a partir de su firma espectral de reflectancia por cada pixel dada una región de interés. Sin embargo, debido a la densidad de datos, técnicas de selección de bandas son

	<b>INFORME FINAL DE TRABAJO DE GRADO</b>	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

utilizadas con el propósito de reducir la dimensión de los datos antes de procesar este tipo de imágenes.

En la obtención del modelo inverso para recuperar las propiedades ópticas de la piel mediante simulación Monte-Carlo entre 490 y 690 *nm*, (Yudovsky & Pilon, 2010) sugieren que las longitudes de onda pueden ser seleccionadas estratégicamente para que coincidan con los picos de absorción de oxi/deoxihemoglobina, pues aumenta la sensibilidad del método inverso en función de la saturación de oxígeno, debido a que exhibe absorciones distintivas y significativas.

También, (Ramanath & Snyder, 2003) realizó la selección de bandas utilizando análisis de componentes independientes, donde la matriz de pesos es evaluada con el propósito de determinar la contribución de cada banda individual y seleccionar aquellas que aportan mayor información en las imágenes hiperespectrales.

### **1.1.3. Planteamiento del problema**

Debido a la utilidad que aportan estos sistemas diagnósticos (los cuales siguen actualmente en desarrollo) y a la carga computacional que supone el análisis de datos hiper/multiespectrales (los cuales pueden ser redundantes entre algunas bandas del espectro visible y del cercano infrarrojo), se requiere de metodologías que permitan realizar una reducción de dimensión ya sea por técnicas de *feature selection* o *feature extraction*. En el caso de este trabajo de grado, se implementa el análisis de componentes principales (PCA) seguido del análisis de componentes independientes (ICA), como método de *feature extraction* para extraer las señales fuente que corresponden a diferentes firmas de absorción, dado un conjunto de reflectancias simuladas con propiedades ópticas aleatorias. Esto se realiza con el propósito de estimar aquellas bandas relevantes en el espectro visible y en el cercano infrarrojo.

	<b>INFORME FINAL DE TRABAJO DE GRADO</b>	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

## **1.2. OBJETIVOS**

### **1.2.1. General**

Estimar las firmas espectrales fuentes correspondientes a la reflectancia difusa de piel humana a partir de un modelo interacción luz–piel y un método de separación ciega de fuentes.

### **1.2.2. Específicos**

- Simular el modelo de interacción luz–piel para procesar las firmas de reflectancia difusa mediante el análisis de componentes principales e independientes.
- Realizar preprocesamiento de los datos mediante el análisis de componentes principales para eliminar la correlación entre los mismos.
- Desarrollar una rutina computacional basada en análisis de componentes independientes para estimar las firmas espectrales fuente a partir de las firmas espectrales pre-procesadas.

## **1.3. ORGANIZACIÓN DE LA TESIS**

En el capítulo 1 se introduce al lector en las generalidades del proyecto, donde se describen la pertinencia, justificación y planteamiento del problema abordados durante la ejecución de este trabajo.

En el capítulo 2 se describen los conceptos utilizados durante la realización del proyecto, que permiten comprender el procedimiento metodológico. En este capítulo, las ecuaciones

	<b>INFORME FINAL DE TRABAJO DE GRADO</b>	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

que rigen el modelo de interacción luz–piel por reflectancia difusa, y los algoritmos utilizados para obtener las componentes principales e independientes son descritos al lector. Además, se explican las propiedades utilizadas para la separación de las señales fuente, las cuales se deducen a partir de los efectos que resultan durante el proceso de mezcla. Este capítulo concluye realizando una comparación entre los dos métodos de separación ciega de fuentes presentados, el método basado en la exploración de proyecciones ortogonales y en el análisis de componentes independientes.

En el capítulo 3 se describe brevemente la metodología, la cual está apoyada por los conceptos que se presentan en el capítulo 2.

En el capítulo 4 se presentan los resultados al lector a medida que se describe como fue realizada cada etapa del proyecto, los cuales van acompañados de una breve discusión que argumenta las condiciones o restricciones bajo las cuales fueron obtenidos.

Finalmente, en el capítulo 5 se presenta las conclusiones generadas por el autor apoyado a partir de los resultados obtenidos, las recomendaciones que no fueron tomadas en cuenta y que fueron descubiertas a medida que fue realizado el proyecto y el trabajo futuro.

## 2. MARCO TEÓRICO

---

### 2.1. MODELO DE INTERACCIÓN LUZ-PIEL POR REFLECTANCIA DIFUSA

En (Zonios & Dimou, 2006), se asume que la piel se compone de un medio homogéneamente turbio, casi infinito, y con geometría unidimensional; con el propósito de representar el modelo de reflectancia difusa en su forma más simple. En este modelo, como se representa en la Figura 1, para un punto  $x_i$  en el cual la luz incide, una parte es absorbida y la otra parte es reflejada.

La parte reflejada emerge del tejido como reflectancia difusa  $R(\lambda)$ , la cual se denota mediante:

$$R(\lambda) = \frac{1}{1 + K \frac{\mu_a(\lambda)}{\mu_s(\lambda)}} \quad (2.1)$$

Donde  $\mu_a$  y  $\mu_s$  corresponden a los componentes de absorción y dispersión, y  $K$  representa un coeficiente de reescalamiento para que coincida con otros modelos más precisos (Zonios & Dimou, 2006).

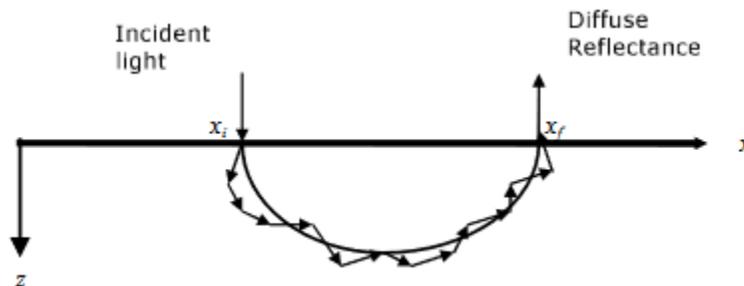


Figura 1. Proceso de reflectancia difusa (Zonios & Dimou, 2006).

	<b>INFORME FINAL DE TRABAJO DE GRADO</b>	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

A continuación se describen las ecuaciones necesarias para obtener  $\mu_a$  y  $\mu_s$  para poder incorporarlas en el modelo (Yudovsky & Pilon, 2010):

$$\mu_{a,back}(\lambda) = (7.84 \times 10^8)\lambda^{-3.255} cm^{-1} \quad (2.2)$$

$$\mu_{a,mel}(\lambda) = (6.60 \times 10^{11})\lambda^{-3.33} cm^{-1} \quad (2.3)$$

Donde  $\mu_{a,back}$  y  $\mu_{a,mel}$  dependen de la longitud de onda  $\lambda$  y determinan los coeficientes de absorción del fondo y de la melanina respectivamente. La absorción en la epidermis se debe principalmente a estas dos componentes:

$$\mu_{a,epi}(\lambda) = \mu_{a,mel}(\lambda)f_{mel} + \mu_{a,back}(\lambda)(1 - f_{mel}) \quad (2.4)$$

Los coeficientes de absorción  $\mu_{a,oxy}$  y  $\mu_{a,deoxy}$  corresponden a la absorción por oxi/deoxyhemoglobina y están dados por:

$$\mu_{a,oxy}(\lambda) = \epsilon_{oxy}(\lambda)C_{heme}SO_2/66500 \quad (2.5)$$

$$\mu_{a,deoxy}(\lambda) = \epsilon_{deoxy}(\lambda)C_{heme}(1 - SO_2)/66500 \quad (2.6)$$

Siendo  $\epsilon_{oxy}$  y  $\epsilon_{deoxy}$  los coeficientes de extinción molar de oxihemoglobina y deoxihemoglobina respectivamente, donde  $C_{heme}$  es el valor promedio de concentración de hemoglobina con un valor típico de 150 g/l.

En el rango visible, la absorción en la sangre  $\mu_{a,blood}$  se debe a la absorción por oxi/deoxihemoglobina. Por lo cual:

$$\mu_{a,blood}(\lambda) = \mu_{a,oxy}(\lambda) + \mu_{a,deoxy}(\lambda) \quad (2.7)$$

	<b>INFORME FINAL DE TRABAJO DE GRADO</b>	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

El coeficiente de absorción de la dermis se compone principalmente por la absorción de la sangre y del fondo, y es obtenido mediante:

$$\mu_{a,derm}(\lambda) = \mu_{a,blood}(\lambda)f_{blood} + \mu_{a,back}(\lambda)(1 - f_{blood}) \quad (2.8)$$

Así, la absorción en la piel puede ser obtenida a partir de las ecuaciones 2.4 y 2.8:

$$\mu_a(\lambda) = \mu_{a,epi}(\lambda) + \mu_{a,derm}(\lambda) \quad (2.9)$$

El coeficiente de dispersión se obtiene con:

$$\mu_s(\lambda) = C \left( \frac{\lambda}{\lambda_0} \right)^{-b} \quad (2.10)$$

Donde el término  $\lambda_0 = 1nm$  es introducido para mantener consistencia entre las unidades.

En la tabla I se presentan los rangos permitidos para las propiedades biológicas de la piel humana, los cuales fueron estimados mediante el modelo inverso desarrollado por (Yudovsky & Pilon, 2010):

Tabla I. Valores frontera para las propiedades biológicas de la piel humana

Propiedad biológica	Símbolo	Límite inferior	Límite superior
Fracción del volumen de melanosomas	$f_{mel}$	1%	10%
Fracción de volumen de sangre	$f_{blood}$	0.2%	7%
Saturación de oxígeno	$S_{O_2}$	0%	100%
Constante de dispersión	$C$	$4.0 \times 10^5 \text{ cm}^{-1}$	$5.5 \times 10^5 \text{ cm}^{-1}$
Potencia de dispersión	$b$	1.10	1.50

	<b>INFORME FINAL DE TRABAJO DE GRADO</b>	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

## 2.2. ANÁLISIS DE COMPONENTES PRINCIPALES

El Análisis de Componentes Principales (en adelante PCA, de *Principal Component Analysis*), es un método que se utiliza principalmente en dos aplicaciones: el reconocimiento de patrones y la compresión de datos (Kong, Hu, & Duan, 2017).

Sea con el fin de poder clasificar la pertenencia de una clase dado un conjunto de características (patrón observado), para acelerar el tiempo de entrenamiento de un clasificador o reducir el espacio requerido para almacenar datos; en tales aplicaciones, es necesario obtener medidas que sean invariantes entre sí y que puedan representar los datos con poca distorsión (Kong et al., 2017).

El proceso utilizado para obtener estas medidas usualmente es denominado como *extracción de características* y consiste en proyectar los datos sobre un subespacio de menor dimensión que pueda describirlos con la menor pérdida de información relevante (Kong et al., 2017).

Geoméricamente, esto consiste en encontrar un sistema de ejes que sea ortogonal y que minimice la distancia de los datos para cada uno de estos, siendo estos ejes las primeras componentes principales las cuales describen la variabilidad de los datos (Ng, n.d.).

### 2.2.1. Pasos para implementar PCA

Sea  $X \in \mathbb{R}^{n \times p}$  la matriz que contiene  $n$  observaciones con  $p$  características iniciales, y  $Z \in \mathbb{R}^{n \times d}$  la matriz que contiene  $d$  características finales, donde  $d \leq p$ . Para reducir la dimensión de los datos con el fin de obtener  $Z$ , se deben seguir los siguientes pasos (Ng, n.d.):

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

1. Centrar los datos y escalarlos al restar la media y dividir por la desviación estándar:

$$\tilde{X} = \frac{X - \bar{X}}{S} \quad (2.11)$$

Donde:

$$S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n |X_i - \bar{X}|^2} \quad (2.12)$$

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad (2.13)$$

2. Calcular la matriz de covarianza  $\Sigma \in \mathbb{R}^{p \times p}$ :

$$\Sigma = \frac{1}{n-1} \tilde{X}^T \tilde{X} \quad (2.14)$$

3. Obtener los vectores y valores propios de  $\Sigma$  mediante la *Descomposición de Valor Singular* (en adelante SVD, de *Singular Value Decomposition*), la cual permite descomponer una matriz de la forma  $X = UDV^T$ , donde cada valor propio en la diagonal principal de la matriz  $D$  tiene asociado un vector propio  $U_i$ .
4. Normalizar las varianzas (valores propios) dividiendo por la traza de  $D$  y calcular la suma acumulativa de la diagonal principal de la matriz resultante. Después seleccionar  $d$  de acuerdo a las siguientes condiciones de  $k$ :

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

- Seleccionar la menor cantidad de  $d$  vectores propios que mantengan el porcentaje  $k$  de varianza retenida (es decir, la suma acumulada que retenga mayor porcentaje de la varianza) cuando  $0 < k < 1$ :
- De lo contrario, cuando  $k > 1$ , entonces seleccionar  $d = k$  vectores propios como nuevo conjunto de características.

Lo anterior se resume en encontrar el valor mínimo para  $d$  que cumpla con la siguiente condición:

$$\frac{\sum_{i=1}^d D_{ii}}{\sum_{i=1}^p D_{ii}} \geq k \quad (2.15)$$

5. Para reducir la dimensionalidad de los datos, se debe seleccionar las primeras  $d$  componentes principales (vectores columna) de  $U$ , donde  $U \in \mathbb{R}^{p \times p}$ . De este modo,  $\tilde{U} \in \mathbb{R}^{p \times d}$  representa el sub-espacio sobre el cual los datos serán proyectados mediante:

$$Z = X\tilde{U} = XW \quad (2.16)$$

Donde  $W = \tilde{U}$ . Esta operación realiza una rotación de los datos en la base definida por los vectores  $U_1, U_2, \dots, U_d$ , los cuales son ortogonales.

### 2.2.2. Observaciones

- El primer paso del algoritmo constituye una etapa de preprocesamiento de los datos, con el fin de centrarlos y hacer que tengan un rango de valores comparables cuando estos tienen diferente escala. En caso contrario no resulta necesario escalar los datos al dividir por la desviación estándar.

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

- Una particularidad de implementar PCA mediante SVD consiste en que los vectores propios (vectores columna) de la matriz  $U$  son retornados en orden descendente de acuerdo a las varianzas de la traza en  $D$ , por lo cual no resulta necesario realizar permutación alguna entre la posición de estos vectores.
- La matriz  $U$  puede constituir un nuevo conjunto de características sin correlación y que, por lo tanto, podría suponer la independencia entre ellas cuando se implementa SVD sobre  $X^T$ . Puede verse un ejemplo de esto para la separación de señales de audio en (Stone, 2004).
- Finalmente, aunque PCA es utilizado para reducir la dimensión de un conjunto de datos o para obtener nuevas características que no tengan correlación entre sí, este también puede ser utilizado como una etapa previa cuando se utiliza en conjunto con otros métodos de optimización. Para realizar el preprocesamiento de los datos mediante *whitening* (también se conoce en la literatura como *sphering*), se debe reemplazar el último paso del algoritmo con:

$$Z = UD^{-1/2}U^T X \quad (2.17)$$

Donde  $Z \in \mathbb{R}^{p \times n}$  representa un nuevo conjunto de datos blanqueados.

### 2.3. ANÁLISIS DE COMPONENTES INDEPENDIENTES

El Análisis de Componentes Independientes (en adelante ICA, de *Independent Component Analysis*) es un método para realizar Separación Ciega de Fuentes (en adelante BSS, de *Blind Source Separation*), siendo comúnmente utilizado para estimar señales fuente subyacentes de un fenómeno observado utilizando solamente un conjunto de datos muestreados. Aunque el método no está restringido en el análisis de datos médicos, estos pueden tener forma de imágenes de resonancia magnética, señales de electrocardiografía (EKG) y

 Institución Universitaria	<b>INFORME FINAL DE TRABAJO DE GRADO</b>	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

electroencefalografía (EEG), de comunicaciones, entre otros. El hecho de que la separación sea “ciega” solo implica que los métodos BSS permiten separar señales, inclusive si no se tiene un conocimiento previo sobre su naturaleza. Sin embargo, ICA es un método que aún es investigado debido a que no se han explorado completamente sus limitaciones y el desempeño que puede presentar en diversas aplicaciones (Stone, 2004)□.

### 2.3.1. El problema de separación entre señales

El proceso de mezcla entre señales puede modelarse mediante (Stone, 2004):

$$\begin{matrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{matrix} = \begin{matrix} a_{11}s_1 + a_{12}s_2 + \dots + a_{1m}s_m \\ a_{21}s_1 + a_{22}s_2 + \dots + a_{2m}s_m \\ \vdots \\ a_{n1}s_1 + a_{n2}s_2 + \dots + a_{nm}s_m \end{matrix} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_m \end{bmatrix} \quad (2.18)$$

Lo cual puede escribirse más sucintamente como:

$$\mathbf{x} = \mathbf{A}\mathbf{s} \quad (2.19)$$

Donde cada elemento de las variables vectoriales  $\mathbf{x}$  y  $\mathbf{s}$  contiene una mezcla  $x_i$  y una fuente  $s_i$  con sus  $n$  observaciones respectivamente. Además, la matriz  $\mathbf{A}$  contiene los coeficientes de mezcla  $a_{ij}$  para  $i = 1, 2, \dots, n$  y  $j = 1, 2, \dots, m$ ; de la cual cada fila constituye un vector de coeficientes que ponderan con la suma de las señales fuente para obtener así una mezcla. De este modo, la separación de las fuentes podría realizarse mediante (Stone, 2004):

$$\mathbf{y} = \mathbf{W}\mathbf{x} \approx \mathbf{s} \quad (2.20)$$

Donde la variable vectorial  $\mathbf{y}$  contiene las señales fuente estimadas a partir de la transformación de los datos mediante el subespacio definido por  $\mathbf{W}$ , donde  $\mathbf{W} = \mathbf{A}^{-1}$ . Sin

	<b>INFORME FINAL DE TRABAJO DE GRADO</b>	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

embargo, los coeficientes de mezcla son desconocidos, y por lo tanto, los coeficientes de separación en  $W$  no pueden ser estimados de esta forma (Stone, 2004).

### 2.3.2. Principios básicos para la separación de señales

Los principios básicos que permiten estimar cada señal fuente  $y_i$  se derivan a partir de los siguientes efectos que surgen durante el proceso de mezcla (Stone, 2004)□:

- **Independencia:**

Suponiendo que las señales fuente son estadísticamente independientes, sus mezclas no lo son. Esto se debe a que cada señal fuente es compartida por todas las mezclas, y las características y atributos resultantes aseguran que estas no pueden serlo.

- **Normalidad:**

Cualquier mezcla entre señales tiene un histograma que tiende a tener forma gaussiana en comparación al de las señales fuente (siempre que estas tengan histogramas muy diferentes).

- **Complejidad:**

La complejidad de cualquier mezcla es mayor o igual a la señal fuente más simple.

Basándose en estas conjeturas, resulta plausible pensar en maximizar o minimizar alguna de estas propiedades con el propósito de estimar tantas señales fuente como sea posible.

	<b>INFORME FINAL DE TRABAJO DE GRADO</b>	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

De este modo, si las señales estimadas son independientes, tienen histogramas con distribución leptocúrtica (con forma de pico) o tienen baja complejidad, es muy probable que estas sean las señales fuente requeridas. Pero si hay menos mezclas que señales fuente a estimar, entonces los métodos BSS difícilmente extraerán dichas fuentes (Stone, 2004) □.

Sin embargo, en la práctica es muy común tener que la cantidad de mezclas es mucho mayor que la cantidad de señales fuente a estimar. Por lo tanto, si se conoce el número de señales fuente a estimar y es menor al número de mezclas, entonces el número de señales a ser extraídas por ICA puede obtenerse mediante el preprocesamiento de las mezclas utilizando PCA (Stone, 2004).

A continuación, se presentan dos formas para implementar BSS mediante ICA (Stone, 2004):

### 2.3.3. Pasos para implementar ICA

Sea  $\mathbf{x} \in \mathbb{R}^{m \times n}$  la variable vectorial que contiene  $m$  mezclas, cada una con  $n$  observaciones; y  $\mathbf{y} \in \mathbb{R}^{k \times n}$  la variable vectorial que contiene  $k$  señales fuente estimadas, donde  $k \leq m$ .

Suponiendo que las mezclas no tienen correlación y, por lo tanto, son independientes entre sí, es posible implementar una separación BSS mediante la búsqueda de proyección ortogonal exploratoria, para lo cual deben seguirse los siguientes pasos:

1. Inicializar un vector de separación  $w$  aleatoriamente y que tenga norma unitaria:

$$w = (\alpha_1 \quad \alpha_2 \quad \dots \quad \alpha_m)^T \quad (2.21)$$

$$w = \frac{w}{|w|} \quad (2.22)$$

 Institución Universitaria	<b>INFORME FINAL DE TRABAJO DE GRADO</b>	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

2. Estimar una señal fuente:

$$y_i = w^T \mathbf{x} \quad (2.23)$$

3. Evaluar la curtosis:

$$K = \frac{\frac{1}{N} \sum_{j=1}^n (y_{ij} - \bar{y}_i)^4}{\left(\frac{1}{N} \sum_{j=1}^n (y_{ij} - \bar{y}_i)^2\right)^2} - 3 \quad (2.24)$$

Si la media es igual a cero y la varianza es unitaria, la curtosis puede escribirse en términos del primer momento como:

$$K = \frac{1}{n} \sum_{j=1}^n y_{ij}^4 - 3 = E[y_i^4] - 3 \quad (2.25)$$

Donde  $E[\dots]$  corresponde al valor esperado.

4. Calcular el gradiente de la curtosis:

$$\frac{\partial K(y_i)}{\partial w} = \frac{\partial}{\partial w} E[(w^T \mathbf{x})^4] - 3 = cE[\mathbf{x}(w^T \mathbf{x})^3] = cE[\mathbf{x}y_i^3] \quad (2.26)$$

Donde  $c = 4$  es una constante de proporcionalidad, la cual es omitida por conveniencia (es decir, se hace  $c = 1$ ).

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

5. Actualizar  $w$  mediante el paso del gradiente ascendente:

$$w = w + \eta \frac{\partial K}{\partial w} = w + \eta E[xy_i^3] \quad (2.27)$$

Donde  $\eta$  representa el factor del paso del gradiente.

6. Normalizar  $w$  para que tenga longitud unitaria en la siguiente iteración (mediante la ecuación 2.22).

7. Repetir desde el paso 2 hasta converger. La condición de parada puede realizarse mediante una o varias de las siguientes condiciones:

- Cuando la tasa de cambio de la curtosis sea inferior a una tolerancia:

$$\frac{\partial K}{\partial w} \leq tol \quad (2.28)$$

Es decir, que la curtosis en la iteración actual no haya tenido un cambio significativo respecto a la iteración anterior.

- Cuando se haya superado un límite permitido de iteraciones.

8. Realizar Ortogonalización de Gram-Schmidt (en adelante GSO, de *Gram-Schmidt Orthogonalization*), para sustraer la señal  $y_i$  estimada del conjunto de mezclas y así generar un nuevo conjunto de mezclas remanente:

$$\mathbf{x} = \mathbf{x} - \frac{E[y_i \mathbf{x}] y_i}{E[y_i^2]} \quad (2.29)$$

	<b>INFORME FINAL DE TRABAJO DE GRADO</b>	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

9. Concatenar  $y_i$  en  $\mathbf{y}$  y repetir el algoritmo hasta extraer tantas  $k$  señales fuente se requiera, siempre que  $k \leq m$ .

El algoritmo anterior se basa en la minimización del efecto de normalidad descrito anteriormente, donde la curtosis es utilizada como una métrica de esta propiedad y como función objetivo en el algoritmo.

Sin embargo, también puede realizarse una separación BSS minimizando o maximizando las demás propiedades que surgen de los efectos al mezclar señales.

Por ejemplo, se puede insertar una función que tenga un modelo específico de función de densidad de probabilidad (en adelante pdf, de *probability density function*). Un ejemplo de esto consiste en utilizar el algoritmo *Infomax* (basado en el algoritmo *Bell-Sejnowski*), en el cual se mide la entropía con el fin de maximizar la propiedad de independencia entre las señales fuente estimadas dado un modelo de pdf específico (Stone, 2004). Este algoritmo puede implementarse siguiendo los siguientes pasos:

1. Inicializar aleatoriamente una matriz  $\mathbf{W} \in \mathbb{R}^{m \times m}$  que contiene los  $m$  vectores de separación  $w_i$ :

$$\mathbf{W} = (w_1 \quad w_2 \quad \dots \quad w_m)^T \quad (2.30)$$

2. Estimar todas las señales fuente:

$$\mathbf{y} = \mathbf{W}^T \mathbf{x} \quad (2.31)$$

3. Evaluar las señales extraídas en la función con pdf modelo:

$$\mathbf{Y} = \tanh(\mathbf{W}^T \mathbf{x}) = \tanh(\mathbf{y}) \quad (2.32)$$

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

4. Evaluar la entropía:

$$h(\mathbf{Y}) = \frac{1}{n} \sum_{i=1}^m \sum_{t=1}^n \ln(1 - \tanh^2(w_i^T \mathbf{x}^t)) + \ln|\mathbf{W}| \quad (2.33)$$

Donde el índice  $t$  en la sumatoria indica que el promedio se realiza para todas las observaciones de cada señal.

5. Calcular el gradiente de la entropía:

$$\frac{\partial h(\mathbf{Y})}{\partial w_{ij}} = \frac{1}{|\mathbf{W}|} + \frac{2}{n} \mathbf{Y} \mathbf{x}^T \quad (2.34)$$

6. Actualizar  $\mathbf{W}$  mediante el paso del gradiente ascendente:

$$\mathbf{W} = \mathbf{w} + \eta \frac{\partial h}{\partial w_{ij}} = \mathbf{W} + \eta \left( \frac{1}{|\mathbf{W}|} + \frac{2}{n} \mathbf{Y} \mathbf{x}^T \right) \quad (2.35)$$

Donde  $\eta$  representa el factor de paso del gradiente.

7. Normalizar  $\mathbf{W}$  para que cada vector tenga longitud unitaria en la siguiente iteración.

$$\mathbf{W} = \frac{\mathbf{W}}{|\mathbf{W}|} \quad (2.36)$$

8. Repetir desde el paso 2 hasta converger. La condición de parada puede realizarse mediante una o varias de las siguientes condiciones:

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

- Cuando la tasa de cambio de la entropía sea inferior a una tolerancia:

$$\frac{\partial h}{\partial w_{ij}} \leq tol \quad (2.37)$$

Es decir, que la entropía en la iteración actual no haya tenido un cambio significativo respecto a la iteración anterior (para todas las señales estimadas).

- Cuando se haya superado un límite permitido de iteraciones.

#### 2.3.4. Observaciones

- El método de separación mediante *ICA Infomax* estima todas las señales fuente al mismo tiempo. Esta técnica supone un tiempo de convergencia menor, comparada con la búsqueda exhaustiva de proyecciones ortogonales.  
Esto se debe a que luego de recuperar una señal fuente, el método basado en proyecciones debe retirarla del conjunto de mezclas mediante GSO, generando así un nuevo conjunto remanente. Por lo tanto, el hecho de tener que ejecutar el algoritmo varias veces para estimar secuencialmente las fuentes lo hace ineficiente en comparación a *ICA Infomax*.

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

### 3. METODOLOGÍA

---

Para simular el modelo de interacción luz-piel y obtener las firmas de reflectancia se realizaron los siguientes pasos:

1. Inicialización de las siguientes constantes:

$\lambda$  entre 250 y 1000  $nm$ , con pasos de 2  $nm$  y  $\lambda_0 = 1 nm$ .

$C_{heme} = 150 g/l$  y  $K = 3$ .

2. Inicialización de las siguientes propiedades ópticas de la piel:

$SO_2$ ,  $f_{blood}$  y  $f_{mel}$  entre los valores límite reportados en la tabla I, y  $C$  y  $b$  con sus valores promedio. Se generan  $M = 3$  valores para cada propiedad.

3. Cargar los datos correspondientes a los coeficientes de extinción molar de oxi/deoxihemoglobina  $\epsilon_{oxy}$  y  $\epsilon_{deoxy}$  (datos tabulados obtenidos de (Jacques & Prahl, 2018)).

4. Calcular los espectros de absorción de la melanina y del fondo  $\mu_{a,mel}$  y  $\mu_{a,back}$  (ecuaciones 2.2 y 2.3).

5. Calcular los siguientes espectros de absorción con la  $m_i$  propiedad óptica generada aleatoriamente:

- Espectro de absorción debido a la sangre  $\mu_{a,blood}$  (ecuación 2.7), la cual depende de  $SO_2$  y de los espectros de absorción por oxi/deoxihemoglobina  $\mu_{a,oxy}$  y  $\mu_{a,deoxy}$  (ecuaciones 2.5 y 2.6).

	<b>INFORME FINAL DE TRABAJO DE GRADO</b>	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

- Espectro de absorción de la dermis  $\mu_{a,derm}$  (ecuación 2.8), el cual depende de  $f_{blood}$  y de  $\mu_{a,blood}$  y  $\mu_{a,back}$  (ecuaciones 2.2 y 2.7).
  - Espectro de absorción de la epidermis  $\mu_{a,epi}$  (ecuación 2.4), el cual depende de  $f_{mel}$  y de  $\mu_{a,mel}$  y  $\mu_{a,back}$  (ecuaciones 2.2 y 2.3).
6. Calcular el espectro de absorción de la piel para el modelo  $\mu_a$  (ecuación 2.9), el cual depende de  $\mu_{a,epi}$  y  $\mu_{a,derm}$  (ecuaciones 2.4 y 2.8).
  7. Calcular el espectro de dispersión del modelo  $\mu_s$  (ecuación 2.10), el cual depende de  $C$  y  $b$ .
  8. Calcular el espectro de reflectancia  $R$  (ecuación 2.1), el cual depende de  $K$ , de  $\mu_a$  y  $\mu_s$  (ecuaciones 2.9 y 2.10).
  9. Repetir desde el paso 5 hasta calcular todos los espectros de reflectancia difusa con todas las  $m_i$  propiedades ópticas generadas aleatoriamente en el paso 2.

Luego, las firmas espectrales obtenidas a partir de la reflectancia difusa simulada son preprocesadas con el propósito de eliminar cualquier correlación que puedan tener los datos, por lo cual se siguen los siguientes pasos:

1. Los espectros de reflectancia adquieren media cero y varianza unitaria, por lo tanto, los datos son escalados (ecuaciones 2.11, 2.12 y 2.13).
2. Se realiza SVD sobre los datos y las primeras  $d = k = 3$  componentes principales son seleccionadas como el nuevo conjunto de mezclas sin correlación entre sí.

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

3. Se obtiene un nuevo conjunto de mezclas  $z = U$  (componentes principales), donde estas adquieren media cero y varianza unitaria.

Después, con el fin de obtener las firmas espectrales subyacentes a las reflectancias simuladas, se implementa ICA *Infomax* con  $x = z = U$  como el nuevo conjunto de mezclas. A continuación, se describen los pasos realizados y los parámetros utilizados:

1. Inicialización de las siguientes constantes:  
500 iteraciones permitidas con paso del gradiente de  $\eta = 0.25$ .
2. Extracción de las dimensiones de las mezclas, donde  $m$  corresponde a la cantidad de mezclas y  $n$  a las observaciones
3. Inicialización de la matriz de separación con la matriz identidad de tamaño  $m = 3$ .
4. El algoritmo ICA *Infomax* (ver sección 2.3.3) es implementado y las firmas espectrales subyacentes de la reflectancia simulada son estimadas.

Finalmente, las firmas espectrales obtenidas mediante ICA son comparadas con las gráficas de los espectros de absorción para discriminar aquellas bandas que generan una distorsión de la forma de onda.

## 4. RESULTADOS Y DISCUSIÓN

Debido a que la simulación del modelo se realizó generando valores aleatorios para las propiedades ópticas  $SO_2$ ,  $f_{blood}$  y  $f_{mel}$ , las gráficas presentadas en este trabajo corresponden a los siguientes valores (los cuales fueron retornados por la función diseñada para evaluar y graficar la reflectancia difusa):

Tabla II. Valores generados para las propiedades biológicas en simulación.

Propiedad biológica	Símbolo	Valor 1	Valor 2	Valor 3
Frac. del vol. de melanosomas	$f_{mel}$	0.0332%	0.0857%	0.0329%
Frac. de vol. de sangre	$f_{blood}$	0.0392%	0.0114%	0.0122%
Saturación de oxígeno	$S_{O_2}$	0.6991%	0.8909%	0.9593%
Constante de dispersión	$C$	$4.5 \times 10^5 \text{ cm}^{-1}$	$4.5 \times 10^5 \text{ cm}^{-1}$	$4.5 \times 10^5 \text{ cm}^{-1}$
Potencia de dispersión	$b$	1.30	1.30	1.30

Con estos parámetros generados por el programa, las firmas obtenidas para la absorción teórica por oxi/deoxihemoglobina son mostradas a continuación:

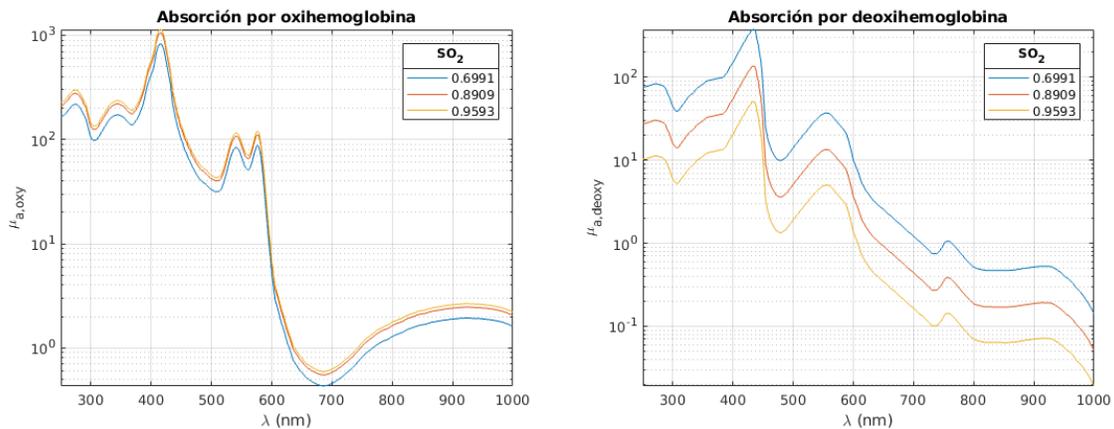


Figura 2. Espectro de absorción por oxihemoglobina (izq.) y por deoxihemoglobina (der). Fuente: autor.

Así, la firma espectral de absorción en la sangre debido a la absorción por oxi/deoxihemoglobina es:

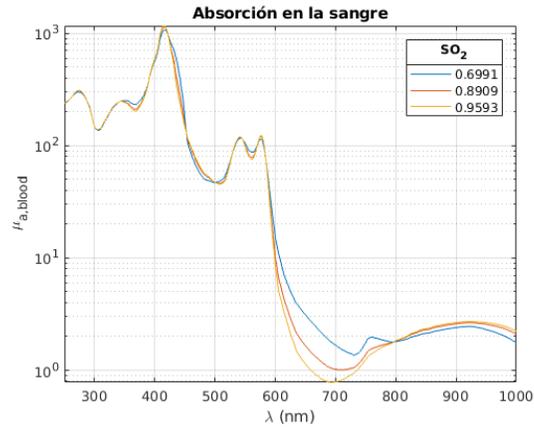


Figura 3. Espectro de absorción en la sangre. Fuente: autor.

Luego, para los componentes de absorción en epidermis y dermis (el cual depende de la variación de absorción en la sangre), los siguientes espectros fueron obtenidos:

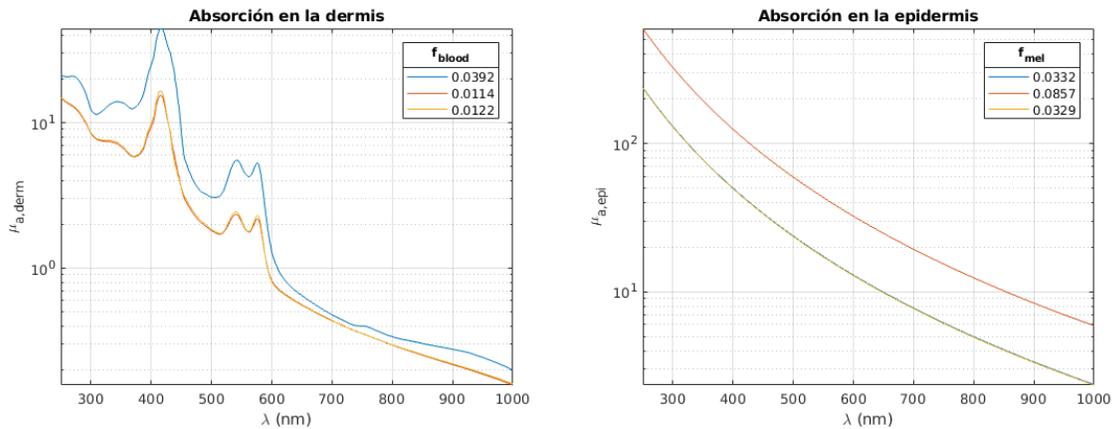


Figura 4. Espectro de absorción en la dermis (izq.) y en la epidermis (der.). Fuente: autor.

A partir de estos espectros se obtuvo la absorción en la piel, y junto con el de dispersión se obtuvo el espectro de reflectancia difusa del modelo teórico:

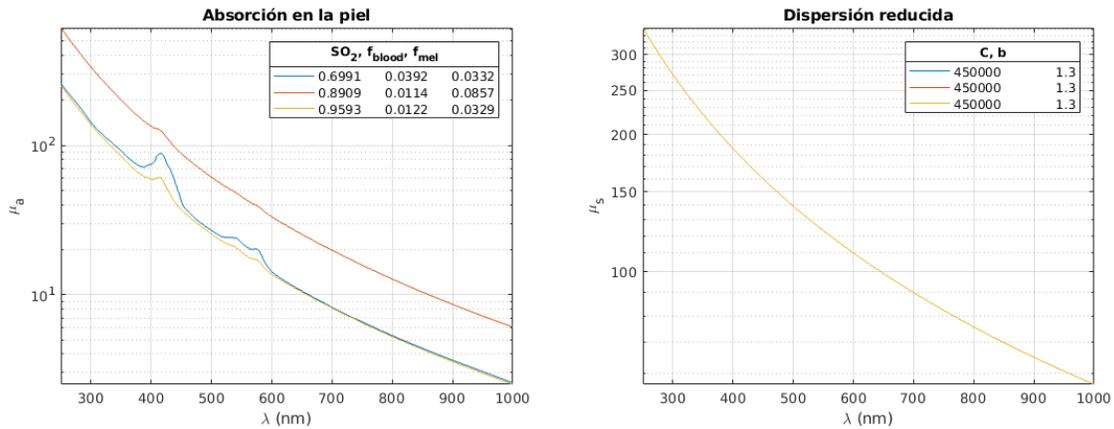


Figura 5. Espectro de absorción (izq.) y de dispersión (der.) en la piel. Fuente: autor.

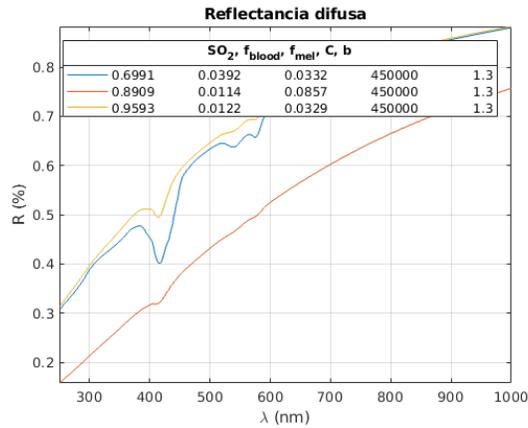


Figura 6. Espectro de reflectancia difusa. Fuente: autor.

Al simular el modelo ante la variación de las propiedades biológicas se observó que los picos presentes en la región comprendida entre  $500$  y  $600\text{nm}$ , así como el de  $420\text{nm}$ , tienen una dependencia a la concentración de melanosomas  $f_{mel}$ , presentándose una atenuación de estos picos en estas longitudes de onda; además de que se presenta una mayor absorción, y por lo tanto menor reflectancia en todo el rango del espectro visible y del cercano infrarrojo para mayores concentraciones de este cromóforo (ver espectro rojo y amarillo en las Figura 5 y Figura 6). También se observa mayor sensibilidad en el espectro de absorción (y por tanto en el de reflectancia) para una mayor fracción de volumen de sangre  $f_{blood}$  con una saturación de oxígeno  $SO_2$  que tienda a un valor medio (alrededor del 50%; ver

espectro azul y amarillo en las Figura 5 y Figura 6), debido a que la absorción en la dermis depende de estos cromóforos, mientras que la absorción en la epidermis depende de la concentración de melanosomas (ver espectro rojo y azul en la Figura 4).

De la Figura 6, los espectros de reflectancia simulados constituyen el conjunto de mezclas utilizado para el análisis mediante PCA e ICA. Luego de implementarse PCA e ICA *Infomax*, las señales fuente fueron estimadas al minimizar el gradiente de entropía. Los resultados obtenidos son mostrados a continuación:

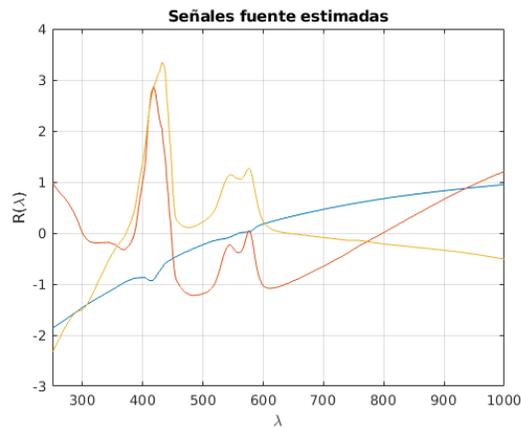


Figura 7. Señales fuente estimadas por ICA. Fuente: autor.

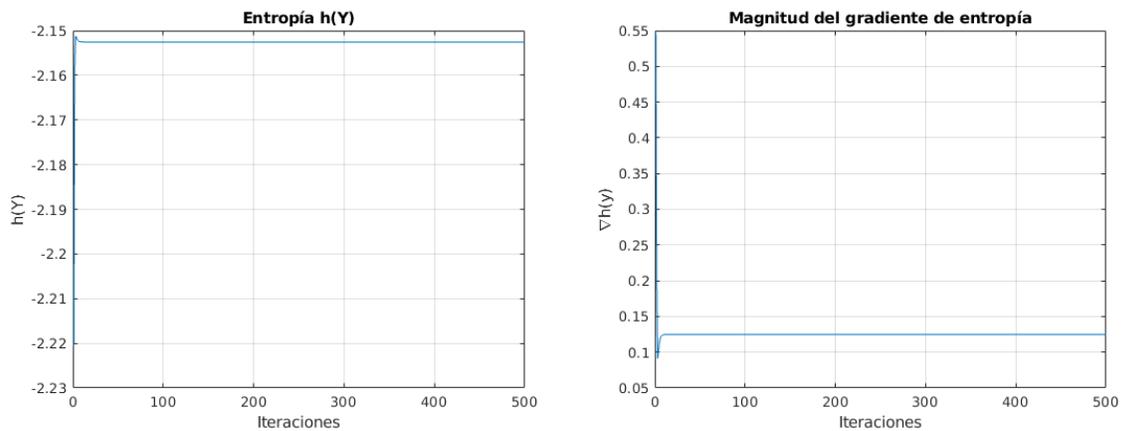


Figura 8. Valor alcanzado para la entropía (izq.) al minimizar su gradiente (der.). Fuente: autor.

	<b>INFORME FINAL DE TRABAJO DE GRADO</b>	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

Luego de haber eliminado la correlación entre las mezclas mediante SVD, al inicializar  $W$  con la matriz identidad y con un paso del gradiente de  $\eta = 0.25$ ; a partir de la Figura 8 se identificó que la separación mediante ICA *Infomax* puede converger en menos de 30 iteraciones utilizando estos parámetros de inicialización.

Sin embargo, las señales fuente obtenidas no coinciden precisamente con los coeficientes de absorción en la dermis, en la epidermis y en la sangre (por oxi/deoxihemoglobina). Esto podría deberse a que en la región del espectro ultravioleta ( $\lambda < 380nm$ ) y del infrarrojo cercano ( $\lambda > 800nm$ ) el rango de los valores de absorción no son comparables debido a la escala, por lo cual al estimar las señales fuente mediante ICA se obtiene una distorsión de la forma de onda cuando se grafica en escala lineal.

Lo anterior resulta verosímil con las limitaciones que presenta ICA y las cuales son manifestadas por (Stone, 2004), indicando que si bien ICA trata de recuperar la forma de onda de las señales fuente, este método no garantiza la energía y el orden en el cual son extraídas.

Por otro lado, esto también podría deberse a la poca densidad de datos utilizados, pues debería obtenerse una mejor estimación de las fuentes para una mayor cantidad de mezclas.

Finalmente, aunque (Yudovsky & Pilon, 2010) indica que las longitudes de onda pueden ser seleccionadas estratégicamente para que coincidan con los picos del espectro de absorción de la piel, este autor se concentra en la región comprendida entre 500 y 600nm del espectro visible para la inversión del modelo mediante simulación Monte-Carlo. De los resultados obtenidos en este trabajo, en la Figura 7 se observa que la forma de onda es coherente en esta región del espectro visible.

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

## 5. CONCLUSIONES, RECOMENDACIONES Y TRABAJO FUTURO

---

### CONCLUSIONES

- Las firmas espectrales fuentes correspondientes a la reflectancia difusa de piel humana son estimadas utilizando ICA como método de separación ciega de fuentes y PCA como etapa de preprocesamiento de los datos. Para validar la separación con ICA *Infomax*, las señales fuente también pueden ser estimadas mediante la búsqueda de proyecciones, con el fin de obtener dos conjuntos de señales comparables al ser graficadas. De esta forma se comprueba que las señales fuente estimadas coinciden para las dos técnicas utilizadas en el método BSS.

Si bien ICA tiene aplicación en muchos campos, esta técnica sigue siendo investigada con el fin de evaluar el desempeño en diferentes aplicaciones y campos del conocimiento.

Entre las fortalezas de ICA se destaca que no requiere conocimiento previo de las señales fuente. Sin embargo, ICA no puede recuperar la energía o el orden de las señales. Esto es una limitación para comparar las señales fuente recuperadas con las componentes de absorción simuladas en el modelo teórico luz–piel.
- Se destaca una dependencia de la reflectancia difusa en función de  $f_{mel}$  y  $SO_2$  principalmente, según lo expuesto en el análisis presentado para los resultados obtenidos durante las simulaciones del modelo de interacción luz–piel. Es decir, cuando  $SO_2$  tiene un valor intermedio (entre 40% y 60% aproximadamente), para una mayor concentración de  $f_{mel}$ , los picos de absorción presentes en la firma

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

espectral de reflectancia se atenúan cada vez más conforme  $f_{mel}$  incrementa. Esto es más evidente en la región del espectro cuando  $500 \text{ nm} < \lambda < 600 \text{ nm}$ .

Igualmente se evidencia que la forma de la forma del espectro de reflectancia cambia conforme a las variaciones de  $SO_2$  entre 40% y 60% y  $f_{blood}$ .

- Realizar el preprocesamiento de los datos mediante el análisis de componentes principales implementando SVD sobre la matriz de covarianza no genera los resultados esperados en el proceso de separación de fuentes. Por otra parte, si bien (Hyvärinen & Oja, 2000) describe el proceso de *whitening* como una etapa de preprocesamiento recomendada para simplificar la estimación de las señales fuente mediante ICA, en el caso particular presentado en este trabajo, los resultados esperados no fueron obtenidos; es decir, no se visualiza algún parecido entre las firmas espectrales estimadas mediante ICA y las simuladas mediante el modelo teórico. El hecho de utilizar SVD sobre las mezclas originales hace que se genere un nuevo conjunto de mezclas sin correlación entre sí, sobre el cual se puede implementar ICA. Probablemente, el procedimiento de *whitening* no haya sido adecuado debido a la poca densidad de datos de prueba utilizados.
- La rutina computacional utilizada para estimar las firmas espectrales fuente a partir de aquellas que fueron pre-procesadas, consiste en una función objetivo a maximizar mediante alguna regla de optimización. En el caso de la búsqueda por proyecciones, la función objetivo maximiza la no-gaussianidad mediante la evaluación de la curtosis para cada una de las estimaciones; mientras que en el caso de ICA *Infomax* la función objetivo se encarga de maximizar la información mutua mediante la evaluación de la entropía. Estas funciones objetivo son maximizadas mediante la regla del gradiente ascendente en ambos casos, con el fin de encontrar los vectores de separación  $w_i$  que las satisfacen.

	<b>INFORME FINAL DE TRABAJO DE GRADO</b>	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

Sin embargo, no es impedimento cambiar la función pdf modelo de  $\mathbf{y}$ , (ecuación 2.32) siendo  $\tanh(\mathbf{W}^T \mathbf{x})$  la función utilizada en este trabajo para la separación mediante *Infomax*.

## RECOMENDACIONES

Debido a que el algoritmo ICA *Infomax* inicializa la matriz de separación  $\mathbf{W}$  aleatoriamente, este puede converger en un mínimo local. Por lo que se recomienda ejecutar el algoritmo varias veces hasta obtener las señales fuente requeridas. Una vez las tres señales fuente hayan sido estimadas satisfactoriamente, el valor con el que haya inicializado  $\mathbf{W}$  puede seguirse utilizando como nuevo valor de inicialización para garantizar que el método pueda converger en la misma solución.

Se recomienda utilizar otros conjuntos de datos con más mezclas con el fin de estimar mejor las señales fuente, ya que el conjunto de datos utilizado consiste de tres mezclas simuladas mediante el modelo teórico luz–piel utilizando parámetros ópticos generados aleatoriamente.

## TRABAJO FUTURO

Para un primer contacto se recomienda probar con el algoritmo FastICA presentado por (Hyvärinen & Oja, 2000) con el propósito de realizar la descomposición después de aplicar *whitening* a los datos utilizando PCA, además de utilizar diferentes funciones pdf modelo que permitan maximizar la separación entre las señales.

Debido a las diversas variantes de ICA encontradas en la literatura, queda como trabajo futuro implementar aquellas técnicas basadas en las medidas que permiten maximizar o

	<b>INFORME FINAL DE TRABAJO DE GRADO</b>	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

minimizar los efectos de las mezclas descritos en la sección 2.3.2. Algunos ejemplos de dichas variantes pueden ser (Hyvärinen, Karhunen, & Oja, 2001):

- Maximizar la no-gaussianidad mediante negentropía.
- Estimación de máxima verosimilitud.
- Minimización de la información mutua.
- Mediante decorrelación no lineal y PCA no lineal.
- Mediante métodos tensoriales.

Entre otros.

También se sugiere probar el método con otras reglas de optimización (con o sin restricción) o combinar el método mediante la implementación de algoritmos híbridos. Por ejemplo, se podría intentar una maximización de la función objetivo que se encargue de estimar las fuentes, mediante la búsqueda por enjambre de partículas. Esto puede llegar a ser útil para estimar diversas fuentes con conjuntos mucho más densos de mezclas, con el fin de reducir el tiempo de procesamiento.

Finalmente, queda como trabajo futuro explorar la capacidad de ICA para la selección de bandas y extracción de características en imágenes multiespectrales de Leishmaniasis cutánea, siendo ICA una técnica potencial para el procesamiento de este tipo de imágenes diagnósticas.

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

## REFERENCIAS

---

- Cheong, W.-F., Prael, S. A., & Welch, A. J. (1990). A Review of the Optical Properties of Biological Tissues. *Ieee Journal of Quantum Electronics*, 26(12), 2166–2185. <https://doi.org/10.1109/3.64354>
- Hyvärinen, A., Karhunen, J., & Oja, E. (2001). *Independent component analysis* (First edit). New York: John Wiley & Sons, Inc. <https://doi.org/10.1002/0471221317>
- Hyvärinen, A., & Oja, E. (2000). Independent Component Analysis: Algorithms and Applications. *Neural Networks*, 13(45), 411–430. [https://doi.org/10.1016/S0893-6080\(00\)00026-5](https://doi.org/10.1016/S0893-6080(00)00026-5)
- Jacques, S., & Prael, S. (2018). Optical Absorption of Hemoglobin. Retrieved September 16, 2018, from <https://omlc.org/spectra/hemoglobin/>
- Kong, X., Hu, C., & Duan, Z. (2017). *Principal Component Analysis Networks and Algorithms*. <https://doi.org/10.1007/978-981-10-2915-8>
- Ng, A. (n.d.). Principal Component Analysis Algorithm - Universidad de Stanford | Coursera. Retrieved September 10, 2018, from <https://www.coursera.org/learn/machine-learning/lecture/ZYIPa/principal-component-analysis-algorithm>
- Ramanath, R., & Snyder, W. E. (2003). Band selection using independent component analysis for hyperspectral image processing. *32nd Applied Imagery Pattern Recognition Workshop, 2003. Proceedings.*, 93–98. <https://doi.org/10.1109/AIPR.2003.1284255>

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

Stone, J. V. (2004). *Independent component analysis* (First). Massachusetts: The MIT Press.

Yudovsky, D., & Pilon, L. (2010). Rapid and accurate estimation of blood saturation, melanin content, and epidermis thickness from spectral diffuse reflectance. *Applied Optics*, 49(10), 1707. <https://doi.org/10.1364/AO.49.001707>

Zonios, G., & Dimou, A. (2006). Modeling diffuse reflectance from semi-infinite turbid media: application to the study of skin optical properties. *Optics Express*, 14(19), 8661. <https://doi.org/10.1364/OE.14.008661>

	<b>INFORME FINAL DE TRABAJO DE GRADO</b>	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

# APÉNDICE

---

## APÉNDICE A: PROGRAMA PRINCIPAL PARA ESTIMAR SEÑALES FUENTE MEDIANTE ICA.

```

%% Simula los espectros teóricos con parámetros ópticos aleatorios. Luego,
%% los espectros son separados mediante ICA y la matriz de separación es
%% optimizada con el fin de reducir la distancia entre los datos de las
%% mezclas estimadas (con  $x = As$ , donde  $A = W^{-1}$ ) y  $s$  fuentes estimadas) y
%% las reflectancias simuladas.
%%

close all; clearvars; clc;
cd('/home/camilo/MATLAB/Proyecto Leishmaniasis');

%% Parte 1: Simulación con parámetros aleatorios.
%% Genera la simulación de los espectros teóricos con el modelo de una capa
%% sami-infinita.
%%

% Longitudes de onda de los espectros teóricos.
lambda = 250:2:1000;

% Genera 3 espectros teóricos en forma aleatoria. Cada fila es un espectro
% y cada columna es un valor observado.
[R_esp, mu_s, mu_a, mu, opt_prop] = ...
    reflectance_spectra_simulation_2018_04_19('random generation', lambda, 3, [],
3, 1);

% La matriz de mezcla inicial proporcionada a ICA corresponde a las
% reflectancias difusas simuladas mediante el modelo de una capa, como
% medida teórica.
X = R_esp;

%% Parte 2: Separación mediante ICA.
%% Estima 3 señales fuente y la matriz de separación mediante ICA
%% Infomax (maximizando entropía).
%%

% Cantidad de señales fuente a estimar.
k = 3;

% Paso del gradiente.
eta = 0.25;

% Cantidad de iteraciones máximas permitidas.
max_iter = 500;

graph = 1;

% Preprocesamiento de datos. Los espectros de reflectancia adquieren media
% cero y varianza unitaria (los datos son escalados).
X = X';

```

	<b>INFORME FINAL DE TRABAJO DE GRADO</b>	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

```

X = X - mean(X);
X = X./std(X);

% Aplica la descomposición de valor singular.
[U, D, ~] = svd(X, 0);

% Obtiene las nuevas mezclas sin correlación entre ellas.
z = U(:,1:k);

% Las mezclas (sin correlación) adquieren varianza unitaria.
z = z - mean(z); z = z./std(z);
z = z';

% Obtiene las dimensiones de las mezclas.
[m, n] = size(z);

% Inicializa la matriz de separación W con la matriz identidad.
W = eye(m);
% W = rand(m);

% Vectores para almacenar el cambio de entropía y de gradiente.
hs = zeros(max_iter,1);
gs = zeros(max_iter,1);

% Extrae las señales fuente maximizando la entropía.
for i = 1:max_iter

    % Estima las señales fuente con máxima entropía.
    y = W*z;

    % Evalúa las señales fuente estimadas en la función pdf modelo.
    Y = tanh(y);

    % Evalúa la entropía.
    h = (1/n)*sum(sum(log(1-Y.^2),2)) + log(abs(det(W)));

    % Calcula la matriz de gradientes @h/@W_ij.
    g = inv(W) - (2/n)*Y*z';

    % Actualiza a W para incrementar h.
    W = W + eta*g;

    % Normaliza la matriz de separación (valores entre 0 y 1);
    W = W./norm(W);

    % Registra el valor de h y la magnitud del gradiente.
    hs(i) = h; gs(i) = norm(g(:));

end

% Grafica los cambios en la magnitud de entropía y su gradiente.
if graph == 1
    figure(); plot(hs); title('Entropía h(Y)');
    xlabel('Iteraciones'); ylabel('h(Y)');
    grid on;

    figure(); plot(gs); title('Magnitud del gradiente de entropía');

```

 <b>Institución Universitaria</b>	<b>INFORME FINAL DE TRABAJO DE GRADO</b>	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

```

xlabel('Iteraciones'); ylabel('\nabla h(y)');
grid on;
end

% Grafica las fuentes estimadas.
figure(); plot(lambda,y);
title('Señales fuente estimadas');
xlabel('\lambda'); ylabel('R(\lambda)');
xlim([250 1000]); grid on;

% Grafica las mezclas.
figure(); plot(lambda, X);
title('Mezclas teóricas (simuladas)');
xlabel('\lambda'); ylabel('R(\lambda)');
xlim([250 1000]); grid on;

% Normalizando las señales estimadas.
y = y';

idx = find(abs(min(y)) > max(y));
y(:,idx) = y(:,idx).*(-1);

y = y - min(y);
y = y./(max(y) - min(y));

y = y';

% Grafica las fuentes estimadas normalizadas.
figure(); plot(lambda,y);
title('Señales fuente normalizadas');
xlabel('\lambda'); ylabel('R(\lambda)');
xlim([250 1000]); grid on;

```

	<b>INFORME FINAL DE TRABAJO DE GRADO</b>	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

## APÉNDICE B: FUNCIÓN PARA SIMULAR LAS REFLECTANCIAS DIFUSAS MEDIANTE EL MODELO TEÓRICO LUZ-PIEL.

```

function [R, mu_s, mu_a, mu, opt_prop] = ...
    reflectance_spectra_simulation_2018_04_19(option, lambda, K, opt_prop, M,
graph)
% Método:                simulación de espectros teóricos de reflectancia
%                        difusa de piel humana.
% Nombre del archivo:    random_spectra_generation_2018_04_18.m
% Autor:                 Camilo Alejandro Bermúdez Mejía.
% Fecha:                 2018-Abr-18.
%
% Descripción:
% Esta función se encarga de simular espectros teóricos de reflectancia
% difusa de la piel humana o de generarlos aleatoriamente.
%
% Parámetros de entrada:
% option:                determina la acción a ejecutar.
%                        'random generation': genera espectros de reflectancia a
%                        partir de propiedades ópticas aleatorias.
%                        'defined simulation': simula espectros de reflectancia con
%                        propiedades ópticas que se pasan como argumento.
% lambda:                vector de longitudes onda sobre el cual simular el modelo
%                        de reflectancia difusa de una capa semi-infinita.
% K:                     constante de reescalamiento de la dispersión reducida.
% opt_prop:              propiedades ópticas definidas en el llamado de la función.
%                        matriz de 'Mx5' que permite obtener M espectros simulados,
%                        debido a las M variaciones de cada parámetro óptico (se
%                        debe seleccionar: option = 'defined simulation').
% M:                     cantidad de espectros aleatorios a generar.
% graph:                 bandera para graficar (con valor 1 grafica).
%
% Parámetros de salida:
% R:                     espectros de reflectancia difusa.
% mu_s:                  espectros teóricos de dispersión reducida.
% mu_a:                  espectros teóricos de absorción.
% mu:                   hipermatriz MxNx5 con los espectros de absorción por
%                       oxi/deoxihemoglobina, absorción en la sangre, dermis y
%                       epidermis.
% opt_prop:              propiedades ópticas generadas aleatoriamente por la
%                       función (se debe seleccionar: 'random generation').

% Carga los espectros teóricos tabulados del coeficiente de extinción molar
% de oxi/deoxihemoglobina.
load model_data e_oxy e_deoxy;

% Necesario para evaluar los coeficientes de extinción molar de
% oxi/deoxihemoglobina en las longitudes de onda pasadas a la función.
aux = 250:2:1000;

% Encuentra los valores de extinción molar que coinciden con las longitudes
% de onda.
if exist('lambda','var')

pos = zeros(1,length(lambda));
    for i = 1:length(lambda)

```

	<b>INFORME FINAL DE TRABAJO DE GRADO</b>	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

```

% Determina las posiciones necesarias para evaluar los espectros de
% extinción molar oxi/deoxihemoglobina.
pos(i) = find(aux == lambda(i));

end

% Evalúa los espectros de extinción molar en las longitudes de onda
% pasadas a la función.
e_oxy = e_oxy(pos); e_deoxy = e_deoxy(pos);

else

% Si el vector de longitudes de onda no fue pasado a la función,
% entonces estará definido por defecto entre 250 y 1000 nm.
lambda = aux;

end

% Limpia las variables auxiliares.
clear aux pos;

% Valor promedio de concentración de hemoglobina.
C_heme = 150;

% Necesario para mantener consistencia entre unidades del coef. de
% dispersión reducida.
lambda_0 = 1;

% Calcula el espectro de absorción por melanosomas. ***
mu_amel = 6.60e11*(lambda.^(-3.33));

% Calcula el espectro de absorción en el fondo.
mu_aback = 7.84e8*(lambda.^(-3.255));

% Determina los parámetros ópticos según el tipo de simulación.
switch option

% Genera propiedades ópticas aleatoriamente.
case 'random generation'

% Genera la saturación de oxígeno.
SO2 = rand(1,M)*(1-0);

% Genera el volumen de sangre en la epidermis.
f_blood = rand(1,M)*(0.07-0.002) + 0.002;

% Genera la fracción de melanosomas.
f_mel = rand(1,M)*(0.1-0.01) + 0.01;

% Genera la constante de dispersión.
C = ones(1,M)*mean([5e5 4e5]);

% Genera la potencia de dispersión.
b = ones(1,M)*mean([1.5 1.1]);

```

	<b>INFORME FINAL DE TRABAJO DE GRADO</b>	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

```

% Retorna los parámetros ópticos generados en una matriz Mx5.
opt_prop = [SO2; f_blood; f_mel; C; b]';

% Los parámetros ópticos proporcionados en la matriz de Mx5 son
% extraídos para simular espectros específicos.
case 'definied simulation'

% Determina la cantidad de espectros teóricos a simular.
[M,~] = size(opt_prop);

% Selecciona la saturación de oxígeno.
SO2 = opt_prop(:,1)';

% Selecciona el volumen de sangre en la epidermis.
f_blood = opt_prop(:,2)';

% Selecciona la fracción de melanosomas.
f_mel = opt_prop(:,3)';

% Selecciona la constante de dispersión.
C = opt_prop(:,4)';

% Selecciona la potencia de dispersión.
b = opt_prop(:,5)';
end

% Prelocaliza variables en memoria.
mu_aoxy = zeros(M,length(lambda)); mu_adeoxy = zeros(M,length(lambda));
mu_ablood = zeros(M,length(lambda)); mu_aderm = zeros(M,length(lambda));
mu_aepi = zeros(M,length(lambda)); mu_a = zeros(M,length(lambda));
mu_s = zeros(M,length(lambda)); R = zeros(M,length(lambda));

% Calcula M espectros teóricos con las propiedades ópticas generadas
% aleatoriamente o con las proporcionadas.
for i = 1:M

% Calcula el espectro de absorción por oxihemoglobina. ***
mu_aoxy(i,:) = e_oxy*(C_heme*SO2(i)/66500);

% Calcula el espectro de absorción por deoxihemoglobina. ***
mu_adeoxy(i,:) = e_deoxy*(C_heme*(1-SO2(i))/66500);

% Calcula el espectro de absorción en la sangre.
mu_ablood(i,:) = mu_aoxy(i,:) + mu_adeoxy(i,:);

% Calcula el espectro de absorción de la dermis.
mu_aderm(i,:) = f_blood(i)*mu_ablood(i,:) + mu_aback*(1-f_blood(i));

% Calcula el espectro de absorción en la epidermis.
mu_aepi(i,:) = mu_amel*f_mel(i) + mu_aback*(1-f_mel(i));

% Calcula el espectro de absorción de la piel.
mu_a(i,:) = mu_aderm(i,:) + mu_aepi(i,:);

% Calcula el espectro de dispersión reducida.
mu_s(i,:) = C(i)*((lambda/lambda_0).^(-b(i)));

```

	<b>INFORME FINAL DE TRABAJO DE GRADO</b>	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

```

% Calcula el espectro de reflectancia difusa.
R(i,:) = 1./(1+K*(mu_a(i,:)./mu_s(i,:)));
end

% Retorna los espectros de absorción teóricos como una hipermatriz.
mu = cat(3, mu_aoxy, mu_adeoxy, mu_ablood, mu_aderm, mu_aepi);

% Grafica los espectros teóricos de absorción y de reflectancia.
if graph == 1

% Grafica los espectros de absorción por oxihemoglobina.
figure(); semilogy(lambda,mu_aoxy);
title('Absorción por oxihemoglobina');
xlabel('\lambda (nm)'); ylabel('\mu_{a,oxy}');
lgd = legend(num2str(round(opt_prop(:,1),4)));
title(lgd,'SO_2'); axis tight; grid on;

% Grafica los espectros de absorción por deoxihemoglobina.
figure(); semilogy(lambda,mu_adeoxy);
title('Absorción por deoxihemoglobina');
xlabel('\lambda (nm)'); ylabel('\mu_{a,deoxy}');
lgd = legend(num2str(round(opt_prop(:,1),4)));
title(lgd,'SO_2'); axis tight; grid on;

% Grafica los espectros de absorción en la sangre.
figure(); semilogy(lambda,mu_ablood);
title('Absorción en la sangre');
xlabel('\lambda (nm)'); ylabel('\mu_{a,blood}');
lgd = legend(num2str(round(opt_prop(:,1),4)));
title(lgd,'SO_2'); axis tight; grid on;

% Grafica los espectros de absorción en la dermis.
figure(); semilogy(lambda,mu_aderm);
title('Absorción en la dermis');
xlabel('\lambda (nm)'); ylabel('\mu_{a,derm}');
lgd = legend(num2str(round(opt_prop(:,2),4)));
title(lgd,'f_{blood}'); axis tight; grid on;

% Grafica los espectros de absorción en la epidermis.
figure(); semilogy(lambda,mu_aepi);
title('Absorción en la epidermis');
xlabel('\lambda (nm)'); ylabel('\mu_{a,epi}');
lgd = legend(num2str(round(opt_prop(:,3),4)));
title(lgd,'f_{mel}'); axis tight; grid on;

% Grafica los espectros de absorción en la piel.
figure(); semilogy(lambda,mu_a);
title('Absorción en la piel');
xlabel('\lambda (nm)'); ylabel('\mu_a');
lgd = legend(num2str(round(opt_prop(:,1:3),4)));
title(lgd,'SO_2, f_{blood}, f_{mel}');
axis tight; grid on;

% Grafica los espectros de absorción por dispersión reducida.
figure(); semilogy(lambda,mu_s);
title('Dispersión reducida');
xlabel('\lambda (nm)'); ylabel('\mu_s');
lgd = legend(num2str(round(opt_prop(:,4:5),4)));
title(lgd,'C, b'); axis tight; grid on;

```

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

```
% Grafica los espectros de reflectancia difusa simuladas.  
figure(); plot(lambda,R);  
title('Reflectancia difusa');  
xlabel('\lambda (nm)'); ylabel('R (%)');  
lgd = legend(num2str(round(opt_prop,4)));  
title(lgd,'SO_2, f_{blood}, f_{mel}, C, b');  
axis tight; grid on;  
  
end  
  
end
```

 Institución Universitaria	<b>INFORME FINAL DE TRABAJO DE GRADO</b>	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

FIRMA ESTUDIANTES         *Cumilo*        

FIRMA ASESOR         *Puly*                 *Cumilo*        

*Entrega informe final trabajo de grado Cumilo Bermúdez.*

FECHA ENTREGA: 27/sep./2018

FIRMA COMITÉ TRABAJO DE GRADO DE LA FACULTAD \_\_\_\_\_

RECHAZADO\_\_\_      ACEPTADO\_\_\_      ACEPTADO CON MODIFICACIONES\_\_\_

ACTA NO. \_\_\_\_\_

FECHA ENTREGA: \_\_\_\_\_

FIRMA CONSEJO DE FACULTAD \_\_\_\_\_

ACTA NO. \_\_\_\_\_

FECHA ENTREGA: \_\_\_\_\_