

Assessment and optimization of new
applied spectroscopic techniques using
functional data analysis: application to non-
invasive sensing of sustained hyperglycemia

by
Aldo Luis Moreno Oyervides.

A dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor of
Philosophy in

Electrical Engineering, Electronics and
Automation.

Universidad Carlos III de Madrid

Supervisors:
Dr. Pablo Acedo Gallardo
and
Dra. María del Carmen Aguilera Morillo

Tutor:
Dr. Pablo Acedo Gallardo

September 2020

This thesis is distributed under license “Creative Commons **Attribution**
– **Non Commercial** – **Non Derivatives**”.



Con mucho amor a mi esposa, “Mi Alma”, quien ha estado a mi lado sin dudar en esta gran aventura: a nuestro hijo amado que ha sido una hermosa bendición en este camino; ambos son un gran impulso para no flaquear, y también con mucho cariño y respecto a mi madre que siempre me ha apoyado y nunca ha dejado de creer en mí.

AGRADECIMIENTOS

Quiero agradecer en primer lugar a los dos grandes pilares de mi investigación doctoral: al Dr. Pablo Acedo por su gran dirección, todo su apoyo y por ayudarme siempre a mantenerme enfocado en el desarrollo de este trabajo; y a la Dra. María del Carmen Aguilera por siempre apoyarme, enseñarme tanto y ser tan paciente conmigo; lo cierto es que este espacio no es suficiente para transmitir toda la gratitud y afecto que siento. Muchas gracias a los dos por todo lo que han hecho por mí y por mi familia.

También quiero agradecer a la Dra. María Durbán por toda la gran disposición que siempre manifestó para ayudarme; de igual forma agradezco mucho a todas aquellas grandes personas y profesionales que me dieron la oportunidad de trabajar con ellos, como son: Dr. Pedro Martín, Dr. Viktor Krozer, Dr. Fernando Larcher, Dra. María José de la Cruz, Dra. Edurne Lecumberri, entre otros, quienes han enriquecido este trabajo con muchas contribuciones invaluable.

No cabe duda de que lo más importante de mi formación doctoral, es toda la experiencia y la amplia gama de conocimientos adquiridos que me llevo de todos aquellos con quienes tuve la maravillosa oportunidad de trabajar durante esta investigación.

Por supuesto, no puedo dejar de mencionar a los compañeros y compañeras del grupo de investigación (SIT por sus siglas en inglés) y el Departamento de Tecnología Electrónica, a quienes he tenido la fortuna de conocer durante mi estancia como son: Dr. Oscar Bonilla, Dr. Julio Posada, Dr. Antonio Souto, Mtro. Andrés Betancourt, Dr. Borja Jerez, Mtra. Dahiana Mojena, Mtro. Dragos Poiana, Mtro. Antonio Varillas, Dra. Martha Ruiz y muchos otros que han estado presente y ayudado en diferentes formas; estoy muy agradecido porque me llevo grandes amistades y muy buenos recuerdos de esta etapa de mi vida.

Quiero también agradecer a mi familia, que siempre están ahí a pesar de la distancia, me animan y me apoyan para seguir adelante; ellos son mi pasado, mi presente y mi futuro, y agradezco a la vida por ponerlos a mi lado, especialmente a mis padres a quienes les debo todo lo que soy, por su confianza y apoyo para ver realizado en mí su máximo anhelo.

Finalmente, pero no menos importante, quiero agradecer profundamente al Consejo Nacional de Ciencia y Tecnología (CONACYT-México) por el apoyo económico brindado durante mi formación doctoral, muchas gracias por darme la oportunidad de superarme académica y profesionalmente.

PUBLISHED AND SUBMITTED CONTENT

JOURNAL PAPERS

- PAPER [A]: **Aldo Moreno-Oyervides**, M. Carmen Aguilera-Morillo, Fernando Larcher, Viktor Krozer & Pablo Acedo. “Advanced Statistical Techniques for Non-Invasive Hyperglycemic States Detection in mice using Millimeter-wave Spectroscopy”. *IEEE Transactions on Terahertz Science and Technology*, Vol. 10, no. 3, p. 237-245, 2020. DOI: 10.1109/TTHZ.2020.2967236. Document content wholly included in chapters 3 and 4. The material from this source included in this thesis is not singled out with typographic means and references.
- PAPER [B]: **Aldo Moreno-Oyervides**, Pedro Martín-Mateos, M. Carmen Aguilera-Morillo, Giacomo Ulisse, María C. Arriba, María Durban, Marcela Del Rio, Fernando Larcher, Viktor Krozer & Pablo Acedo. “Early, Non-Invasive Sensing of Sustained Hyperglycemia in Mice Using Millimeter-Wave Spectroscopy”. *Sensors*, Vol. 19, no. 15, p. 3347, 2019. DOI: 10.3390/s19153347. Document content wholly included in chapters 3 and 4. The material from this source included in this thesis is not singled out with typographic means and references.
- PAPER [C]: Pedro Martín-Mateos, Fabian Dornuf, Blanca Duarte, Bernhard Hils, **Aldo Moreno-Oyervides**, Oscar Elias Bonilla-Manrique, Fernando Larcher, Viktor Krozer & Pablo Acedo. “In-vivo, non-invasive detection of hyperglycemic states sin animal models using mm-wave spectroscopy”. *Scientific Reports*, Vol. 6, p. 34035, 2016. DOI: 10.1038/srep34035. Document content partly included in chapter 4. The material from this source included in this thesis is not singled out with typographic means and references.

CONFERENCE CONTRIBUTIONS

1. **Aldo Moreno-Oyervides**, M. Carmen Aguilera-Morillo, María José de la Cruz Fernández, Edurne Lecumberri, Viktor Krozer & Pablo Acedo, “Development and evaluation of new spectroscopic techniques for biomedical applications by Functional Data Analysis,” in V Congreso de Jóvenes Investigadores de la RSME (Real Sociedad Matemática Española, 2020). Document content partly included in chapters 4 and 5. The material from this source included in this thesis is not singled out with typographic means and references.
2. **Aldo Moreno-Oyervides**, Pedro Martín-Mateos, M. Carmen Aguilera-Morillo, Giacomo Ulisse, Fernando Larcher, Viktor Krozer & Pablo Acedo, “Non-invasive detection and monitoring of sustained glycemic fluctuations using mm-wave spectroscopy,” in European Microwave Week (EuMA, 2018). Document content partly included in chapter 4. The material from this source included in this thesis is not singled out with typographic means and references.
3. **Aldo Moreno-Oyervides**, M. Carmen Aguilera-Morillo, María Durban, Pedro Martín-Mateos, Fernando Larcher & Pablo Acedo, “Modelización de datos espectrales complejos mediante Análisis de Datos Funcionales,” in XXXVII Congreso Nacional de Estadística e Investigación Operativa (SEIO, 2018). Document content partly included in chapter 4. The material from this source included in this thesis is not singled out with typographic means and references.
4. Pedro Martín-Mateos, **Aldo Moreno-Oyervides**, Giacomo Ulisse, Blanca Duarte, Fernando Larcher, Viktor Krozer & Pablo Acedo, “Monitoring the evolution of hyperglycemia in mice using mm-wave spectroscopy,” in 42nd International Conference of Infrared, Millimeter, and Terahertz Waves (IRMMW-THz, 2017). Document content partly included in chapter 4. The material from this source included in this thesis is not singled out with typographic means and references.

OTHER RESEARCH MERITS

JOURNAL PAPERS

- Dimitrios C. Zografopoulos, Antonio Ferrato, José Francisco Algorri, Pedro Martín-Mateos, Braulio García-Cámara, **Aldo Moreno-Oyervides**, Viktor Krozer, Pablo Acedo, Ricardo Vergaz, José Manuel Sánchez-Pena & Romeo Beccherelli. “All-Dielectric Silicon Metasurface with Strong Subterahertz Toroidal Dipole Resonance”. *Advanced Optical Materials*, Vol. 7, no. 19, p. 1900777, 2019.

CONFERENCE CONTRIBUTIONS

- José Francisco Algorri, Dimitrios C. Zografopoulos, Antonio Ferrato, Pedro Martín-Mateos, Braulio García-Cámara, **Aldo Moreno-Oyervides**, Viktor Krozer, Pablo Acedo, Romeo Beccherelli, José Manuel Sánchez-Pena, Ricardo Vergaz, “All-Dielectric Metasurfaces with Toroidal Multipole Resonances at sub-THz,” in 13th International Congress on Artificial Materials for Novel Wave Phenomena (Metamaterials, 2019).
- **Aldo Moreno-Oyervides**, Jesús Palací López, Pedro Martín-Mateos, Frederik Walla, Rubén Criado, Viktor Krozer & Pablo Acedo, “Use of functional principal components analysis in CW subTHz spectroscopy for hydrocarbon emulsified water assessment,” in 42nd International Conference of Infrared, Millimeter, and Terahertz Waves (IRMMW-THz, 2017).

LIST OF ACRONYMS

AGEs	Advanced Glycation End-products
AUC	Area Under Curve
B-spline	Basis spline
BGL	Blood Glucose Level
BLUPs	Best Linear Unbiased Predictors
BSS	Blind Signal Separation
CCR	Correct Classification Rate
CW	Continuous Wave
DM	Diabetes Mellitus
DPMSE	Discrete Penalized Mean Square Error
FDA	Functional Data Analysis
FLM	Functional Linear Model
GFLM	Generalized Functional Linear Model
GUI	Graphic User Interface
FLoR	Functional Logistic Regression
FN	False Negative
FP	False Positive
FPC	Functional Principal Component
FPCA	Functional Principal Component Analysis
FPG	Fasting Plasma Glucose
FPCLoR	Functional Principal Component Logistic Regression
GCV	Generalized Cross Validation
iid	independent and identically distributed

ICA	Independent Component Analysis
K-L expansion	Karhunen-Loève expansion
LFPCA	Longitudinal Functional Principal Component Analysis
LFPCR	Longitudinal Functional Principal Component Regression
LPFR	Longitudinal Penalized Functional Regression
LOOCV	Leave-One-Out Cross Validation
mm-wave	Millimeter-wave
MSE	Mean Square Error
MDA	Multivariate Data Analysis
OGTT	Oral Glucose Tolerance Test
OLS	Ordinary Least Squares
P-splines	Penalized splines
PCA	Principal Component Analysis
REML	Restricted Maximum Likelihood
RMS	Root Mean Square
RMSE	Root Mean Square Error
ROC	Receiver Operating Characteristic
RTE	Radiation Transfer Equation
STZ	Streptozotocin
TNR	True Negative Rate
TPR	True Positive Rate

ABSTRACT

Millimeter-wave and THz spectroscopy is nowadays a huge research field largely focused on the development of new and improved instruments and techniques for practical applications such as medical diagnostics, security screenings, and industrial control processes; mainly due to its non-invasive and remote scope for matter elucidation and characterization. Such instruments are based on spectroscopic techniques that, broadly speaking, irradiate a target sample using electromagnetic radiation at different frequencies (wavelengths) to measure its optical properties as frequency-dependent functions. Thus, the use of spectroscopic techniques leads to the acquisition of spectral data containing the information about the interaction between the sample under evaluation and the applied electromagnetic radiation. In this situation, the processing and analysis of the spectral data are important tasks when developing new instruments based on spectroscopic techniques (applied spectroscopy systems), and one of the most challenging scenarios appears in applications in which the sample includes multiple species with very similar optical properties, as often happens in biomedical applications. Consequently, statistical methods are required, not only to extract the desired information from the spectral data, but also as part of the calibration process of the technique/instrument.

In this doctoral thesis, a novel non-invasive approach using mm-wave spectroscopy for in-vivo detection and monitoring of sustained hyperglycemia, typically associated with Diabetes Mellitus (DM), is evaluated through several experimental tests including the use of animal models and a pilot clinical study on humans (Type 1 DM patients). The experimental tests were carried out using a W-band spectrometer built specifically for this study, capable of acquiring the reflection and transmission spectra from in-vivo and non-invasive measurements performed on the animals and humans. The spectral data collected from these tests have been processed and analyzed using Functional Data Analysis (FDA) methods due to their suitability for the abovementioned challenging spectroscopic scenarios in the framework of biomedical applications. FDA techniques have allowed us to study the spectral response measured within the W-band in unsupervised and supervised settings, providing an interpretation of the different interrogation

frequencies contribution at all the analysis stages, even for a longitudinal analysis performed on spectral data collected at different times during the pilot clinical study with type 1 DM patients. It is important to note that in the conception of the diagnostic we have used a non-targeted spectral profiling approach that allowed us to consider collectively the spectral features of all the sample constituents (avoiding the necessity of individual metabolites identification) thus providing with a wider perspective about the applicability of the proposed spectroscopic technique for in-vivo sensing of hyperglycemia.

The experimental results of this thesis demonstrate that simple transmission-type spectrometers in the W-band in combination with the right statistical analysis tools show great potential for the further development of a non-invasive diagnostic tool for in-vivo sensing of sustained glycemia in humans. This would mean a clear breakthrough in Diabetes Mellitus diagnostic and management as it could substitute the current standard tool in medical practice for DM diagnosis and monitoring, as it is the invasive HbA1c test.

RESUMEN

Hoy en día, la espectroscopia de ondas milimétricas y THz comprenden un gran campo de investigación que se enfoca en gran parte al desarrollo de técnicas e instrumentos nuevos y mejorados en aplicaciones prácticas, tales como: diagnósticos médicos, controles de seguridad y el control de procesos industriales; esto se debe principalmente a su alcance para elucidar y caracterizar la materia de forma remota y no invasiva. Dichos instrumentos se basan en técnicas espectroscópicas que, en términos generales, irradian una muestra objetivo utilizando radiación electromagnética a diferentes frecuencias (longitudes de onda), para medir sus propiedades ópticas en función de las frecuencias utilizadas. Siendo ésta la razón por la que el uso de técnicas espectroscópicas resulta en la adquisición de datos espectrales, los cuales contienen la información sobre la interacción entre la muestra evaluada y la radiación electromagnética aplicada. En esta situación, el procesamiento y análisis de los datos espectrales son tareas importantes a la hora de desarrollar nuevos instrumentos basados en técnicas espectroscópicas (sistemas de espectroscopía aplicada), y uno de los escenarios más desafiantes se encuentra en aplicaciones donde la muestra incluye múltiples especies con propiedades ópticas muy similares entre sí, como suele ocurrir en las aplicaciones biomédicas. En consecuencia, se requieren métodos estadísticos, no solo para extraer la información deseada de los datos espectrales, sino también como parte del proceso de calibración de la técnica o el instrumento.

En esta tesis doctoral, se evalúa un enfoque novedoso que utiliza espectroscopía de ondas milimétricas para la detección y monitorización in vivo y de forma no invasiva de la hiperglucemia sostenida, típicamente asociada a la Diabetes Mellitus (DM), en donde se incluyen varias pruebas experimentales realizadas con modelos animales y un estudio clínico piloto en los seres humanos (pacientes con DM tipo 1). Las pruebas experimentales se realizaron mediante un espectrómetro de banda W que se construyó específicamente para esta investigación, capaz de adquirir los espectros de reflexión y transmisión a partir de las mediciones in vivo y no invasivas realizadas sobre las muestras biológicas. Todos los datos espectrales obtenidos durante estas pruebas, fueron procesados y analizados utilizando métodos de Análisis de Datos Funcionales (ADF)

debido a su idoneidad para abordar los escenarios espectroscópicos complejos que se encuentran muy comúnmente en el marco de aplicaciones biomédicas. Las técnicas del ADF nos permitieron estudiar la respuesta espectral medida en la banda W bajo los entornos de aprendizaje supervisado y no supervisado, proporcionando una interpretación de la contribución de las frecuencias medidas en todas las etapas del análisis, incluso para un análisis longitudinal realizado con datos espectrales que fueron observados a diferentes tiempos durante el estudio clínico piloto con pacientes de DM tipo 1. También es importante señalar que referente a la concepción de los diagnósticos presentados, se utilizó un enfoque de perfilamiento espectral no específico, el cual nos permitió considerar de forma colectiva las características espectrales de todos los constituyentes de la muestra (sin tener que identificar metabolitos individualmente). Además, este enfoque nos proporcionó una perspectiva más amplia a la hora de estudiar la aplicabilidad de la técnica espectroscópica propuesta para la detección in vivo de hiperglucemia.

Los resultados experimentales obtenidos en esta tesis, demuestran que un simple espectrómetro capaz de medir la transmitancia de la muestra biológica en la banda W, combinado con las herramientas de análisis estadístico adecuadas, proporciona un enfoque potencial para el desarrollo de una herramienta de diagnóstico no invasiva enfocada a la detección in vivo de la glucemia sostenida en humanos. Esto significaría un claro avance en el manejo y en el diagnóstico de la diabetes, ya que podría sustituir la prueba invasiva de HbA1c, que es la herramienta estándar utilizada actualmente en la práctica médica para el diagnóstico y seguimiento de la Diabetes.

TABLE OF CONTENTS

AGRADECIMIENTOS	IV
PUBLISHED AND SUBMITTED CONTENT	VI
OTHER RESEARCH MERITS	VIII
LIST OF ACRONYMS	IX
ABSTRACT	XI
RESUMEN	XIII
TABLE OF CONTENTS	XV
1. Introduction	1
1.1. The spectroscopy: fundamentals and challenges in biomedical engineering	2
1.2. The spectral data analysis in the development process of applied spectroscopy systems	10
1.3. Thesis overview and objectives	14
2. Statistical tools for spectral data analysis	18
2.1. Functional Data Analysis (FDA)	20
2.1.1. Definitions and theoretical assumptions	21
2.2. Main summary statistics for functional data	23
2.3. Approximation of Functional data	24
2.3.1. Basis expansion.....	26
2.3.1.1. B-spline bases.....	27
2.3.2. Regression splines.....	28
2.3.2.1. Leave-One-Out Cross Validation (LOOCV)	30
2.3.3. Penalized splines (P-splines)	31
2.3.3.1. Generalized Cross Validation (GCV)	32
2.4. Functional Principal Component Analysis	33
2.4.1. Approximation of the FPCA solution	35
2.4.2. Orthogonal representation of a Stochastic Process	36

2.5. Functional Linear Regression	37
2.5.1. Functional Logit Model.....	39
2.5.1.1. FLoR model estimation by FPCs	40
2.5.1.2. Functional parameter interpretation.....	41
2.6. Longitudinal Functional Data Analysis	43
2.6.1. Longitudinal FPCA (LFPCA).....	44
2.6.2. Longitudinal scalar-on-function regression analysis	47
3. Millimeter-wave spectroscopy for non-invasive sensing of sustained hyperglycemia	51
3.1. Diabetes Mellitus: technological advances and challenges for non-invasive sensing of blood glucose level.....	53
3.2. THz, Microwave and Millimeter- wave radiation in biomedical engineering.....	57
3.3. Mm-wave spectroscopy instrument used for experiments ..	59
4. Use of W-band spectroscopy for in-vivo non-invasive assessment of hyperglycemic states using animal models	64
4.1. Description of the experiments using animal models	66
4.1.1. Animal models used in the experiments.....	67
4.1.2. Experiment A.....	68
4.1.3. Experiment B	71
4.1.4. Experiment C.....	73
4.2. Blind exploration and characterization of the measured spectral response.....	74
4.2.1. Interpreting FPCs' scores.....	77
4.2.2. Interpreting FPCs' loading functions	83
4.3. Evaluation of the consistency and robustness of the measured spectral response	84
4.3.1. Multi-test analysis: performance and robustness of the proposed approach	87
4.3.2. Validation of the consistency of the applied spectroscopic technique and analysis of the FPCLoR model	90
4.4. Sensitivity of the non-invasive approach to detect changes in sustained glycemia	93
4.5. Discussion and conclusions	97

5. Use of W-band spectroscopy for in-vivo and non-invasive assessment of hyperglycemic states in Humans: Pilot clinical study	101
5.1. Design of the pilot clinical test	103
5.2. First results in humans: Scalar-on-function regression.....	109
5.3. Analyzing inter and intra-subject variations within the W-band spectrometer measurements	113
5.3.1. LFPCA on longitudinal spectral data	114
5.3.2. Longitudinal regression analysis on HbA1c value.....	118
5.4. Discussion and conclusions	120
6. General conclusions and discussion	123
7. References	131
8. Appendix	157
8.1. Ethical issues	158
8.1.1. Animal experiments	158
8.1.2. Pilot clinical study.....	158
8.2. Informed consents	161
8.2.1. Controls.....	161
8.2.2. Patients	164

1. INTRODUCTION

1.1. THE SPECTROSCOPY: FUNDAMENTALS AND CHALLENGES IN BIOMEDICAL ENGINEERING

The origins of the spectroscopy dates from the 17th century, being the systematic studies made by Isaac Newton about the white light decomposition (rainbow effect) considered as the first main contribution in this field [1]. Early in the 19th century, Joseph von Fraunhofer invented the first spectroscope and discovered a set of dark lines in the visible spectrum (~ 400 THz - ~ 790 THz), known as Fraunhofer lines [2]. Subsequently, Gustav Robert Kirchhoff and Robert Wilhelm Bunsen constructed a flame spectroscope to study the light emission properties of chemical components [3], [4], and their experiments allowed them to relate the Fraunhofer lines to the absorption properties of some gases in the atmosphere of the sun. Kirchhoff and Bunsen stated that each chemical element has a set of characteristic spectral lines (fixed frequencies) in which can either emit or absorb energy [5]. In 1913, Niels Bohr analyzed the hydrogen spectrum combining the quantum theories proposed by Albert Einstein and Max Planck with the Rutherford model, resulting in the Bohr model [6], [7]. The Bohr model postulates that electrons exist only in quantified energy orbitals (states of constant energy) and they can be promoted to higher energy orbitals (excited state) through photon absorption. Similarly, the electrons fall back to lower energy orbitals (electron relaxation) by emitting photons, with the lowest possible energy orbital called ground state. But the electrons only can absorb or emit photons carrying an amount of energy exactly equal to the energy difference between the initial and final state of the transition. This theory provided an explanation to the already observed characteristic spectral lines associated to chemical components (absorption and emission spectrum) since the photon energy is determined by the frequency of the electromagnetic wave as follows (Planck-Einstein relation)

$$E = h\nu, \tag{1-1}$$

where E is the photon energy, h is the Planck's constant (6.625×10^{-34} J·s), and ν is the frequency. Thus, the transition of an electron from an initial state E_1 to a higher energy state E_2 by the photon energy absorption satisfies the relation

$$E_2 - E_1 = nh\nu, \quad n \in \mathbb{N}. \quad (1-2)$$

Let us to note that the frequency ν and the wavelength λ of an electromagnetic wave are related as follows

$$\nu = \frac{c}{\lambda}, \quad (1-3)$$

where c is the speed of light in vacuum (3×10^8 m/s). The electromagnetic waves (electromagnetic radiation) include all types of energy travelling through space at the constant velocity of c , being the visible light a portion of the electromagnetic spectrum [8].

In other words, when a molecule interacts electromagnetic waves only the frequencies satisfying Eq. (1-2) will produce molecular electronic transitions resulting in a partial or complete absorption of the radiant energy. This wavelength dependent absorption coefficient is seen every day on the characteristic colors of matter; the observed color corresponds to the not absorbed wavelengths by the chromophores of molecules within the visible light. Then, as illustrated in Figure 1-1 where the absorption spectra for several gases are shown, an absorption spectrum is a wavelength dependent function with its peaks matching with the absorbing frequencies of a determined chemical compound observed over a frequency range.

Let us note that besides the molecular electronic transitions (electronic energy), there are other rotational and vibrational molecular energy levels (related to the absorption of photons) that are also characterized by their dependence on the wavelength [9, Ch. 1]. The energy absorption by molecules within the millimeter-wave and THz radiations is determined by the rotational and vibrational energy levels rather than molecular electronic transitions.

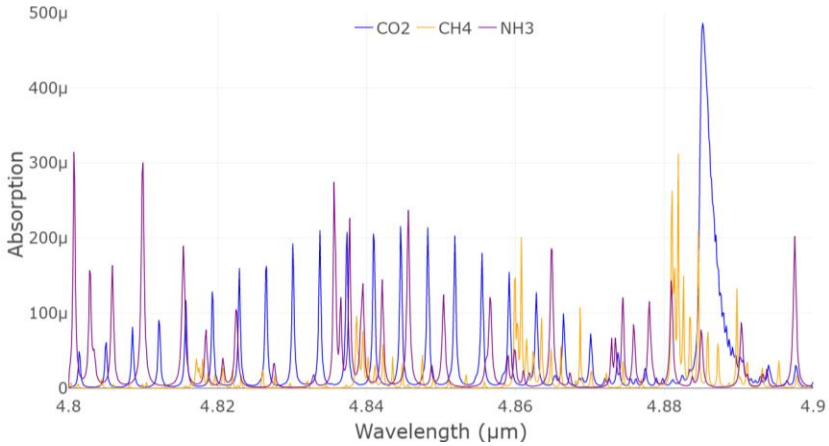


Figure 1-1. Absorption spectra of several gases: carbon dioxide (CO₂), methane (CH₄) and ammonia (NH₃) in the 4.8 μm to 4.9 μm range [10]. Different concentrations were considered for viewing purposes.

The relationship between a chemical specie and the amount of energy absorbed when is irradiated by electromagnetic waves (absorbance) was studied by Johann Heinrich Lambert in 1760 and later by August Beer in 1852 providing the Beer-Lambert law [11], [12, Ch. 1]. This law demonstrated (under certain conditions) that absorbance A is directly related to the concentration of the chemical specie as follows

$$A = \varepsilon \ell \rho = \alpha \ell, \tag{1-4}$$

where ε is the wavelength dependent molar extinction coefficient, ℓ is the length crossed by the electromagnetic wave, ρ is the concentration of the given chemical specie, and α is known as the absorption coefficient. Thus, the absorption coefficient indicates the number of photons (amount of energy) absorbed per unit distance. In cases in which there are multiple species (without chemical reactions between them) absorbing at a given wavelength, the total absorbance equals to the sum of the individual absorbances [12, Ch. 1]. The relation shown in Eq. (1-4) was empirically determined by measuring the attenuation of the intensity of an electromagnetic wave (at a specific frequency) travelling through an absorbing media (known as transmittance). The transmittance T and the absorbance A are related by the Beer-Lambert law as follows

$$A = -\log_{10} T, = -\log_{10} \frac{I_{out}}{I_{in}}, \tag{1-5}$$

where I_{in} is the intensity (radiant energy) of the electromagnetic wave that reaches the absorbing media, and I_{out} is the radiant energy that passes through it without being absorbed.

Over the years more precise theories have been formulated describing the absorption phenomenon and providing solutions for cases in which Eq. (1-4) is no longer valid (e.g. due to stronger intermolecular interactions or scattering effects described below), but, those findings were fundamental pieces in the development of the spectrochemical analysis (absorption and emission spectroscopy) [7], [13], [14]. The characteristic and unique behavior of chemical species observed from their interaction with electromagnetic radiation, making identifiable most of the chemical species (as a fingerprint), has attracted increasing attention to the spectroscopy field for the development of non-invasive exploratory tools (spectroscopic techniques) of matter. Today, the spectroscopy field spreads over the whole electromagnetic spectrum involving different types of radiation such as microwaves, Terahertz (THz), infrared (IR), ultraviolet (UV), X-Rays, among others [15]–[19]. The linear spectroscopic techniques are commonly focused to measure the spectrum resulting from the irradiation of a sample under study (gaseous, liquid, or solid) by electromagnetic waves at different frequencies within a frequency interval of interest (spectral interrogation).

In addition to the atomic absorption and emission of energy, there are other natural phenomena associated to the interaction of electromagnetic radiation with matter such as reflection, refraction, and scattering, among others [20]–[25]. These phenomena are widely studied and used in the spectroscopy field because they are related to the natural properties of the matter and are wavelength dependent too [26]–[32]. In Figure 1-2 are shown simply examples of absorption, reflection, refraction and scattering phenomena above mentioned. (A) Absorption: the light beam enters the liquid and is strongly attenuated until it fades. (B) Reflection: the light is returned when hits the water surface retaining the imaging geometry (specular reflection), when the light is reflected in all directions is called diffuse reflection. (C) Refraction: the pencil seems to have two different positions (appears to be broken) because the speed and direction of the light change into the water. Another example of refraction is the rainbow effect (light dispersion). (D) Scattering: the photons of the light beam are deviated (scattered) from its straight trajectory in random directions when hits the particles, when the photon is repeatedly scattered by different molecules is called multiple scattering. This phenomenon is

also observed in the sky; the electromagnetic waves corresponding to those frequencies that human eye perceives like blue light are scattered (more than the other wavelengths) by gases in Earth atmosphere (Rayleigh scattering). Let us observe here that specular reflection and refraction are considered as particular cases of coherent scattering from a large number of molecules when the incident wave interacts with “optically smooth” surfaces [25], [32], [33].

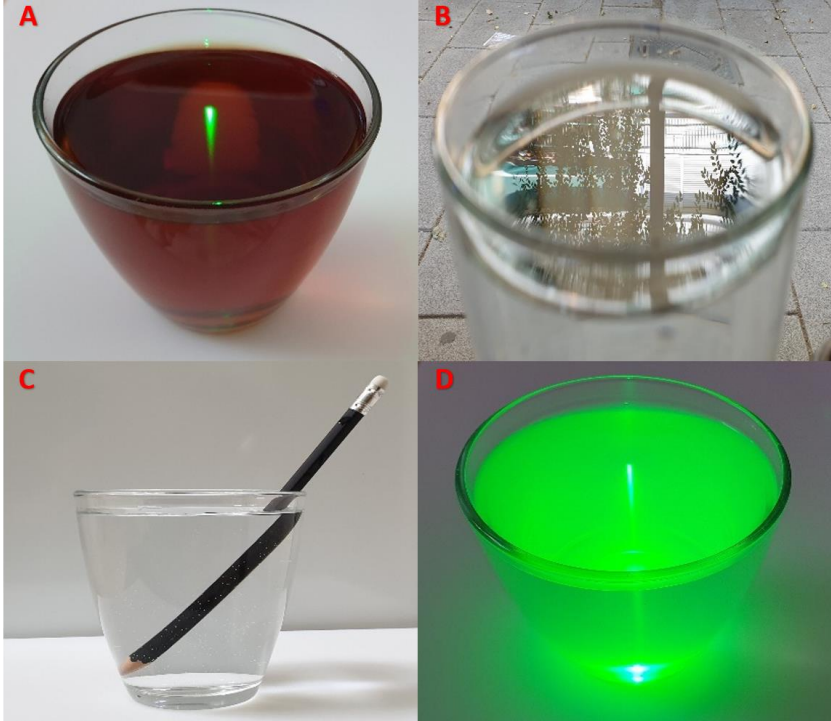


Figure 1-2. Photographs illustrating in the simplest way the absorption (A), reflection (B), refraction (C), and scattering (D) phenomena.

The reflection, refraction and scattering effects are inherently related to the refractive index η of matter [24], [34] which is commonly defined as follows

$$\eta_i = \frac{c}{u_i}, \quad (1-6)$$

where c is the speed of the electromagnetic waves in vacuum and u_i is the phase velocity (speed of propagation) of the electromagnetic wave in a given medium, being u_i dependent on the natural properties of the medium. As shown by Fresnel laws [35], [36, Ch. 1] (Augustin-Jean Fresnel; 1821) the refractive index is strongly related to the reflection and transmission coefficients of matter, and it can be also used to determine the change in direction of the incident wave when it passes from one medium to another (transmitted wave) according to its angle of incidence on the interface surface between both mediums (generally true for optically smooth surfaces). Such relation between the refractive index and the change in wave direction was stated in the Snell's law [36, Ch. 1], [37] (Willebrord Snel van Royen; 1637) which in its most basic form (non-absorbing media) is given by

$$n_1 \sin \theta_1 = n_2 \sin \theta_2 , \quad (1-7)$$

where n_1 and n_2 are the respective refractive indices of each medium, θ_1 is the angle of incidence, θ_2 angle of refraction. Let us observe that the refractive index is complex valued, with the imaginary and real parts determined by the molar extinction coefficient and the refractive properties of the medium, respectively [38]. Both properties (absorption and refraction) are interlinked by the Kramers-Kronig relations (Ralph de Laer Kronig, 1926; Hendrik Anthony Kramers, 1927) [39]–[42], which demonstrate that the refractive properties of matter are directly associated with its characteristic absorption spectral lines. For some cases, the Kramers-Kronig analysis has been proposed to characterize the complex refractive index from reflectance spectra [43], [44]. However, the mathematically modelling of complex refractive index becomes a challenging task for highly absorbent and scattering media [45]–[48]. The spectroscopy applied to biomedical engineering is an example of such cases, where the samples usually involve a large number of components with strong interactions between molecules (liquids) and exhibiting similar absorption properties, as illustrated in Figure 1-3. As it can be seen, absorption spectrums from oxyhemoglobin and deoxyhemoglobin show a continuum level of absorption over the frequency interval with wider and less pronounced peaks compared to the narrow spectral features observed in Figure 1-1 for gases. This kind of spectral responses are very characteristic for biological tissues. In particular, the spectral interrogation of in-vivo biological samples will be the target of the applied spectroscopic technique addressed throughout this thesis.

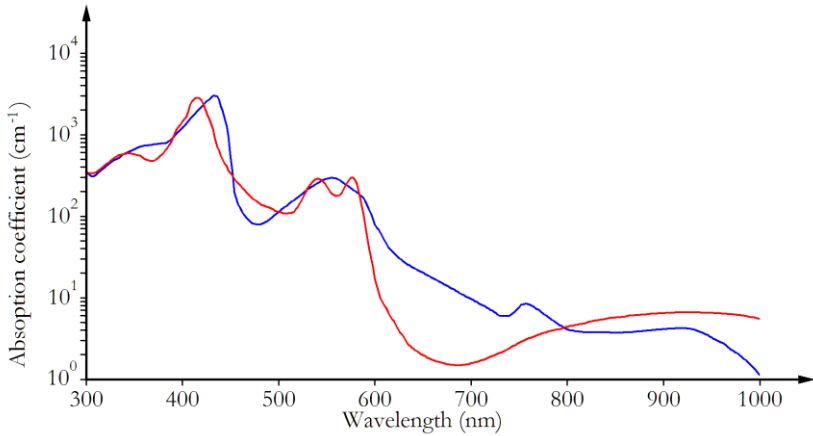


Figure 1-3. Absorption spectrums of oxyhemoglobin (red line) and deoxyhemoglobin (blue line) blood in the 300 nm to 1000 nm range [49]

In addition to the absorption coefficient, the propagation of electromagnetic radiation through biological media is also determined by scattering effects (scattering coefficient) due to their inhomogeneities. The scattering coefficient indicates the number of photons scattered per unit distance. There are many factors influencing the measured characteristics of scattering such as the wavelength, complex refractive index, concentration, geometrical thickness of the medium, and the size, shape and orientation of molecules, among many others. It has been proven that scattering measurements provide structural and functional information about tissue and cells that can be used to detect and monitor pathologies in humans [50], [51], being the detection and monitoring of cancer cells one of the more promising biomedical spectroscopic applications.

The strong absorption of many biological tissues (poor penetration depth in tissue) makes many spectroscopic techniques only possible for in-vivo measurements by reflectance approaches (reflection and backscattering). However, it has been shown that transmittance measurements provide much more information about absorption and scattering characteristics of tissue [52]. This issue is addressed in chapter 4 by comparing reflection and transmission profiles from in-vivo measurements of different metabolic conditions in mice.

As shown by the Radiation Transfer Equation (RTE) theory [53], [54], the propagation of electromagnetic radiation through biological media is mainly determined by the combination of the abovementioned

effects, this is, the absorption coefficient $\delta_a(\lambda)$, and a transport scattering coefficient $\delta'_s(\lambda) = \delta_s(1 - g)$ with δ_s being the wavelength dependent scattering coefficient and g an anisotropy factor of scattering which is defined as the average cosine of the scattering angles. However, the great diversity and structural complexity of biological tissues make quite difficult to mathematically modelling the cooperative effects of absorption and scattering therefore, in practice, many assumptions have to be done to simplify the mathematics, with exact solutions achieved only in a small number of particular cases [50], [53], [55], [56]. New methods based on Monte Carlo simulations have been proposed for the numerical solution of the radiation transfer problems, allowing to incorporate more real conditions in the model plus other advantages over analytical methods [50], [54], [56], [57], but, usually they require high computational resources with large simulation times, they are not valid for high accuracy problems, and are impractical for high in-depth tissue optical analysis. An extensive review of the theoretical background and current spectroscopic methods for biomedical diagnostics, and optical properties of tissues among other biological objects can be found in [50], [55], [58].

In general, modelling the interaction between electromagnetic radiation and biological systems is an extremely hard task, and this thesis does not intend to address the above-mentioned radiation transfer theories and the current analytical or numerical methods commonly used to address such issue. On the contrary we will follow the also very common approach of new spectroscopy systems applied to biomedical engineering that rely on performing the spectral interrogation on the biological media (within a frequency range of interest), and then, try to correlate the spectral responses measured at each frequency with a specific biological component (or components) of interest. Such approach is hindered due to the superposition of similar spectral responses from different components (as shown in Figure 1-3), besides instrumental noise and other interferences associated to environmental and physiological factors (depending on the spectroscopic technique). Therefore, statistical methods are often needed for modelling and classification of measured spectral data. This thesis focuses to the analysis and interpretation of spectral data obtained from multi-species structures exhibiting complex spectral features. As described in chapter 4, a non-targeted spectral

profiling approach is proposed to explore and evaluate new applied spectroscopic techniques in biomedical engineering.

1.2. THE SPECTRAL DATA ANALYSIS IN THE DEVELOPMENT PROCESS OF APPLIED SPECTROSCOPY SYSTEMS

As introduced in previous section, there are numerous spectroscopic techniques focused to measure the relative energy that is whether emitted, absorbed, reflected (including backscattering) or transmitted (affected by refraction, scattering and absorption) by a sample as a function of frequency (wavelength). The use of spectroscopic techniques for spectral interrogation of such sample under study leads to the acquisition of spectral data for further processing and analysis. The measured spectral data usually correspond to the intensity (amplitude) or phase (phase shift with respect to the incident wave) of the reflected and/or transmitted signals resulting from the spectral interrogation. Then, amplitude and phase parameters are used to find a characteristic relationship that allows for modelling one or more sample constituents (or other sample properties). The eventual goal is to obtain a complete spectroscopy system (incorporating the spectroscopic interrogation technique, the electronics and the software for data acquisition and processing) capable of autonomously carrying out measurements and automatically provide precise results (like a medical diagnosis or estimate the concentration of one or more sample constituents).

However, to achieve this, the realization of experimental tests for the evaluation and calibration of the spectroscopic technique are necessary, as shown in Figure 1-4, which illustrates the general process for the development of applied spectroscopy systems. The difficulty of this process can be coined to the complexity of the sample, relatively simpler for samples exhibiting sharp spectral responses (easy to model and without superposition effects), and to the spectroscopic technique (robustness against external factors and the signal-to-noise ratio). In this sense, the application of spectroscopic techniques to biomedical engineering leads to the more complex multi-species scenarios in which spectral responses of many sample constituents are overlapped and there are strong interferences from instrumental, environmental and physiological factors. Consequently, statistical tools are needed for the

processing and analysis of spectral data in order to make the development process of applied spectroscopy systems as efficient as possible.

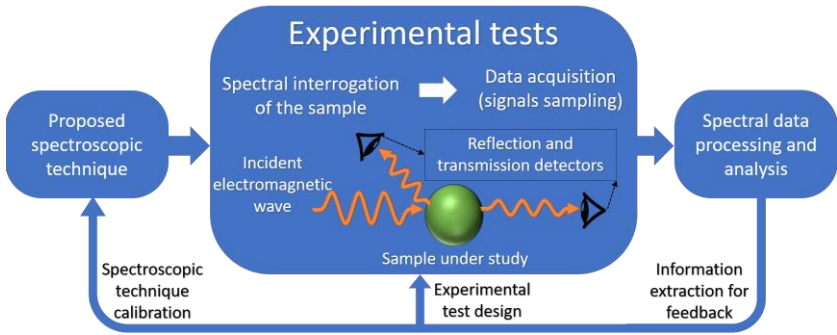


Figure 1-4. Scheme illustration of the general process for development and evaluation of applied spectroscopy systems.

As depicted in Figure 1-4, the statistical analysis of the spectral data can be used not only to validate and calibrate the spectroscopic technique, but also to improve the design of the experimental tests. Furthermore, as it will be seen throughout this doctoral thesis, the spectral data analysis can provide useful information to optimize the spectral interrogation process and setup of the spectroscopic technique. This thesis focuses on showing the relevance of such feedback process and how the use of adequate statistical methods can maximize the information extracted from the spectral data analysis, especially in the early stages of the development process of an applied spectroscopy system.

The first step in the experimental evaluation of a spectroscopic technique is to explore the spectral response of the sample in the frequency interval of interest. As highlighted above, in the case of biomedical applications, the overlap of the spectral features from sample constituents prevents their direct isolation and quantification. The complexity of the spectral features of multi-species structures (such as biological tissues) force the spectral interrogation of broader frequency intervals. Besides this, the high resolution in frequency (and time) of modern spectroscopic instruments and currently available technologies for data collection and storage have made that spectral interrogation usually collects a large number of frequencies (or time points).

Consequently, a common feature in the structure of spectral data sets is that the number of measured frequencies p is much larger compared

to the number of observations n (small sample populations), for example in biological experiments in which access to the sample is quite limited. Nowadays, the standard methods used in the analysis of spectral data consist of Multivariate Data Analysis (MDA) that analyzes the measured frequencies as individual variables [59]–[62]. The analysis of spectral data under the MDA framework, considering the structure features previously mentioned, have two important consequences that appear immediately. The first one is that the spectral data analysis results in high-dimensional statistical problems such as the so called “large p , small n ” problem [63]–[65], basically, as the number of variables increases (measured frequencies) much more observations are required. The second is the well-known multicollinearity problem [66], [67], [68, Ch. 37] since the variables of spectral data sets might be highly linearly correlated, depending on the frequency resolution and the spectral features of the sample. Such correlation is originated, for example when multiple measured frequencies are comprising the same spectral peak (highly likely for broad spectral peaks), which results in strong dependencies between frequencies. These multicollinearity and high-dimensional problems cause that many classical statistical methods for regression analysis become inadequate or not feasible (over-fitted or inefficient regression models, inaccurate computations or without solution, etc.), requiring more dedicated statistical analysis [69]–[72]. Dimension reduction methods [73]–[76] are often used prior to any modelling efforts, but since frequencies contribution is somehow summarized, the resulting regression analysis are not directly interpretable in terms of the original measured frequencies. An alternative approach that has been also proposed to deal with both high-dimensional and multicollinearity without losing original semantics of measured variables consist of variable selection methods [77]–[80]. The variable selection methods are focused on finding specific wavelengths (within the whole frequency range) that provide the best regression results without redundant information (uncorrelated wavelengths), however, in spectroscopy applications, this is still a challenging task (leading to high computational loads in many cases) and selecting optimal interval variables (frequency sub-intervals) [81]–[83] has been proposed as a more practical and stable approach. In fact, functional representation, which is the first step in Functional Data Analysis (FDA), has been proposed as a pre-processing step to facilitate the variable selection problem and to improve the statistical significance of the obtained results [84]. In any case, as will be seen in chapter 2, summary statistics are required in multivariate statistical analysis of spectral data for longitudinal settings (the samples are

measured repeatedly over a period of time), losing a lot of available information within the measured frequency range and interpretation capacity.

As seen before, a spectrum contains a set of values measured describing the evolution of a parameter (amplitude or phase) over a frequency range. This underlying continuous nature of spectral interrogation process generating the spectral data is completely bypassed when measured frequencies are treated as “independent variables” (MDA framework). In this sense, an alternative and more adequate statistical approach for spectral data analysis has emerged from the statistic field, called Functional Data Analysis (FDA) [85], [86] already mentioned above. As illustrated in Figure 1-5 the FDA framework, described in detail in chapter 2, deals with spectral data as continuous trajectories (discretized at a sequence of frequency points) of a wavelength dependent function being advantageous in many ways: it solves high-dimensional problems, handles multicollinearity problems without losing interpretability of regression analysis in terms of the original measured frequencies, allows to consider all the available information in longitudinal analysis, and enhances visualization of spectral data.

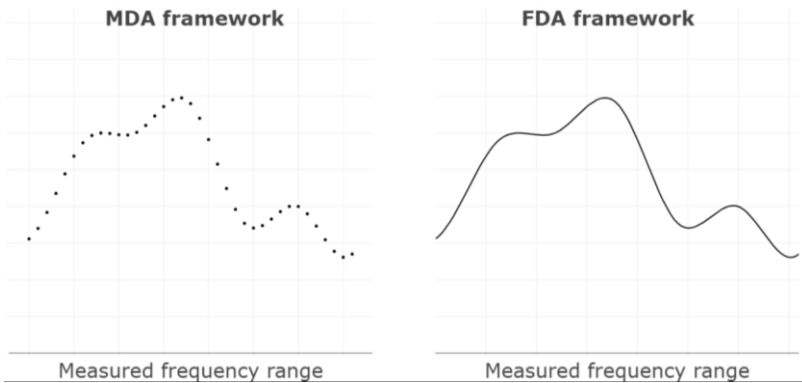


Figure 1-5. On the left, a spectrum consisting of a set of values measured at different frequencies (as acquired in practice) is illustrated; the MDA considers the measured frequencies as a set of independent variables. In contrast, the FDA analyzes the continuous trajectories (functional data) over the whole frequency range, estimated from the measured spectra (on the right).

The greatest strength of FDA is the perspective used to address the spectral data analysis. Spectral data modelling is based on trends within the approximated continuous trajectories (the shape of the wavelength

dependent function) instead of a set of values corresponding to the discretely measured frequencies that are easily affected by undesirable interferences. As will be seen, the estimation of the continuous trajectories (functional data) from the measured spectra (vectors) is based on very flexible approximation methods that can be used in spectroscopy to reduce the effects of undesirable interferences on spectroscopic measurements. Furthermore, the estimated continuous trajectories facilitate the study of dynamics of the spectral responses within the frequency range of interest, such as their derivative functions [87]–[89], that in some cases can be used to annihilates noise factors in measured spectra. The spectral data analysis carried out throughout this thesis is addressed under the FDA framework since provides a set of more suitable and powerful statistical tools for spectral data analysis, can be applied to a wide range of spectroscopy applications and being particularly promising in biomedical engineering and applications [89].

1.3. THESIS OVERVIEW AND OBJECTIVES

The work presented in this thesis dissertation stems from two main objectives: firstly to assess the applicability of a novel non-invasive approach, based on millimeter-wave (mm-wave) spectroscopy, for the diagnosis and follow-up of Diabetes Mellitus (DM) through the in-vivo sensing of sustained hyperglycemia and, secondly, to motivate and boost the use of FDA methods for spectral data analysis in the development, assessment and calibration of new applied spectroscopy systems, especially for biomedical applications. Regarding to the first objective, the proposed spectroscopic technique was initially tested by performing spectral interrogation on animal models of different glycemic states, and subsequently, a pilot diagnostic validation study was carried out on humans, referred to as “pilot clinical study” in the remainder of the document. All the spectral data collected from the in-vivo measurements were processed and analyzed under the framework of the FDA to show the current available FDA methods and their potential for spectral data modelling and interpretation. The introduced FDA methods can be applied to any spectroscopy application in which the collected data is measured over a continuum, which is the most common scenario in spectroscopy field, e.g. spectral data measured over time or frequency domains.

The remainder of this thesis dissertation is organized in five more chapters as described below.

First, the theoretical background and basic concepts needed in the framework of the FDA are introduced in chapter 2. A set of statistical tools for continuous trajectories approximation from measured spectra (functional data estimation) are described in detail and their applications in spectroscopy approaches are discussed. Different FDA methods based on unsupervised and supervised approaches, which will be applied in the analysis carried out in chapters 4 and 5, are also introduced.

In chapter 3, the relevant clinical concepts associated to DM are introduced, and a brief review of the state of the art in DM diagnosis and monitoring is done. The need for new non-invasive technologies to improve current medical procedures in diagnosis and follow-up of DM, enhancing patients' quality of life and reducing the global incidence rate, is motivated. Then, the potential of Terahertz (THz), microwave, and mm-wave spectroscopy for biomedical applications is discussed and the mm-wave spectroscopic technique proposed for in-vivo sensing of sustained hyperglycemia is described. Considering that there is no a priori information of the spectral features of biological sample constituents within the measured frequency range and mathematical modelling is for now impractical, the in-vivo measurements were performed under a non-targeted spectral profiling approach [90], [91] focused to characterize the clinical pathology (DM) in terms of the classification of the spectral response from the biological samples (healthy and diabetic animals) over the whole frequency range, thus, the sum of the spectral features from all sample constituents are considered instead of targeting a specific metabolite (non-targeted metabolomics).

In chapter 4, the most relevant experimental results obtained from the assessment and validation of the proposed spectroscopic technique using animal models are presented. The content of this chapter includes the published results in papers A, B and C. The animal experiments were conducted on different mice strains involving healthy cases (controls) and two different mice models representing sustained hyperglycemic cases: mild and full-blown diabetes. The in-vivo measurements consist of spectral interrogation over the W-band, within the mm-wave range, directly performed on a fold of skin from the back of the mice. The experimental protocols and the mice models employed in each experimental test for in-vivo measurements using the W-band

spectrometer are described in detail. All this research work was carried out in collaboration with the Epithelial Biomedicine Division of the Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT) and following the European and Spanish laws and regulations for housing and care of laboratory animals. This chapter shows how the extraction of information from the spectral data analysis interpretation not only allows to reformulate a simpler design of the spectroscopic technique but also to design the next experimental tests with clear and specific objectives.

In chapter 5, the spectroscopic technique proposed for in-vivo non-invasive sensing of hyperglycemic metabolism is tested on humans. The excellent performance of the proposed approach achieved during animal experiments encourage the realization of a pilot clinical study focused on type 1 DM patients. The in-vivo measurements using the W-band spectrometer were directly performed on a skin fold in the first interdigital space (between the thumb and index finger) of the right hand. All this research work was carried out in collaboration with the Instituto de Investigación Sanitaria Fundación Jiménez Díaz (Servicio de Endocrinología y Nutrición) and following the International and European ethical and safety principles regarding human experimentation. This chapter validates the results obtained from animal experiments on humans and studies the relation between the non-invasive measure and the mean value of the Blood Glucose Level (BGL) in body for most recent three months. Therefore, the pilot clinical study was designed as a longitudinal study and lasted about twelve months, with seven months between the first and the last measurements. The longitudinal spectral data collected throughout the entire pilot clinical study (subjects were measured at three times with three months at least between each time) were analyzed using novel FDA methods to show the scope and advantages of FDA framework in spectroscopy applied to biomedical applications.

In chapter 6, the main conclusions extracted from this doctoral thesis are listed. The use of FDA methods in spectral data analysis and the obtained results from experimental assessments of the proposed spectroscopic technique are discussed. Future considerations for following experimental assessments in the development and calibration of the proposed spectroscopic approach for non-invasive sensing of sustained hyperglycemia in diabetics are also revised.

Finally, the corresponding section including all cited references throughout this thesis dissertation and an appendix including the ethical issues corresponding to the animal experiments and the pilot clinical trial, and the informed consents containing all the relevant information provided to the volunteers prior they participation in the pilot clinical study are shown.

2. STATISTICAL TOOLS FOR SPECTRAL DATA ANALYSIS

We have previously stressed the intrinsic continuous nature underlying the spectral data raised from spectral interrogation throughout a continuous frequency interval, although experimental and practical constraints allow only for the obtaining of spectral information at given sampling points. One of the main objectives of this thesis is to show the challenges and exciting possibilities of addressing these measured spectral data (at discrete sampling points) by using Functional Data Analysis (FDA), which is a current topic in statistical mathematics of continuous signals. In FDA, the sample trajectories discretized in frequency (vectors containing the sampling information points) are converted into continuous curves over the measured frequency range, obtaining a functional data set. It is important to note that the set of sampling points in which spectral data is observed does not need to be regular and can differ between observations. However, the spectral data sets analyzed in this thesis were measured on a set of equally spaced frequencies. In contrast to a finite collection of sampled values, the continuous representation (functional data) obtained from the spectral data is a rich source of information that permits us to evaluate many aspects of the spectral response in the whole measured frequency band (trends, derivatives, etc.). Other advantages of the use of FDA are data noise reduction by curve smoothing methods, improves the intrinsic patterns detection in data throughout the continuous domain, and better flexibility for longitudinal and high dimensional spectral data analysis.

In this chapter, the necessary FDA theory and basic concepts needed for further statistical analysis of spectral data used in this work are firstly introduced. Subsequently, some functional data approximation approaches and FDA methods for unsupervised and supervised spectral data analysis will be briefly described.

The spectral data collected from different experiments designed to assess the proposed non-invasive approach for detection and monitoring of hyperglycemic metabolism will be analyzed using both supervised and unsupervised FDA methods. The aim is to extract relevant information from the spectral data analysis allowing for a feedback between the spectroscopic technique and the measured spectral response. Such

feedback from the measured spectral response is essential for assessment and optimization of the applied spectroscopic technique.

2.1. FUNCTIONAL DATA ANALYSIS (FDA)

FDA is a very active branch of research in the Statistics field, devoted to solving statistical problems involving functional variables. In a functional variable its observations are functions that represent the evolution of a scalar variable over a continuum. The aim of this research topic is to provide statistical tools for the analysis and modelling of functional data sets (curves), defined and sampled on a continuous domain, that represents a population. As we shall see throughout this thesis, many of the statistical methods developed in the FDA framework are extended versions of well-established Multivariate Data Analysis (MDA) methods. A very comprehensive collection of information about FDA basic methods, computational aspects, and some practical applications can be found in the pioneer books of Ramsay and Silverman [92], [93] and Ramsay et al. [94].

A functional data set $x_1(t), x_2(t), \dots, x_n(t)$ can be seen as a set of observations of a functional variable. Usually, such observations are sample functions (also called realizations, sample paths, trajectories, among other names) of a stochastic process $\{\mathcal{X}(t): t \in \mathcal{T}\}$. A stochastic process is formally defined [95, pp. 201–202] *as a family of random variables* $\{\mathcal{X}(t, \omega): t \in \mathcal{T}, \omega \in \Omega\}$ *indexed by some real set* \mathcal{T} *(discrete or continuous) and defined on a common probability space* $(\Omega, \mathcal{A}, \mathcal{P})$. If all the random variables of the random process take only real values, then, it is known as a real stochastic process. The sample function $x_i(t)$ is an observation (outcome) of the stochastic process and contains a single value of each of the indexed random variables of the stochastic process. Thus, the sample functions represent different states of the stochastic process. Usually, the sample functions describe the evolution of a property or characteristic of a system under study, which is observed over \mathcal{T} , a frequency interval in our case, but can also be time, geographic location, probability, etc.

Throughout this thesis, and for the purposes of the FDA, only stochastic processes holding the three following hypotheses will be considered: (H_1) the stochastic processes are of second order, (H_2) they are continuous in quadratic mean, and (H_3) their trajectories are square-

integrable functions. The corresponding definitions implicated by these assumptions on the stochastic processes are detailed below.

2.1.1. DEFINITIONS AND THEORETICAL ASSUMPTIONS

Let us consider a probability space $(\Omega, \mathcal{A}, \mathcal{P})$, and let $\mathcal{L}^2(\Omega)$ be the space of real random variables \mathcal{X} on Ω with finite second order moments, so that

$$E[|\mathcal{X}|^2] = \int_{\Omega} |\mathcal{X}(\omega)|^2 d\mathcal{P}(\omega) < \infty, \forall \mathcal{X} \in \mathcal{L}^2(\Omega). \quad (2-1)$$

Thus, the natural scalar product associated to the space $\mathcal{L}^2(\Omega)$ is given by the bilinear form

$$\begin{aligned} \mathcal{L}^2(\Omega) \times \mathcal{L}^2(\Omega) &\rightarrow \mathbb{R} \\ (\mathcal{X}, \mathcal{Y}) &\rightarrow E[\mathcal{X}\mathcal{Y}] = \int_{\Omega} \mathcal{X}(\omega)\mathcal{Y}(\omega)d\mathcal{P}(\omega), \end{aligned} \quad (2-2)$$

providing to $\mathcal{L}^2(\Omega)$ with a Hilbert space structure.

A stochastic process $\{\mathcal{X}(t): t \in \mathcal{T}\}$ defined on $\mathcal{T} \times \Omega$, with their observations given by the function

$$\begin{aligned} \mathcal{X}(\omega): \mathcal{T} &\rightarrow \mathbb{R} \\ t &\rightarrow \mathcal{X}(t, \omega) = x(t), \end{aligned} \quad (2-3)$$

is of second order if $\mathcal{X}(t) \in \mathcal{L}^2(\Omega)$, i.e. a real stochastic process.

Related to a second order stochastic process, the following functions can be defined:

- Mean function

$$\begin{aligned} \mu: \mathcal{T} &\rightarrow \mathbb{R} \\ t &\rightarrow \mu(t) = E[\mathcal{X}(t)] = \int_{\Omega} \mathcal{X}(t, \omega)d\mathcal{P}(\omega). \end{aligned} \quad (2-4)$$

- Covariance function

$$\begin{aligned} \mathcal{C}: \mathcal{T} \times \mathcal{T} &\rightarrow \mathbb{R} \\ (t, s) &\rightarrow \mathcal{C}(t, s), \end{aligned} \quad (2-5)$$

where

$$C(t, s) = \int_{\Omega} [(\mathcal{X}(t, \omega) - \mu(t))(\mathcal{X}(s, \omega) - \mu(s))] d\mathcal{P}(\omega). \quad (2-6)$$

Then, a second order stochastic process is continuous in quadratic mean if

$$\lim_{h \rightarrow 0} E [(\mathcal{X}(t+h) - \mathcal{X}(t))^2] = 0, \forall t \in \mathcal{T}, \quad (2-7)$$

which implies the continuity of the covariance function in $\mathcal{T} \times \mathcal{T}$.

On the other hand, the continuity in quadratic mean of the stochastic process it is not a sufficient condition for the continuity of their sample functions. Therefore, the Kolmogorov continuity theorem is considered, which implies that if a second order stochastic process is continuous in quadratic mean there exists another process stochastically equivalent, whose sample paths are square-integrable functions.

A stochastic process $\{\mathcal{X}(t): t \in \mathcal{T}\}$ verifying these three hypotheses:

H_1 The stochastic process is of second order,

H_2 The stochastic process is continuous in quadratic mean,

H_3 The trajectories of the stochastic process are square-integrable functions, i.e. belongs to the Hilbert space $\mathcal{L}^2(\mathcal{T})$,

may be seen as a random functional variable defined on $\mathcal{L}^2(\mathcal{T})$:

$$\begin{aligned} \mathcal{X}: \quad \Omega &\rightarrow \mathcal{L}^2(\mathcal{T}) \\ \omega &\rightarrow \mathcal{X}(\omega): \mathcal{T} \rightarrow \mathbb{R} \\ &\quad t \rightarrow \mathcal{X}(t, \omega), \end{aligned} \quad (2-8)$$

with an associated covariance operator defined as

$$\begin{aligned} \mathcal{C}: \quad \mathcal{L}^2(\mathcal{T}) &\rightarrow \mathcal{L}^2(\mathcal{T}) \\ f &\rightarrow \mathcal{C}(f), \end{aligned} \quad (2-9)$$

verifying that

$$[\mathcal{C}(f)](t) = \int_{\mathcal{T}} C(t, s) f(s) ds. \quad (2-10)$$

Let us observe that \mathcal{C} is a Hilbert-Schmidt operator, whose kernel is the covariance function $C(t, s)$ of the process. Furthermore, since $C(t, s)$

is a continuous function in $\mathcal{T} \times \mathcal{T}$, the covariance operator \mathcal{C} is both bounded and continuous in the Hilbert space $\mathcal{L}^2(\mathcal{T})$. The main properties of the covariance operator \mathcal{C} as defined here are [96]:

- i. \mathcal{C} is a compact operator on $\mathcal{L}^2(\mathcal{T})$.
- ii. \mathcal{C} is a self-adjoint operator (also called Hermitian operator), i.e. $\langle \mathcal{C}(f)|g \rangle = \langle f|\mathcal{C}(g) \rangle \forall f, g \in \mathcal{L}^2(\mathcal{T})$.
- iii. \mathcal{C} is a positive-definite operator, i.e. $\langle \mathcal{C}(f)|f \rangle \geq 0 \forall f \in \mathcal{L}^2(\mathcal{T})$.

The theoretical framework introduced until now paves the formulation for the statistical methods that will be used in further analysis of the spectral data. Therefore, in the remaining of this thesis, we will consider a functional variable \mathcal{X} , whose observations are realizations of a second order stochastic process $\mathcal{X} = \{\mathcal{X}(t): t \in \mathcal{T}\}$, continuous in quadratic mean, and with their sample functions belonging to the Hilbert space $\mathcal{L}^2(\mathcal{T})$ defined by

$$\mathcal{L}^2(\mathcal{T}) = \left\{ f: \mathcal{T} \rightarrow \mathbb{R}: \int_{\mathcal{T}} f^2(t) dt < \infty \right\}, \quad (2-11)$$

with the usual scalar product given by

$$\langle f, g \rangle = \int_{\mathcal{T}} f(t)g(t) dt \quad \forall f, g \in \mathcal{L}^2(\mathcal{T}). \quad (2-12)$$

2.2. MAIN SUMMARY STATISTICS FOR FUNCTIONAL DATA

As in the MDA, summary statistics such as sample mean, sample covariance, between others, can be defined for functional data [92, pp. 11–16] in the framework described above. Let us consider a random sample $\{x_i(t): t \in \mathcal{T}, i = 1, \dots, n\}$ of a functional variable \mathcal{X} as a set of independent and equally distributed realizations of a continuous second order stochastic process $\{\mathcal{X}(t): t \in \mathcal{T}\}$. Then, the sample mean function is defined as follows

$$E[\mathcal{X}(t)] = \bar{x}(t) = n^{-1} \sum_{i=1}^n x_i(t) \quad \forall t \in \mathcal{T}, \quad (2-13)$$

providing the mean value estimated for all sample functions at every $t \in \mathcal{T}$ (pointwise). Likewise, the variance function is given by

$$\text{var}[\mathcal{X}(t)] = (n - 1)^{-1} \sum_{i=1}^n [x_i(t) - \bar{x}(t)]^2, \quad (2-14)$$

and the square root of the variance function is the known standard deviation function.

Other widely used concept is the sample covariance function, which provides a measure of dependence of the sample curves across different argument values and is defined as follows

$$\hat{C}(t, s) = (n - 1)^{-1} \sum_{i=1}^n (x_i(t) - \bar{x}(t))(x_i(s) - \bar{x}(s)), \quad (2-15)$$

$$\forall t, s \in \mathcal{T},$$

and the associated correlation function is

$$\text{corr}[\mathcal{X}(t), \mathcal{X}(s)] = \frac{\hat{C}(t, s)}{\sqrt{\text{var}[\mathcal{X}(t)]\text{var}[\mathcal{X}(s)]}} \quad (2-16)$$

2.3. APPROXIMATION OF FUNCTIONAL DATA

Although during the last decades resolution for measuring systems has been considerable improved, natural phenomena (that can almost always considered as continuous in space and time) can only be recorded at discrete times and positions (digital acquisition), i.e. sample curves are observed (measured) at a finite number of points $\{x_{ij}: i = 1, \dots, n; j = 1, \dots, m_i\}$, being n the number of observations and m_i the total number of points in which the i -th sample curve has been measured $\{t_1, t_2, \dots, t_{m_i} \in \mathcal{T}\}$. Therefore, the first step in the FDA is converting the discrete measured values $(x_{i1}, x_{i2}, \dots, x_{im_i})$ for the observations (raw functional data) into a true functional form $x_i(t)$ of the sample curves. This leads to the first challenge for the researcher performing FDA, rising important questions about the raw data (measured data): Were data measured with error? How big is the error introduced by the measuring system? Is the measurement process expected to have a smooth or rough

response throughout \mathcal{T} ?. These questions are key for the researcher to make necessary and important decisions during the functional data approximation. As we shall see throughout this thesis, the smoothness degree in the approximated sample curves has a big influence in the statistical analysis and data interpretation. Therefore, the “smoothness” in the approximate sample curves should be defined by the researcher according to the goal of the experiment and his *a priori* knowledge about the nature of the measure.

For cases in which data is assumed to be measured without error:

$$x_{ij} = x_i(t_{ij}), i = 1, \dots, n, j = 1, \dots, m_i, \quad (2-17)$$

where x_{ij} is the j -th measured value for the i -th observation, and $x_i(t_{ij})$ is the true function observed at the sampling point t_{ij} , the functional data can be estimated using polynomial interpolation methods, among others [97]. In practice, however, and especially in applications with many external factors introducing noise to the measurement, such as in spectroscopy, it is mostly assumed that data is observed with an error. In this thesis, data analysis will be performed by assuming that measurement data includes some error, adding an error ϵ_{ij} term, representing noise, to Eq. (2-17) follows:

$$x_{ij} = x_i(t_{ij}) + \epsilon_{ij}, 1, \dots, n, j = 1, \dots, m_i. \quad (2-18)$$

In Figure 2-1, the interpolation (Eq. (2-17)) and smoothing (Eq. (2-18)) approaches are illustrated by fitting a set of random values simulated within the continuous interval $\mathcal{T} = [0,20]$. As can be seen, the interpolated curve (dashed blue line) cross exactly at the raw data points (measured values). The interpolation can lead to overfitting of data, and for spectroscopic techniques where usually some noise is introduced in measurements, the interpolated spectral data result in rough curves exhibiting strong frequency oscillations. This kind of approximations not only hinder the spectral data interpretation, but also the affects the robustness in regression analysis due to the included random noise in the fitted curves. On the other hand, the smoothing approach (green line) provides a flexible approximation of the raw data points that allows to reduce the noise contribution in measurements and highlight the main trends in the fitted curves.

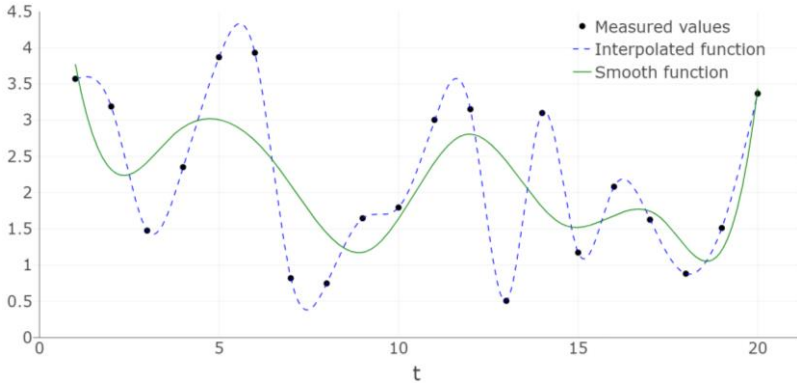


Figure 2-1. Illustration of an interpolated function vs a smooth function from a set of values discretely observed over continuum (raw data).

For both cases, interpolation and smoothing, the functional form can be estimated by considering the sample curves as linear combinations of a basis of functions, which is known as basis expansion methods.

2.3.1. BASIS EXPANSION

Let us consider a set of sample functions $\{x_i(t) : t \in \mathcal{T}, i = 1, \dots, n\}$, which are related to a functional variable \mathcal{X} , belonging to a finite dimension space generated by an orthogonal basis $\{\phi_1(t), \phi_2(t), \dots, \phi_p(t)\}$, in this case we can express $x_i(t)$ as follows:

$$x_i(t) = \sum_{k=1}^m a_{ik} \phi_k(t), t \in \mathcal{T}, i = 1, \dots, n, \quad (2-19)$$

where $x_i(t)$ is the estimated sample curve, a_{ik} are the basis coefficients and $\phi_k(t)$ are the basis functions considered for the basis expansion.

This is the fundamental equation for FDA but there are two further points that must be addressed: How to estimate the basis coefficients? and, which basis functions should we use? Regarding to the second question, basis functions should be considered according to the data characteristics. Then, once the appropriate basis is chosen, the simplest approach to estimate the basis coefficients is by Ordinary Least Squares (OLS), but we will come back to this later.

This thesis focuses to those approaches based on B-spline bases, since they provide with easy computability of polynomials with great flexibility and fit for changing local behaviour. B-splines will be addressed in more detail below, but a more complete review and comparison of existing methods for the approximation of smooth curves with B-spline bases can be found in [98]–[100]. Other basis functions commonly used in practice are the Fourier basis for periodic data, and wavelet bases for data with discontinuities or strong local behavior. For more detailed information about these basis functions, among others, the reader can refer to [92], [101].

2.3.1.1. B-SPLINE BASES

A B-spline bases of degree h and order $h + 1$ generates a space of splines with the same characteristics (order and degree). A spline of degree h is a function that is basically constructed by joining smoothly polynomials of the same degree end-to-end (piecewise polynomials) at a set of defined points $\tau_l \in \mathcal{T}$ called knots (also known as break points or nodes), so that the adjacent polynomials must match in their derivatives up to order $h - 1$.

Let us consider a set of knots $(\tau_0, \tau_1, \dots, \tau_{s-1}, \tau_s)$ such that $\tau_{l-1} < \tau_l$ for $l = 0, \dots, s$, such that the interval \mathcal{T} is divided into s subintervals, with the two outside knots (τ_0, τ_s) defining the interval \mathcal{T} and the internal knots $(\tau_1, \dots, \tau_{s-1})$ defining the s subintervals. Then, for an extended partition $\tau_{-3} < \tau_{-2} < \tau_{-1} < \tau_0 < \dots < \tau_s < \tau_{s+1} < \tau_{s+2} < \tau_{s+3}$ of \mathcal{T} , the B-spline basis of order $h + 1$ is defined iteratively defined by [102]

$$B_{l,h+1}(t) = \frac{t - \tau_{l-2}}{\tau_{l+h-2} - \tau_{l-2}} B_{l,h}(t) + \frac{\tau_{l+h-1} - t}{\tau_{l+h-1} - \tau_{l-1}} B_{l+1,h}(t) \quad (2-20)$$

$$h = 1, 2, \dots; l = -1, 0, \dots, s - h + 4,$$

with

$$B_{l,1}(t) = \begin{cases} 1 & \tau_{l-2} \leq t \leq \tau_{l-1}, l = -1, 0, 1, \dots, s + 4. \\ 0 & \text{in other case} \end{cases} \quad (2-21)$$

In Figure 2-2, two examples of B-spline bases defined on the continuum $[0,1]$ with a basis dimension equals 10 are shown: B-spline basis functions of order 2 (at the top), and a B-spline basis functions of

order 4 (at the bottom). The basis dimension and the number of knots of a B-spline bases are related by the following formula [102]

$$\text{number of basis functions} = \text{order} + \text{number of knots} - 2. \quad (2-22)$$

In particular, the Cubic B-spline bases provides us with the expected smoothness fit for our data. A cubic B-spline bases generates the space of splines of order 4, with piecewise polynomials of degree 3 joining up smoothly in a set of knots with continuity in their derivatives up to order 2. Let us observe that, according to Eq.(2-22), the dimension of the cubic B-spline bases equals to the number of internal knots plus 4 (order) or equivalently $s + 3$.

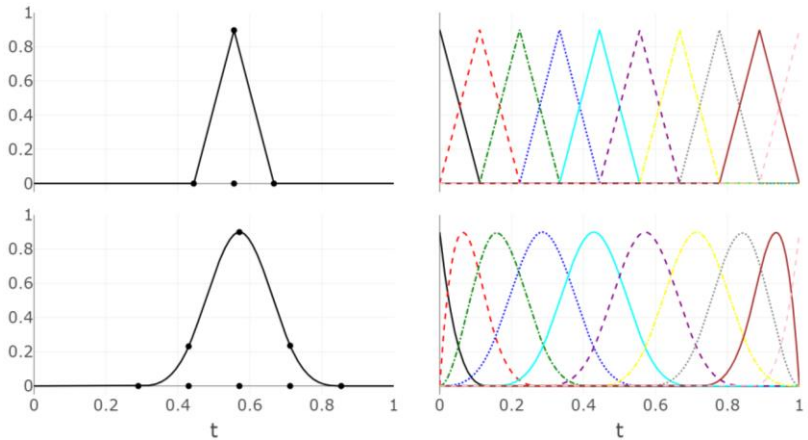


Figure 2-2. B-spline basis of order 2 and 4 are shown at the top and bottom, respectively.

2.3.2. REGRESSION SPLINES

As mentioned above, once the basis functions are defined, the simplest approach to estimate the basis coefficients \mathbf{a}_i of Eq.(2-19) is by least square criterion. For the case of B-spline bases, such approximation is known as Regression splines.

Let us rewrite Eq. (2-19) in its matrix form $x_i(t) = \mathbf{a}_i' \boldsymbol{\phi}(t)$, where $\mathbf{a}_i = (a_{i1}, \dots, a_{im})'$ are a vector containing the basis coefficients and $\boldsymbol{\phi}(t) = (\phi_1(t), \dots, \phi_m(t))'$ are the basis functions, which in our case,

will be a cubic B-spline bases. In the simplest way, the basis coefficients \mathbf{a}_i are obtained by OLS method, minimizing the Mean Squared Error (MSE) [103]

$$MSE(\mathbf{a}_i|x_i) = (\mathbf{x}_i - \Phi_i \mathbf{a}_i)'(\mathbf{x}_i - \Phi_i \mathbf{a}_i), \quad (2-23)$$

where $\mathbf{x}_i = (x_{i1}, \dots, x_{im_i})'$ are the measured values of the i -th observation, and $\Phi_i = (\phi_k(t_{ij}))_{m_i \times m}$ is a matrix comprising the basis functions observed at arguments t_{ij} . Thus, taking partial derivatives of Eq. (2-23) with respect to \mathbf{a}_i , and setting derivative equals to zero as follows

$$\frac{\partial}{\partial \mathbf{a}_i} (\mathbf{x}_i - \Phi_i \mathbf{a}_i)'(\mathbf{x}_i - \Phi_i \mathbf{a}_i) = 0, \quad (2-24)$$

the estimate of \mathbf{a}_i which minimizes the mean squared error is given by

$$\hat{\mathbf{a}}_i = (\Phi_i' \Phi_i)^{-1} \Phi_i' \mathbf{x}_i. \quad (2-25)$$

Then, the fitted values for the x_i measured values are given by the vectors

$$\hat{\mathbf{x}}_i = \Phi_i \hat{\mathbf{a}}_i = \Phi_i (\Phi_i' \Phi_i)^{-1} \Phi_i' \mathbf{x}_i, \quad (2-26)$$

and the fitted curves (functional form) are given by

$$\hat{x}_i(t) = \hat{\mathbf{a}}_i' \phi(t) \quad \forall i = 1, \dots, n. \quad (2-27)$$

An important observation working with the regression splines, is that smoothness of the fitted curve is directly determined by the number m of basis functions (dimension of the orthogonal basis), which is related to the number of knots by Eq. (2-22). The influence of the number of knots in the smoothness of the fitted curves is illustrated in Figure 2-3, which shows the fitted curves from a set of measured values using different number of knots. If the number of knots is too large, an overfitting of the data might occur, introducing noise in the fitted curve. In contrast, if too few knots are provided, relevant information could be omitted by the fitted curve, which is known as underfitting.

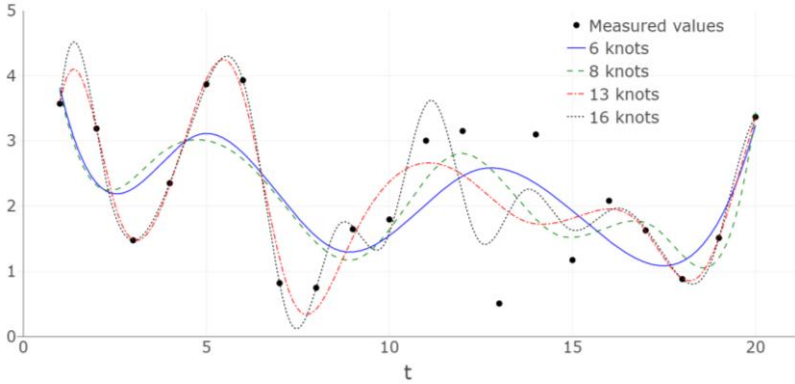


Figure 2-3. Fitted curves by regression splines using cubic B-spline basis defined on 6, 8, 15 and 18 equally spaced knots.

Therefore, defining the number of knots is a very important decision applying regression splines. In this sense, researcher must define the number of knots according the desired fit for the data under study. For cases in which researcher does not want to highlight a specific shape for estimated curves or has no prior knowledge about the data, the number of knots can be chosen by using the Leave-One-Out Cross Validation (LOOCV) method (see below).

2.3.2.1. LEAVE-ONE-OUT CROSS VALIDATION (LOOCV)

The LOOCV is a resampling method very helpful for choosing the ideal number of knots which best fits the data. The LOOCV provides with an error measure based on the estimation of the Root Mean Square Error (RMSE) for the fitted curves at each measured point t_j as follows

$$CV(q) = \frac{1}{n} \sum_{i=1}^n CV_i(q), \quad (2-28)$$

where

$$CV_i(q) = \sqrt{\sum_{j=0}^{m_i} (x_{ij} - \hat{x}_{ij}^{-j})^2 / (m_i + 1)}, \quad (2-29)$$

with q as the number of knots, m_i as the total number of points t_{ij} in which the i -th observation has been measured, and x_{ij}^{-j} being the values of the corresponding fitted curve for the i -th observation at points t_{ij} avoiding the j -th point in the iterative estimation process.

The LOOCV method is computationally expensive and for cases in which data is observed in a large set of points t_j might not be the best choice. This problem is alleviated by an alternative approach for sample curves approximation called Penalized splines (P-splines).

2.3.3. PENALIZED SPLINES (P-SPLINES)

The P-splines are an alternative approach to regression splines reducing the dependency between the number of knots and the smoothness of the fitted curve. The P-splines comes from a discrete penalty approach based on d -order differences between adjacent B-splines coefficients [104]. It is also important to note that for P-splines estimation, unlike the regression splines, the knots must be equally distributed along the continuum \mathcal{T} . The basis coefficients a_i for P-splines are estimated by adding a penalty term, involving a smoothing parameter, in the least square's formula shown in Eq. (2-23). Hence, the basis coefficients are computed to minimize the Discrete Penalized Mean Squares Error (DPMSE) as

$$DPMSE_d(a_i|x_i) = (x_i - \Phi_i a_i)'(x_i - \Phi_i a_i) + \lambda a_i' P_d a_i, \quad (2-30)$$

where λ is the smoothing parameter, and $P_d = (\Delta^d)' \Delta^d$ with Δ^d as the matrix of d -order differences. In practice, $d = 2$ is the most usual value, and then, the matrix of 2-order differences is given by [98]

$$\Delta^2 = \begin{pmatrix} 1 & -2 & 1 & 0 & 0 & \dots \\ 0 & 1 & -2 & 1 & 0 & \dots \\ 0 & 0 & 1 & -2 & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}_{(m-2) \times m}. \quad (2-31)$$

Finally, the basis coefficients a_i are estimated by

$$\hat{a}_i = (\Phi_i' \Phi_i + \lambda P_d)^{-1} \Phi_i' x_i. \quad (2-32)$$

Let us observe that the smoothness of the fitted curve according to Eq. (2-32) is controlled and directly proportional to the smoothing

parameter λ , and the number and location of the knots are not so determinant factors as in regression splines case. In practice, Ruppert's law of thumb is usually considered to select the number of knots: one knot for every 4 or 5 observations up to a maximum of 40 knots [105]. Nevertheless, working with a small sample size, the Ruppert's law it is not always the best way to define the number of knots. On the other hand, if we want to highlight characteristic shapes at specific points t_j , the flexibility provided by the regression splines for the arbitrary location of the knots, might represent a desired scenario. In this sense, the knots can be strategically fixed at the most contributing frequencies in terms of sample curves variability. Regarding the smoothing parameter, we will introduce the Generalized Cross Validation (GCV) method to find the optimal value of λ .

2.3.3.1. GENERALIZED CROSS VALIDATION (GCV)

The GCV method [106] will be considered in this thesis as selection criteria choosing the optimal value for the smoothing parameter λ . The GCV method estimate λ so that minimizes the following expression [98]

$$GCV(\lambda) = \frac{1}{n} \sum_{i=1}^n GCV_i(\lambda), \quad (2-33)$$

where

$$GCV_i(\lambda) = \frac{(m_i + 1)^{-1} MSE_i}{[(m_i + 1)^{-1} tr(I - H_i)]^2}, \quad (2-34)$$

$$MSE_i = \frac{1}{n} \sum_{j=0}^{m_i} (x_{ij} - \hat{x}_{ij})^2, \quad (2-35)$$

and $H_i = \Phi_i(\Phi_i' \Phi_i + \lambda P_d)^{-1} \Phi_i'$.

Both methods LOOCV and GCV can be used estimating the λ value, being the last one simpler in computational terms. A simulation study comparing both approaches was developed in [98, pp. 30–34]. The P-spline approach as well as LOOCV and GCV are useful when researcher does not seek to highlight any special feature or tend by the fitted curves.

2.4. FUNCTIONAL PRINCIPAL COMPONENT ANALYSIS

A challenge developing new applied spectroscopic techniques in biomedical engineering is that experimental tests usually involve small sample size due to high costs, limit access or difficult handling of the sample (in our case, mice). The sample size is an important factor to take into account in the statistical analysis, and for a small sample size regression models are not feasible. An alternative approach is to addressing this kind of studies as unsupervised learning problems, being the Blind Signal Separation (BSS) techniques such as Independent Components Analysis (ICA) [107] and Principal Component Analysis (PCA) [108], the most commonly used working with a set of mixed signals. The unsupervised techniques are characterized by providing a mathematical modelling of data without a priori information of the sample population.

The Functional Principal Component Analysis (FPCA) is the extended version of the well-suited multivariate PCA to the framework of the FDA [92, Ch. 6]. The FPCA provides an easy way of looking the main sources of variance contained in the sample curves by synthesizing such variability into a small set of uncorrelated functions, known as Functional Principal Components (FPCs). Obtaining a reduced set of uncorrelated variables that efficiently summarize the contained variance in original data allows for dimension reduction and can be used to prevent multicollinearity problems in regression analysis.

For the FPCA formulation, we will consider a set of sample functions $\{x_i(t): t \in \mathcal{T}, i = 1, \dots, n\}$ of a random functional variable \mathcal{X} , meeting the three hypothesis (H_1, H_2 and H_3) established in section 2.1.1, and it will be assumed without loss of generality that the sample functions are centered, implying that $\bar{x}(t) = 0$, i.e. the sample mean, defined in Eq. (2-13), equals 0. Working with the centered sample functions ensures that maximizing the variance of the principal component is equivalent to maximizing their sample variance. Then, the FPCs are estimated as generalized linear combinations of the sample curves, uncorrelated and with maximum variance. In general, the j -th principal component scores are given by

$$\xi_j = \int_{\mathcal{T}} x_i(t)\psi_j(t)dt, \quad (2-36)$$

where f_j is the weight function or loading associated to the j -th principal component. The loadings are computed such that maximize the most important modes of variation of the sample functions under orthonormality conditions as follows

$$\left\{ \begin{array}{l} \max_{\psi} \text{var} \left[\int_{\mathcal{T}} x_i(t)\psi(t)dt \right] \\ \text{s. t. } \|\psi\|^2 = 1 \text{ and } \langle \psi_j, \psi_l \rangle = 0, \forall j \neq l \end{array} \right., \quad (2-37)$$

with the estimated principal components scores ξ_j verifying that

$$E[\xi_j] = 0, \quad \text{var}[\xi_j] = \lambda_j, \quad \text{cov}[\xi_j, \xi_l] = 0. \quad (2-38)$$

As demonstrated in [109, pp. 15–19], the solution for Eq. (2-37) is given by the eigenfunctions associated to the spectral decomposition of the sample covariance operator, defined on Eq. (2-9). The Spectral Theorem provides a spectral decomposition of the sample covariance operator (compact, self-adjoint and positive linear operators), such that

$$[\mathcal{C}(f_j)](t) = \lambda_j \psi_j(t), \quad (2-39)$$

with $\{\lambda_j\}$ being a sequence of positive non-null numbers in decreasing order, known as eigenvalues, and $\{\psi_j\}$ being an orthonormal basis associated to the eigenvalues, known as eigenfunctions. Then, the estimated eigenvalues and eigenfunctions of \mathcal{C} , are the corresponding variances and loadings of the FPCs.

As will be seen further on, the scores of the FPCs can be used for clustering analysis on the sample curves either by scatterplots, in the simplest way, or using clustering algorithms based on partitioning or hierarchical methods, among others [110]. The data clustering occurs when some observations are somehow related, and such relation is not shared by all the observations in the population. The relations between resulting clusters using the scores of the FPCs can be studied by interpreting their corresponding loadings.

2.4.1. APPROXIMATION OF THE FPCA SOLUTION

The FPCA until now has been developed in terms of continuous functions, but it is well known that in practice, involved computations must be performed in matrix form. To this end, the FPCA will be estimated by expressing the loadings of the FPCs in terms of the basis functions employed in the basis expansion of the sample curves (see section 2.3). Thus, as shown in Eq. (2-19), the corresponding loading to the j -th FPC can be expressed as a linear combination of the basis functions

$$f_j(t) = \sum_{k=1}^m b_{jk} \phi_k(t), \quad (2-40)$$

where b_{jk} are the corresponding basis coefficients. Considering basis expansion of both sample curves Eq. (2-19) and the loadings Eq. (2-40) in their matrix form as shown in Eq. (2-27), the variance of the FPCs defined in Eq. (2-36) can be expressed as follows

$$\text{var}[\xi] = \text{var} \left[\int_{\mathcal{T}} x_i(t) f(t) dt \right] = b' \Psi V \Psi b, \quad (2-41)$$

where $V = n^{-1} A' A$, being $A = (a_{ij})_{n \times m}$ a matrix containing the basis coefficients of the sample curves and $\Psi_{m \times m} = \int_{\mathcal{T}} \phi_i(t) \phi_j(t) dt$ a matrix containing the inner products between the basis functions. Consequently, as formulated in [111], the FPCA problem defined in Eq. (2-37) is equivalent to the multivariate PCA problem on the matrix $A \Psi^{1/2}$ and the basis coefficients of the loading functions associated to the estimated FPCs are given by $F = \Psi^{1/2} U$, with F being the matrix comprising the basis coefficients of the loading functions and U is a matrix whose columns are the associated eigenvectors to the sample covariance matrix of $A \Psi^{1/2}$, estimated as follows:

$$n^{-1} \Psi^{1/2} A' A \Psi^{1/2} u = \lambda u. \quad (2-42)$$

See [111] for more details.

2.4.2. ORTHOGONAL REPRESENTATION OF A STOCHASTIC PROCESS

Other relevant result is that the theoretical framework behind the FPCA allows the applicability of the Karhunen-Loève (K-L) expansion presented in [96] as an harmonic analysis technique for functional data. The K-L expansion supports that the sample functions can be expressed in terms of the FPCs, and in fact, it can be realized using a truncated number q of the FPCs as follows:

$$\hat{x}_i(t) = \bar{x}(t) + \sum_{j=1}^q \xi_{ij} \psi_j(t). \quad (2-43)$$

where $\hat{x}_i(t)$ is the approximated function by the K-L expansion, and $\bar{x}(t)$ is the sample mean function (Eq. (2-13)). Such reconstruction will be the best linear approximation of the sample curves $x_i(t)$, in the least squares sense, and the total explained variance corresponds to the sum of the contributed variances by the q first FPCs used

$$\text{var}(x_i^q) = \sum_{i=1}^q \lambda_i. \quad (2-44)$$

Such K-L reconstruction is illustrated in Figure 2-4, where two interpolated sample curves, corresponding to two spectrums chosen arbitrarily from a spectral data set, are accurately approximated, as shown in Eq. (2-43), using the first two principal components accumulating a 99.69% of explained variance of the spectral data set.

The K-L reconstruction is an interesting and powerful statistical tool that can be used for generating (limitless) synthetic sample functions [112], [113]. In this way, new sample data can be simulated without the need to experimental replications. This approach can be implemented, from a moderate size sample, by fitting a probability model (distribution of probability) to the corresponding scores of the principal components used in the K-L reconstruction. Then, new score values can be simulated, considering the fitted probability model, to synthesize sample functions according to Eq. (2-43). In this sense, the generated curves will preserve the original variability captured by the FPCs from the measured sample data.



Figure 2-4. Two interpolated sample curves (black lines) with their corresponding K-L reconstruction using the first two principal components with an accumulative explained variance of 99.69%.

2.5. FUNCTIONAL LINEAR REGRESSION

As in the MDA framework, linear regression models have been constructed in the FDA framework (functional regression) for relating one or more variables (as covariates or predictors) to a response variable. The Functional Linear Model (FLM) can be generally categorized into three types of functional regression depending on the nature of the predictors or the response variable [94], [114]: modeling a functional response variable by functional predictors (Function-on-function regression), modeling a functional response variable by scalar predictors (Function-on-scalar regression), and modeling a scalar response variable by functional predictors (Scalar-on-function regression). The Scalar-on-function regression is the most common scenario in functional regression problems and can be addressed by linear, non-linear, and non-parametric approaches [115]. In particular, the supervised statistical analysis performed on the spectral data presented in this thesis is based on Scalar-on-function linear regression.

Let us consider a sample data consisting on an independent and identically distributed (iid) sample of random pairs $(x_i(t), y_i), i = 1, \dots, n$, where $\{x_i(t): t \in \mathcal{T}, i = 1, \dots, n\}$ are a set of observations of a random functional variable $\mathcal{X} = \{\mathcal{X}(t): t \in \mathcal{T}\}$, meeting the three hypothesis (H_1, H_2 and H_3) established in section 2.1.1, and $\{y_i: i = 1, \dots, n\}$ the associated sample of the continuous (Gaussian) scalar

response variable Y . Then, the FLM for Scalar-on-function regression is expressed as

$$y_i = \alpha + \int_{\mathcal{T}} x_i(t)\beta(t)dt + \epsilon_i, \quad i = 1, \dots, n, \quad (2-45)$$

where α is a scalar intercept, $\beta(t)$ is the functional parameter of the regression model, and ϵ_i are iid errors with zero mean. As observed in [115], the functional parameter β can be naturally interpreted, being the locations t with largest $|\beta(t)|$ the most influential to the response variable. The FLM is also extended for the case of non-Gaussian response variables, known as Generalized Functional Linear Model (GFLM) [116], relating the functional predictor and the response variable by using a link function $g(\cdot)$, modelled as follows

$$g\{E(y_i)\} = \alpha + \int_{\mathcal{T}} x_i(t)\beta(t)dt. \quad (2-46)$$

A classic example of a link function $g(\cdot)$ is the logistic link commonly used modelling a Bernoulli variable (binary response variable) by the Functional Logit Model. The Functional Logit Model will be addressed in more detail below.

In order to estimate the FLM, a common general approach is assuming the basis expansions, as defined in Eq. (2-19), of both the sample functions $x_i(t)$ and the functional parameter $\beta(t)$ in Eq. (2-45) such that

$$\hat{Y} = \alpha + \int_{\mathcal{T}} \mathcal{X}(t)\hat{\beta}(t)dt = \alpha\mathbf{1} + A\Psi b, \quad (2-47)$$

where $\mathbf{1}$ is a unit vector of length n , $A_{n \times m}$ is the matrix containing the basis coefficients of the sample functions, $\Psi_{m \times m}$ a matrix containing the inner products between the basis functions, and $b_{m \times 1}$ is a vector containing the basis coefficients of the functional parameter. Let us observe here that the number of basis functions used in the basis expansion of sample functions and the functional parameter do not necessarily need to be the same. Then, Eq. (2-47) implies that the FLM can be reduced in a standard multiple regression problem with $[\mathbf{1} \ A\Psi]$ as the design matrix and $(\alpha, b_1, \dots, b_m)$ as the regression parameters.

2.5.1. FUNCTIONAL LOGIT MODEL

A very common approach developing new spectroscopic techniques for medical diagnosis is the statistical classification problem. This kind of problems are widely modeled by logistic regression in many research fields such as sociology [117]–[119], finance [120]–[122] and medicine [123]–[126], since allows to model the probability of occurrence of an event from a set of predictors, and the predictors contribution can be quantitatively studied in terms of the odds ratio [127], as will be shown later.

The Functional Logistic Regression (FLoR) is the extended version of the multivariate logit model in the FDA context, and it is a particular case of the GFLM in which the sample functions $x_i(t)$ are related to a qualitative variable with two possible outcomes $y_i \in \{0,1\}$, with 1 indicating the occurrence of the event. Let us consider, without loss of generality, a set of sample functions $\{x_i(t): t \in \mathcal{T}, i = 1, \dots, n\}$ of a centered random functional variable \mathcal{X} , and a binary scalar response $\{y_i: i = 1, \dots, n\}$ associated to them. Then, the FLOR model is formulated as follows

$$y_i = \pi_i + \epsilon_i, \quad i = 1, \dots, n, \quad (2-48)$$

where $y_i \in \{0,1\}$ is the corresponding outcome to the i -th observation $x_i(t)$ of the centered functional variable $\mathcal{X}(t)$, ϵ_i are iid errors with zero mean, and π_i is the expectation of Y given $x_i(t)$ modelled as

$$\begin{aligned} \pi_i &= P[Y = 1 | \{x_i(t): t \in \mathcal{T}\}] \\ &= \frac{\exp\{\alpha + \int_{\mathcal{T}} x_i(t)\beta(t) dt\}}{1 + \exp\{\alpha + \int_{\mathcal{T}} x_i(t)\beta(t) dt\}}, i = 1, \dots, n, \end{aligned} \quad (2-49)$$

with α a real parameter (intercept) and $\beta(t)$ a functional parameter. Equivalently, considering Eq. (2-49), the FLOR model can be seen as a FGLM with the logit transformation as the link function as follows

$$\begin{aligned} l_i &= \ln \left[\frac{\pi_i}{1 - \pi_i} \right] \\ &= \alpha + \int_{\mathcal{T}} x_i(t)\beta(t) dt, \quad i = 1, \dots, n. \end{aligned} \quad (2-50)$$

2.5.1.1. FLOR MODEL ESTIMATION BY FPCs

As stated earlier, a common approach estimating Eq. (2-50) considers the basis representation of the sample functions $x_i(t)$ and the functional parameter $\beta(t)$, resulting in standard multivariate regression on the design matrix $[1 A\Psi]$, however, the way in which the matrix $A\Psi$ is obtained leads to multicollinearity problems in the FLOR [128]. This problem is alleviated by using the K-L expansion (see Eq. (2-43)) for the sample functions and estimating the functional parameter in terms of the loadings associated to the FPCs. This alternative approach, proposed in [128] as Functional Principal Component Logistic Regression (FPCLoR), is based on introducing a set of FPCs as predictors for the FLOR model estimation. Considering a reduced number of FPCs, estimated according to Eq. (2-36) and Eq. (2-37), as uncorrelated predictors not only prevents multicollinearity between covariates of the FLOR model but also reduces the dimensionality of the regression problem.

In the FPCLoR model, Eq. (2-50) can be equivalently expressed in terms of a multivariate logit model with the scores of the first q FPCs as predictors as follows

$$l_i = \alpha + \sum_{j=1}^q \xi_{ij}\gamma_j, \quad i = 1, \dots, n, \quad (2-51)$$

with its matrix form given by

$$L = \alpha\mathbf{1} + \Gamma\gamma, \quad (2-52)$$

where $L = (l_1, \dots, l_n)$, $\mathbf{1}$ is a unit vector of length n , α is the intercept, $\Gamma = (\xi_{ij})_{n \times q}$ is a matrix comprising the columns of the first q FPCs, and γ is the vector containing the model coefficients.

Regarding the estimation of the functional parameter $\beta(t)$, let us replace ξ_{ij} in Eq. (2-51) by Eq. (2-36) as follows

$$\begin{aligned}
l_i &= \alpha + \sum_{j=1}^q \left(\int_{\mathcal{T}} x_i(t) \psi_j(t) dt \right) \gamma_j \\
&= \alpha + \int_{\mathcal{T}} x_i(t) \left(\sum_{j=1}^q \psi_j(t) \gamma_j \right) dt, \quad i = 1, \dots, n,
\end{aligned} \tag{2-53}$$

and according to Eq. (2-50), an expression of $\beta(t)$ is obtained in Eq. (2-53), in terms of the loading functions $\psi_j(t)$ associated to the first q FPCs, and the model coefficients γ_j . Then, considering the basis expansion of $\beta(t)$ and $f_j(t)$, the basis coefficients $\beta = (c_1, \dots, c_p)'$ of the basis expansion of the functional parameter can be estimated by $\hat{\beta} = F_{(m \times q)} \mathcal{Y}_{(q \times 1)}$, where F is a matrix whose columns correspond to the basis coefficients of the loading functions associated to the first q FPCs (see 2.4.1).

The number q of the FPCs to be taken into account in the FPCLoR model estimation is a fundamental question addressed in [128], with the variance of the estimated functional parameter as the criterium for model selection. In [98, Ch. 3], a double-GCV procedure for selection of optimum number q is presented and studied for different estimation approaches of the FPCLoR model.

2.5.1.2. FUNCTIONAL PARAMETER INTERPRETATION

A significant additional contribution of the FLoR model, compared to other regression models, is the interpretation of the functional parameter $\beta(t)$, which is also achieved in the FPCLoR model. The estimated functional parameter represents the discriminating relation between the response variable Y and the functional predictor $\mathcal{X}(t)$, providing a qualitative interpretation of such relation. Furthermore, as mentioned above, such relation can be also quantitatively studied in terms of the odds ratio [127]. The odds of a determined outcome “A” is interpreted as the chances of A occurs and is defined as the ratio between the probability of occurrence of $\mathcal{P}(A)$ and its complement $1 - \mathcal{P}(A)$, i.e. the probability of absence of A. Then, the odds ratio is used as a measure of the relationship between the odds of two outcomes [129], and can be used to quantify the influence of a factor associated to the outcome A, e.g.

considering that A is the probability of being diagnosed as diabetic and the sugar consumption is a determining factor (predictor), then, the odds ratio can be used to compare the odds of being diabetic according to the sugar consumption.

Let us consider the logit transformation l_i for a specific sample observation $x_i(t)$ as shown in Eq. (2-50), and l_i^* be the resulting logit transformation for the same $x_i(t)$ increased in a constant way by a factor K within a period $[t_0, t_0+h] \subseteq \mathcal{T}$ as follows $x_i(t) \rightarrow x_i^*(t) = x_i(t) + I_{\Delta(t)=h}(t) \cdot K$, with

$$I_{\Delta(t)=h}(t) = \begin{cases} 1 & \text{si } t \in [t_0, t_0+h] \\ 0 & \text{other wise} \end{cases}, \quad (2-54)$$

such that

$$l_i^* = \alpha + \int_{\mathcal{T}} [x_i(t) + (I_{\Delta(t)=h}(t) \cdot K)] \beta(t) dt. \quad (2-55)$$

Then, the odds ratio for l_i^* and l_i is given by

$$\ln \left[\frac{\frac{\pi_i^*}{1 - \pi_i^*}}{\frac{\pi_i}{1 - \pi_i}} \right] = \int_{t_0}^{t_0+h} K \cdot \beta(t) dt. \quad (2-56)$$

This means, that a constant increment in K units in a fixed interval for $x_i(t)$ increases the odds of $y = 1$ against $y = 0$ by a factor of the same magnitude. In spectroscopy applications, the factor K may concern to a variation in the response of the optical properties such as reflection and transmission of the sample under study either in the whole measured frequency band or at specific sub frequency intervals. The interpretation of the functional parameter of the FPCLoR model $\beta(t)$ in terms of the odds ratio is very useful since allows to quantitatively study the impact of the measured spectral response in diagnosis of a disease when developing new biomedical spectroscopic techniques. Furthermore, interpretation can be used to detect major contributing frequencies for diagnosis since can be estimated for specific frequency intervals within the measured frequency band.

As it can be inferred, the smoothness of the fitted sample curves, addressed in section 2.3, can have important effects on the regression relation, and consequently, influences the interpretability of the FPCLoR model. This issue will be discussed later during the spectral data analysis.

2.6. LONGITUDINAL FUNCTIONAL DATA ANALYSIS

So far, only statistical methods for analysis of functional data involving independent sample functions has been introduced. The functional data becomes structurally dependent between observations when sample population is repeatedly observed during the study. This scenario naturally emerges in longitudinal studies, based on spectroscopic techniques, in which the sample population is repeatedly assessed at different times over a period. A common approach in longitudinal studies is to understand better the evolution of chronic diseases, their risk factors, and find indicators for their progress prediction. Some examples of the applications of longitudinal studies are Alzheimer's disease [130], [131], Diabetes Mellitus [132], [133] and Multiple Sclerosis [134], [135]. In this sense, the longitudinal spectral data usually consist of profiles or images collected on the same subjects over several visits.

Traditionally, the longitudinal spectral data is analyzed by combining summary statistics and mixed effects models [136]–[138] but this strategy may miss important sources of variability within subjects profiles or images. Similarly, the FDA methods described above ignores the longitudinal structure in the collected spectral data from longitudinal studies. These two aspects have led to the development of novel FDA methods for Longitudinal Functional Data Analysis (LFDA).

In the LFDA, the longitudinal spectral data is addressed as longitudinal functional data by considering the time variable associated to the sample functions. The LFDA focuses not only on the study possible of relations between the measured spectral response and the target pathology, but also on the variations or trends of the measured spectral response within dependent observations, that may be related to different states of the target pathology at the evaluation times.

In this section, some statistical methods for LFDA will be briefly described. Firstly, a variance decomposition method for longitudinal functional data, called Longitudinal Functional Principal Component Analysis (LFPCA), based on mixed effects model and eigenanalysis is

introduced. Then, different approaches for longitudinal scalar-on-function regression, in which the scalar response variable and the functional predictor are measured longitudinally, will be mentioned. Additionally, the regression approaches will allow to consider both functional and scalar predictors simultaneously.

2.6.1. LONGITUDINAL FPCA (LFPCA)

As can be deduced, the LFPCA extends the FPCA to the LFDA. The main idea behind the LFPCA, introduced in [139], is to extract the main differences between subjects' average profiles (between-subjects variation) and the subjects profiles evolution over time.

Let us consider a set of sample functions $\{x_{ij}(t): t \in \mathcal{T}, i = 1, \dots, n, j = 1, \dots, J_i\}$, with n as the total of subjects and J_i the number of visits recorded for the i -th subject. Let us observe that LFPCA addresses the case in which longitudinal functional data involves some subjects with at least three visits, and cases in which $J_i \leq 2$ were fully addressed by the multi-level case [140], [141]. A significant advantage of LFPCA working with experimental measurements such as clinical trials is that subjects has not to be necessarily assessed at the same number of visits or at the same visits (unbalanced data). Other advantage is that LFPCA allows to work with sample curves that have missing values.

In order to estimate the LFPCA, the longitudinal functional data is modelled by a functional random intercept and random slope model [139] as follows

$$x_{ij}(t) = \eta(t, T_{ij}) + B_{i,0}(t) + B_{i,1}(t)T_{ij} + U_{ij}(t) + \varepsilon_{ij}(t), \quad (2-57)$$

where $\eta(t, T_{ij})$ is an overall fixed mean surface, T_{ij} is the standardized time of visit j for subject i , $B_{i,0}(t)$ and $B_{i,1}(t)$ are the functional intercept and functional slope for subject i , respectively, $U_{ij}(t)$ is a visit-specific functional deviation from the functional trend of subject i , and $\varepsilon_{ij}(t)$ is white noise error with variance σ^2 . In addition, $B_i(t) = \{B_{i,0}(t), B_{i,1}(t)\}$, $U_{ij}(t)$ and $\varepsilon_{ij}(t)$ are assumed to be centered square-integrable and mutually uncorrelated random processes on \mathcal{T} . Thus, the bivariate process $B_i(t)$ captures the between-subjects variation, $U_{ij}(t)$

captures the within-subject variation among visits and $\varepsilon_{ij}(t)$ captures random uncorrelated variation within each sample curve.

Then, the principal components scores are estimated as zero-mean uncorrelated random variables using the eigen decomposition of the corresponding covariance operators \mathcal{C}_U and \mathcal{C}_B to the random processes $U(t)$ and $B(t)$ respectively. In general, the principal component scores related to the between-subjects variability, captured by the $B(t)$ process are given by

$$\xi_{ik} = \int_{\mathcal{T}} B_{i,0}(t)\psi_k^0(t)dt + \int_{\mathcal{T}} B_{i,1}(t)\psi_k^1(t)dt, \quad (2-58)$$

where $\psi_k^B(t) = \{\psi_k^0(t), \psi_k^1(t)\}'$ are the ordered eigen functions of \mathcal{C}_B , corresponding to the non-null decreasing eigenvalues $\{\lambda_k\}$. Similarly, the principal component scores related to the within-subject variability, captured by the $U(t)$ process are given by

$$\zeta_{ijk} = \int_{\mathcal{T}} U_{ij}(t)\psi_k^U(t)dt, \quad (2-59)$$

where $\psi_k^U(t)$ are the eigen functions of \mathcal{C}_U corresponding to the non-null decreasing eigenvalues $\{v_k\}$.

Similarly to the FPCA, the functions $\psi_k^B(t) = \{\psi_k^0(t), \psi_k^1(t)\}'$ and $\psi_k^U(t)$ can be seen as the loadings of the principal components and correspond to the largest variation modes in processes $B(t)$ and $U(t)$, respectively. The estimated subject-specific scores ξ_{ik} and ζ_{ijk} allow to study the possible relations between the measured spectral and the status of the target pathology over visits or other collected variables during the clinical trial.

Estimates of the principal component scores ξ_{ik} and ζ_{ijk} are obtained by considering finite-dimensional approximations of the $B(t)$ and $U(t)$ processes (K-L decomposition) as follows

$$\begin{aligned}
B_{i,0}(t) &= \sum_{k=1}^{N_B} \xi_{ik} \psi_k^0(t), \\
B_{i,1}(t) &= \sum_{k=1}^{N_B} \xi_{ik} \psi_k^1(t), \\
U_{ij}(t) &= \sum_{k=1}^{N_U} \zeta_{ijk} \psi_k^U(t).
\end{aligned} \tag{2-60}$$

Thus, the functional random intercept and random slope model shown in Eq. (2-57) can be expressed as a linear mixed model as follows

$$x_{ij}(t) \approx \eta(t, T_{ij}) + \sum_{k=1}^{N_B} \xi_{ik} V'_{ij} \psi_k^B(t) + \sum_{k=1}^{N_U} \zeta_{ijk} \psi_k^U(t) + \varepsilon_{ij}(t), \tag{2-61}$$

with $V_{ij} = (\mathbf{1}, T_{ij})'$.

The number of principal components N_B and N_U used in Eq. (2-60) are chosen by considering the total proportion of explained variability captured by the K-L reconstruction. The explained variance can be interpreted in terms of the estimated eigen values $\{\hat{\lambda}_k\}$ and $\{\hat{\nu}_k\}$, only if the time variable T_{ij} is standardized to have zero mean and unit variance as follows

$$\frac{\sum_{k=1}^{N_B} \hat{\lambda}_k + \sum_{k=1}^{N_U} \hat{\nu}_k + \hat{\sigma}^2}{\int_0^1 \text{var}[x_{ij}(t)] dt}. \tag{2-62}$$

For simplicity, and as stated in [139], the LFPCA approximation process for principal component scores estimation according to Eq. (2-61) is summarized in five steps:

1. The overall fixed mean surface $\eta(t, T_{ij})$ is estimated using a bivariate smoother in t and T under the independence assumption $x_{ij}(t) = \eta(t, T_{ij}) + \varepsilon_{ij}(t)$ considering P-splines smoothing with Restricted Maximum Likelihood (REML) estimating for the smoothing parameter [142].

2. The covariance functions associated to the random processes $B_i(t) = \{B_{i,0}(t), B_{i,1}(t)\}$ and $U_{ij}(t)$ are estimated from the residuals $x_{ij}(t) - \hat{\eta}(t, T_{ij})$ using linear multiple regression on the cross-products $x_{ij}(t)x_{ij'}(s)$ with interaction effects to consider the time variable T .
3. Bivariate smoothing is applied to the “raw” covariance functions estimated from previous step and, the estimate of σ^2 is based on the residuals between the raw and smoothed covariance functions from U process.
4. Spectral decomposition of smoothed \mathcal{C}_U and \mathcal{C}_B are obtained by eigenanalysis, and their estimated eigenfunctions are used as the basis functions for K-L reconstructions of $B_{i,0}(t)$, $B_{i,1}(t)$ and $U_{ij}(t)$ processes as shown in Eq. (2-60).
5. Finally, the principal component scores ξ_{ik} and ζ_{ijk} are estimated as the Best Linear Unbiased Predictors (BLUPs) of the linear mixed model shown in Eq. (2-61).

For more details about the LFPCA theory and estimation, the reader can refer to [139].

2.6.2. LONGITUDINAL SCALAR-ON-FUNCTION REGRESSION ANALYSIS

In this section, two novel models for longitudinal scalar-on-functions regression will be briefly introduced: Longitudinal Penalized Functional Regression (LPFR) and Longitudinal Functional Principal Components Regression (LFPCR), with two different approaches for the LFPCR model. These models address the regression problem in which a scalar variable is related to a functional variable, and both are measured longitudinally. Additionally, it allows to consider more than one functional predictor, that can be observed on different domains, and multivariate scalar predictors can be included in the model’s estimation. This is a very useful and interesting approach analyzing data from clinical studies in which both spectral data and clinical variables can be related to the response variable. The LPFR and LFPCR models estimation is based on the mixed-model framework [143]–[146].

Let us consider a sample data set with following structure $[y_{ij}, x_{ij1}, x_{ij2}, \dots, x_{ijK}, W_{ij}]$, $i = 1, \dots, n$, $j = 1, \dots, J_i$, where n is the total of subjects, J_i the total of visits recorded for subject i , $x_{ijl}(t) \in \mathcal{L}^2[0,1]$, $1 \leq l \leq K$ are functional predictors (not necessarily observed over the same domain), W_{ij} is a row vector of scalar predictors, and y_{ij} the associated scalar response. Then, the LPFR model [147] is given by

$$g\{E(y_{ij})\} = W_{ij}\gamma + Z_{ij}b_i + \sum_{l=1}^K \int_0^1 x_{ijl}(t)\beta_l(t)dt \quad (2-63)$$

where γ is a vector containing the standard fixed-effects coefficients related to the scalar predictors W_{ij} , Z_{ij} is a vector modelling random intercepts (accounting for repeated observations), b_i is a vector containing the standard random-effects coefficients, and β_l are the functional coefficients related to the functional predictors. Thus, the scalar coefficients γ and the functional coefficients β_l are fixed effects that do not vary among visits and the subject-specific effects are modelled by the component $Z_{ij}b_i$.

The approximation process for LPFR model, shown in Eq. (2-63), is summarized in two steps:

1. K-L reconstruction of the functional predictors is obtained by using the spectral decomposition of their corresponding covariance operators \mathcal{C}_l as follows

$$x_{ijl}(t) = \mu_l(t) + \sum_{k=1}^{N_x} \xi_{ijkl} \psi_{lk}(t)dt, \quad (2-64)$$

where $\mu_l(t)$ is the mean function estimated for the l -th functional predictor, $\{\psi_{l1}(t), \dots, \psi_{lN_x}(t)\}$ are the eigenfunctions corresponding to the spectral decomposition of \mathcal{C}_l with associated eigenvalues $\lambda_{l1} \geq \lambda_{l2} \geq \dots \geq \lambda_{lN_x}$, and $a_{ijkl} = \int_0^1 (x_{ijl}(t) - \mu_l(t))\psi_{lk}(t)dt$ are the principal component scores corresponding to the l -th functional predictor.

2. Then, basis expansions (see section 2.3.1) of the functional coefficients $\beta_l(t)$ using spline basis $\{\phi_{l1}(t), \phi_{l2}(t), \dots, \phi_{lN_\beta}(t)\}$ are considered, such that

$$\beta_l(t) = \phi'_l(t)c_l, \quad (2-65)$$

$$\text{with } c_l = (c_{l1}, \dots, c_{lN_\beta})'.$$

Thus, the functional term in Eq. (2-63) can be rewritten as follows

$$\begin{aligned} \int_0^1 x_{ijl}(t)\beta_l(t)dt &= \alpha_l + \Gamma_l \left[\int_0^1 \psi_l(t)\phi'_l(t)dt \right] c_l \\ &= \alpha_l + \Gamma_l \Psi_l c_l \end{aligned} \quad (2-66)$$

where $\alpha_l = \int_0^1 \mu_l(t)\beta_l(t)dt$, $\Gamma_l = (\xi_{ijl})_{I \times N_x}$ with $I = \sum_{i=1}^n J_i$ is a matrix comprising the principal component scores for all observations, I corresponding to the l -th functional predictor, and Ψ_l is a $(N_x \times N_\beta)$ -dimensional matrix containing the inner products between the eigenfunctions ψ_l and the basis functions ϕ_l .

Thus, the LPFR model can be estimated by using standard mixed effects models, with the terms α_l incorporated in the overall model intercept. For more details the reader can refer to [147].

As can be seen in Eq. (2-64), the LPFR model basically decompose the functional predictors $\mathcal{X}_l(t)$, as in the FPCA case, without consider the longitudinal information associated to the sample functions $x_{ijl}(t)$ that may lead to omit relevant sources of variation. Therefore, two alternative approaches for LFPCR were proposed in [148] based on LFPCA for variability decomposition of the functional predictors.

The first approach, which is the more intuitive case, directly uses the functional principal components scores estimated by LFPCA, with the LFPCR model given by

$$g\{E(y_{ij})\} = \alpha(T_{ij}) + W_{ij}\gamma + Z_{ij}b_i + \sum_l^K \Gamma_l^B \theta_l + \sum_l^K \Gamma_l^U \delta_l, \quad (2-67)$$

where components $W_{ij}\gamma$ and $Z_{ij}b_i$ are the same introduced in Eq. (2-63), Γ_l^B and Γ_l^U are the principal components scores estimated for $B(t)$ and

$U(t)$ processes from the l -th functional predictor, and $\alpha(T_{ij})$ is a time-varying intercept estimated by considering $\eta(t, T_{ij})$ (see (2-57)).

The second approach directly replaces the components in Eq. (2-67), associated to the principal component scores, with the $B(t)$ and $U(t)$ processes as follows

$$\begin{aligned} \sum_l^K \Gamma_l^B \theta_l &\Rightarrow \sum_l^K \int U_{ijl}(t) \beta_l^U(t) \\ \sum_l^K \Gamma_l^U \delta_l &\Rightarrow \sum_l^K B_{il}(t) \beta_l^B(t) \end{aligned}, \quad (2-68)$$

with $B_i(t) = \{B_{i,0}(t), B_{i,1}(t)\}$ and $U_{ij}(t)$ processes obtained by K-L reconstruction, as shown in Eq. (2-60), and the functional coefficients $\beta_l^U(t)$ and $\beta_l^B(t)$ estimated using P-splines expansions with the smoothing parameter estimated via REML.

The LFPCR approaches are studied and compared together with the LPFR model in [148]. One of the LFPCR model advantages over the LPFR model is that it adds the natural interpretation provided by the LFPCA on the functional predictors. Other advantage is that LFPCR model can be applied when the response variable does not change from visit to visit, for instance, when y_{ij} indicates the presence or absence of a disease that does not change among visits. Both LFPCR models can be estimated analogously to the LPFR model using the generalized additive mixed models. For more details see [147], [148].

3. MILLIMETER-WAVE
SPECTROSCOPY FOR NON-
INVASIVE SENSING OF
SUSTAINED
HYPERGLYCEMIA

As briefly introduced in the first chapter, applied spectroscopy takes advantage of the spectral fingerprint of molecules to characterize the spectral response of one or more targeted components in a sample under study. In particular, such characterization is a hard task in high-complexity, multi-species, scenarios where there is no a-priori information about the sample or its constituents in terms of their spectral response (absorption, reflexion and transmission properties), and accurate mathematical modelling is practical impossible due to their so complex integration. This kind of scenarios are commonly found when developing applied spectroscopic techniques in biomedical engineering in which the spectral interrogation is performed on biological samples with numerous metabolites and substances resonating at different partially overlapping frequencies when probed with electromagnetic waves. Additionally, there are always interferences associated to natural physiological process (body temperature, body fluid shifts and transpiration, among others) affecting adversely the spectral response.

The superposition of the spectral features from metabolites and other substances such as water and fat, besides several interferences such as physiological processes, environmental factors, and intrinsic spectroscopic instrument noise, make it very difficult to isolate and quantify (i.e. to estimate the concentration) a specific metabolite within the biological sample accurately, especially in in-vivo applications. Therefore, a non-targeted spectral profiling approach is proposed in this thesis for the evaluation of applied spectroscopic techniques in biomedical engineering, being particularly useful at first stages of the development process. As will be seen, a non-targeted spectral profiling approach can have great potential to explore and evaluate the applicability of a spectroscopic technique within a frequency range in which the target sample has not been characterized, as often happens in biomedical engineering.

Thus, in this work, and similarly to the principle applied in non-targeted metabolomics (see [149]–[151]), the biological sample is to be interrogated throughout a widely frequency range of interest, and the collected spectral data is to be analysed to study possible relations between

the measured spectral response and a targeted clinical pathology (comparing the spectra associated to the target pathology with control cases), without the necessity to quantify an specific metabolite involved in the biological sample. This approach is exploited under the FDA framework, that allows to address the spectral profiles as single entities, highlighting their main trends, and provides flexible handling of undesired factors contributions (noise) in the measured spectral response.

Under this two-fold approach (non-targeted spectral profiling combined to FDA), in this thesis, a mm-wave spectroscopic technique is proposed and experimentally assessed as a novel non-invasive approach for in-vivo detection and monitoring of DM. Basically, the proposed system interrogates the biological sample with a tunable source spanning the W-band (75 GHz – 110 GHz) using a Continuous Wave (CW) spectrometer. Thus, a W-band spectral profile is obtained from the biological sample under study involving the spectral features of all components of the biological sample that will be processed, analyzed, and interpreted by using FDA methods.

In this chapter, we briefly describe the current state-of-the-art in DM and sustained hyperglycemia non-invasive detection. Subsequently some advantages of the use of THz and mm-wave radiation for biomedical engineering are described and, finally, the proposed mm-wave spectroscopic instrument for in-vivo and non-invasive, detection and monitoring of sustained hyperglycemic metabolism is described in detail.

3.1. DIABETES MELLITUS: TECHNOLOGICAL ADVANCES AND CHALLENGES FOR NON-INVASIVE SENSING OF BLOOD GLUCOSE LEVEL

DM is a very complicated metabolic disorder affecting a great part of the world population, with almost 500 millions of cases around the world and an expected rate of increase of 51% in the years to come [152]. This chronic disease encompasses a group of complex metabolic conditions characterized by the continuous presence of high blood glucose levels (BGL), known as sustained hyperglycemia, due to the inability of diabetic patients to produce or use insulin adequately [153]. So far, there is not any kind of cure for DM, and diabetic patients follow-up

is based on the continuous monitoring of BGL. It is well known that medium/long-term exposure to sustained hyperglycemia on the body leads to the formation of Advanced Glycation End-products (AGEs) [154], [155], resulting from non-enzymatic chemical bonding of free sugars to proteins or lipids (glycation), which plays an important role in the development of several and irreversible physiological consequences such as early aging, cardiomyopathy, nephropathy, retinopathy, and neuropathy, among others [156]–[161]. The AGEs are irreversible compounds that accumulate in the body, but the generation of these metabolomic end-products go through two previous main stages: the Schiff base (early stage) and Amadori products (intermediate stage), being both reversible steps in the AGEs formation process [154]. Therefore, sustained hyperglycemia detection in its earliest stages is critical not only for DM diagnostics, but also for metabolomic control and supervision of the patients, facilitating an early intervention to regulate the carbohydrate metabolism, and consequently, reducing health complication risks.

Nowadays, the well-established methods used for diagnosis and follow-up of DM are invasive (blood samples are required) based on enzymatic reactions [162]. Some common procedures performed at hospitals include Fasting Plasma Glucose (FPG) measurements, Oral Glucose Tolerance Tests (OGTT), and glycohemoglobin tests (A1c or HbA1c) [163]–[166], that are conducted under well-defined protocols and standards for diagnostic criteria [167]. These methods involve the two physiological parameters currently used in medical practice for diabetes treatment: the instantaneous BGL indicating the concentration of free glucose in blood at the time of the measurement (FPG and OGTT), and glycated hemoglobin (HbA1c) presence providing an indicator of the average blood glucose content over the preceding three months [164]. The glycated hemoglobin is an example of AGEs resulting from the glycation of red blood cells (erythrocytes), with an average lifespan of three months. The HbA1c test was approved in 2010 by the American Diabetes Association as a diagnostic criteria for DM [168]. In contrast to the instantaneous BGL, the HbA1c measure is a more stable and reliable parameter in diabetes treatment practice because it is not significantly affected by peaks of glucose presence in blood associated to many physiological processes such as digestion, exercise, or strong emotions [169]. However, HbA1c test can be only used for long-term glycemic control due to his poor time resolution (depends on the lifespan of red

blood cells) and requires complex equipment only available at specialized laboratories, being less accessible in low- and middle-income countries.

The instantaneous BGL is the most common parameter used in DM control introduced in 1962 with the first glucose enzyme electrode [170], thus paving the way for development of conventional glucometers [171]. The glucometers are widely used at hospitals and allow diabetic patients to non-continuous self-monitoring of BGLs through finger pricking with a lancet to extract full blood drops. Other recent minimally invasive technologies currently available in the market are the continuous glucose monitoring devices consisting on a wireless receiver, a transmitter, and a tiny subcutaneous sensor that estimate the BGL by the glucose content in the interstitial fluid [172]. Some disadvantages of continuous glucose monitoring devices are that requires frequent calibration using standard glucometers and provides shifted-in-time BGL measurements of approximately 5 minutes [173]. In general, these invasive and minimally-invasive methods for the BGL sensing have significant drawbacks such as discomfort, they are time-consuming, are based on consumable materials (resulting in high follow-up costs in long-term use), can be painful (tissue damage), and may cause infections. All the drawbacks mentioned above explain somehow the unwillingness of an important number of diabetic patients to follow completely medical recommendations in DM care [174].

Therefore, there is a worldwide effort to develop new non-invasive technologies with the aim to provide more efficient methods for instantaneous BGL monitoring, improve the patients' comfort and to alleviate the drawbacks mentioned above. An extensive revision of current proposed technologies for non-invasive sensing of BGL, involving thermal, electrical, and optical methods, can be found in [175]–[177].

Some of these non-invasive techniques are based on transdermal approaches (considered also for some authors as minimally invasive), such as bioimpedance spectroscopy and reverse iontophoresis, their actual performance being considerable affected by variations in tissue water content and perspiration. Side effects as skin irritation has been also shown by reverse iontophoresis. Besides interstitial fluid evaluation, which is the target of these minimally invasive technologies for BGL monitoring including reverse iontophoresis, there are other proposed technologies of wearable sensors focused to continuous BGL monitoring (or other DM biomarkers) by using different biological fluids such as sweat, breath, saliva, aqueous humour, etc. [178].

Among the most studied and developed non-invasive optical approaches are the infrared technologies. The near-infrared spectroscopy (~ 215 THz - ~ 400 THz) offers relatively low-cost materials and allows for deeper detection of glucose concentrations (low water absorption of near-infrared radiation) but shows high scattering in tissue and significant interferences by proteins, fat, hemoglobin, among others biological substances that have similar absorption properties to glucose molecule [177], [179]. On the other hand, even though the mid-infrared spectroscopy (~ 30 THz - ~ 120 THz) exhibits less scattering and more specific absorption bands, the strong water absorption coefficient in mid-infrared band (poor penetration depth of tissue) makes necessary the use of powerful sources such as quantum cascade lasers, resulting in more complex and expensive implementations [180]. Some other approaches such as photoacoustic, photothermal, occlusion, and Raman techniques has been proposed to improve infrared technologies deficiencies [181]–[186]. Other proposed methods are fluorescence [187], [188], optical polarimetry [189], [190], optical coherence tomography [191], [192] and metabolic heat conformation [193], [194].

In general, all the non-invasive technologies proposed for BGL sensing are influenced by physiological variability and environmental conditions, affecting their accuracy and consistency, and although the extent of the problem varies among these technologies, there is still no completely viable solution compared to current enzyme-based standard methods, especially for cases in which accurate and frequent monitoring of BGL is a serious matter.

Besides the research lines above mentioned, more recent approaches for non-invasive sensing of BGL has been emerging based on THz time-domain spectroscopy [195], [196] and microwaves spectroscopy [197]–[199]. These approaches are motivated by their good interaction properties with biological media mentioned above, but they are still far from an operational diagnostic. In particular, the mm-wave spectroscopy (30 GHz - 300 GHz) has shown great potential for the non-invasive sensing of BGL, with in-vivo blood glucose monitoring in animal models reported using Ka-band frequencies (27 GHz - 40 GHz) [200], [201], and the detection of glucose spikes in humans during an intravenous glucose tolerance test using transmission measurements at 60 GHz [202].

In this thesis the application of mm-wave spectroscopy (W-band frequencies: 75 GHz – 110 GHz) is studied as an alternative approach for

non-invasive sensing of hyperglycemia. It is important to note here that the actual target of the thesis is not to obtain a diagnostic for the instantaneous value of the BGL (as most of the diagnostics enumerated before), but to evaluate sustained hyperglycemia conditions associated to the physiological changes that early accumulation of AGEs have on the tissues. In this sense, the system performance will be ultimately tested against HbA1c measurements that, as discussed before, it is a more stable and reliable parameter in diabetes treatment practice.

Finally, it is worth mention that the selection of the W-band for this diagnostic is associated to some of the attractive advantages of such frequencies for this kind of diagnostics as they are that the interaction region is well defined, the probing location is small due to tighter focusing capabilities or smaller waveguide dimensions, the dispersion effects are less pronounced, and can lead to very compact sensing equipment.

3.2. THz, MICROWAVE AND MILLIMETER- WAVE RADIATION IN BIOMEDICAL ENGINEERING

Nowadays, the development of new technologies in biomedical industry is a huge research field with a substantial growth, aiming not only to provide more efficient tools for medical diagnosis, treatments, and procedures, but also to enhance patients' quality of life and well-being. However, a complete settlement of new techniques to develop minimally or non-invasive medical procedures in diagnosis and follow-up of existing diseases is a current barrier to break in the twenty-first century. In this sense, increasing attention is being paid to microwave [197], [203]–[206], millimeter-wave [207]–[209] and THz [210]–[214] spectroscopy as very promising non-invasive approaches for biomedical applications.

The microwaves region refers to the frequencies from 300 MHz to 30 GHz, millimeter waves (mm-waves) correspond to frequencies from 30 GHz to 300 GHz, and the THz region lies between the millimeter waves and infrared band, being typically defined as 100 GHz – 10 THz.

Among the unique properties that make microwave, mm-wave and THz regions well-suited for health-care systems applications is their non-ionizing character. In both cases, the electromagnetic radiation implies very low photon energy below 10 meV, which is known not to be capable of modifying atoms structure or cause other kind of chemical damages to

molecules, unlike X Rays. However, biosafety of THz and microwaves radiation is still a controversial question since some studies have reported evidence of chromosome lesions, among other physiological hazards using specific modes of irradiation [205], [212], [214], [215].

The THz radiation is characterized for exhibiting strong absorption by water, which is a double-edged sword: biological substances are mainly composed of water, limiting THz waves penetration depth biologic media, but such high absorption also provides THz radiation with a very effective contrast capacity to differentiate tissues with different water content, which usually happens between diseased and normal tissue. An example of such contrast capacity is the interest in THz imaging for cancer diagnosis since presence of tumors often implies an increased blood supply and water content in the affected tissues. Besides that, it has been also proven that structural changes in affected tissues can be also identified using THz radiation by eliminating water content in tissue [210]. Other advantages of THz radiation are that scattering losses in biological tissues are negligible, offers excellent time and spatial resolutions, and the photon energy of THz waves coincides with rotational and vibrational transitions of biomolecules. An overview of THz radiation and several biological/biomedical applications such as vessel imaging, corneal tissue sensing, identification of different pathogenic bacteria, detection of dental caries, detection of DNA fragments in aqueous solutions, among others can be found in [212]–[214]. However, the development of compact and cost-effective mm-wave and THz systems with good operating performance is still a common current challenge to make them suitable for healthcare application and clinical practice.

Similarly, microwaves can also polarize biological substances, can achieve greater penetration depth in biological tissue than THz radiation, and many microwaves modalities are better suited for reliable, compact, and relatively cost-effective implementations. The microwaves have also shown great potential for biomedical applications taking advantage of the different electrical properties (permittivity and conductivity) of biological substances and tissues (primarily related to water content), with several applications such as microwave imaging of the heart, brain, bones and breast cancer diagnostics, blood glucose monitoring, brain stroke and heartbeat detection, sensing of blood flow and pressure, etc. [197], [203], [216]. Other applications of microwaves in medicine are based on hyperthermia induction and ablation for medical therapies such as the

treatment of cardiac arrhythmias, liver, benign prostate hypertrophy, angioplasty, brain tumors, etc. [205], [217]–[220].

In general, microwaves, mm-waves, and THz radiation are very promising developing spectroscopic techniques in biomedical approaches with good sensibility to differentiate biological tissues and substances, offering different balances between penetration depth and resolution. However, there is still much work to be done to achieve the complete settlement of effective technologies, based on THz, microwave and mm-wave radiations, in the medicine field.

3.3. MM-WAVE SPECTROSCOPY INSTRUMENT USED FOR EXPERIMENTS

This work focuses on the assessment of an alternative approach for non-invasive sensing of hyperglycemia, typically associated to DM, by spectral interrogation using mm-waves (W-band). As mentioned above, high frequencies such as mm-waves and THz radiation reduce dispersion effects in tissue and are very potential developing compact spectroscopic devices, besides their biosafety, which is very attractive in biomedical applications. The small waveguide dimensions and the small dispersion in tissue allows for small probing areas. However, frequencies closer to THz radiation exhibit poor penetration depth in tissue due to the water absorption and poor signal-to-noise ratio requiring longer measurement times that are usually not compatible with in-vivo monitoring. In this sense, the W-band offers a good balance between penetration depth in tissue, interaction volume, and signal-to-noise ratio.

In order to study the capabilities of W-band for non-invasive sensing of hyperglycemia, a CW mm-wave spectrometer spanning the whole W-band from 75 GHz to 111 GHz was put together specifically for this project in collaboration with Professor Viktor Krozer's research group at Goethe University in Frankfurt. The spectrometer, designed and mounted in Frankfurt, uses different multiplication chains to reach the W-band frequencies from a Ku-band generator (12 GHz - 18 GHz). The spectrometer allows to collect transmission and reflection signals from spectral interrogation performed to the biological media.

A simplified block diagram of the setup used in the measurements is shown in Figure 3-1, and explained in the following lines. A frequency

sweep from 12.5 GHz to 18.5 in steps of 0.25 GHz is generated using a synthesized signal generator APSIN20G (AnaPico, Zurich, Switzerland) with 0.3 MHz frequency difference between the outputs SG1 and SG2. The output signal SG1 is directly connected (SMA) to an AFM6 Active Frequency Multiplier (Radiometer Physics GmbH, Meckenheim, Germany) that increase the frequency signal by a factor of six, resulting in a frequency sweep from 75 GHz to 111 GHz in steps of 1.5 GHz. The AFM6 Frequency Multiplier is realized in a waveguide housing and exhibits a WR10 waveguide output, which is fed to a dual directional coupler. The coupled arms of the dual directional coupler define the reference and reflection channels, respectively, with two HMR6 subharmonic mixer receivers (Radiometer Physics GmbH, Meckenheim, Germany) one at each coupled port. The thru branch of the coupler is connected to a waveguide probe and the incident wave is sent through the biological media (sample). Then, the signal that traveled through the biological media is directed towards a subharmonic mixer receiver, similarly as in the reference and reflection ports, by using a second rectangular waveguide probe positioned after the sample. The probes are straight cuts of a WR10 waveguide tapered on the outside directly in contact with the sample. The outputs of the subharmonic mixer receivers deliver an intermediate frequency of IF = 1.8 MHz and are connected to a data acquisition unit (Handyscope HS4-10, TiePie engineering, Sneek, Netherlands), which digitizes the measured signals (reference, reflection, and transmission) with a sampling rate of 10 MHz for further filtering and processing using LabVIEW software [221]. All the acquisition routines were developed in the Sensors and Instrumentation Techniques Group of the Universidad Carlos III de Madrid.

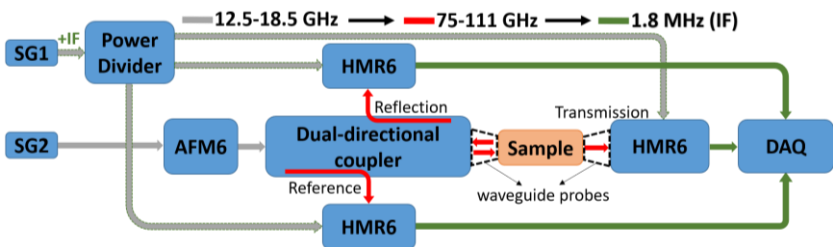


Figure 3-1. Block diagram of the mm-wave spectroscopy system used for non-invasive assessment of biological media. See text for details.

A photography of the mm-wave spectroscopic instrument involved in the spectroscopy system described above is shown in Figure 3-2, identifying the different components enumerated above. The mm-wave instrument can be made very compact and it is based on potentially low-cost electronic technology.

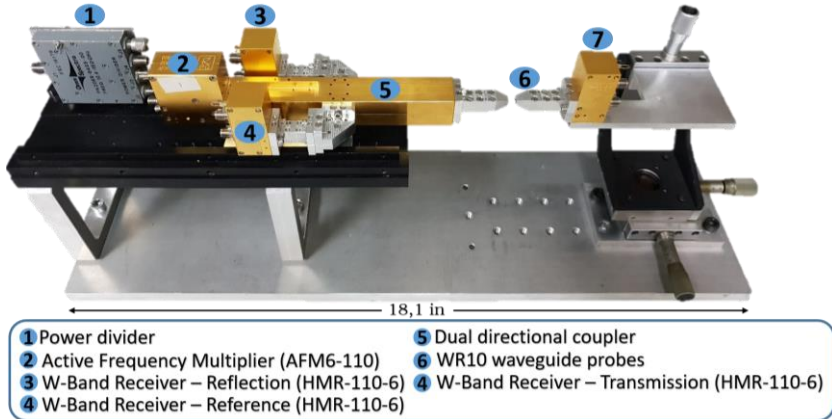


Figure 3-2. Photography of the mm-wave spectroscopic instrument used for spectral interrogation within the W-band.

A flowchart summarizing the LabVIEW routine used for that acquisition and pre-processing is shown in **¡Error! No se encuentra el origen de la referencia..** All the reference, reflection, and transmission signals measured at each frequency point (spectral data acquisition) are averaged (18 periods sampled) to reduce the effects of random noise in the sampled signal. Then, all the signals are filtered through a narrow band filter with the intermediate frequency (1.8 MHz) as the center frequency. The LabVIEW program runs in real-time and implements a multi-channel lock-in amplifier for the estimation of the amplitude of the sampled signals improving the signal-to-noise ratio [222], [223]. Simultaneously, the phase of the sampled signals is estimated by tone identification using the fast Fourier Transform. The obtained phase from the signals is used to estimate the phase shift of the transmitted and the reflected waves with respect to the reference signal (measured from incident wave). The measurement is repeated 5 times at each frequency point and the mean value and the standard deviation are estimated for the amplitude and phase parameters. The collected spectral data, including the corresponding standard deviations, are exported as a .txt file.

The complete measurement process, covering 25 frequencies equally distributed over the W-band takes around 45 seconds. The measurement time is mainly limited by the control electronics (signal generator and LabVIEW program for signal acquisition and processing) rather than the mm-wave instrument.

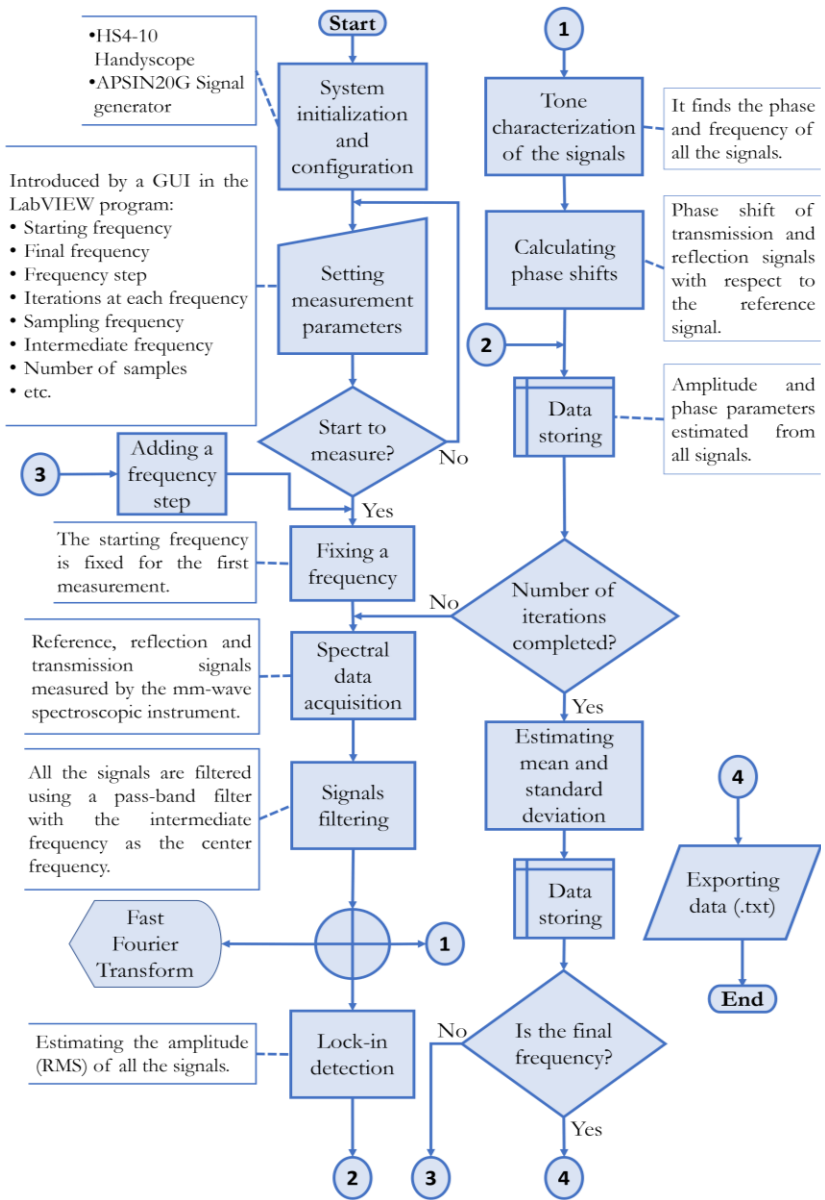


Figure 3-3. Flowchart summarizing the LabVIEW program for data acquisition and pre-processing. The numbers within the circles indicate branch connectors.

4. USE OF W-BAND
SPECTROSCOPY FOR IN-
VIVO NON-INVASIVE
ASSESSMENT OF
HYPERGLYCEMIC STATES
USING ANIMAL MODELS

In this thesis dissertation, a non-invasive approach for in-vivo detection and monitoring of DM by spectral interrogation in the W-band is assessed using the system described in the previous chapter. The aim is to study possible relations between the measured W-band spectral response and sustained hyperglycemia condition, typically associated to DM. To this end, three experiments were carried out using mice as animal models representative of normoglycemia and different sustained hyperglycemic conditions. As outlined in chapter 3, the experiments were designed under a non-targeted spectral profiling approach, adopting a strategy of comparison between the W-band profiles to detect variations in the measured spectral response that qualitatively differ between the glycemic states involved in the sample population. The spectral data sets collected from the experiments will be processed using FDA methods introduced in chapter 2. In this way, the W-band spectral profiles are analyzed as single continuous responses, which contains all the spectral features of biological components constituting the sample.

In this fourth chapter, all the experiments related to the animal models and the corresponding statistical analysis developed for the assessment of the applied spectroscopic technique will be presented. Firstly, the sample populations and the experimental protocols common to the three experiments, that were carried out sequentially, will be described. Later, the most relevant findings obtained from the experiments by the FDA performed on the collected spectral data sets will be shown. In the first experiment, from now on referred to as “Experiment A”, the diagnostic technique is assessed by exploring the contained variability in the spectral response measured achieved by FPCA. The obtained results from the spectral data sets, corresponding to amplitude and phase of both transmission and reflection measurements, are compared to find the indicator that best discriminates the sustained hyperglycemia condition. Then, the FPCA results are interpreted with two aims in mind: first, to study possible influences of known variables involved in the experiment on the measured spectral response and, second, to characterize the spectral response associated to the sustained hyperglycemia discrimination. A second experiment, herein after referred to as “Experiment B”, was carried out after the first one with the aim to

evaluate the consistency and robustness of the diagnostic by a validation test and a multi-test analysis using a classification model. The classification model is estimated by FPCLoR on the measured spectral data. Then, the FPCLoR model is interpreted to study, qualitatively and quantitatively, the relation between the spectral response and the sustained hyperglycemia discrimination. Additionally, two different approximations of continuous spectral profiles using regression splines and P-splines are compared. Both approaches, with a different smoothness degree, are used to show the main trade-off between the classification rate and interpretation of the fitted model. The idea behind this scenario is to highlight typical choices and decisions to be made by the researcher in the FDA, and the exciting possibilities working with functional data. Finally, a third experiment, referred to as “Experiment C”, was carried out to evaluate the feasibility of the applied spectroscopic technique for DM monitoring. To achieve this, diabetes was induced on a group of healthy mice to observe their evolution over a period of two weeks. A classifier obtained from the first experiment (different sample population) is used to predict the condition of diabetized mice at each measurement day separately, to test the sensitivity of the proposed approach to detect different glycemic states.

4.1. DESCRIPTION OF THE EXPERIMENTS USING ANIMAL MODELS

The experiments presented in this chapter were carried out in collaboration with the Epithelial Biomedicine Division of the Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT).

All the experimental procedures involved in the mice experiments were carried out according to European and Spanish laws and regulations (see Appendix for more details on the corresponding Ethical Issues in section 8.1). The animals were purchased from Elevage-Janvier (Le Genest-Saint-Isle, France), treated and housed individually in pathogen-free conditions at the Laboratory Animals Facility (Spanish registration number 28079-21 A) of the CIEMAT.

4.1.1. ANIMAL MODELS USED IN THE EXPERIMENTS

In the experiments different mice strains exhibiting varied physical features between them such as hair color and skin types were used, as illustrated in Figure 4-1. The C57BL6/J mouse has a dark brown hair and it is widely used for genetic modifications to model human diseases (being diabetes on of such diseases). The BALB/C is an albino mouse and is one of the most popular inbred strains used in animal experimentations. The NMRI-Foxn1^{nu}/Foxn1^{nu} is a hairless mouse, commonly known as “nude”, and it is an immunodeficient animal model (T cell deficient) widely used in biomedical research.

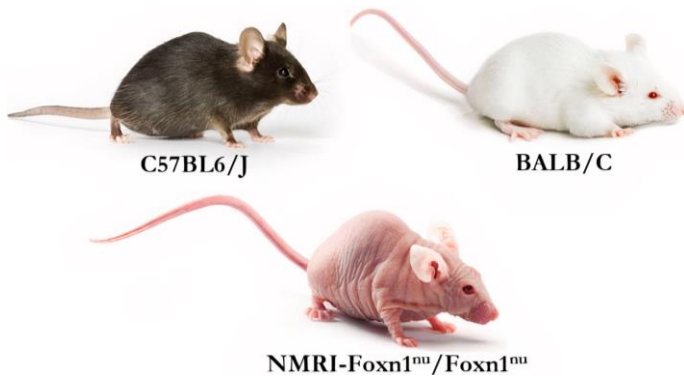


Figure 4-1. Images of the main mice strains showing their physical features.

Regarding to the non-targeted spectral profiling approach, different normal/pathological conditions were considered in the experiments and implemented using the abovementioned animals. The two main groups used in the experiments involved all the mice models: normoglycemia and hyperglycemia conditions. The normoglycemia condition is represented by healthy mice exhibiting normal and stable BGLs, with an expected glucose level of 100 mg/dl. Within the hyperglycemia condition two types of pathologies with different associated sustained hyperglycemic states were considered: overweight mice by overeating, hereinafter referred to as “obese mice”, representing mild diabetes, and diabetic mice representing a full-blown diabetes (“diabetic mice”). To achieve these two conditions, mice were genetically modified to have spontaneous mutations leading to deficiency of Leptin (Lep^{ob}/Lep^{ob}), the “obese mice” and Insulin resistance (Lep^{db}/Lep^{db}) [224], the “diabetic mice”. Additionally, a third

type of hyperglycemic animals were obtained through drug-induced diabetes on healthy mice, hereinafter referred to as “diabetized mice”. This third condition was considered as an alternative and complementary condition of full-blown diabetes due to its strongest effects in mice (compared to genetically modified “diabetic mice”). The induction process for the diabetized mice consisted of three intraperitoneal injections of Streptozotocin (Sigma-Aldrich, Inc., St. Louis, MO, USA) with concentrations of 0.1 mg/g, 0.1 mg/g, and 0.15 mg/g on alternate days during a period of 5 days. The Streptozotocin (STZ) is a compound designed to be especially toxic to pancreatic islet insulin-producing β -cells [225]. These obese, diabetic and diabetized mice are widely recognized animal models undergoing sustained hyperglycemia [226], i.e. they exhibit higher BGL than normoglycemic mice under normal conditions (meals, exercise, sleep, etc.), being the less aggressive condition for the obese mice.

4.1.2. EXPERIMENT A

The first experiment was designed to explore the influences of different glyceic states and varied physical features on the W-band measured spectra. Therefore, all animal models introduced in section 4.1.1 were considered in the experiment.

The non-invasive assessments of hyperglycemia in the first experiment was performed on a sample population of twenty mice with ten normoglycemic cases and ten hyperglycemic cases. The group of hyperglycemic cases involves five obese mice, two diabetic mice, and three diabetized mice. The group of normoglycemic cases consist of eight healthy mice, and two obese mice treated with precise and continuous doses of human leptin using implantable 28-day-lasting micro osmotic pump (ALZET Osmotic Pumps, California, USA) [227]. The mice strain among other characteristics of the sample population are detailed in Table 4-1.

Table 4-1. Sample population measured in experiment A.

Condition	Mice strain	Variation	Treatment	Expected Glucose level	Age	Qty.
Normoglycemia	NMRI-Foxn1 ^{nu} /Foxn1 ^{nu}	-	-	100 mg/dl	1 month	4
	C57BL6/J	-	-	100 mg/dl	6 months	2
	C57BL6/J	Lep ^{ob} / Lep ^{ob}	Leptin-pump	100 mg/dl	6 months	2
	BALB/C	-	-	100 mg/dl	6 months	2
Hyperglycemia	C57BL6/J	Lep ^{ob} / Lep ^{ob}	-	>150 mg/dl	6 months	5
	C57BL6/J	Lep ^{db} / Lep ^{db}	-	>250 mg/dl	6 months	2
	NMRI-Foxn1 ^{nu} /Foxn1 ^{nu}	-	STZ-induced diabetes	>400 mg/dl	1 month	3

Let us to note that the mice with hair were not shaved for the non-invasive assessment by the mm-wave spectroscopic instrument. Instead, mice hair at the measurement location was regularly cut off to ensure and facilitate the positioning of the skin between the probes. Besides this, no additional special treatments of the skin were considered. Prior to the non-invasive assessment, mice were anesthetized to prevent excessive movement and self-harm risks during the measuring process, which takes around 45 seconds. The mice were induced into a skeletal muscle relaxation state (around 30 minutes) using a standard rodent anesthesia (ketamine-medetomidine), administered by an intraperitoneal injection five minutes before taking the measurement.

The spectroscopic measurements were carried out directly on a fold of the skin on the mice back, as shown in Figure 4-2. The probes of the spectrometer instrument are brought into direct contact with the skin fold, without applying to much pressure, to ensure the propagation of the signal through the skin and underlying layers, resulting in an interaction between the generated waves and the skin fold during the spectroscopic measurement. As seen in section 3.3, the probes of the spectrometer instrument consist of previously aligned standard rectangular WR10 waveguides tapered on the outside to hold the skin fold. A separation of ~1 mm between probes was carefully adjusted as the minimum required to hold the skin of the mice without infringing damage or pain on them. It should be noted that the pressure of the two waveguide straights on the

mice skin was not of prime importance in the non-invasive assessment procedure.

The measuring process for spectral interrogation consists of a frequency sweep across the W-band (75 GHz - 111 GHz) in steps of 1.5 GHz by using the mm-wave instrument described in section 3.3. The amplitude and phase measurements of both reflection and transmission coefficients of the biological sample were continuously acquired during the whole measuring process.



Figure 4-2. Photograph taken during the measuring process for two different mice strains.

It is important to note that no special calibration procedures were performed to the spectroscopy instrument previous the measurements. One of the most important challenges was adjusting an appropriate reference of the signal power generator to ensure both a good dynamic range for the measured signals at harmonic receivers and to avoid receiver's saturation. To this end, several tests were carried out using a quartz cuvette (5 mm) containing water to simulate the expected high absorption level of the biological tissue. The reference level for transmission signal was dictated by the highest absorption level and the receiver noise floor. Then, the output power of the signal generator was adjusted to obtain a frequency response as flat as possible at the transmission port.

Additionally, almost simultaneously to the non-invasive assessment, the BGL was measured on the mouse tail by an Accu-Chek Aviva Nano (Hoffmann-La Roche, Basilea, Switzerland), as shown in Figure 4-3, and the skin-fold thickness used in the measurement was determined by a Mitutoyo Digital Caliper (Mitutoyo Corp., Kanagawa, Japan).

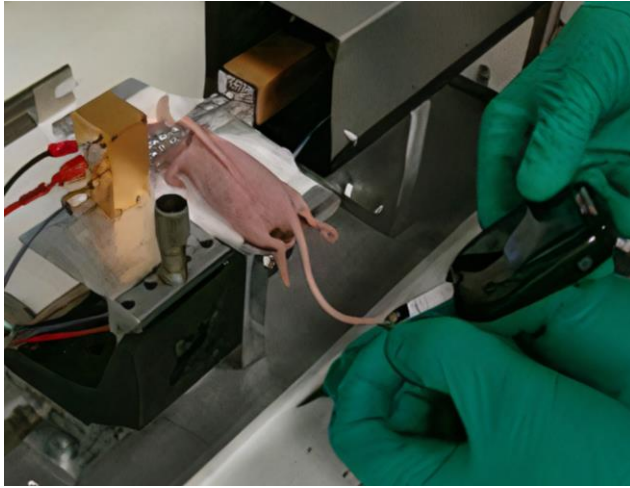


Figure 4-3. Photograph taken during the BGL measurement by blood extraction from the tail of the mouse.

4.1.3. EXPERIMENT B

The second experiment was focused to validate the obtained results from Experiment A, and to assess the robustness of the applied spectroscopic technique. Therefore, a new sample population of thirty-three mice was measured using the proposed non-invasive approach. In this experiment, only one mice strain was considered (C57BL6/J). The new sample population consisted of eighteen healthy mice as normoglycemic cases, and fifteen hyperglycemic cases: nine leptin deficient (Lep^{ob}/Lep^{ob}) mice undergoing a mild diabetes, and six insulin resistant (Lep^{db}/Lep^{db}) mice undergoing a full-blown diabetes. The sample population measured in Experiment B is detailed in Table 4-2.

Table 4-2. Sample population measured in experiment B.

Condition	Mice strain	Variation	Treatment	Expected Glucose level	Age	Qty.
Normoglycemia	C57BL6/J	-	-	100 mg/dl	8 weeks	18
Hyperglycemia	C57BL6/J	Lep^{ob}/Lep^{ob}	-	>150 mg/dl	5 weeks	9
	C57BL6/J	Lep^{db}/Lep^{db}	-	>250 mg/dl	6 weeks	6

As in the previous experiment (Experiment A), the non-invasive assessment was performed directly on a skin fold on the back of the mice (see Figure 4-2) and the mice hair was regularly cut off to facilitate the handling of their skin, but no special procedures were conducted on the measurement skin area prior the non-invasive assessment. The BGL was measured on the mouse tail by an Accu-Chek Aviva Nano (Hoffmann-La Roche, Basilea, Switzerland), almost simultaneously to the non-invasive assessment (see Figure 4-3).

Let us to note that Experiment A and Experiment B were conducted separately with sixteen months elapsed between them and using a totally different sample of mice. During that time, many changes in hardware and software of the mm-wave spectroscopy system were made to improve its functionality. Therefore, the output power of the signal generator, previously calibrated in experiment A (see section 4.1.2), had to be recalibrated to obtain a flat frequency response at the transmission port without the biological sample. This is an important issue as that even changes were incorporated in the hardware and software the results obtained were consistent and allowed proper comparison of the results between the two experiments as discussed below.

Also, considering the harmful effects of the injected anesthesia (ketamine-medetomidine) on mice observed in previous experiment (experiment A), the experimental protocol was modified to replace the injected anesthesia with other less aggressive method. In this second experiment, excessive movement of the mice during the non-invasive assessment was prevented by using inhaled anesthesia. As shown in Figure 4-4, previous to the measurement process, each mouse was introduced into an induction chamber with isoflurane mixed with oxygen (3% - 4% of isoflurane concentration) until anesthetic takes effect. Once the mouse is “sedated”, it was placed on a table with its snout into a supplying mask to continuously inhale a lower anesthesia concentration (1.5% - 3% of isoflurane concentration) during the whole measurement process.

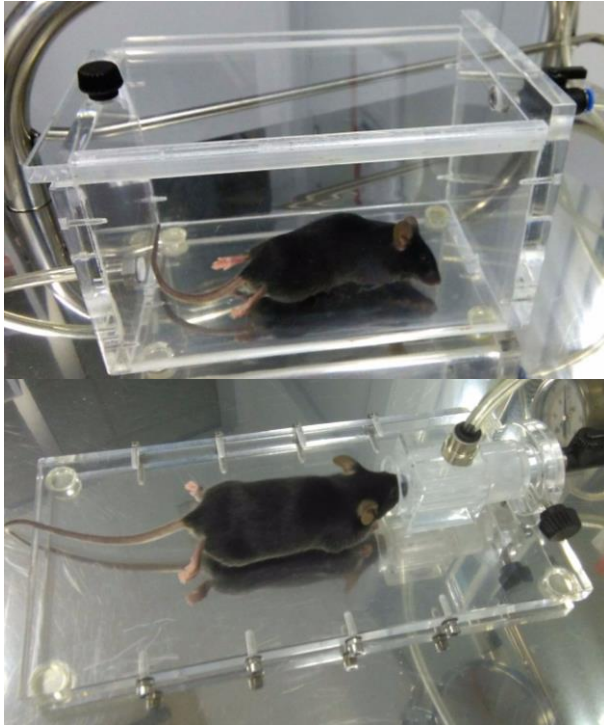


Figure 4-4. Photograph of the mouse into the anesthetic induction chamber (at the top), and of the same mouse placed on the table for continuous administration of the anesthesia via a mask during the measuring process (at the bottom).

4.1.4. EXPERIMENT C

The third experiment focused on the feasibility of the proposed non-invasive approach to detect gradual changes in the glyceic state of the mice. The experiment included the study of the measured spectral response from over time on transition cases from normoglycemia to hyperglycemia. This experiment was performed under the same conditions and proceedings described for Experiment B (see 4.1.3). The sample population used in the experiment is detailed in Table 4-3. Let us observe that most of the mice employed in the experiment were taken from the sample population measured in Experiment B, including the diabetized mice that were identified as normoglycemic cases before the diabetes induction process.

Table 4-3. Sample population measured in Experiment C.

Condition	Mice strain	Variation	Treatment	Expected Glucose level	Age	Qty.
Normoglycemia	C57BL6/J	-	-	100 mg/dl	8 weeks	6
Transition cases	C57BL6/J	-	STZ-induced diabetes	>400 mg/dl	8 weeks	16
Hyperglycemia	C57BL6/J	Lep ^{db} /Lep ^{db}	-	>250 mg/dl	6 weeks	6

The experiment lasted nineteen days and a total of 4 measurements were performed. The measurements started on the last day of the treatment (animals receiving the final dose) and finished 14 days hereafter. The six diabetic mice and six healthy mice were also measured simultaneously to diabetized mice as references. During the experiment, six diabetized mice died before the third measurement day (ten days after treatment) due to the adverse effects of the STZ drug [228].

4.2. BLIND EXPLORATION AND CHARACTERIZATION OF THE MEASURED SPECTRAL RESPONSE

In this section, the measured spectral data obtained from the non-invasive assessment of hyperglycemia on the sample population employed in Experiment A (see 4.1.2), will be analyzed using FPCA. The sample information for each measurement consists of four sets of twenty observations containing the complex information of amplitude and phase for the reflection at the top skin layers and the transmission through the biological media (fold of skin). Each observation was measured at twenty-five frequencies equally spaced across the W-band (75 GHz - 111 GHz). In Figure 4-5, the measured spectra (raw data) for examples of each of the glycemic cases from sample population of Experiment A are shown, with normoglycemia and hyperglycemia conditions identified by dotted blue lines and dashed red lines, respectively. It is apparent how the transmittance of the fold of skin of the hyperglycemic mice is considerably higher than normoglycemic mice, providing with a relatively good “a priori” classification. However, FPCA was estimated for each of the

spectral data sets to explore all the captured variability by the non-invasive assessment.

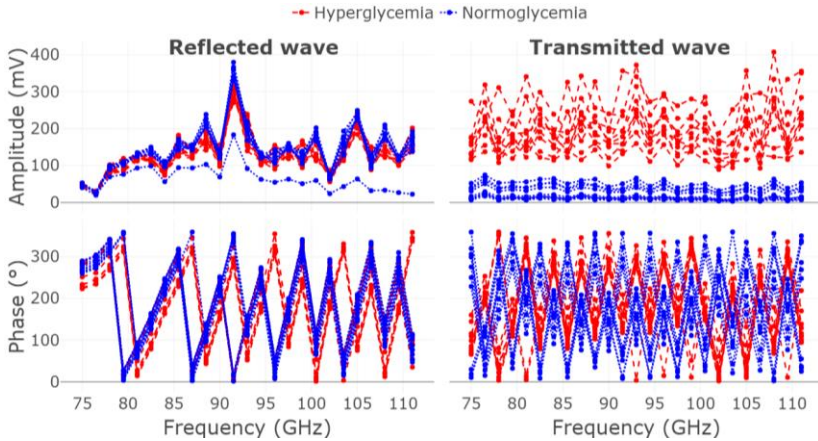


Figure 4-5. Amplitude and phase spectra measured from both reflected waves by the top skin layers (left column) and transmitted waves through the fold of skin (right column). Normoglycemic and hyperglycemic cases are identified by dotted blue lines and dashed red lines, respectively.

In spectroscopy applications, as shown at the bottom of the Figure 4-5, the instantaneous phase is typically wrapped within the interval π and $-\pi$. The phase unwrapping is a very common problem due to the 2π discontinuities, that occurs when an extreme value of the interval $[\pi, -\pi]$ is reached. Basically, phase unwrapping is a process for solving problems of ambiguity in the measured phase of the signals (instantaneous phase). Then, in its simplest way, reconstructing the continuous phase variation can be done by adding or subtracting multiples of 2π to the instantaneous phase at the frequencies in which discontinuities occurs. The unwrapped phase, shown in Figure 4-6, was estimated before performing FPCA to remove the periodic behavior exhibited by the spectral data corresponding to the instantaneous phase. In this way, we prevent influence the FPCA by the excessive variability in the spectra associated to the 2π discontinuities in the instantaneous phase spectra.

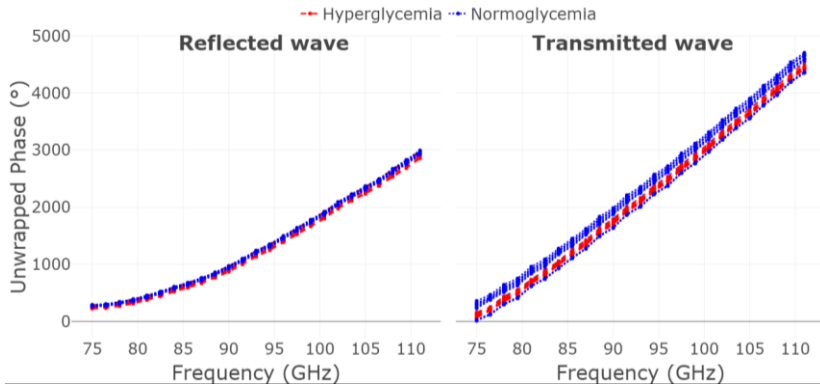


Figure 4-6. Unwrapped phase of reflected (left panel) and transmitted (right panel) waves. Normoglycemic and hyperglycemic cases are identified by dotted blue lines and dashed red lines, respectively.

All the data processing and analysis were developed using the statistical free software R [229], and the FPCA was estimated by using the package “*fda*” available in the library of R [230]. The *fda* packages includes a wide range of tools for functional data analysis and processing, such as the basis representation, curves registration, functional linear models, etc. (see [94]).

The functional data sets obtained from amplitude and phase of the reflected and transmitted waves are shown in Figure 4-7. The functional data was approximated by cubic regression splines, as shown in Eq. (2-19), with the basis coefficients estimated as shown in Eq. (2-23). The cubic B-splines were defined on 7 and 18 equally spaced knots, for amplitude and unwrapped phase spectra, respectively.

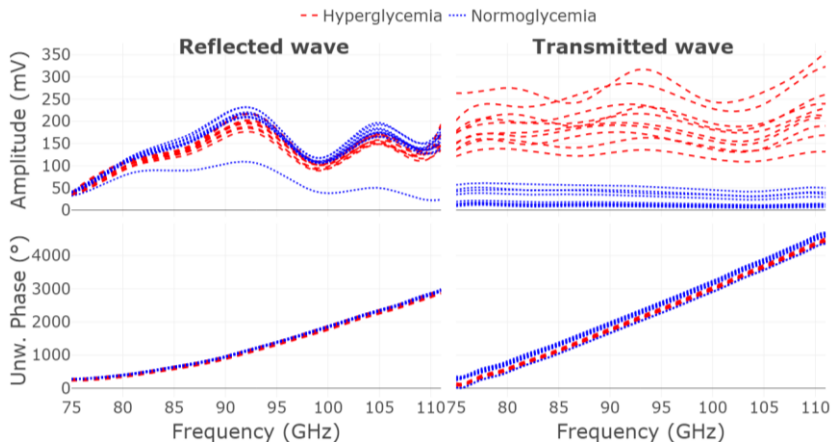


Figure 4-7. Approximated sample curves from amplitude and phase of both reflected waves by the top skin layers (left column) and transmitted waves through the fold of skin (right column). Normoglycemic and hyperglycemic cases are identified by dotted blue lines and dashed red lines, respectively.

4.2.1. INTERPRETING FPCs' SCORES

A FPCA was performed separately on each of the approximated functional data sets shown in Figure 4-7. The percentages of the explained variance by the estimated FPCs for each functional data set are summarized in Table 4-4.. In all cases, the cumulative explained variance by the first two FPCs is above the 97%, i.e. more than the 97% of the contained variability in the sample curves is retained and explained by the first two FPCs. Here is important to note that, in some cases, the captured variance by the first FPCs might be related with no significant information about the current study. Let us observe that the loading functions associated to the FPCs are estimated as the orthonormal set that best describes variance in the data, but, this does not mean that there is no other orthonormal set meeting the same objective less efficiently, i.e. no largest variability is captured by the first FPC. The VARIMAX rotation is a widely used tool in multivariate PCA, also extended to the FPCA, which allows to find other orthonormal sets (loading functions) decomposing variability in the sample curves in different proportions than the estimated by the original FPCA [92, pp. 95–99]. This is a very useful approach that allows to look for different sources of variation in data that might be initially overshadowed by the largest variance estimated in terms of the

FPCA problem, especially when the first FPC originally accounts for large proportion of explained variance. In our case, VARIMAX rotation did not provide additional relevant information than the one obtained directly by the FPCA.

Table 4-4. Explained variance by the FPCs estimated on the approximated functional spectral data sets.

	Parameter	FPC1	FPC2
Reflected wave	Amplitude	94.91 %	2.74 %
	Unwrapped Phase	99.89 %	0.09 %
Transmitted wave	Amplitude	99.69 %	0.18 %
	Unwrapped Phase	99.92 %	0.06 %

In Figure 4-8, the estimated scores of the first FPC are plotted versus the estimated scores of the second FPC for each functional data set obtained from Experiment A. The scatterplots of the scores provide an easy way to look for natural clustering of the sample curves according to the captured variance by the estimated FPCs. From this figure, we can see how the zero axis of the first FPC potentially discriminates normoglycemia from hyperglycemia, as discrimination rule, using the amplitude of the reflected wave and both parameters (amplitude and phase) of the transmitted wave. Then, as expected, the amplitude of the transmitted wave provides a clearer separation between both conditions with no misclassified cases. It is also equally worth noting that, other clusters can be glimpsed from the measured parameters of the transmitted wave, which implies the second FPC is providing additional information about the sample. In general, we can say that the transmitted wave provides a potentially better indicator than the reflected wave. This is predictable because transmitted wave propagates through deeper skin layers as compared to the reflected wave, travelling throughout the whole biological media. Therefore, hereinafter, the analysis will be focused only on the transmitted wave.

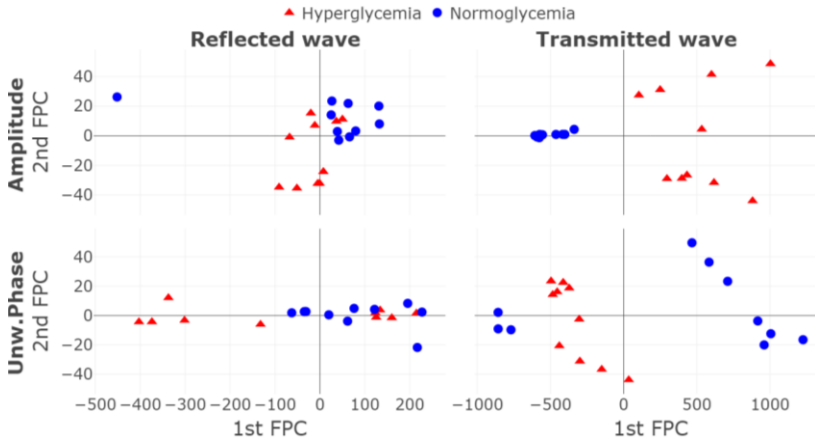


Figure 4-8. Scatterplots of the scores of the first FPC vs scores of the second FPC estimated for amplitude (top line) and unwrapped phase (bottom line) of reflected (left column) and transmitted (right column) waves. Hyperglycemic and normoglycemic cases are identified by red triangles and blue circles, respectively.

In Figure 4-9, the different mice strains and pathologies involved in the sample mice population are identified. In general, we can observe the scores of the phase parameter exhibits strong clustering related to the different mice strains. This implies that the phase parameter is affected (it is more sensitive) by the anatomic varieties between the mice strains: type of skin, hair density, hair color, skin-fold thickness, etc. Unlike the phase parameter, the second FPC estimated from the amplitude separates obese mice from diabetic and diabetized mice, and all normoglycemic mice are grouped together. As before, considering the zero axis of the second FPC as the discrimination rule, obese mice are potentially discriminated between diabetic cases (either genetically diabetic or diabetized) with no misclassified cases. This is very interesting because, as depicted in Table 4-1, diabetic and diabetized mice are expected to have higher BGL than obese mice. This result implies that the amplitude parameter is not only capable to discriminate hyperglycemia condition from normoglycemia condition, but it is also able to distinguish between different hyperglycemic states.

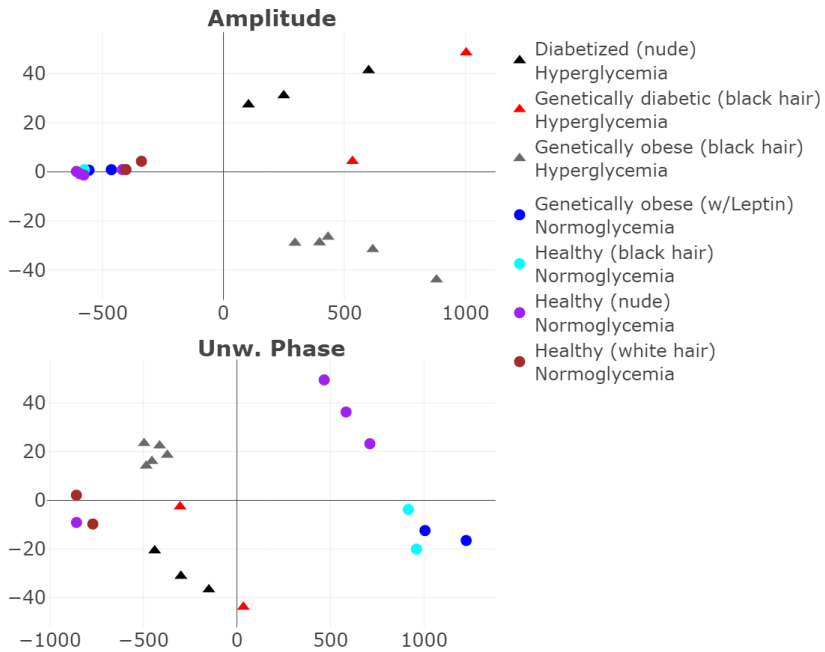


Figure 4-9. Scatterplots of the scores of the first FPC vs scores of the second FPC estimated for amplitude (left panel) and unwrapped phase (right panel) of the transmitted wave. Hyperglycemic and normoglycemic cases are identified by triangles and circles, respectively, and the different types of mice has been identified by colors.

Let us remark that, as shown in Figure 4-9, the obese mice treated with human leptin in Experiment A (see section 4.1.2) were successfully detected as normoglycemic by both FPCs, being grouped together to the healthy mice. This also implies that the non-invasive assessment is capable of detecting changes taking place in obese mice associated to the leptin treatment, which normalize their glycemia.

Then, amplitude of the transmitted wave, directly related to the transmission coefficient of the biological media, has shown to be less sensitive to the outer skin layers among other physical features between mice strains, and provides a potential indicator for the non-invasive discrimination of hyperglycemic states.

Additionally, the skin-fold thickness (μm) of the mice employed in Experiment A (see section 4.1.2) was measured, and the instantaneous BGL (mg/dl) was measured almost simultaneously to the non-invasive assessment during all experiments (Experiment A, B and C). In Figure

4-10, we evaluate the influence of the BGL and skin-fold thickness measurements on the measured spectral response by using the scores associated to the first two FPCs. It is worthwhile to note that the measured BGL values of normoglycemic mice are higher compared to the expected BGL, as shown in Table 4-1. Such increment in the BGL values of normoglycemic mice was attributed to the stress caused on the animals during the measuring process and the effect of the injected anesthesia (see section 4.1.2). Then, we can observe that obese mice are discriminated as hyperglycemic by the first two FPCs, even though they have similar BGL values to the healthy mice during the non-invasive assessment. In fact, obese mice treated with human leptin were detected as normoglycemic since their BGL values were normalized during the last 28 days previous the non-invasive assessment. This means that hyperglycemia discrimination depends on the sustained glycemic state rather than the instantaneous BGL value at the time of the measurement (as provided by a glucometer). From Figure 4-10, it can be also noted that, in some cases, the fold of skin of obese mice is twice as thick as the diabetic mice, which implies that hyperglycemia discrimination by the first FPC is not affected by the skin-fold thickness. On the other hand, an inverse relation may be noted between the skin-fold thickness and the clusters obtained by second FPC. However, we can observe that such relation is poor between observations since strong clustering is present in the scores even though the measured values of the skin-fold thickness are homogeneously distributed.

All the obtained results until now, from analyzing the scores variability estimated by the FPCA, are clear evidence that the amplitude parameter, related to the transmission coefficient of the biological sample, is strongly affected by the hyperglycemia condition in mice. From the same analysis, we experimentally prove that such parameter provides a clear and robust indicator for the hyperglycemia discrimination, being not affected by the biological varieties between the mice strains such as the hair types, skin types or the skin-fold thickness.

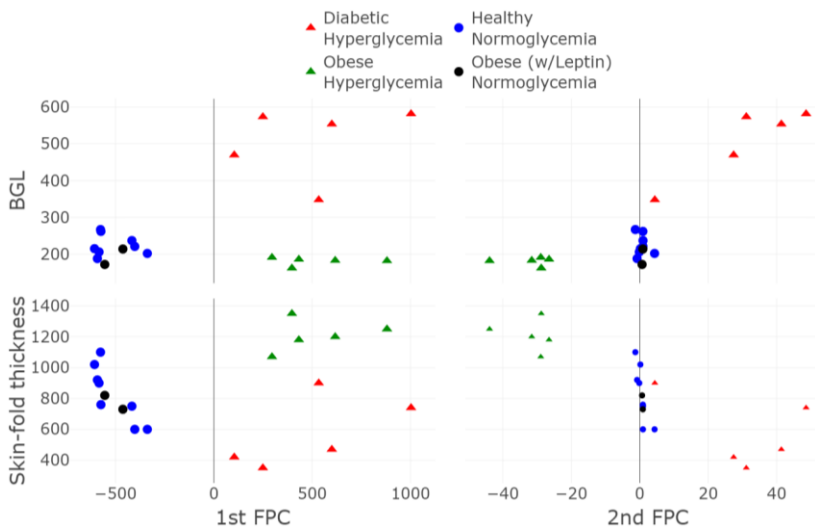


Figure 4-10. Scatterplots of the BGL (mg/dl) and skin-fold thickness (μm) plotted versus the scores of the first two FPCs are shown at the top line and bottom line, respectively. Hyperglycemia and normoglycemia conditions are identified by triangles and circles, respectively. Diabetic cases (genetically diabetic and diabetized mice), obese mice, healthy mice and obese mice treated with leptin are identified by red, green, blue, and black, respectively.

Moreover, no correlations were found between the instantaneous BGL of the mice, measured simultaneously to the non-invasive assessment, and the indicator obtained from the amplitude parameter. This supports that such indicator is sensitive to sustained glycemic states in the mice rather than their instantaneous BGL.

Once we have demonstrated that the proposed spectroscopic approach can discriminate the pathology under study, we move one step further and start asking ourselves how to build an actual diagnostic from this system. In this sense one of the main issues is what is the actual frequency range that is needed for proper classification, i.e. Which are the frequencies that contribute most to discrimination? and, Which is the minimum frequency interrogation span that is needed for proper discrimination? These questions are paramount to obtain a compact and low-cost instrument and will be addressed qualitatively by interpreting the corresponding loading functions of the FPC's. In this sense, the potential of the statistics techniques described in chapter 2 for biomedical instrument design is demonstrated.

4.2.2. INTERPRETING FPCs' LOADING FUNCTIONS

A significant advantage working with FPCA is the interpretation of the loading functions $f_j(t)$ associated to the FPCs. Until now, all the information obtained has been deduced by comparing the computed scores of the FPCs (see Eq. (2-36)), and the available information of the sample mice population, but additional and more detailed information about the achieved discrimination can be obtained by interpreting the corresponding loading functions. The loading functions allows for qualitatively study the relation between the spectral response and the scores variability, which is very useful to understand the relation between clusters.

In Figure 4-11, the loading functions corresponding to the estimated FPCs are shown on the top line, and the same loading functions are plotted as perturbation of the sample mean function on the bottom line. In order to show more clearly the effects of the loading functions on the sample mean, the loading functions were multiplied by the standard deviation of their associated scores and added (+) and subtracted (-) to the mean function of the measured spectral response.

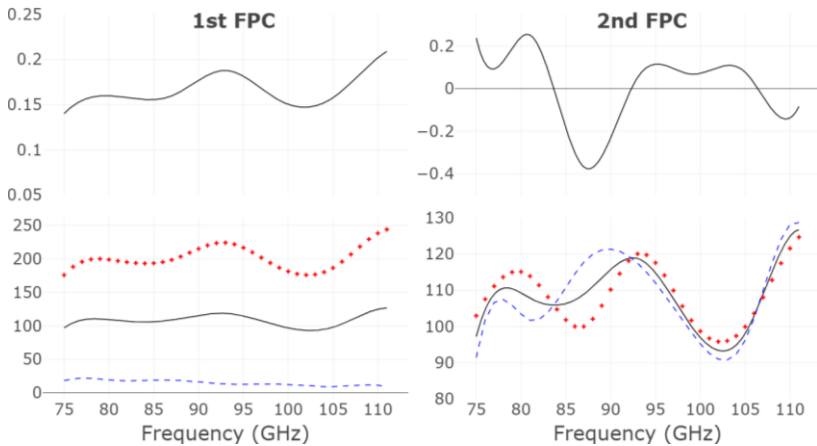


Figure 4-11. (Top) Loading functions associated to the first FPC (left column) and second FPC (right column), and the mean function \pm the corresponding loading function multiplied by standard deviation of their associated scores are shown (bottom).

As we could see, the second FPC allows for discrimination between hyperglycemic states (mild diabetes and full-blown diabetes), and such discrimination is directly related to those frequencies in which the loading function takes negative values, being more determining the frequency interval between 84 GHz - 93 GHz. The observed behavior in the loading function remarks a slope change in the spectral response between obese and diabetic mice, as shown by the loading effect on the sample mean function. Thus, interpreting the loading functions of the FPCs allows to characterize the spectral response for normoglycemic, obese and diabetic mice.

Furthermore, considering the frequencies contribution, observed in Figure 4-11, for the discrimination achieved by the FPCs, we can deduce that the frequency interval between 84 GHz - 93 GHz not only provides a clear discrimination between normoglycemia and sustained hyperglycemia conditions, but also allows to discriminate a mild diabetes from a full-blown diabetes. Then, interpreting the loading functions we are able to identify those frequencies as the main contributions for the discrimination between all the sustained glycemic states involved in the mice sample population.

Identifying major contributing frequencies is a very important issue when developing actual spectroscopic medical diagnostics. As mentioned in chapter 1, the wavelength selection problem has been widely studied in the MDA with the aim to reduce the number of variables involved in the regression model, and to optimize the frequency interval for spectral interrogation [231]–[234]. In this way, the noisy frequencies can be eliminated, reducing the dimension of the learning problem, and simplifying the spectroscopic instrument.

4.3. EVALUATION OF THE CONSISTENCY AND ROBUSTNESS OF THE MEASURED SPECTRAL RESPONSE

In this section, a classification model for sustained hyperglycemia prediction is obtained and analyzed to study the consistency and robustness of the W-band measured transmission spectra. A second experiment described in section 4.1.3 (Experiment B) was carried out to assess normoglycemia and hyperglycemia conditions on a new sample of

mice (independent from sample population employed in Experiment A) using the diagnostic technique. The classification model, based on FPCLoR introduced in 2.5.1, is estimated and tested by a multi-test analysis with hundreds of iterations using the spectral data obtained from Experiment B. As will be shown later, the multi-test analysis allows to evaluate the performance and robustness of the fitted model. Additionally, the results obtained from both approximations, regression splines and P-splines of the sample curves will be compared. Then, the estimated FPCLoR model is used to predict sustained hyperglycemia on the sample population measured in Experiment A (described in 4.1.2). In this way, we are able to validate the consistency of the W-band measured transmission spectra for sustained hyperglycemia detection, which is important to support the claim that the non-invasive assessment by the mm-wave spectrometer provides a characteristic response associated to the sustained hyperglycemia condition in mice.

From now on, exclusively spectral data corresponding to the transmission amplitude measurements will be considered since has been previously proved that such parameter is less sensitive to the outer skin layers among other biological features associated to the different mice strains involved in the sample mice population (see section 4.1.1).

The measured spectra (raw data) from sample mice population of Experiment B is shown in Figure 4-12, identifying normoglycemia and hyperglycemia conditions by dotted blue lines and dashed red lines, respectively.

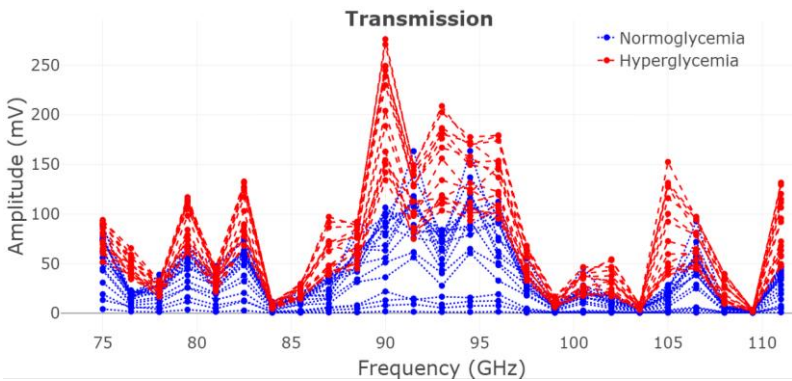


Figure 4-12. Amplitude spectra of the transmitted wave through the fold of skin. Normoglycemic and hyperglycemic cases are identified by dotted blue lines and dashed red lines, respectively.

In contrast to Experiment A, hyperglycemic and normoglycemic mice shows the same absorption level at several frequencies. This fact is attributed to the age difference; hyperglycemic mice from Experiment B are almost five months younger than hyperglycemic mice from Experiment A (see Table 4-1 and Table 4-2). The presence of sustained medium/long-term hyperglycemia led to dehydration problems and could reduce the thickness of the skin layers, resulting in higher transmittance levels for older mice. Then, considering the different absorption levels of the hyperglycemic mice between experiments, and that output power of signal generator was recalibrated due to hardware and software changes in the mm-wave spectroscopy system between experiments (see section 4.1.3), the transmission amplitude was normalized according to the mean of each sample previous the statistical analysis. This fact also demonstrated the robustness of the instrument because, as we can see in this section, we will be able to properly classify the different metabolic states regarding these aspects.

For comparison purposes, two functional data sets with different smoothness degree were estimated using regression splines and P-splines. Figure 4-13 shows the approximated curves for a normoglycemic and hyperglycemic case using both approaches. The cubic regression splines (top panel) were defined on 17 knots strategically positioned to reproduce most of the variability in raw data. On the other hand, the P-splines (bottom panel) were defined on the same number of knots (17 knots), equally spaced over the W-band, with the smoother parameter $\lambda = 0.11$ selected by the GCV method (see section 2.3.3.1).

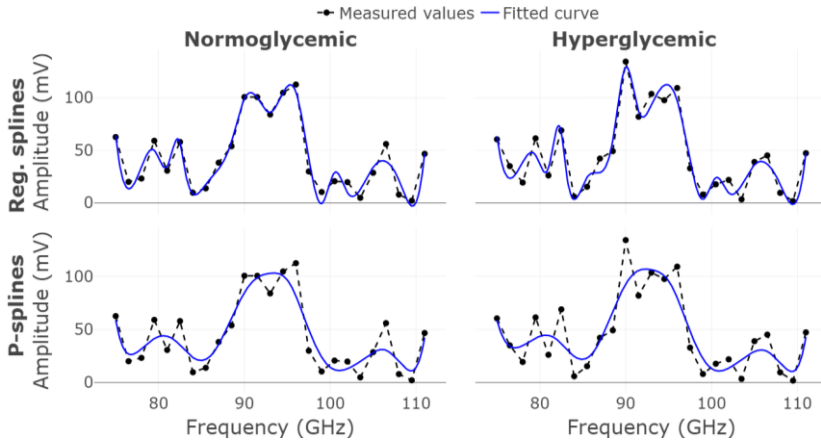


Figure 4-13. Estimated sample curves from the measured transmission amplitude for a normoglycemic (left panel) and a hyperglycemic mouse (right panel) by cubic regression splines (at the top) and P-splines (at the bottom).

4.3.1. MULTI-TEST ANALYSIS: PERFORMANCE AND ROBUSTNESS OF THE PROPOSED APPROACH

As mentioned above, the mice sample population of thirty-three mice corresponding to the Experiment B was used for the multi-test analysis. The multi-test analysis consisted of estimate a FPCLoR model using the 80% of the sample curves (training sample) and predict the condition (outcome) on the remaining 20% (test sample) repeatedly for one hundred times. Both subgroups, the training and test sample, were selected randomly at each iteration preserving the original proportion of the classes within the global group: 54% of the cases are normoglycemic and 46% are hyperglycemic.

In the literature on logit regression, measures such as the Area Under Curve (AUC), the True Positive Rate (TPR), the True Negative Rate (TNR), and the Correct Classification Rate (CCR) are very useful parameters to quantify the goodness of fit and predictive capability of the classification model. The TPR and TNR, also known as sensitivity and specificity, respectively, are commonly used in medical diagnostics [235]. The TPR, TNR and CCR values are estimated taking into account the confusion matrix, shown in Table 4-5, as follows:

$$TPR = \frac{TP}{TP + FN}, TNR = \frac{TN}{TN + FP}, CCR = \frac{TP + TN}{TP + TN + FP + FN}$$

Table 4-5. Confusion matrix

		True condition	
		Hyperglycemia	Normoglycemia
Predicted condition	Hyperglycemia	Hyperglycemia correctly classified (True Positive TP)	Normoglycemia misclassified (False Positive FP)
	Normoglycemia	Hyperglycemia misclassified (False Negative FN)	Normoglycemia correctly classified (True Negative TN)

The Receiver Operating Characteristic (ROC) curve [236] shows the inverse relation between the sensitivity and the specificity (sensitivity vs $1 - \text{specificity}$) varying the diagnostic criterion (cut-off value to assign $y = 1$) of the test, and the AUC (which refers to the area under the estimated ROC curve defined by the integral equation $\int_a^b f(x)dx$), can be approximated by numerical integration algorithms [237], provides an effective measure of the diagnostic accuracy of the predictive model.

In are summarized the obtained results of the multi-test analysis for both approximations: regression splines and P-splines (see Figure 4-13). As it can be seen, the mean value of AUC is above 0.95 for both cases, being slightly higher for the model based on regression splines. The achieved AUC values indicate an excellent discrimination capability of the inter-condition (normoglycemia and hyperglycemia), with very robust results since the estimated standard deviation is under 0.02. Also, a very good classification rate on new observations (prediction) was achieved for both cases, evaluated by the CCR value, with a mean value of 82% of correctly predicted cases by the FPCLoR model based on P-splines and increased in a 10% by the regression splines. These results suggest that the FPCLoR model obtained from regression splines is the best in terms of the prediction capabilities, being considerable more accurate to detect the sustained hyperglycemia in mice, as depicted by the TPR values shown in Table 4-6.

Table 4-6. Summary of the multi-test analysis.

Parameter	Regression splines		P-splines	
	Mean	Std. Dev.	Mean	Std. Dev.
AUC	0.99	0.01	0.96	0.02
CCR	0.92	0.08	0.82	0.11
TPR	0.91	0.18	0.74	0.23
TNR	0.93	0.11	0.87	0.18

Nevertheless, as mentioned above, a significant advantage of working with the FPCLoR is the interpretation of the estimated functional parameter $\beta(t)$, which is a potential source of information to calibrate and optimize the applied spectroscopic technique. Here, we can emphasize that the lack of smoothness in approximated sample curves is reflected in the discriminating functional parameter. Figure 4-14 shows one of the functional parameters estimated for both approaches, based on regression splines (left panel) and P-splines (right panel). Comparing both beta functions, we can see that the corresponding to the FPCLoR model based on regression splines exhibits strongest oscillation versus frequency and such variability makes its interpretation very difficult. As shown in Functional parameter interpretation, the FPCLoR is interpreted in terms of the odds ratio, which is estimated by the integral of the beta function (see Eq. (2-56)). The odds ratio, previously introduced in section 2.5.1.2, is a statistic tool that allows us to measure the influence of the W-band measured transmission spectra in diagnosis by introducing variations in the beta function estimated by the FPCLoR model.

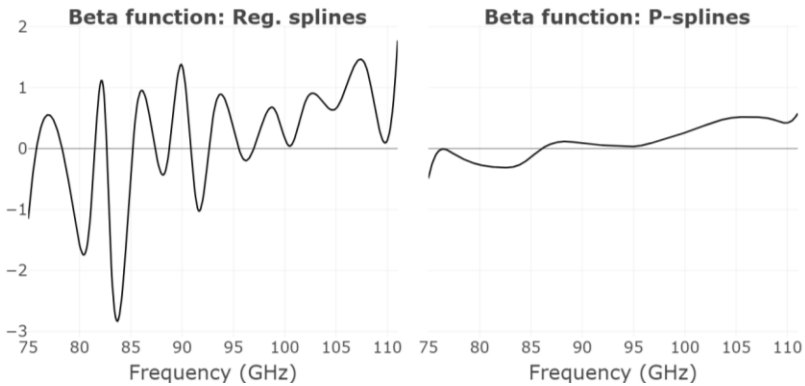


Figure 4-14. Functional parameter $\beta(t)$ estimated for FPCLoR on the functional data approximated by regression splines (left panel) and P-splines (right panel)

Then, for a noisy beta function, as the obtained from regression splines, the odds ratio must be estimated considering very short frequency intervals so that the integral of the beta function does not tends to zero, which implies an estimated odds ratio close to one. This minimizes the frequencies contribution for discrimination and a lot of information is lost since too few frequencies are considered in the estimation of the odds ratio. In contrast, a smoother beta function, as the obtained from P-splines, allows to identify wider frequency intervals estimating the odds ratio, and consequently, an easier and clearer interpretation from the FPCLoR model can be obtained. Therefore, we will work with the P-splines since the resulting functional parameter offers a much better interpretation of the FPCLoR model, being such feedback very important when developing applied spectroscopic techniques, and both FPCLoR models provide excellent multi-test results.

4.3.2. VALIDATION OF THE CONSISTENCY OF THE APPLIED SPECTROSCOPIC TECHNIQUE AND ANALYSIS OF THE FPCLoR MODEL

Two sample mice populations evaluated separately at different experiments were employed to validate the consistency of the spectral response, measured by the mm-wave spectroscopic instrument, for the non-invasive detection of sustained hyperglycemia. This time, the FPCLoR model was estimated using all the spectral data from Experiment B as the training sample, consisting on eighteen normoglycemic mice plus fifteen hyperglycemic mice, and a group of twenty mice with normoglycemia and hyperglycemia conditions proportionately distributed (see section 4.1.2) were used as the test sample. Functional data from both spectral data sets were estimated by using P-splines defined on seventeen equally spaced knots as shown in Figure 4-13 (bottom line). The ROC curve computed for the fitted FPCLoR model is shown in Figure 4-15, with the $AUC = 0.95$. Furthermore, testing the predictive capabilities of the fitted on the test sample, we obtain a CCR of 100% with all the new observations correctly classified. These results probe the consistency of the proposed non-invasive approach for sustained hyperglycemia detection in mice, supporting the reliability of the mm-wave spectroscopic technique.

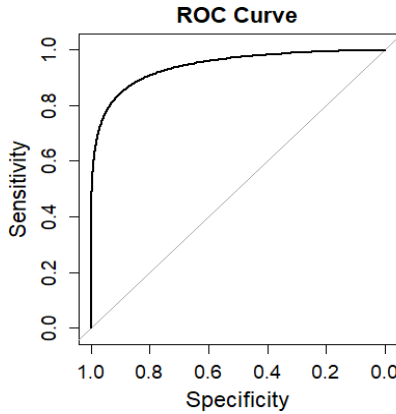


Figure 4-15. ROC curve of the fitted FPCLoR model.

Then, as one of the main objectives of this work, the next step is to show how to analyze the functional parameter estimated for the fitted FPCLoR model in terms of the odds ratio so we can extract information to optimize the diagnostic system for its eventual use in clinical practice. From Figure 4-16, it can be notice that the beta function varies from negative to positive values versus frequency with a zero crossing at 86 GHz, indicated by the red line. Then, the W-band is subdivided into two frequency intervals, which inversely relates the measured spectral response and the sustained hyperglycemia discrimination. This effect is illustrated by computing and interpreting the odds ratio for both sub-frequency intervals, considering a constant increase in the transmission amplitude of 0.3 mV ($K = 0.3$) for the measured spectral response, as shown in Eq. (2-56). For the first frequency interval the estimated odds ratio $OR_{75-86}^{0.3} = 0.516$ indicates that such a constant increment in the transmitted amplitude for the frequencies under 86 GHz reduce the possibilities of being diagnosed with sustained hyperglycemia to one half. In contrast, the estimated odds ratio for the second frequency interval $OR_{86-111}^{0.3} = 6.52$ indicates that the possibilities of being diagnosed with sustained hyperglycemia is six-fold. In this way, we are able to measure the relationship between the measured spectral response and the diagnosis.

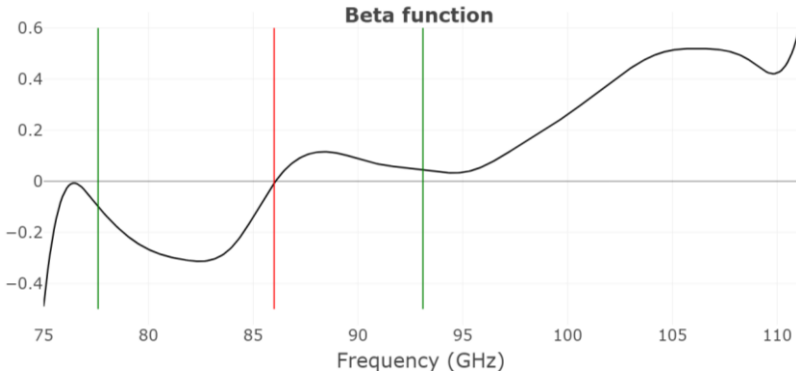


Figure 4-16. Discriminating function $\beta(t)$ for sustained hyperglycemia discrimination estimated by the FPCLoR model. The red line indicates a sign change for the function at the frequency of 86 GHz. The green lines indicate the frequencies, chosen arbitrarily, as the lower and upper limits for a reduced frequency interval.

Then, considering the observed relation between the spectral response and the discrimination of sustained hyperglycemia, we will consider a reduced frequency band around the frequency (~ 86 GHz) in which the beta function crosses zero, with the lower and upper frequency limits chosen arbitrarily to be 78 GHz and 93 GHz, respectively, indicated by the green lines in Figure 4-16. The delimited frequency range 78 GHz - 93 GHz provides two regions inversely relating the spectral response to the sustained hyperglycemia condition, which enhance the achieved discrimination. The reduction of an operating frequency range represents a drastic improvement in measurement time, and it can also have an important impact in the spectroscopic instrument complexity, demonstrating the utility of these tools for biomedical instrumentations design.

However, since less frequencies are considered, less information is provided to the regression model, and that may affect the achieved discrimination. Therefore, the sustained hyperglycemia detection was re-evaluated by repeating the multi-test analysis shown in previous section, and the validation test, considering only the contribution of the reduced frequency range (78 GHz - 93 GHz). All the sample curves were approximated by P-splines defined on eleven equally spaced knots with $\lambda = 0.528$ chosen by GCV method (see section 2.3.3.1). Table 4-7 summarizes the obtained results re-evaluating the multi-test analysis and validation test using the measured spectral response corresponding to the reduced frequency range.

Table 4-7. Obtained results for multi-test analysis and validation test considering only frequencies from 78 GHz and 93 GHz.

	Parameter							
	AUC		CCR		TPR		TNR	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Multi-test analysis	1	0	0.91	0.1	0.94	0.14	0.92	0.16
Validation test	1		0.95		0.90		1	

The results of the multi-test analysis were improved when compared to those shown in Table 4-6, and the CCR of the validation test decreased only 5% compared to the used of the whole frequency range, even though the frequency band was reduced to the half. These results show that the reduced frequency band originates the major contribution for sustained hyperglycemia discrimination within the W-band, and the remaining frequencies introduce some noise to the discrimination analysis.

4.4. SENSITIVITY OF THE NON-INVASIVE APPROACH TO DETECT CHANGES IN SUSTAINED GLYCEMIA

As shown in section 4.2, the non-invasive approach for in-vivo hyperglycemia detection was capable to differentiate between a mild diabetes and a full-blown diabetes, suggesting that the measured spectral response in the W-band can be used to monitor different glycemetic states. Such results motivate the realization of a third experiment focused to test the sensitivity of the proposed non-invasive approach to detect changes in metabolism of mice related to gradual increments in glycemia due to an uncontrolled diabetes, i.e. to be able to study the temporal evolution of the metabolomic condition (longitudinal study).

In this section, the evolution of the spectral response from a group of sixteen diabetized mice (see section 4.1.1) measured in Experiment C (see section 4.1.4) is studied during their transition from normoglycemia to hyperglycemia condition. The condition of the diabetized mice is predicted at each measurement day using a classifier based on the first FPC estimated from the FPCA performed on section 4.2. Using the

classifier, which is obtained from a different sample population, allows to test the capability of the proposed approach for prediction of different glycemic states on new cases, besides the sustained hyperglycemia detection proved in section 4.3. Additionally to the diabetized mice, six diabetic mice with full-blown diabetes and six healthy mice were measured simultaneously as reference for hyperglycemia and normoglycemia conditions respectively (see Table 4-3). As described in section 4.1.4 (Experiment C), the sample population (sixteen diabetized mice, six healthy mice, and six diabetic mice) was evaluated at four different days along fourteen days between the first and last measurement. Due to the fast metabolomic processes in mice, a duration of fourteen days should give a clear indication of glycemic conditions in the animals. Measurements had started when the diabetes induction process was completed, hence, the first measurement corresponded to the final dose day (the fifth day after the applications of the first dose of STZ).

As in previous section, the sample curves were estimated using P-splines defined on 17 equally spaced knots over the W-band, with the smoother parameter $\lambda = 0.11$. The obtained functional data sets at each day are shown in Figure 4-17. As depicted in the figure, transmittance level of diabetized mice increase over time until reach similar values to those measured for diabetic mice.

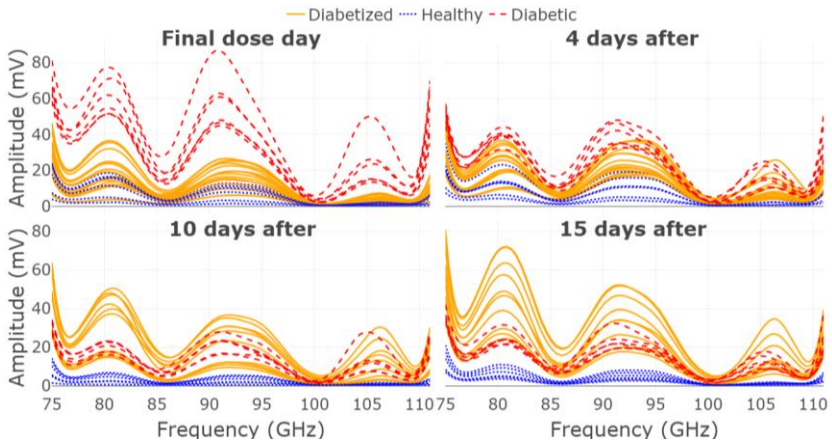


Figure 4-17. Approximated sample curves from amplitude of the transmitted waves through the fold of skin at the four measurement days. Healthy, diabetic and diabetized cases are identified by dotted blue lines, dashed red lines, and orange lines, respectively.

The hyperglycemic state in diabetized mice at each measurement day was blindly predicted using the loading function associated to the first FPC estimated from the FPCA performed on amplitude spectra in section 4.2 (which explains 99% of the total variance of the data). As before, the transmitted amplitude was normalized according the mean of each sample data set to compensate for variations in amplitude ranges of spectral data sets between experiment A and experiment C. The mean value of the predicted scores by groups, with the corresponding standard deviation, for all measurement days are shown in Figure 4-18. In this figure, we can clearly see that the diabetized mice evolved from scores between the diabetic and healthy mice towards the diabetic region of the positive scores. Such an evolution was confirmed by the evolution of the measured blood glucose levels for the diabetized mice at the same days, shown in the Figure 4-19. Comparing the described trajectories for diabetized mice by the scores and BGLs, it can be clearly deduced that the sensing approach can closely follow the evolution from normoglycemia to hyperglycemia condition right from the beginning of the test.

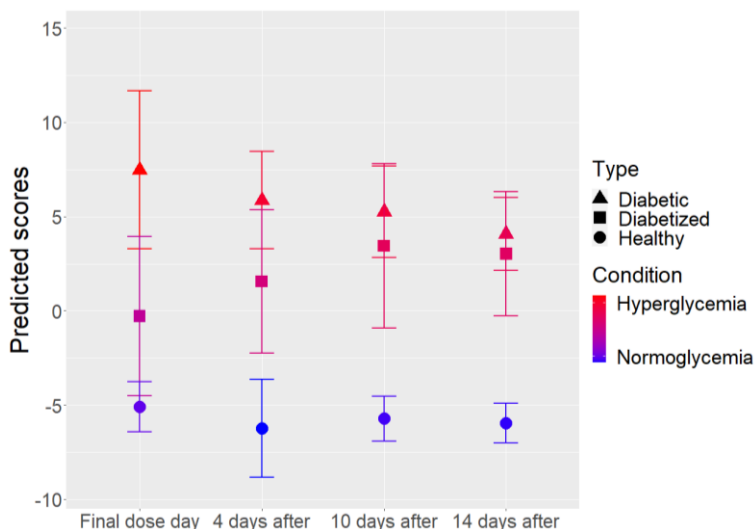


Figure 4-18. The mean score determined for each group with the corresponding standard deviation are shown as a function of the measurement days. The healthy, diabetized, and diabetic group are identified by circle, square, and triangle, respectively.

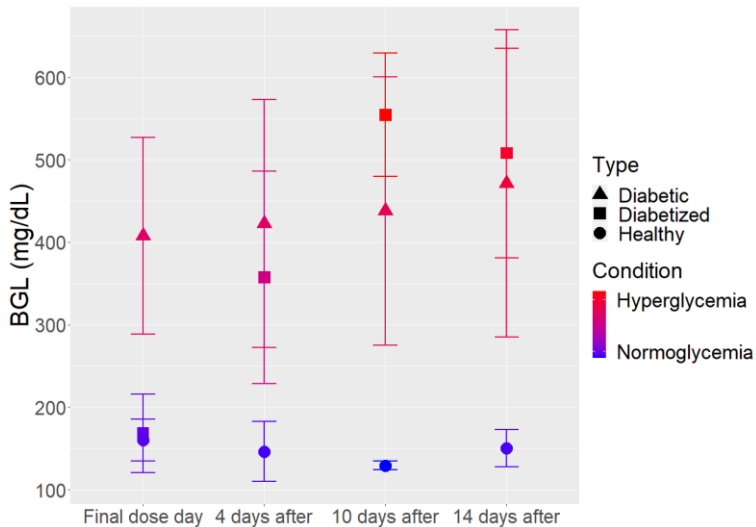


Figure 4-19. The mean value of the instantaneous BGLs measured for each group with the corresponding standard deviation are shown as a function of the measurement days. The healthy, diabetized, and diabetic group are identified by circle, square, and triangle, respectively.

It is important to remark that the biological response of each mouse to the diabetes induction process may vary as some mice are more resistant to the drug (STZ). This biological variation is also captured by the proposed non-invasive approach, with some diabetized mice already identified as hyperglycemic at the final dose day, even though the measured BGL is still very similar to the healthy mice. This suggests that the proposed non-invasive approach allows for the prompt detection of diabetes effects in sustained hyperglycemia. The remaining cases have a slower transition accounting for different speeds of development of the induced diabetes in mice. Nevertheless, most of the diabetized mice were predicted as hyperglycemic four days after the final dose day, and the average predicted scores of diabetized and diabetic mice at the last two days were very similar, specially at day fourteen.

The obtained results indicate that the measured spectral response in the W-band allows to detect changes in mice metabolism with four days or less of sustained hyperglycemia, and a full-blown diabetes in mice two weeks after the induction. Moreover, normal BGLs measured on the diabetized mice at the final dose day support the obtained results in section 4.2 demonstrating that the proposed non-invasive approach is sensitive to

the effects of sustained hyperglycemia in metabolism rather to the instantaneous BGLs.

4.5. DISCUSSION AND CONCLUSIONS

In this chapter, a new non-invasive approach for hyperglycemia detection, based on spectral interrogation in the W-band, has been tested on animal models using the FDA methods introduced in chapter 2. A first experiment was carried out by non-invasively assessing a sample of twenty mice of different strains representing three glycemic states (normoglycemia, mild diabetes and full-blown diabetes). The measurements were performed directly on a fold of skin in their back, and the instantaneous BGL and the skin-fold thickness were measured almost simultaneously. The amplitude and phase parameters of the reflection at the top skin layers and the transmission through the fold of skin were analyzed separately. The sample curves estimated from the measured spectra were analyzed by interpreting the FPCA results. The computed FPC's scores from the reflection and transmission parameters allows to detect hyperglycemia condition, and to distinguish other biological features associated to the different mice strains involved in the experiment. However, it could be concluded that the amplitude parameter from transmission is more adequate for hyperglycemia discrimination since appears to be not affected by the biological varieties of the different mice strains involved in the sample mice population. The FPCA on amplitude parameter of transmission provides a robust indicator not only for hyperglycemia detection, but also for discrimination between a mild diabetes and full-blown diabetes in mice. In addition, it was experimentally proven that such discrimination of the hyperglycemic states does not depend on the instantaneous BGL or the skin-fold thickness. Another interesting observation in the achieved discrimination is that obese mice suffering hyperglycemia condition for five months were detected has normoglycemic after on month of having normoglycemia by a human leptin treatment. At the view of all the obtained results, it is clear that the proposed approach is rather robust against surface skin properties, in contrast to optical methods, which are strongly affected by the skin characteristics such as the hair, subepidermal fat, skin thickness or melanin content, among others. Besides that, the fact that the proposed approach is not directly monitoring the instantaneous BGL points to other metabolites associated to sustained hyperglycemia in metabolism, such as

the AGEs. This supports that the non-invasive assessment is sensitive to detect metabolic changes related to sustained hyperglycemic states.

On the other hand, we were able to show that the relation between the spectral response and the discrimination achieved can be quantitatively studied by interpreting the loading functions of the FPCs. In this way, the FPCA was able to characterize the spectral responses associated to the discrimination of the sustained hyperglycemic states.

The obtained results from the first experiment encouraged us to perform two more experiments with the aim to validate the hyperglycemia detection on mice, and to study the sensitivity of the proposed approach to detect different sustained hyperglycemic states.

In the second experiment, a new sample population of mice consisting on eighteen normoglycemic mice and fifteen hyperglycemic mice was non-invasively assessed by the applied spectroscopic technique. The measured amplitude, related to the transmission coefficient of the biological sample, was used to estimate a predictive model for classification of sustained hyperglycemia by using FPCLoR. A multi-test analysis, with hundred iterations, was developed to test the performance of the predictive model. During the multi-test analysis two different smoothness degrees were considered for the fitted sample curves to show the tradeoffs between them in the FDA and highlight the importance of obtaining smooth sample curves. The multi-test analysis shows that the FPCLoR models estimated from both approaches provide excellent performance in terms of the goodness of fit and prediction capabilities, with an AUC of the ROC curve > 0.95 and a CCR $> 80\%$. It should be noted here that, regarding on the sustained hyperglycemia discrimination, the obtained results from the multi-test analysis demonstrates that such discrimination do not depend on the choice of a particular splines or fitting function. On the other hand, we show that the lack of smoothness in the fitted sample curves affects the interpretations and analysis of the discriminating functional parameter estimated by the FPCLoR model. Then, the non-invasive assessment for sustained hyperglycemia detection was validated using the estimated FPCLoR model to predict the condition of the twenty mice measured in the first experiment, with a 100% of correct classification rate. Such results prove that the amplitude of the coefficient transmission in the W-band provides a robust and reliable indicator for sustained hyperglycemia prediction, with no prior calibration required for the amplitude parameter.

On the other hand, to remark the great potential of the presented statistical methods in the development of applied spectroscopy systems to build medical diagnostics, the estimated functional parameter of the FPCLoR model was analyzed and interpreted in terms of the odds ratio. This analysis allows to obtain a quantitative measure relating the measured spectral response and the sustained hyperglycemia detection. Such measure can be considered for further calibration of the applied spectroscopic technique. Moreover, by the same analysis, we were able to identify a narrow frequency interval between 78 GHz and 93 GHz as the potential contributing frequencies for the achieved discrimination of sustained hyperglycemic states, previously highlighted by interpreting the FPCA performed in the first experiment. From the obtained feedback by the statistical analysis we experimentally prove that the operating frequency band of the spectroscopic instrument can be substantially reduced without affects the performance of the predictive model. This is very important not only for optimization of the applied spectroscopic technique, but also for further in deep analysis of the condition because it could lead to identify novel biomarkers indicative of metabolic alterations related to the development of diabetes.

Finally, it was experimentally proven that the proposed non-invasive approach is able to monitor temporal changes in sustained glycemia associated to uncontrolled diabetes in a group of sixteen mice with a drug-induced diabetes. Using a classifier obtained from the first experiment we were able to clearly track the evolution of diabetized mice from normoglycemia to hyperglycemia, enabling the prompt detection of hyperglycemia condition. These results support the great potential of the proposed approach not only for sustained hyperglycemia detection but also for its early detection and monitoring with a temporal resolution of few days. Although the minimum length of time required for the method to detect changes in metabolism associated to different sustained glycemic states is still not determined, the results indicate that different sustained hyperglycemic states related to uncontrolled diabetes can be detected with only four days of occurrence.

In summary, using the FDA to study the measured spectral response by the spectroscopic instrument we were able to extract significant information of the tested non-invasive approach for hyperglycemia detection. We experimentally probe that the amplitude parameter of transmission provides a robust indicator (not affected by skin thickness, subepidermal fat or other biological varieties among the mice strains) with

high potential for the detection of sustained glycemic states in mice. This simplifies the overall spectroscopy system resulting in a simple transmission-type spectrometer with limited frequency bandwidth of operation, which significantly reduces the complexity and cost implementation of the spectroscopic instrument. These results show the great potential of FDA in spectral data analysis for characterization and optimization of applied spectroscopic techniques. It was also demonstrated that the proposed approach can be potentially used to monitor changes in sustained glycemia associated to diabetes with a resolution of four days between measurements. The spectroscopic system provides results in seconds, do not require frequent calibration, it is not based on consumable materials, and can be made very compact and realized cost-efficiently. Therefore, we consider that the proposed approach has great potential in developing a new non-invasive technique for diabetes monitoring not depending on the instantaneous BGL.

5. USE OF W-BAND
SPECTROSCOPY FOR IN-
VIVO AND NON-INVASIVE
ASSESSMENT OF
HYPERGLYCEMIC STATES IN
HUMANS: PILOT CLINICAL
STUDY

As described in section 3.1, sustained hyperglycemic metabolism leads to the irreversible formation of AGEs, (free sugars in blood that bind permanently to proteins and lipids), that have strong effects on many body organs and tissues. This pathological condition is typical in diabetic patients and is the underlying cause of all the complications associated to DM. In this sense, the appearance of the HbA1c test, which provides a measure of the glycated hemoglobin associated to sustained hyperglycemia, is a recent indicator that has improved the diagnosis and monitoring of DM (since 2010). The HbA1c measurement is a more stable indicator than BGL measurement, and consequently, more convenient for medical treatment decisions in diabetes care. Nevertheless, the HbA1c test can only be used for long-term diabetes monitoring and control due to the fact that the general consensus is that the HbA1c is a good indicator of the average hyperglycemic metabolism over a period three months [238].

The observed characteristics of the non-invasive sustained hyperglycemia measurement approach, described and experimentally assessed using animal models in chapter 4, indicate that the proposed system offers a comparable measure of the metabolomic state of the animal to that provided by the HbA1c test, but the lead times required to detect the sustained hyperglycemia condition are much shorter compared to the periods required for the HbA1c test. For this reason, a preliminary clinical study with human patients was proposed to compare the non-invasive measure of sustained hyperglycemia using mm-wave spectroscopy and the HbA1c standard measurement protocol in type 1 DM patients. The objective of this study, in collaboration with the Instituto de Investigación Sanitaria Fundación Jiménez Díaz (Servicio de Endocrinología y Nutrición) was to validate the actual performance of the developed diagnostic in humans.

In this chapter the spectral profiles in the W-band corresponding to type 1 DM patients obtained from a pilot clinical test will be analyzed. The sample population was assessed repeatedly during three different visits over a period of seven months. In line with the previous work, the spectral data analysis is performed using FDA methods, but with the HbA1c value as the response variable of interest.

The collected data sets will be firstly analyzed separately (by visit) using scalar-on-function regression to study the linear relation between the W-band spectral response and the HbA1c value. Then, the spectral data obtained from all visits will be analyzed simultaneously using LFDA methods. More detailed information about the sample population and the measurement protocol, as well as the obtained statistical results related to the pilot clinical test, will be presented throughout this chapter.

5.1. DESIGN OF THE PILOT CLINICAL TEST

The clinical study protocol was developed in conformity with international and national standards established for realization of human experiments, and safeguarding confidentiality of participants (see section 8.1.2). All the participants gave written informed consent after having understood about the objectives and protocol of the clinical study in the recruitment stage (see section 8.2).

A total of thirty subjects were included in the sample population: ten healthy volunteers (non-diabetics subjects) and twenty volunteer patients with clinically diagnosed type 1 DM. Only type 1 DM was considered in the pilot clinical study since affects homogeneously to the population, in contrast to type 2 DM which is more frequent in older adults. The study subjects recruited for the pilot clinical study were adults ranging from 25 to 79 years old and the gender distribution was 60 % women and 40 % men.

The diagnostic criteria considered for inclusion of type 1 DM patients in the pilot clinical study are the following:

- a. The presence of carinal signs of DM such as polyuria, polydipsia, and unexplained weight loss, with plasma glucose ≥ 200 mg/dl or diagnosis of diabetic ketoacidosis.
- b. FPG (≥ 8 hrs) ≥ 126 mg/dl.
- c. Plasma glucose ≥ 200 mg/dl after two hours in the OGTT (75g of glucose).
- d. HbA1c ≥ 6.5 % (according to the National Glycohemoglobin Standardized Program and standardized by the Diabetes Control and Complications Trial).

The exclusion criteria were diagnosis of type 2 DM or monogenic diabetes, pregnancy, severe renal or hepatic insufficiency, poor short-term prognosis (< 6 months), treatment with glucocorticoids or immunosuppressive medications, or who have been hospitalized in the past three months. Two different types of metabolic control were equivalently included within the diabetic group: patients with good metabolic control (HbA1c < 7 %) and patients with poor metabolic control (HbA1c > 8 %), according to their clinical records at the time of the recruitment. The sample population related to the pilot clinical test is summarized in Table 5-1.

Table 5-1. Sample population evaluated in the pilot clinical test.

Condition	HbA1c	Quantity
Non-diabetic	-	10
Diabetics with good metabolic control	< 7 %	10
Diabetics with poor metabolic control	> 7 %	10

The general outline of the pilot clinical test is shown in Table 5-2 indicating the tests carried out at each visit. The evaluation process was repeated at three visits within a period of seven months, with three months between visits 1 and 2, and four months between visits 2 and 3. All the evaluations were scheduled to match with regular medical appointments at the Hospital Universitario Fundación Jiménez Díaz.

At each visit, subjects were non-invasively assessed using the mm-wave spectroscopy system (see section 3.3) to obtain their spectral response within the W-band. The frequency step was halved to increase the measured frequencies (49 frequencies), resulting in a frequency sweep from 75 GHz to 111 GHz in steps of 0.75 GHz.

Table 5-2. General outline of the pilot clinical test.

Visit 0	<ul style="list-style-type: none"> • Recruitment (compliance assessment of selection criteria). • Supplementing informed consent forms.
Visit 1	<ul style="list-style-type: none"> • Blood analysis (hemogram and general biochemistry). • Non-invasive assessment using the mm-wave spectroscopic instrument. • Fasting BGL (FBGL) measurement. • Skin-fold thickness (SFT) measurement.
Visit 2	<ul style="list-style-type: none"> • Blood analysis (hemogram and general biochemistry). • Non-invasive assessment using the mm-wave spectroscopic instrument. • Fasting BGL (FBGL) measurement. • Skin-fold thickness (SFT) measurement.
Visit 3	<ul style="list-style-type: none"> • Blood analysis (hemogram and general biochemistry). • Non-invasive assessment using the mm-wave spectroscopic instrument. • Fasting BGL (FBGL) measurement. • Skin-fold thickness (SFT) measurement.

As shown in Figure 5-1, the non-invasive measure was performed on a skin fold located in the first interdigital space (between the thumb and index finger) of the right hand. The first interdigital space was chosen because it is highly vascularized [239], [240], the skin is relatively thin and is a comfortable location to perform the measurement. Additionally, an elbow support was incorporated to the structure of the mm-wave spectroscopic instrument for subjects' comfort during the non-invasive assessment and reduce arm movements and fatigue. No special indications concerning the non-invasive assessment were given to the participants, and all the patients continued their medical treatments at all times.

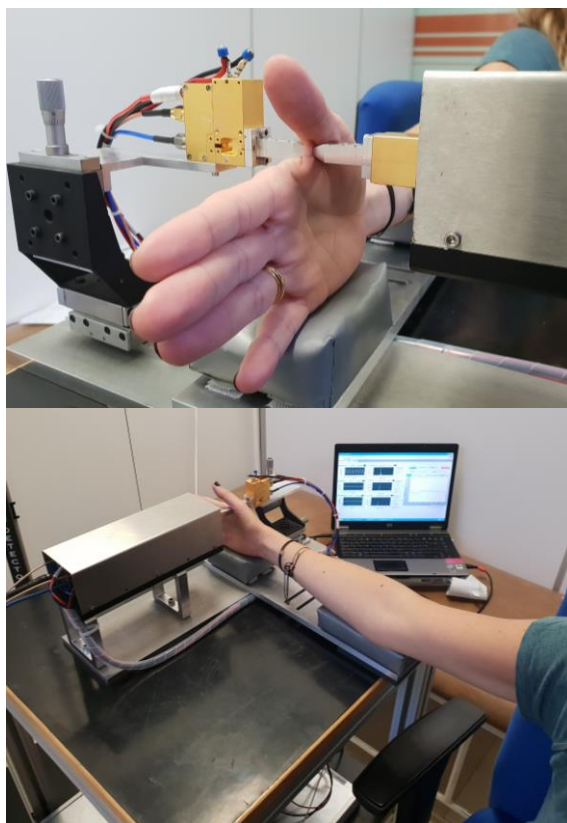


Figure 5-1. Photographs taken during the non-invasive assessment of sustained hyperglycemia using the mm-wave spectroscopic instrument installed at the Hospital Fundación Jiménez Díaz.

Besides the non-invasive assessment, other relevant physiological variables were collected such as age, weight, height, Body Mass Index (BMI), Skin-fold thickness (SFT), Systolic Blood Pressure (SBP) and Diastolic Blood Pressure (DBP). The clinical variables considered in the blood analysis are total hemoglobin (HGB), Glycated Hemoglobin (HbA1c), Triglycerides (TG), total Cholesterol (Chol), High-density Lipoprotein (HDL), Low-density Lipoprotein (LDH), Glomerular Filtration Rate (GFR), Thyroid-stimulating Hormone (TSH), Aspartate Transaminase (AST), and Alanine Transaminase (ALT). All the clinical variables considered during the statistical analysis of spectral data collected for pilot clinical test are listed and briefly described in Table 5-3.

Table 5-3. Clinical/physiological variables considered in the statistical analysis of spectral data.

Variable	Description	Medical Relevance
Age	-	-
Weight	-	-
Height	-	-
Body Mass Index (BMI)	Rate between weight (Kg) and the square of height (m ²).	Overweight (fatness) indicator.
Systolic Blood Pressure (SBP)	Pressure within arteries during active cardiac contraction.	Associated to cardiovascular diseases, arterial aneurysm, kidney failure, hormonal abnormalities, sepsis, cardiogenic shock, among others.
Diastolic Blood Pressure (DBP)	Pressure within arteries during heart relaxation (between heart beats).	Indicator of mean BGL of the last three months used for sustained hyperglycemia detection and DM diagnosis and control.
Glycated Hemoglobin (HbA1c)	Hemoglobin chemically linked to sugar.	Indicator of mean BGL of the last three months used for sustained hyperglycemia detection and DM diagnosis and control.
Fasting Blood Glucose Level (FBGL)	Free sugar in blood after an overnight fast.	Indicator for DM diagnosis and control.
Creatinine (Cr)	Waste product generated by the daily function of muscles.	Biomarker associated to renal insufficiency and kidney failure.
Glomerular Filtration Rate (GFR)	The glomerular filtration is the process of blood filtering by kidneys.	Used as biomarker to test kidneys function.
Aspartate Transaminase (AST)	Important enzymes in the amino acid metabolism, and are found in many organs and tissue such as liver, hearth, kidneys, muscle tissue, etc. The ALT enzymes are found primarily in liver.	Commonly used as biomarkers to test liver function by estimating the AST/ALT ratio.
Alanine Transaminase (ALT)	Important enzymes in the amino acid metabolism, and are found in many organs and tissue such as liver, hearth, kidneys, muscle tissue, etc. The ALT enzymes are found primarily in liver.	Commonly used as biomarkers to test liver function by estimating the AST/ALT ratio.
Total Cholesterol (Chol)	Organic substance necessary for many body processes.	Associated to cardiovascular diseases and kidneys failure.

High-density Lipoprotein (HDL)	Lipoproteins which transport cholesterol to the liver from different body tissues. Commonly known as “good” cholesterol.	
Low-density Lipoprotein (LDL)	Lipoprotein which transport and deliver fats around the body. Commonly known as “bad” cholesterol.	
Triglycerides (TG)	They are the main form of fats in body.	
Thyroid-stimulating hormone (TSH)	Glycoprotein hormone that regulates the thyroid hormone production.	Used as biomarker for thyroid disorders.
Total Hemoglobin (HGB)	Protein contained in red blood cells which transports oxygen, and other gases, between lungs and the rest of the body.	Used as biomarker for anemia, lungs and hearth diseases, among others.

As could be expected, the first test using the mm-wave spectrometer described in chapter 3 in human tissue exhibited higher absorption (thicker skin and higher water content) than the observed during the mice experiments, therefore, an amplifying stage for the transmitted signal was required. The average SFT value measured for mice was $394 \pm 83 \mu\text{m}$ and for humans was $1.85 \pm 2 \text{ mm}$. The transmission signal was amplified using a W-LNA low noise amplifier (Radiometer Physics GmbH, Meckenheim, Germany) connected at the input of the subharmonic mixer receiver corresponding to the transmission measure. The W-LNA provides a typical gain of 40 dB with a noise figure of around 4 dB. Then, the output power of the signal generator was calibrated to obtain a flat frequency response at the transmission port (around $250 \text{ mV} \pm 5 \text{ mV}$) using a RF attenuator (17 dB) at the input of the active frequency multiplier (see Figure 3-2) to compensate for human tissue absorption. A separation of 1.8 mm between the probes of the mm-waves spectroscopic instrument was fixed for calibration tests and to hold the skinfold of the participants during the non-invasive assessments at the visits (see Figure 5-1).

5.2. FIRST RESULTS IN HUMANS: SCALAR-ON-FUNCTION REGRESSION

As described above, a sample population of 30 subjects was recruited and non-invasively measured using the mm-wave spectroscopic instrument described in section 3.3 at three different visits, as detailed in previous section. The collected spectral data include measurements from twenty type 1 DM patients and ten controls (participants without DM). After the 3 visits a total of eighty-three measurements were obtained distributed as shown in Table 5-4 where some absences are indicated.

Table 5-4. Record of visits during the pilot clinical test.

	Visit 1	Visit 2	Visit 3
Total attendance	30	27	26
Absences	0	1 control 2 patients	3 controls 1 patient

The approximated sample curves from amplitude spectra of transmission collected at each visit are shown in Figure 5-2 identifying the control cases, type 1 DM patients with good metabolic control, and type 1 DM with poor metabolic control by dotted blue lines, dashed orange lines and red lines, respectively. The sample curves were approximated using P-splines defined on 35 equally spaced knots over the W-band, with the smoother parameter $\lambda = 0.3$ chosen by the GCV method (as discussed in section 2.3.3). As seen in Figure 5-2, spectra from all cases (non-diabetics, diabetics with good and poor metabolic control) are mixed and, in contrast to the transmittance observed in animal models, there is no apparent differentiation between the spectra for the different populations. Unlike the case for the animal models used previously, glycemia in human subjects is continuously controlled (medication) reducing the effects of sustained hyperglycemia in tissue and metabolism, therefore, similar transmission properties were expected between subjects, whether diabetics or non-diabetics. In fact, no differentiation in the W-band measured transmission spectra between diabetics with good and poor metabolic control is appreciated at first sight.

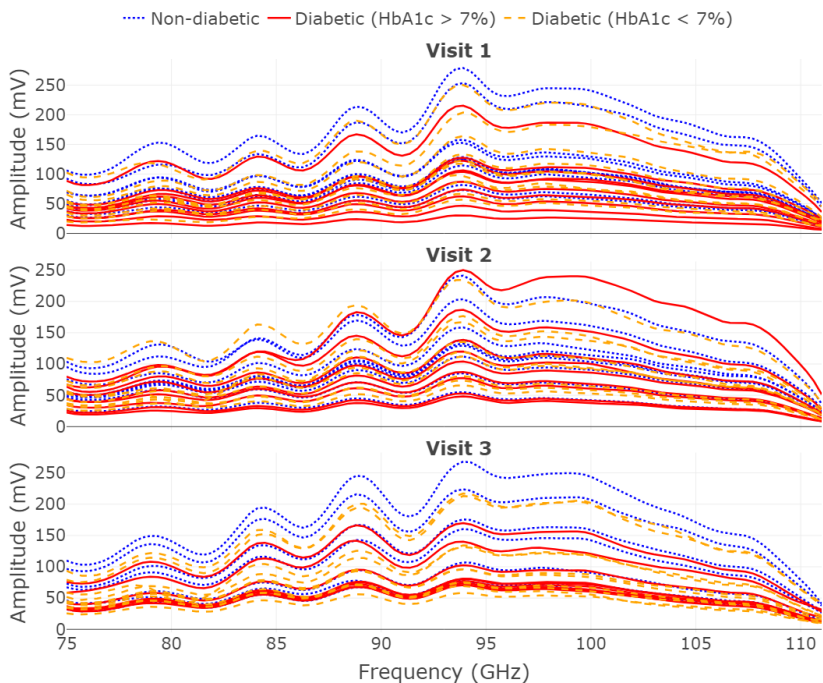


Figure 5-2. Approximated sample curves from amplitude spectra of the transmitted wave through the fold of skin of all the subjects at each visit. Non-diabetic subjects, type 1 DM patients with good metabolic control, and type 1 DM patients with poor metabolic control are identified by dotted blue lines, dashed orange lines and red lines, respectively.

When dealing with this type of measurements in spectroscopy the typical approach uses summary statistics for analyzing the spectral data without consider repeated observations and the intrinsic longitudinal nature of the observations. In this approach the relation between the W-band measured transmission spectra and the HbA1c value would be studied at each visit separately, without consider all the available information (longitudinal data) of the sample population into the statistical analysis. However, when dealing with such complex experimental scenarios, such as biological ones, there are too many parameters (biological variability, environmental, or, even, instrumental) affecting the measured spectral response and it is extremely difficult to control them completely (if eventually possible). Therefore, usually different results are obtained when spectral data sets are analyzed separately (each visit independently), even though all measurements come from repeated

observations done under the same circumstances (sample individuals, spectroscopic instrument and experimental protocol), leading to an inconclusive or poor interpretation of spectral data analysis.

To illustrate this point, a scalar-on-function regression model (see section 2.5) was estimated for each visit recorded during the pilot clinical test. Thus, the W-band measured transmission spectra measured at each visit is used to predict the corresponding HbA1c values. The R-squared coefficient, also known as determination coefficient, is a goodness-of-fit summary statistic that measures the proportion of variability of the response variable (HbA1c value) explained by the fitted model [241], [242]. As shown in Eq. (2-47), the number of basis functions used for the estimation of the functional parameter by scalar-on-function regression can differs to the number of basis functions used in the sample curves approximation. In order to compare the analysis results of the three visits, the functional parameter of the regression models was estimated considering the same number of basis functions, and the best fit for HbA1c values at the three visits was achieved using seventeen cubic B-spline basis functions.

The scatter plots of predicted HbA1c values by the scalar-on-function regression model versus the measured HbA1c values at each visit are shown in Figure 5-3. Let us observe that, as depicted by the HbA1c values of diabetics, the initial distinction of good ($\text{HbA1c} < 7\%$) and poor ($\text{HbA1c} > 7\%$) metabolic control, according to their medical records, was not continued during the pilot clinical test. Therefore, such distinction between diabetics will be not considered for further analysis.

As shown in Figure 5-3, different results were obtained from the estimated regression models, with the best fit ($R\text{-squared} = 0.88$) achieved on the last visit. The results of the last visit indicate that a good linear relation can be achieved between the W-band measured transmission spectra and the HbA1c value. Nevertheless, the inconsistent results between visits shows that there are uncontrolled factors involved in the non-invasive assessment that strongly affects the HbA1c value prediction. Such variation in the obtained results might be caused by changes in the calibration of the spectroscopic instrument or changes in metabolism of subjects between visits (within-subject variation among visits). Besides this, other interferences in the non-invasive assessment such as physiological, environmental or instrumental noise cannot be dismissed.

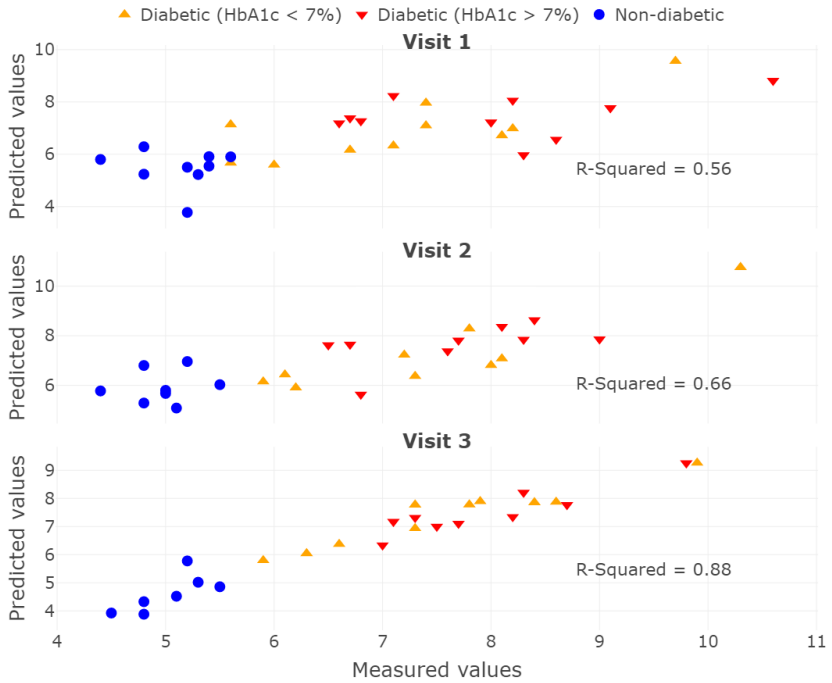


Figure 5-3. HbA1c values versus predicted values by the regression model at each visit. Non-diabetics, diabetics with good metabolic control and diabetics with poor metabolic control are identified by blue circles, orange down-pointing triangles and red triangles, respectively.

The changes in the W-band measured transmission spectra among visits can be clearly seen in the functional parameter $\beta(t)$ estimated by the regression model at each visit (see Figure 5-4). It can be seen that the functional parameter exhibits different shapes between visits, providing inconsistent information about the relation between the W-band measured transmission spectra and the HbA1c value.

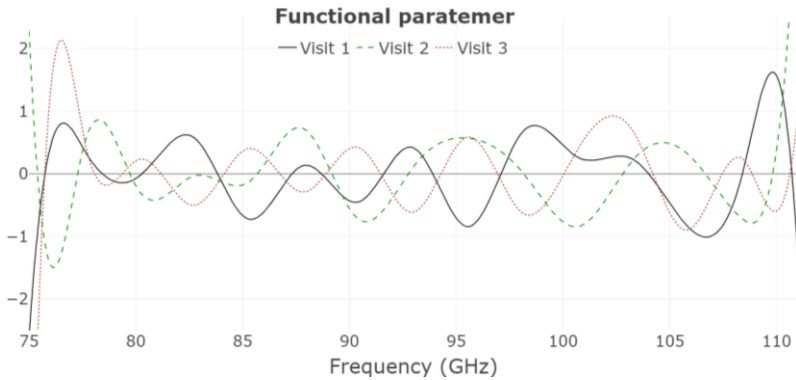


Figure 5-4. Functional parameter $\beta(t)$ estimated by the regression model at each visit.

To address this problem in the evaluation of the proposed diagnostic we are going to use the Longitudinal Functional Data Analysis LFDA techniques introduced in section 2.6. The use of these tools will allow us not only to consider the between-subjects variability in the statistical analysis, but also to take into account the within-subject variation when subjects are repeatedly observed. Thus, the captured variability in spectral data is decomposed into three parts: inter-subject variability (between subjects), intra-subject variability (within subject) and noise. The part of the variance estimated as noise can be seen as the variability in spectral data not shared in either of the first two modes (inter and intra-subject). In this way, the intra-subject variability extraction may improve the previously obtained results since it will allow to consider much more information in the statistical analysis (larger sample size) and can be used to identify the factors that are introducing changes in W-band measured transmission spectra between visits.

5.3. ANALYZING INTER AND INTRA-SUBJECT VARIATIONS WITHIN THE W-BAND SPECTROMETER MEASUREMENTS

In this section, the spectral data collected during the pilot clinical test (eighty-three observations) will be analyzed simultaneously using novel LFDA techniques introduced in section 2.6. Firstly, the longitudinal spectral data will be decomposed performing LFPCA (see section 2.6.1) to study possible influences of the available clinical variables in the intra-

subject variation among visits depicted in previous section. Then, longitudinal scalar-on-function regression methods (see section 2.6.2) are used to study the relation between the W-band measured transmission spectra and HbA1c value considering the inter and intra-subject variation components along the time (three different visits). For both cases, the functional data analyzed corresponds to the sample curves already shown in Figure 5-2.

5.3.1. LFPCA ON LONGITUDINAL SPECTRAL DATA

The LFPCA was performed estimating the first four FPCs associated to the inter-subject variability $B_i(t)$ and intra-subject variability $U_{ij}(t)$ extracted from the longitudinal spectral data. A total of 95.8% of variability contained in the longitudinal spectral data was captured by the eight estimated FPCs, with a 54.8% related to the inter-subject variation, a 41% related to the intra-subject variation, and a 0% of variability estimated as noise.

The scores ξ_i (see Eq. (2-58)) of the first FPC of $B_i(t)$ versus the scores ζ_{ij} (see Eq. (2-59)) of the first principal component of $U_{ij}(t)$, corresponding to the inter and intra-subject variability, respectively, are shown in Figure 5-5. The first FPC related to the inter-subject variability, which explains 44.7% of variability in longitudinal spectral data (estimated according to Eq. (2-62)), discriminates most of the observations between diabetics and non-diabetics. The achieved discrimination between non-diabetics and diabetics among visits considering the inter-subject variability, extracted by the LFPCA, supports that the W-band spectrometer measurements are capable to detect biological changes related to type 1 DM in humans, as it was able to do in the animal experiments.

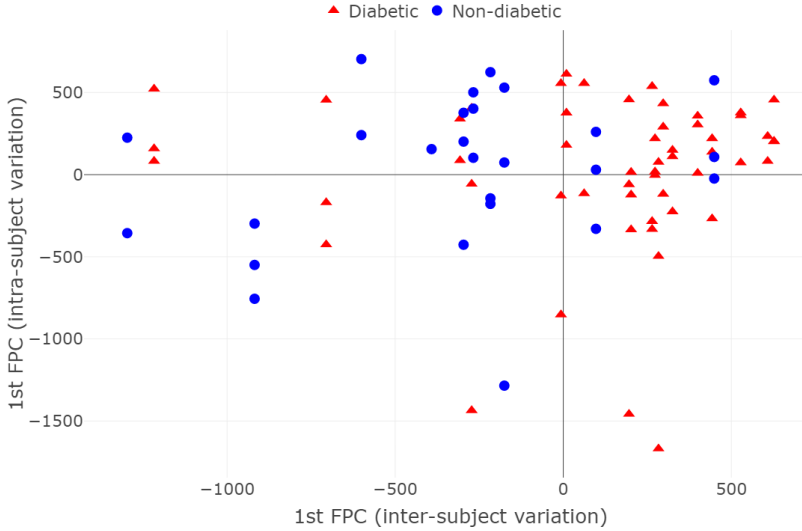


Figure 5-5. Scatterplot of the scores of the first FPCs corresponding to the inter-subject variability $B_i(t)$ and intra-subject variability $U_{ij}(t)$ estimated by the LFPCA. Non-diabetics and diabetics are identified by blue circles and red triangles, respectively.

The W-band obtained spectra associated to type 1 DM detection in humans is characterized by the loading functions (random functional intercept (ψ_1^0) and random functional slope (ψ_1^1)), corresponding to the first FPC of the inter-subject variation $B_i(t)$. The loading functions ψ_1^0 and ψ_1^1 are plotted at the top of the Figure 5-6. At the bottom, the loading functions multiplied by twice the standard deviation of the FPC were added (+) and subtracted (-) to the estimated time-constant mean function of the W-band measured transmission spectra. It is interesting to note that, similarly to the results obtained in section 4.2.2, the determining factors for discrimination between non-diabetics and diabetics are the transmittance level of tissue over the W-band (characterized by the random functional intercept) and a significant change of slope from 81 GHz to 89 GHz (characterized by the random slope function). These findings suggest that both characteristic responses captured by the W-band spectrometer are strongly related to changes in metabolism or tissue in both sustained hyperglycemic mice models (used in chapter 4) and Type 1 DM patients, being less pronounced in the last ones.

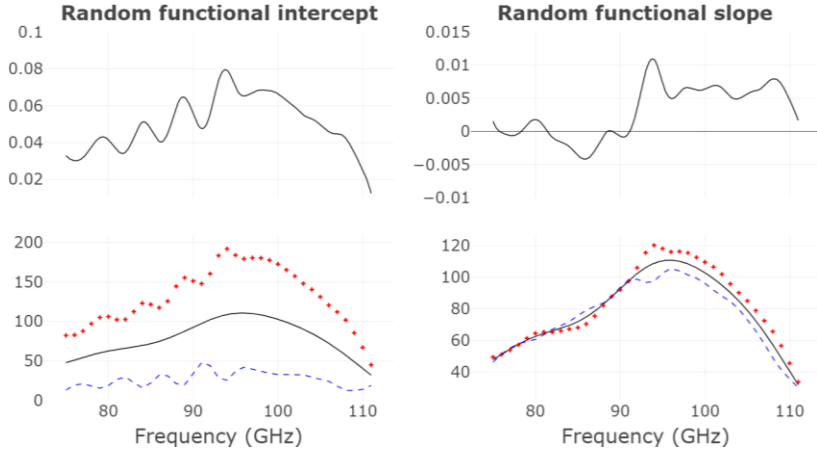


Figure 5-6. (Top) Loading functions associated to the first FPC of the inter-subject variation estimated by the LFPCA, and the estimated time-constant mean function \pm the corresponding loading function multiplied by twice the standard deviation of the associated FPC (bottom).

On the other hand, the first four FPCs ζ_{ij} related to the intra-subject variation among visits, which accumulate 41% of variance in longitudinal spectral data, were used to study possible relations between collected clinical variables and the evolution of the W-band measured spectra among visits. In Figure 5-7 a correlation matrix involving the most relevant clinical variables and the first four FPCs ζ_{ij} estimated by the LFPCA is shown. The correlation matrix contains the Pearson's correlation coefficients (r), which measures the linear correlation between two variables (W_1, W_2) [243], [244], estimated as follows

$$r = \frac{\sum(W_1 - \mu_{W_1})(W_2 - \mu_{W_2})}{\sqrt{\sum(W_1 - \mu_{W_1})^2 \sum(W_2 - \mu_{W_2})^2}} \quad (5-1)$$

where W_1 and W_2 are the random variables, and μ_{W_1} and μ_{W_2} are their sample means, respectively. The correlation coefficient can take values between -1 and $+1$, the magnitude $|r|$ indicates the strength of the relation between both random variables, and the sign of r indicates how the random variables are related; a positive value indicates that the random variables are directly proportional, and negative values indicate that the random variables are inversely proportional.

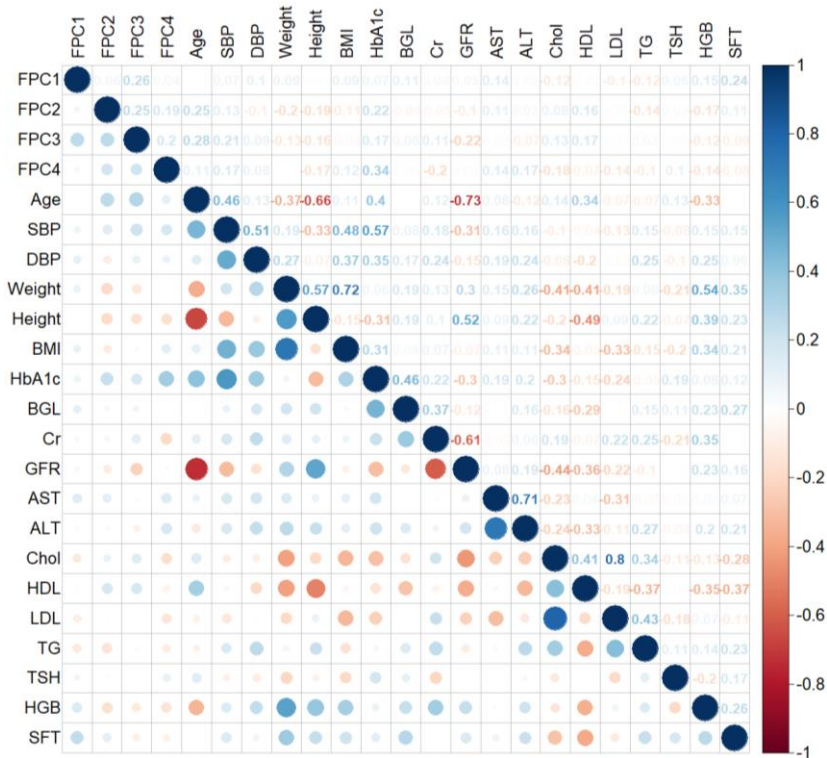


Figure 5-7. Correlation matrix involving the first four FPCs related to the intra-subject variation among visits and the more relevant clinical variables (described in Table 5-3).

As can be seen in the first four rows of the correlation matrix shown in Figure 5-7, no strong correlations were found between the first four FPCs related to the intra-subject variation and collected clinical variables. The greatest correlation associated to the FPCs ($r = 0.34$) is between FPC4 and HbA1c. This means that the intra-subject variation components captured by LFPCA on longitudinal spectral data are not linearly related to the collected clinical variables. However, considering that the intra-subject variation among visits is estimated as variation modes shared between all subjects (diabetics and non-diabetics), it becomes apparent that such variation is not determined by the presence of diabetes in subjects and is mostly related to other biological factors not considered into the pilot clinical test or could be related to systematic noise introduced in the measurements among visits, such as changes in calibration of spectroscopic instrument or environmental conditions.

Additionally, and as a curiosity, we must note that the statistical analysis performed on the different clinical variables obtained for the different patients can provide with further relevant clinical information. In this sense, some strong linear relations between clinical variables can be highlighted from the correlation matrix estimated above, for example, the strong correlation between LDL and the total cholesterol ($r = 0.8$), which is higher than the observed between HDL and total cholesterol ($r = 0.41$), implying that the total cholesterol in blood is mostly affected by the LDL. Also, we can see that GFR exhibits a strong inverse relation ($r = -0.73$) with Age, implying that kidney function deteriorates with age, and it exhibits a good inverse correlation with creatinine since higher content of creatinine in blood is indicator of renal diseases. Weight and BMI show a strong positive correlation ($r = 0.72$) due to their relation, and finally AST and ALT production in body it seems to be closely related with a correlation value of ($r = 0.71$).

5.3.2. LONGITUDINAL REGRESSION ANALYSIS ON HbA1C VALUE

As mentioned above, the main objective of the pilot clinical test is to assess the capability of the non-invasive approach to predict the HbA1c value. The obtained results in Figure 5-3 show that, even though a good linear relation can be achieved between the W-band measured transmission spectra and the HbA1c value, such relation is significantly affected among visits due to uncontrolled variations in the W-band measured transmission spectra (shown in Figure 5-4). Nevertheless, as shown in Figure 5-5, such variation among visits can be improved by considering the intra-subject variation into the statistical analysis. Therefore, the relation between the W-band measured transmission spectra and the HbA1c value will be studied using longitudinal regression models (LPFR and LFPCR) described in section 2.6.2. The main difference between LPFR and LFPCR is that the first one estimates scalar random intercepts to compensate the intra-subject variation among visits and the functional parameter $\beta_l(t)$ is estimated considering all the sample curves (including repeated observations) as a fixed effect across visits. By contrast, the second one is based on the variance decomposition estimated by the LFPCA (inter and intra-subject variation).

In Figure 5-8 is shown the scatter plot of the HbA1c values against the predicted values by the fitted LPFR model, as introduced in Eq. (2-64), considering the longitudinal spectral data shown in Figure 5-2. Let us remark that no scalar predictors were considered in the model estimation. As it can be seen, the LPFR model provides an excellent fit of the HbA1c values measured from all visits (R-squared = 0.98), with the RMSE for prediction equals 0.24. These are very promising results indicating that the W-band spectra acquired has great potential for non-invasive Type 1 DM monitoring, being strongly related to the biomarker of HbA1c test.

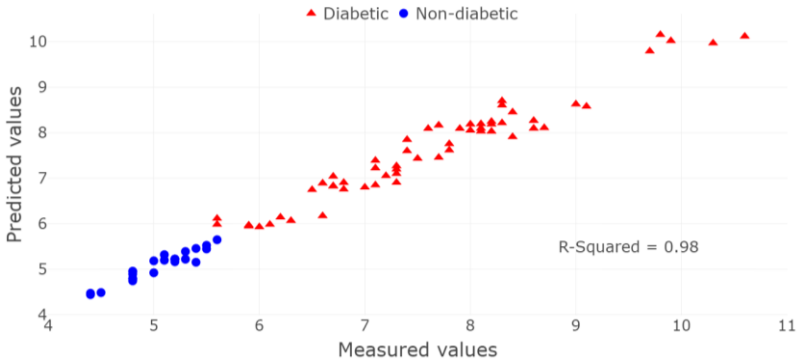


Figure 5-8. HbA1c values versus predicted values by the LPFR model. Non-diabetics, diabetics are identified by blue circles, and red triangles, respectively.

The measurement frequencies contribution for such relation can be qualitatively evaluated observing the functional coefficient $\beta_l(t)$ estimated by the LPFR model and plotted in Figure 5-9. As it can be seen, the functional parameter splits the W-band into two frequency ranges with the zero crossing at 94.8 GHz, indicating that transmittance for frequencies below of 95 GHz is directly proportional to the HbA1c value, and the transmittance measured for all other frequencies has inversely relation.

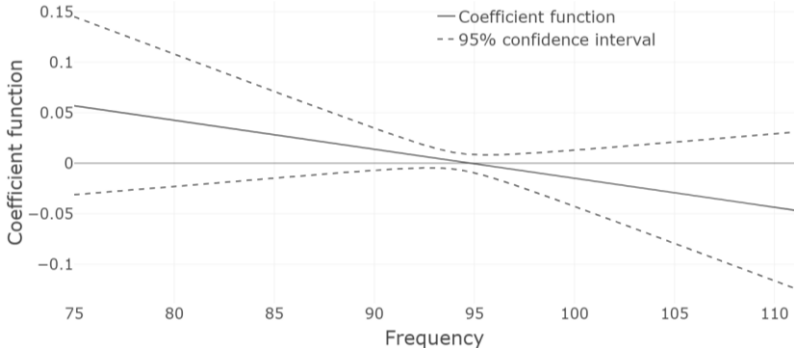


Figure 5-9. Functional coefficient $\beta_l(t)$ estimated by the LPFR model for HbA1c regression, and their pointwise 95% confidence bands (dashed line).

Additionally, in order to validate the results achieved by LPFR, the LFPCR model has also been estimated. The comparison of the performance of both models can be seen in Table 5-5. An excellent fit is also achieved by the LFPCR model (scores-based approach) with very similar results even though the regression models are based on different estimation methods. The consistent results obtained in Table 5-5 demonstrates the relation between the W-band spectral response and the HbA1c values.

Table 5-5. Results obtained from fitted LPFR and LFPCR models on HbA1c values.

Model	R-squared	RMSE
LPFR	0.9753	0.2408
LFPCR (scores-based)	0.9771	0.2318

5.4. DISCUSSION AND CONCLUSIONS

In this chapter, a pilot clinical test was performed to assess the applicability of the non-invasive approach (based on the transmittance of the biological sample within the W-band) for hyperglycemia detection and monitoring in Type 1 DM patients. A group of thirty volunteers (ten non-diabetics and twenty diabetics) were non-invasively measured using the proposed spectroscopic technique, which has previously been assessed using animal models in chapter 4. Considering the observed features of the non-invasive measure during the animal experiments, the pilot clinical

study was focused to study the relation between the measured spectral response and the HbA1c test. The measurements were performed directly on a fold of skin located in the first interdigital space (between the thumb and index finger) of the right hand. The subjects were measured at three different visits (regular medical appointments) over a period of seven months with a blood analysis and other clinical variables collected at each visit.

Firstly, the functional data sets approximated from each visit were analyzed separately performing scalar-on-function regression to study the relation between the HbA1c test and the W-band measured transmission spectra. The obtained results provided evidence that there is a linear relation between the W-band measured transmission spectra and the HbA1c values, but the goodness-of-fit achieved by the regression analysis is considerably affected due to changes in the W-band measured transmission spectra among visits.

The LFDA allows to analyze changes in the measured spectral response when the sample is repeatedly observed at different times (intra-subject variation), as in longitudinal studies. LFPCA was performed to decompose the variance in longitudinal functional data (spectra measured from all visits) into inter- and intra-subject variability. The first four FPCs (with an accumulative explained variance of 41%) associated to the intra-subject variability were used to study possible influences of collected clinical variables in the spectral response evolution across visits. Although no significant correlations were found between the estimated components capturing the intra-subject variation and the collected clinical variables, the first component of the inter-subject variability (with a 44.7% of explained variance) allows to discriminate most of the observations between diabetics and non-diabetics. Such discrimination implies that, in contrast to the intra-subject variation, the inter-subject variations in the measured spectral response is mostly determined by the presence of type 1 DM in subjects. The characteristic spectral responses associated to the detection of type 1 DM patients were qualitatively studied by interpreting the corresponding loading functions. Such interpretation shows that the achieved discrimination of type 1 DM patients is related to spectral responses within the W-band that are quite similar to those observed in the animal models studied in chapter 4. This suggests that sustained hyperglycemia consistently affects the spectral response of both mice and humans within the W-band, which might be associated to the spectral

features of a biological indicator strongly related to hyperglycemic metabolism.

On the other hand, even though no significant correlations were found between the intra-subject variability, estimated by the LFPCA, and the collected clinical variables, the achieved discrimination by the inter-subject variability indicate that the relation between the HbA1c value and the non-invasive measure can be significantly improved by considering both inter- and intra-subject variability in the regression analysis. Therefore, taking advantage of the longitudinal regression analysis, the HbA1c values measured from all visits were fitted using LPFR and LFPCR (scores-based approach) obtaining a R-squared value above 0.97 and a RMSE below 0.24. The longitudinal regression analysis not only improve the regression results on the HbA1c values, but also allows to interpret the relation between the measured spectral response and the HbA1c value since provided a much more smooth functional parameter $\beta_l(t)$ than those estimated by the scalar-on-function regression model.

In summary, these preliminary results from the pilot clinical study indicate that the proposed spectroscopic approach provide a non-invasive measure, based on the transmittance properties of the biological sample, that can be potentially used for in-vivo detection and monitoring of sustained hyperglycemia in humans. Nevertheless, there is still a lot of further experimental work to identify and evaluate the predominant factors (biological, environmental, or instrumental) associated to the observed intra-subject variation in order to calibrate a spectroscopy system that can be used as a non-invasive diagnostic tool in DM medical proceedings.

6. GENERAL CONCLUSIONS AND DISCUSSION

In this thesis a mm-wave spectroscopic technique has been proposed and experimentally evaluated as a new non-invasive approach for in-vivo sensing of sustained hyperglycemia, typically associated to DM. The proposed non-invasive approach has been tested in different experimental settings involving animal models and a pilot clinical study on humans. The development and evaluation processes have been addressed under a two-pronged strategy: a non-targeted spectral profiling approach and the analysis of the collected spectral data using FDA techniques. The main conclusions gathered throughout this thesis are reviewed and discussed below.

A first experimental test (Experiment A) was designed according to the non-targeted spectral profiling approach to explore the reflection and transmission spectral responses within the W-band (measured by the proposed spectroscopic technique) of mice models which represent different glycemic conditions (normoglycemia, hyperglycemia from a mild diabetes, and hyperglycemia from a full-blown diabetes). Two significant conclusions were extracted from such exploratory test by using unsupervised analysis (FPCA) on measured spectral data: the amplitude parameter of transmission (related to the transmittance of the biological media) provides the best indicator (unaffected by biological variations between mice strains) for discrimination of the different glycemic states, and such indicator is strongly associated to sustained glycemia rather than the instantaneous BGL of the mice at the time of the non-invasive assessment. Such discrimination between the different glycemic conditions (normoglycemia, mild diabetes, and full-blown diabetes) was also achieved by modelling the mice skin phenotypes using a full model calibration procedure, typical for microwave network analysis [245]. In contrast to the approach used here, the proposed mathematical model requires amplitude and phase information of reflection and transmission measurements, and special procedures for spectral data calibration are needed.

Let us observe here that the obtained results from the exploratory test provided a clearer and more specific perspective about the applicability of the proposed spectroscopic approach for in-vivo sensing

of hyperglycemia, which determined the design of the following stages of the diagnostic development process. Thus, the next two experimental tests allowed us to validate the detection of sustained hyperglycemia on a new sample of mice (Experiment B) using a classification model (FPCLoR) and to validate the sensitivity of the proposed approach to detect changes in sustained glycemia (with at least four days of occurrence) tracking gradual transitions from a normoglycemia to a hyperglycemia condition in a set of healthy mice with a drug-induced diabetes by studying the evolution of their corresponding FPC's scores over a period of two weeks (Experiment C).

A significant contribution of the FDA to the spectral data analysis performed to the W-band transmission spectra collected from the experiments with animal models is that we were able to identify and validate a sub-interval of frequencies retaining the most relevant spectral features associated to sustained hyperglycemia detection in mice. These findings were achieved by visually inspecting the loading functions and the functional parameter estimated from FPCA and FPCLoR, respectively, performed on the approximated functional data. Furthermore, we showed that the relationship between the measured spectral response (by frequency intervals) and sustained hyperglycemia prediction can be quantitatively studied in terms of the odds ratio by analyzing the functional parameter of the FPCLoR model. This is very interesting since enables the calibration of a spectroscopic technique, proposed for a medical diagnostic, without the need to relate the measured spectral response to a specific component of the biological sample.

The extracted information from the spectral data analysis during the animals experiments indicated that the proposed non-invasive approach could be optimized according to the achieved detection of sustained hyperglycemic states in mice, simplifying the spectroscopic technique from a W-band spectrometer designed for reflection and transmission complex measurements to a simple transmission-type spectrometer with a significantly reduced interrogation frequency band-width.

Regarding to the non-invasive spectroscopic approach under evaluation, the excellent capabilities to detect sustained hyperglycemia and track different levels of sustained hyperglycemia achieved during the previously discussed experimental tests provided strong evidence that the non-invasive measurement (based on the transmittance of the biological media) could be used to monitor sustained glycemia in humans, similarly

to the HbA1c test, but with a better time resolution. The HbA1c test is widely recognized in diagnosis and follow-up of DM as an indicator of the average BGL in the body during the last two to three months. Therefore, a pilot clinical test was performed to study the capability of the described non-invasive approach to predict the HbA1c value on non-diabetic volunteers and type 1 DM patients.

The pilot clinical test was designed as a longitudinal study (subjects were measured repeatedly at different times) due to the desired evaluation of the time resolution for the test, resulting in a total of three visits over a period of seven months. This scenario allowed us to show the scope of the FDA methods that enables the longitudinal spectral data analysis in which all the collected spectral data during the three visits were simultaneously analyzed in terms of the original contributions of the measured frequencies. Firstly, the spectral data collected from each visit were analyzed separately by scalar-on-function regression, with the HbA1c value as the response variable and the W-band measured transmission spectra corresponding to amplitude parameter as a functional predictor. The non-longitudinal regression analysis showed that a good linear relation can be achieved between the non-invasive measure and the HbA1c value (with an R-squared value of 0.88 for the first visit), but such relation was significantly affected due to changes in the measured spectral response among visits, with the R-squared value ranging from 0.56 to 0.88. Such variations were visually assessed by comparing the functional parameters estimated for the three visits, which exhibited completely different shapes from each other. These results clearly indicated that, in contrast to the measurements obtained with the animal experiments, the interferences raised from biological, environmental or instrumental factors were strong enough to considerably affect the performance of the non-invasive approach. Such difficulties were expected since, compared to the animal models, humans exhibit stronger absorption, much bigger biological variability between subjects, and a slower metabolism.

In view of the observed variations in the measured spectral response among visits, the longitudinal spectral data was analyzed using LFDA methods which consider the inter and intra-subject variability. A LFPCA was performed for a blindly exploration of the inter- and intra-subject variation captured by the longitudinal spectral data. The inter-subject variation allowed to discriminate most of the type 1 DM patients from the non-diabetics, similarly to the discrimination achieved during animal experiments. Once again, taking advantage of the interpretability of the

corresponding loading functions, we realized that the achieved discrimination of type 1 DM in humans and the detection of sustained hyperglycemia in the mice models were determined by quite similar characteristic spectral responses within the W-band. This clearly indicates that there is a biological component (or components), with characteristic spectral lines at the identified frequencies within the W-band, which is closely related to the presence of sustained hyperglycemia. Besides this, the estimated scores associated to the intra-subject variability (capturing the variation of the spectral response among visits) were used to study possible influences of several clinical variables collected during the pilot clinical test, including the HbA1c value, but no significant correlations were found. Let us remark here that LFPCA is a very powerful tool for longitudinal spectral data analysis that not only allows to study possible correlations between the estimated scores from the inter and intra-subject variability decomposition and other measured variables (either biological, environmental or instrumental), but also allows to study the temporal evolution of the spectral response using surface graphics, which is very useful for longitudinal studies of progressive diseases.

Then, two different longitudinal regression methods (LPFR and LFPCR) were used in order to model all the HbA1c values in terms of the longitudinal spectral data by considering the inter and intra-subject variation which improved the prediction of the HbA1c values with an R-squared value above 0.97 and a RMSE below 0.24. These results obtained from the two models that are estimated through different approaches are very promising since demonstrates that the measured spectral response can provide an excellent indicator for sustained glycemia closely related to the HbA1c value.

Thus, on one hand, it has been shown in this thesis that the non-targeted spectral profiling approach allows to gain perspective in the exploration and characterization of spectral responses obtained from spectral interrogation performed on multi-species scenarios exhibiting complex spectral features, such as biological media. This approach allowed us to evaluate the applicability of the proposed spectroscopic technique for in-vivo sensing of hyperglycemia by considering collectively the spectral responses of all the components involved in the biological sample. This is a very interesting alternative approach in the development of new spectroscopy systems applied to biomedical engineering for two reasons: no a-priori information about the spectral features of the sample constituents is required, and it is not necessary to model an isolated

metabolite (which is quite difficult in biological media) to associate the measured spectral response to the pathology under study. This approach is useful not only for cases in which the diagnostic does not require the quantification of a specific metabolite, but also to identify characteristic spectral responses associated to the pathology under study that might lead to determination of relevant biomarkers by specialized analysis of the biological media composition (from the inverse problem perspective).

On the other hand, we have shown that the FDA framework covers most of the classical multivariate statistical methods widely used for unsupervised and supervised learning problems involving spectroscopy data. FDA is advantageous for the spectral data analysis in many ways, but one of the most significant benefits that makes it very suitable for the development of applied spectroscopy systems is the ongoing interpretation of the spectral data. However, as discussed in chapter 4, an amenable interpretation of spectral data using FDA methods is subject to the smoothness degree used in the approximation of the functional data and regression analysis. In this regard, approximating smooth sample curves from heterogeneous spectral features might be a challenging task using a fixed overall smoothness degree. These cases can be addressed using the approximation methods presented in this thesis by manually fixing the knots at strategic locations over the measured frequency range. Nevertheless, such a task requires considerable efforts if researcher does not know which spectral peaks are truly associated to the optical properties of the sample. The adaptive smoothing approaches can be very useful in this situations, such as the adaptive P-splines [246], [247], that allows for different local penalizations within the measured frequency range (varying smoothing parameter).

It has been shown throughout this doctoral thesis that the FDA framework is well suited for spectral data analysis, providing powerful and amenable statistical methods for different experimental settings such as the longitudinal studies. The FDA methods enable naturally the direct interpretation of frequencies contribution at all time for supervised and unsupervised analysis of spectral data (not usually possible for MDA methods). As shown throughout this research work, the interpretation of the measured spectral data plays a key role in the evaluation process of applied spectroscopic techniques. Therefore, we support that FDA methods have a great scope for processing and analysis of spectral data measured over a continuous domain.

Finally, as the most relevant result of this thesis, the obtained results from the pilot clinical study prove that the transmittance properties of the biological media measured indirectly by the spectroscopic technique using the W-band are closely related to sustained glycemia in humans. But the spectroscopic technique requires further calibration to compensate the intra-subject variation introduced in the measured spectral response. Even though an excellent fit for the HbA1c value was achieved by the longitudinal regression models, such approach limits the prediction capabilities of the non-invasive diagnostic to individuals previously used for calibration, requiring repeated measurements on each subject until proper calibration is achieved. Therefore, further experimental tests need to be made in order to identify and reduce the interferences introduced by the intra-subject variation in the measured spectral response. An important step prior the next experimental test in humans is to obtain a prototype of the final spectroscopic instrument to ensure the stability of the spectroscopic instrument response within the W-band and reduce the instrumental interferences in the non-invasive assessments. Besides this, other factors not considered in the pilot clinical trial, such as the skin moisture, temperature of the targeted biological sample, humidity and temperature of environment, should be registered during the measurements in order to study their influence in the intra-subject variation. Other interesting experimental test that should be also considered for a next clinical study is to carry out the non-invasive assessments more frequently (at least weekly), in order to compare the non-invasive measure with the average BGL of subjects (using multiple BGL readings per day) over shorter time periods. This will enable us to study the time resolution offered by the proposed non-invasive approach to track changes in sustained glycemia on humans.

Nevertheless, the statistical results obtained throughout this thesis work demonstrate that the proposed spectroscopic technique is very promising for the development of a non-invasive approach to sustained hyperglycemia monitoring since it could potentially be less costly and time consuming in diabetes management compared to current medical procedures based on BGL and HbA1c invasive measurements. The non-invasive instrument can be made very compact and mass-produced using semiconductor processes, and its painless, non-invasive operation can eliminate infection risks and ongoing expenses for consumables. All these improvements in medical proceedings for diabetes management and diagnosis would have a great positive impact in early diabetes detection,

patient's quality of life, and in risk mitigation of potentially diabetic individuals.

7. REFERENCES

- [1] N. C. Thomas, “The early history of spectroscopy,” *Journal of Chemical Education*, vol. 68, no. 8. Division of Chemical Education , pp. 631–634, 1991, doi: 10.1021/ed068p631.
- [2] A. I. Shapiro, H. Peter, and S. K. Solanki, “The Sun’s Atmosphere,” in *The Sun as a Guide to Stellar Physics*, Elsevier, 2019, pp. 59–85.
- [3] J. W. Robinson, *Atomic spectroscopy*, 2nd ed. CRC Press, 1996.
- [4] N. Ulrich, “Atomic Emission and Absorption,” in *Encyclopedia of Food Sciences and Nutrition*, 2nd ed., Academic Press, 2003, pp. 5441–5447.
- [5] S. R. Koirtyohann, “A history of atomic absorption spectroscopy,” *Spectrochim. Acta Part B At. Spectrosc.*, vol. 35, no. 11–12, pp. 663–670, 1980, doi: 10.1016/0584-8547(80)80006-1.
- [6] H. Kragh, *Niels Bohr and the quantum atom: The Bohr model of atomic structure 1913-1925*. OUP Oxfords, 2012.
- [7] P. Davidovits, “Atomic Physics,” in *Physics in Biology and Medicine*, Elsevier, 2013, pp. 227–245.
- [8] G. Butcher, *Tour of the Electromagnetic Spectrum*, 3rd ed. National Aeronautics and Space Administration , 2016.
- [9] J. Mohan, *Organic spectroscopic: Principles and Applications*, 2nd ed. Alpha Science Int’l Ltd., 2004.
- [10] I. E. Gordon *et al.*, “The HITRAN2016 molecular spectroscopic database,” *J. Quant. Spectrosc. Radiat. Transf.*, vol. 203, pp. 3–69, Dec. 2017, doi: 10.1016/j.jqsrt.2017.06.038.
- [11] D. F. Swinehart, “The Beer-Lambert law,” *Journal of Chemical Education*, vol. 39, no. 7. Division of Chemical Education , pp. 333–335, 1962, doi: 10.1021/ed039p333.
- [12] N. V. Tkachenko, *Optical Spectroscopy: Methods and Instrumentations*. Elsevier, 2006.
- [13] Lajunen. L.H.J. and p. Perämäki, *Spectrochemical Analysis by Atomic Absorption and Emission*, 2nd ed. The Royal Society of Chemistry, 2004.
- [14] D. N. Sathyanarayana, *Electronic Absorption Spectroscopy and Related Techniques*. Universities Press, 2001.

- [15] S. Y. Tang, Z. N. Xia, Y. J. Fu, and Q. GOU, “Advances and Applications of Microwave Spectroscopy,” *Chinese Journal of Analytical Chemistry*, vol. 36, no. 8. Chinese Academy of Sciences, pp. 1145–1151, 01-Aug-2008, doi: 10.1016/S1872-2040(08)60061-4.
- [16] S. L. Dexheimer, *Terahertz Spectroscopy: Principles and Applications*. CRC Press, 2007.
- [17] H. Günzler and H.-U. Gremlich, *IR Spectroscopy: An Introduction*. WILEY-VCH, 2002.
- [18] Y. Ozaki and S. Kawata, *Far- and Deep-Ultraviolet Spectroscopy*. Springer, Tokyo, 2015.
- [19] B. K. Agarwal, *X-Ray Spectroscopy: An Introduction*, 2nd ed. Springer-Verlag Berlin Heidelberg, 1991.
- [20] J. Lekner, *Theory of reflection of electromagnetic and particle waves*. Springer Science & Business Media, 2013.
- [21] T. Zhang, “Refraction of light in media,” *Sci. China, Ser. G Physics, Mech. Astron.*, vol. 50, no. 5, pp. 591–600, Oct. 2007, doi: 10.1007/s11433-007-0066-2.
- [22] J. M. Cowley, *Diffraction Physics*, 3rd ed. Elsevier, 1995.
- [23] R. G. Newton, *Scattering Theory of Waves and Particles*, 2nd ed. Springer Science & Business Media, 2013.
- [24] D. Attwood, *X-Rays and Extreme Ultraviolet Radiation: Principles and Applications*. Cambridge University Press, 1999.
- [25] C. F. Bohren and D. R. Huffman, *Absorption and Scattering of Light by Small Particles*. John Wiley & Sons, 2008.
- [26] C. D. Tran and M. Xu, “Dual-Wavelength Photothermal Refraction Spectrometry for Small-Volume Samples,” *Appl. Spectrosc.*, vol. 43, no. 6, pp. 1056–1061, Aug. 1989, doi: 10.1366/0003702894203886.
- [27] S. E. Bialkowski, “Using an optical novelty filter to enhance contrast in photothermal refraction spectrometry,” in *Proceedings of the 10th international conference on photoacoustic and photothermal phenomena*, 2009, vol. 463, no. 1, pp. 67–71, doi: 10.1063/1.58166.
- [28] E. T. Stathopoulou, V. Psycharis, G. D. Chryssikos, V. Gionis, and

- G. Theodorou, “Bone diagenesis: New data from infrared spectroscopy and X-ray diffraction,” *Palaeogeogr. Palaeoclimatol. Palaeoecol.*, vol. 266, no. 3–4, pp. 168–174, Sep. 2008, doi: 10.1016/j.palaeo.2008.03.022.
- [29] M. Kessler, K. Frank, J. Hoper, D. Tauschek, and J. Zundorf, “Reflection spectrometry,” in *Advances in Experimental Medicine and Biology*, 1992, vol. 317, pp. 203–212, doi: 10.1007/978-1-4615-3428-0_21.
- [30] J. W. Rabalais, *Principles and Applications of Ion Scattering Spectrometry: Surface Chemical and Structural Analysis*. New York: Wiley, 2003.
- [31] F. M. Mirabella, *Internal Reflection Spectroscopy Theory and Applications*, 1st ed. CRC Press, 1992.
- [32] G. Kortüm, *Reflectance Spectroscopy: Principles, Methods, Applications*. Springer Science & Business Media, 2012.
- [33] M. Jonasz and G. R. Fournier, *Light Scattering by Particles in Water: Theoretical and Experimental Foundations*. Elsevier, 2007.
- [34] S. Singh, “Refractive Index Measurement and its Applications,” *Phys. Scr.*, vol. 65, no. 2, pp. 167–180, Jan. 2002, doi: 10.1238/physica.regular.065a00167.
- [35] I. Popescu, P. Sterian, and M. Dobre, “The photon wave function and the Fresnel formulas,” *Rom. Reports Phys.*, vol. 62, no. 2, pp. 360–368, 2010.
- [36] M. Barnoski, *Introduction to Integrated Optics*. Springer Science & Business Media, 2012.
- [37] D. Drosdoff and A. Widom, “Snell’s law from an elementary particle viewpoint,” *Am. J. Phys.*, vol. 73, no. 10, pp. 973–975, Oct. 2005, doi: 10.1119/1.2000974.
- [38] Z. Deng *et al.*, “Determination of continuous complex refractive index dispersion of biotissue based on internal reflection,” *J. Biomed. Opt.*, vol. 21, no. 1, p. 015003, Jan. 2016, doi: 10.1117/1.jbo.21.1.015003.
- [39] J. E. Chamberlain, “On a relation between absorption strength and refractive index,” *Infrared Phys.*, vol. 5, no. 4, pp. 175–178, Dec. 1965, doi: 10.1016/0020-0891(65)90020-5.

- [40] J. M. Carcione, F. Cavallini, J. Ba, W. Cheng, and A. N. Qadrouh, “On the Kramers-Kronig relations,” *Rbeol. Acta*, vol. 58, no. 1–2, pp. 21–28, Feb. 2019, doi: 10.1007/s00397-018-1119-3.
- [41] J. M. Carcione, *Wave fields in real media: Wave propagation in anisotropic, anelastic, porous and electromagnetic media*, 2nd ed. Elsevier, 2007.
- [42] C. Labuda and I. Labuda, “On the mathematics underlying dispersion relations,” *Eur. Phys. J. H*, vol. 39, no. 5, pp. 575–589, Oct. 2014, doi: 10.1140/epjh/e2014-50021-1.
- [43] K. Yamamoto and A. Masui, “Complex Refractive Index Determination of Bulk Materials from Infrared Reflection Spectra,” *Appl. Spectrosc.*, vol. 49, no. 5, pp. 639–644, May 1995, doi: 10.1366/0003702953964048.
- [44] P. Grosse and V. Offermann, “Analysis of reflectance data using the Kramers-Kronig Relations,” *Appl. Phys. A Solids Surfaces*, vol. 52, no. 2, pp. 138–144, Feb. 1991, doi: 10.1007/BF00323731.
- [45] D. J. Segestein, “The complex refractive index of water,” University of Missouri, 1981.
- [46] J. F. Lodenquai, “Electromagnetic wave propagation in media with complex refractive indices and transmission through absorbing films,” *Am. J. Phys.*, vol. 59, no. 3, pp. 248–254, Mar. 1991, doi: 10.1119/1.16571.
- [47] S. A. Kovalenko and S.A., “Descartes-Snell law of refraction with absorption,” Інститут фізики напівпровідників імені В.Є. Лашкарьова НАН України, 2001.
- [48] J. Shen, H. Yu, and J. Lu, “Light propagation and reflection-refraction event in absorbing media,” *Chinese Opt. Lett. Vol. 8, Issue 1, pp. 111-114*, vol. 8, no. 1, pp. 111–114, Jan. 2010.
- [49] P. Martín-Mateos, “New spectroscopic techniques and architectures for environmental and biomedical applications,” Universidad Carlos III de Madrid, 2015.
- [50] V. V. Tuchin, *Handbook of Optical Biomedical Diagnostics, Second Edition, Volume 1: Light-Tissue Interaction*. SPIE, 2016.
- [51] O. C. Marina, C. K. Sanders, and J. R. Mourant, “Correlating light scattering with internal cellular structures,” *Biomed. Opt. Express*, vol. 3, no. 2, p. 296, Feb. 2012, doi: 10.1364/boe.3.000296.

- [52] Z. Wang, "Tissue refractive index as marker of disease," *J. Biomed. Opt.*, vol. 16, no. 11, p. 116017, Nov. 2011, doi: 10.1117/1.3656732.
- [53] A. T. Lovell, J. C. Hebden, J. C. Goldstone, and M. Cope, "Determination of the transport scattering coefficient of red blood cells," in *Optical Tomography and Spectroscopy of Tissue III*, 1999, vol. 3597, pp. 175–182, doi: 10.1117/12.356795.
- [54] R. Graaff, M. H. Koelink, F. F. M. de Mul, W. G. Zijlstra, A. C. M. Dassel, and J. G. Aarnoudse, "Condensed Monte Carlo simulations for the description of light transport," *Appl. Opt.*, vol. 32, no. 4, p. 426, Feb. 1993, doi: 10.1364/ao.32.000426.
- [55] V. V. Tuchin, *Tissue Optics: Light Scattering Methods and Instruments for Medical Diagnosis*. Society of Photo-Optical Instrumentation Engineers (SPIE), 2015.
- [56] S. Avrillier, E. Tinet, and E. Delettre, "Monte Carlo simulation of collimated beam transmission through turbid media," *J. Phys.*, vol. 51, no. 22, pp. 2521–2542, Nov. 1990, doi: 10.1051/jphys:0199000510220252100.
- [57] C. Zhu and Q. Liu, "Review of Monte Carlo modeling of light transport in tissues," *J. Biomed. Opt.*, vol. 18, no. 5, p. 050902, May 2013, doi: 10.1117/1.jbo.18.5.050902.
- [58] T. Durduran, R. Choe, W. B. Baker, and A. G. Yodh, "Diffuse Optics for Tissue Monitoring and Tomography," *Rep. Prog. Phys.*, vol. 73, no. 7, 2010, doi: 10.1088/0034-4885/73/7/076701.
- [59] P. Geladi, "Chemometrics in spectroscopy. Part 1. Classical chemometrics," *Spectrochimica Acta - Part B Atomic Spectroscopy*, vol. 58, no. 5. Elsevier, pp. 767–782, 30-May-2003, doi: 10.1016/S0584-8547(03)00037-5.
- [60] K. Varmuza and P. Filzmoser, *Introduction to Multivariate Statistical Analysis in Chemometrics*. CRC Press, 2016.
- [61] H. Mark and J. Workman, *Chemometrics in Spectroscopy*. Elsevier, 2018.
- [62] M. E. Keating, H. Nawaz, F. Bonnier, and H. J. Byrne, "Multivariate statistical methodologies applied in biomedical Raman spectroscopy: Assessing the validity of partial least squares

- regression using simulated model datasets,” *Analyst*, vol. 140, no. 7, pp. 2482–2492, Apr. 2015, doi: 10.1039/c4an02167c.
- [63] I. M. Johnstone and D. M. Titterington, “Statistical challenges of high-dimensional data,” *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.*, vol. 367, no. 1906, pp. 4237–4253, Nov. 2009, doi: 10.1098/rsta.2009.0159.
- [64] M. G. Madden and A. G. Ryder, “Machine learning methods for quantitative analysis of Raman spectroscopy data,” in *Opto-Ireland 2002: Optics and Photonics Technologies and Applications*, 2003, vol. 4876, p. 1130, doi: 10.1117/12.464039.
- [65] L. O. Jimenez and D. A. Landgrebe, “Supervised classification in high-dimensional space: Geometrical, statistical, and asymptotical properties of multivariate data,” *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.*, vol. 28, no. 1, pp. 39–54, 1998, doi: 10.1109/5326.661089.
- [66] D. E. Farrar and R. R. Glauber, “Multicollinearity in Regression Analysis: The Problem Revisited,” *Rev. Econ. Stat.*, vol. 49, no. 1, pp. 92–107, 1967, doi: 10.2307/1937887.
- [67] A. Alin, “Multicollinearity,” *Wiley Interdiscip. Rev. Comput. Stat.*, vol. 2, no. 3, pp. 370–374, May 2010, doi: 10.1002/wics.84.
- [68] M. P. Allen, *Understanding Regression Analysis*. Springer, Boston, MA, 1997.
- [69] B. Mei and Z. Wang, “An efficient method to handle the ‘large p, small n’ problem for genomewide association studies using Haseman–Elston regression,” *J. Genet.*, vol. 95, no. 4, pp. 847–852, Dec. 2016, doi: 10.1007/s12041-016-0705-3.
- [70] S. Chakraborty, “Bayesian multiple response kernel regression model for high dimensional data and its practical applications in near infrared spectroscopy,” *Comput. Stat. Data Anal.*, vol. 56, no. 9, pp. 2742–2755, Sep. 2012, doi: 10.1016/j.csda.2012.02.019.
- [71] S. Chakraborty, M. Ghosh, and B. K. Mallick, “Bayesian nonlinear regression for large p small n problems,” *J. Multivar. Anal.*, vol. 108, pp. 28–40, Jul. 2012, doi: 10.1016/j.jmva.2012.01.015.
- [72] M. Roozbeh, “Generalized ridge regression estimator in high dimensional sparse regression models,” *Stat. Optim. Inf. Comput.*,

- vol. 6, no. 3, pp. 415–426, Aug. 2018, doi: 10.19139/soic.v6i3.581.
- [73] L. Li, “Dimension reduction for high-dimensional data,” *Methods in molecular biology (Clifton, N.J.)*, vol. 620. Humana Press, Totowa, NJ, pp. 417–434, 2010, doi: 10.1007/978-1-60761-580-4_14.
- [74] T. Howley, M. G. Madden, M.-L. O’Connell, and A. G. Ryder, “The Effect of Principal Component Analysis on Machine Learning Accuracy with High Dimensional Spectral Data,” in *Applications and Innovations in Intelligent Systems XIII*, Springer London, 2007, pp. 209–222.
- [75] J. R. Riba Ruiz, T. C. Parelló, and R. Cantero Gómez, “Comparative study of multivariate methods to identify paper finishes using infrared spectroscopy,” *IEEE Trans. Instrum. Meas.*, vol. 61, no. 4, pp. 1029–1036, Apr. 2012, doi: 10.1109/TIM.2011.2173048.
- [76] R. M. Balabin, R. Z. Safieva, and E. I. Lomakina, “Gasoline classification using near infrared (NIR) spectroscopy data: Comparison of multivariate techniques,” *Anal. Chim. Acta*, vol. 671, no. 1–2, pp. 27–35, Jun. 2010, doi: 10.1016/j.aca.2010.05.013.
- [77] Y. Saeys, I. Inza, and P. Larrañaga, “A review of feature selection techniques in bioinformatics,” *Bioinformatics*, vol. 23, no. 19. Oxford Academic, pp. 2507–2517, 01-Oct-2007, doi: 10.1093/bioinformatics/btm344.
- [78] F. Liu, Y. He, and L. Wang, “Determination of effective wavelengths for discrimination of fruit vinegars using near infrared spectroscopy and multivariate analysis,” *Anal. Chim. Acta*, vol. 615, no. 1, pp. 10–17, May 2008, doi: 10.1016/j.aca.2008.03.030.
- [79] F. Westad and H. Martens, “Variable Selection in near Infrared Spectroscopy Based on Significance Testing in Partial Least Squares Regression,” *J. Near Infrared Spectrosc. Vol. 8, Issue 2, pp. 117-124*, vol. 8, no. 2, pp. 117–124, Mar. 2000.
- [80] Z. Xiaobo, Z. Jiewen, M. J. W. Povey, M. Holmes, and M. Hanpin, “Variables selection methods in near-infrared spectroscopy,” *Analytica Chimica Acta*, vol. 667, no. 1–2. Elsevier, pp. 14–32, 14-May-2010, doi: 10.1016/j.aca.2010.03.048.
- [81] J. Sun, W. Yang, M. Feng, Q. Liu, and M. S. Kubar, “An efficient

- variable selection method based on random frog for the multivariate calibration of NIR spectra,” *RSC Adv.*, vol. 10, no. 28, pp. 16245–16253, Apr. 2020, doi: 10.1039/d0ra00922a.
- [82] E. Andries and S. Martin, “Sparse Methods in Spectroscopy: An Introduction, Overview, and Perspective,” *Appl. Spectrosc. Vol. 67, Issue 6*, pp. 579–589, vol. 67, no. 6, pp. 579–589, Jun. 2013.
- [83] L. L. Wang *et al.*, “A selective review and comparison for interval variable selection in spectroscopic modeling,” *Chemometrics and Intelligent Laboratory Systems*, vol. 172. Elsevier B.V., pp. 229–240, 15-Jan-2018, doi: 10.1016/j.chemolab.2017.11.008.
- [84] F. Rossi, D. François, V. Wertz, and M. Verleysen, “A functional approach to variable selection in spectrometric problems,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2006, vol. 4131 LNCS-I, pp. 11–20, doi: 10.1007/11840817_2.
- [85] Z. Barati, I. Zakeri, and K. Pourrezaei, “Functional data analysis view of functional near infrared spectroscopy data,” *J. Biomed. Opt.*, vol. 18, no. 11, p. 117007, Nov. 2013, doi: 10.1117/1.jbo.18.11.117007.
- [86] H. G. Müller, “Functional modelling and classification of longitudinal data,” *Scandinavian Journal of Statistics*, vol. 32, no. 2. John Wiley & Sons, Ltd, pp. 223–240, 01-Jun-2005, doi: 10.1111/j.1467-9469.2005.00429.x.
- [87] C. Goutis, “Second-derivative functional regression with applications to near infra-red spectroscopy,” *J. R. Stat. Soc. Ser. B (Statistical Methodol.*, vol. 60, no. 1, pp. 103–114, Feb. 1998, doi: 10.1111/1467-9868.00111.
- [88] A. Mas and B. Pumo, “Functional linear regression with derivatives,” *J. Nonparametr. Stat.*, vol. 21, no. 1, pp. 19–40, Jan. 2009, doi: 10.1080/10485250802401046.
- [89] H. Sørensen, J. Goldsmith, and L. M. Sangalli, “An introduction with medical applications to functional data analysis,” *Stat. Med.*, vol. 32, no. 30, pp. 5222–5240, Dec. 2013, doi: 10.1002/sim.5989.
- [90] S. Naz, M. Vallejo, A. García, and C. Barbas, “Method validation strategies involved in non-targeted metabolomics,” *J. Chromatogr.*

- A*, vol. 1353, pp. 99–105, Aug. 2014, doi: 10.1016/j.chroma.2014.04.071.
- [91] X. Zhang, X. Zhu, C. Wang, H. Zhang, and Z. Cai, “Non-targeted and targeted metabolomics approaches to diagnosing lung cancer and predicting patient prognosis,” *Oncotarget*, vol. 7, no. 39, pp. 63437–63448, 2016, doi: 10.18632/oncotarget.11521.
- [92] J. Ramsay and B. W. Silverman, *Functional Data Analysis*, 2nd ed. Springer-Verlag New York, 2005.
- [93] J. O. Ramsay and B. W. Silverman, *Applied Functional Data Analysis*, 1st ed. Springer-Verlag New York, 2002.
- [94] J. O. Ramsay, G. Hooker, and S. Graves, *Functional Data Analysis with R and MATLAB*, 1st ed. Springer-Verlag New York, 2009.
- [95] M. K. Ochi, *Applied probability and stochastic processes: In engineering and physical sciences*, 1st ed. Wiley-Interscience, 1990.
- [96] J.-C. Deville, “Méthodes statistiques et numériques de l’analyse harmonique,” *Ann. Insee.*, no. 15, pp. 3–101, 1974, doi: 10.2307/20075177.
- [97] B. Obsieger, *Numerical Methods IV: Interpolation and Shape Functions*, 1st ed. Boris Obsieger, 2015.
- [98] M. C. Aguilera-Morillo, “Penalized estimation methods in functional data analysis,” Universidad de Granada, 2013.
- [99] R. L. Eubank, *Nonparametric Regression and Spline Smoothing*, 2nd ed. CRC Press, 1988.
- [100] Y. Wang, *Smoothing Splines: Methods and Applications*, 1st ed. Chapman and Hall/CRC, 2011.
- [101] F. A. Ocaña, O. Valenzuela, and A. M. Aguilera, “A Wavelet Approach to Functional Principal Component Analysis,” in *COMPSTAT*, 1998, pp. 413–418, doi: 10.1007/978-3-662-01131-7_57.
- [102] J. R. and C. de Boor, “A Practical Guide to Splines.,” *Math. Comput.*, vol. 34, no. 149, p. 325, 1980, doi: 10.2307/2006241.
- [103] A. M. Aguilera and M. C. Aguilera-Morillo, “Comparative study of different B-spline approaches for functional data,” *Math. Comput. Model.*, vol. 58, no. 7–8, pp. 1568–1579, Oct. 2013, doi:

- 10.1016/j.mcm.2013.04.007.
- [104] P. H. C. Eilers and B. D. Marx, “Flexible smoothing with B-splines and penalties,” *Stat. Sci.*, vol. 11, no. 2, pp. 89–121, 1996, doi: 10.1214/ss/1038425655.
- [105] D. Ruppert, “Selecting the Number of Knots for Penalized Splines,” *J. Comput. Graph. Stat.*, vol. 11, no. 4, pp. 735–757, 2002, doi: 10.1198/106186002853.
- [106] P. Craven and G. Wahba, “Smoothing noisy data with spline functions - Estimating the correct degree of smoothing by the method of generalized cross-validation,” *Numer. Math.*, vol. 31, no. 4, pp. 377–403, Dec. 1978, doi: 10.1007/BF01404567.
- [107] A. Hyvärinen and E. Oja, “Independent component analysis: algorithms and applications,” *Neural networks*, vol. 13, no. 4–5, pp. 411–430, 2000, doi: 10.1016/S0893-6080(00)00026-5.
- [108] I. T. Jolliffe and J. Cadima, “Principal component analysis: A review and recent developments,” *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.*, vol. 374, no. 2065, p. 20150202, 2016, doi: 10.1098/rsta.2015.0202.
- [109] A. M. Aguilera, “Métodos de Aproximacion de Estimadores en el ACP de un Proceso Estocástico,” Universidad de Granada, 1993.
- [110] A. Kassambara, *Practical Guide to Cluster Analysis in R: Unsupervised Machine Learning*, 1st ed. STHDA, 2017.
- [111] F. A. Ocaña, A. M. Aguilera, and M. Escabias, “Computational considerations in functional principal component analysis,” *Comput. Stat.*, vol. 22, no. 3, pp. 449–465, Sep. 2007, doi: 10.1007/s00180-007-0051-2.
- [112] S. P. Huang, S. T. Quek, and K. K. Phoon, “Convergence study of the truncated Karhunen–Loeve expansion for simulation of stochastic processes,” *Int. J. Numer. Methods Eng.*, vol. 52, no. 9, pp. 1029–1043, Nov. 2001, doi: 10.1002/nme.255.
- [113] M. C. Aguilera-Morillo, A. M. Aguilera, F. Jiménez-Molinos, and J. B. Roldán, “Stochastic modeling of Random Access Memories reset transitions,” *Math. Comput. Simul.*, vol. 159, pp. 197–209, May 2019, doi: 10.1016/j.matcom.2018.11.016.
- [114] J. S. Morris, “Functional Regression,” *Annu. Rev. Stat. Its Appl.*, vol.

- 2, no. 1, pp. 321–359, Apr. 2015, doi: 10.1146/annurev-statistics-010814-020413.
- [115] P. T. Reiss, J. Goldsmith, H. L. Shang, and R. T. Ogden, “Methods for scalar-on-function regression,” *Int. Stat. Rev.*, vol. 85, no. 2, pp. 228–249, 2017, doi: 10.1111/insr.12163.
- [116] G. M. James, “Generalized Linear Models with Functional Predictors,” *J. R. Stat. Soc. Ser. B (Statistical Methodol.*, vol. 64, no. 3, pp. 411–432, 2002.
- [117] M. Weitzman, S. Gortmaker, and A. Sobol, “Racial, Social, and Environmental Risks for Childhood Asthma,” *Am. J. Dis. Child.*, vol. 144, no. 11, pp. 1189–1194, Nov. 1990, doi: 10.1001/archpedi.1990.02150350021016.
- [118] C. G. Victora, J. P. Vaughan, and P. G. Smith, “Social and environmental influences on child mortality in Brazil: Logistic regression analysis of data from census files,” *J. Biosoc. Sci.*, vol. 18, no. 1, pp. 87–102, 1986, doi: 10.1017/S0021932000006520.
- [119] P. T. Kempe *et al.*, “Predictors of course in obsessive-compulsive disorder: logistic regression versus Cox regression for recurrent events,” *Acta Psychiatr. Scand.*, vol. 116, no. 3, pp. 201–210, Sep. 2007, doi: 10.1111/j.1600-0447.2007.00997.x.
- [120] C. R. Hayhoe, L. Leach, P. R. Turner, and Mo, “Discriminating the number of credit cards held by college students using credit and money attitudes *,” *J. Econ. Psychol.*, vol. 20, no. 6, pp. 643–656, Dec. 1999, doi: 10.1016/S0167-4870(99)00028-8.
- [121] M. Bensic, N. Sarlija, and M. Zekic-Susac, “Modelling small-business credit scoring by using logistic regression, neural networks and decision trees,” *Intell. Syst. Accounting, Financ. Manag.*, vol. 13, no. 3, pp. 133–150, Jul. 2005, doi: 10.1002/isaf.261.
- [122] E. ul Hassan, Z. Zainuddin, and S. Nordin, “A Review of Financial Distress Prediction Models: Logistic Regression and Multivariate Discriminant Analysis,” *Indian-Pacific J. Account. Financ.*, vol. 1, no. 3, pp. 13–23, Jul. 2017, doi: 10.32890/IPJAF.2017.1.3.15.
- [123] D. L. Langer, T. H. van der Kwast, A. J. Evans, J. Trachtenberg, B. C. Wilson, and M. A. Haider, “Prostate cancer detection with multi-parametric MRI: Logistic regression analysis of quantitative

- T2, diffusion-weighted imaging, and dynamic contrast-enhanced MRI,” *J. Magn. Reson. Imaging*, vol. 30, no. 2, pp. 327–334, Aug. 2009, doi: 10.1002/jmri.21824.
- [124] A. Direkvand-Moghadam, A. Khosravi, and K. Sayehmiri, “Predictive factors for preeclampsia in pregnant women: A univariate and multivariate logistic regression analysis,” *Acta Biochim. Pol.*, vol. 59, no. 4, pp. 673–677, Nov. 2012, doi: 10.18388/abp.2012_2109.
- [125] D. J. Hernandez *et al.*, “Predicting the outcome of prostate biopsy: comparison of a novel logistic regression-based model, the prostate cancer risk calculator, and prostate-specific antigen level alone,” *BJU Int.*, vol. 103, no. 5, pp. 609–614, Mar. 2009, doi: 10.1111/j.1464-410X.2008.08127.x.
- [126] T. D. Valenzuela, D. J. Roe, S. Cretin, D. W. Spaite, and M. P. Larsen, “Estimating Effectiveness of Cardiac Arrest Interventions,” *Circulation*, vol. 96, no. 10, pp. 3308–3313, Nov. 1997, doi: 10.1161/01.CIR.96.10.3308.
- [127] M. Escabias, A. M. Aguilera, and M. J. Valderrama, “Modeling environmental data by functional principal component logistic regression,” *Environmetrics*, vol. 16, no. 1, pp. 95–107, Feb. 2005, doi: 10.1002/env.696.
- [128] M. Escabias, A. M. Aguilera, and M. J. Valderrama, “Principal component estimation of functional logistic regression: discussion of two different approaches,” *J. Nonparametr. Stat.*, vol. 16, no. 3–4, pp. 365–384, Jun. 2004, doi: 10.1080/10485250310001624738.
- [129] J. M. Bland and D. G. Altman, “The odds ratio,” *BMJ*, vol. 320, no. 7247, p. 1468, May 2000, doi: 10.1136/bmj.320.7247.1468.
- [130] J. M. Schott, C. Frost, D. G. MacManus, F. Ibrahim, A. D. Waldman, and N. C. Fox, “Short echo time proton magnetic resonance spectroscopy in Alzheimer’s disease: a longitudinal multiple time point study,” *Brain*, vol. 133, no. 11, pp. 3315–3322, 2010, doi: 10.1093/brain/awq208.
- [131] S. Van Duijn, R. J. A. Nabuurs, S. G. Van Duinen, R. Natté, M. A. Van Buchem, and A. Alia, “Longitudinal monitoring of sex-related in vivo metabolic changes in the brain of Alzheimer’s disease transgenic mouse using magnetic resonance spectroscopy,” *J.*

- Alzheimer's Dis.*, vol. 34, no. 4, pp. 1051–1059, Jan. 2013, doi: 10.3233/JAD-122188.
- [132] W. T. Wang, P. Lee, H. W. Yeh, I. V. Smirnova, and I. Y. Choi, “Effects of acute and chronic hyperglycemia on the neurochemical profiles in the rat brain with streptozotocin-induced diabetes detected using in vivo 1H MR spectroscopy at 9.4 T,” *J. Neurochem.*, vol. 121, no. 3, pp. 407–417, May 2012, doi: 10.1111/j.1471-4159.2012.07698.x.
- [133] P. Lee, W. Leong, T. Tan, M. Lim, W. Han, and G. K. Radda, “In Vivo hyperpolarized carbon-13 magnetic resonance spectroscopy reveals increased pyruvate carboxylase flux in an insulin-resistant mouse model,” *Hepatology*, vol. 57, no. 2, pp. 515–524, Feb. 2013, doi: 10.1002/hep.26028.
- [134] S. Llufriu *et al.*, “Magnetic resonance spectroscopy markers of disease progression in multiple sclerosis,” *JAMA Neurol.*, vol. 71, no. 7, pp. 840–847, Jul. 2014, doi: 10.1001/jamaneurol.2014.895.
- [135] J. Orije *et al.*, “Longitudinal monitoring of metabolic alterations in cuprizone mouse model of multiple sclerosis using 1H-magnetic resonance spectroscopy,” *Neuroimage*, vol. 114, pp. 128–135, Jul. 2015, doi: 10.1016/j.neuroimage.2015.04.012.
- [136] H. P. A. Van Dongen, E. Olofsen, D. F. Dinges, and G. Maislin, “Mixed-Model Regression Analysis and Dealing with Interindividual Differences,” *Methods Enzymol.*, vol. 384, pp. 139–171, Jan. 2004, doi: 10.1016/S0076-6879(04)84010-2.
- [137] T. P. Garcia and K. Marder, “Statistical Approaches to Longitudinal Data Analysis in Neurodegenerative Diseases: Huntington’s Disease as a Model,” *Current Neurology and Neuroscience Reports*, vol. 17, no. 2. Current Medicine Group LLC 1, p. 14, 01-Feb-2017, doi: 10.1007/s11910-017-0723-4.
- [138] N. M. Laird and J. H. Ware, “Random-Effects Models for Longitudinal Data,” *Biometrics*, vol. 38, no. 4, p. 963, Dec. 1982, doi: 10.2307/2529876.
- [139] S. Greven, C. Crainiceanu, B. Caffo, and D. Reich, “Longitudinal functional principal component analysis,” *Electron. J. Stat.*, vol. 4, pp. 1022–1054, 2010, doi: 10.1214/10-EJS575.

- [140] C. M. Crainiceanu, A. M. Staicu, and C. Z. Di, “Generalized multilevel functional regression,” *J. Am. Stat. Assoc.*, vol. 104, no. 488, pp. 1550–1561, Dec. 2009, doi: 10.1198/jasa.2009.tm08564.
- [141] C. Z. Di, C. M. Crainiceanu, B. S. Caffo, and N. M. Punjabi, “Multilevel functional principal component analysis,” *Ann. Appl. Stat.*, vol. 3, no. 1, pp. 458–488, Mar. 2009, doi: 10.1214/08-AOAS206.
- [142] T. Krivobokova and G. Kauer mann, “A note on penalized spline smoothing with correlated errors,” *J. Am. Stat. Assoc.*, vol. 102, no. 480, pp. 1328–1337, Dec. 2007, doi: 10.1198/016214507000000978.
- [143] W. W. Stroup, *Generalized Linear Mixed Models*, 1st ed. CRC Press, 2012.
- [144] D. M. Berridge and R. Crouchley, *Multivariate Generalized Linear Mixed Models Using R*, 1st ed. CRC Press, 2011.
- [145] S. N. Wood, *Generalized additive models: an introduction with R*. Chapman & Hall/CRC, 2006.
- [146] C. E. McCulloch and S. R. Searle, *Generalized, Linear, and Mixed Models*. John Wiley & Sons, Inc. , 2000.
- [147] J. Goldsmith, C. M. Crainiceanu, B. Caffo, and D. Reich, “Longitudinal penalized functional regression for cognitive outcomes on neuronal tract measurements,” *J. R. Stat. Soc. Ser. C Appl. Stat.*, vol. 61, no. 3, pp. 453–469, May 2012, doi: 10.1111/j.1467-9876.2011.01031.x.
- [148] J. Gertheiss, J. Goldsmith, C. Crainiceanu, and S. Greven, “Longitudinal scalar-on-functions regression with application to tractography data,” *Biostatistics*, vol. 14, no. 3, pp. 447–461, 2013, doi: 10.1093/biostatistics/kxs051.
- [149] G. J. Patti, O. Yanes, and G. Siuzdak, “Metabolomics: the apogee of the omics trilogy,” *Nat. Rev. Mol. Cell Biol.*, vol. 13, no. 4, pp. 263–269, Apr. 2012, doi: 10.1038/nrm3314.
- [150] Y. Gu *et al.*, “A non-targeted metabolomics study on different glucose tolerance states,” *Int. J. Diabetes Dev. Ctries.*, vol. 39, no. 3, pp. 478–485, Jul. 2019, doi: 10.1007/s13410-018-0662-x.

- [151] N. Friedrich, “Metabolomics in diabetes research,” *J. Endocrinol.*, vol. 215, no. 1, pp. 29–42, 2012, doi: 10.1530/JOE-12-0120.
- [152] International Diabetes Federation, “IDF Diabetes Atlas,” 2019.
- [153] American Diabetes Association, “Diagnosis and classification of diabetes mellitus,” *Diabetes Care*, vol. 37, no. SUPPL.1, pp. S81–S90, Jan. 2014, doi: 10.2337/dc14-S081.
- [154] V. P. Singh, A. Bali, N. Singh, and A. S. Jaggi, “Advanced glycation end products and diabetic complications,” *Korean Journal of Physiology and Pharmacology*, vol. 18, no. 1. Korean Physiological Soc. and Korean Soc. of Pharmacology, pp. 1–14, 2014, doi: 10.4196/kjpp.2014.18.1.1.
- [155] R. Singh, A. Barden, T. Mori, and L. Beilin, “Advanced glycation end-products: A review,” *Diabetologia*, vol. 44, no. 2. Springer, pp. 129–146, 2001, doi: 10.1007/s001250051591.
- [156] D. M. Nathan, “Long-Term Complications of Diabetes Mellitus,” *New England Journal of Medicine*, vol. 328, no. 23. pp. 1676–1685, 10-Jun-1993, doi: 10.1056/NEJM199306103282306.
- [157] K. G. M. M. Alberti and P. Z. Zimmet, “Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: Diagnosis and classification of diabetes mellitus. Provisional report of a WHO consultation,” *Diabet. Med.*, vol. 15, no. 7, pp. 539–553, 1998, doi: 10.1002/(SICI)1096-9136(199807)15:7<539::AID-DIA668>3.0.CO;2-S.
- [158] M. Brownlee, “The pathobiology of diabetic complications: A unifying mechanism,” in *Diabetes*, 2005, vol. 54, no. 6, pp. 1615–1625, doi: 10.2337/diabetes.54.6.1615.
- [159] P. M. Trief *et al.*, “Natural history of cardiovascular disease in patients with diabetes: role of hyperglycemia,” *Diabetes Care*, vol. 30, no. 5, pp. 1266–1268, 2007.
- [160] P. Gkogkolou and M. Böhm, “Advanced glycation end products: Keyplayers in skin aging?,” *Dermato-Endocrinology*, vol. 4, no. 3. pp. 259–70, 01-Jul-2012, doi: 10.4161/derm.22028.
- [161] B. Duran-Jimenez *et al.*, “Advanced glycation end products in extracellular matrix proteins contribute to the failure of sensory nerve regeneration in diabetes,” *Diabetes*, vol. 58, no. 12, pp. 2893–

2903, Dec. 2009, doi: 10.2337/db09-0320.

- [162] A. H. Free, “Enzymatic Determinations of Glucose,” *Adv. Clin. Chem.*, vol. 6, no. C, pp. 67–96, Jan. 1963, doi: 10.1016/S0065-2423(08)60238-3.
- [163] K. P. Peterson, J. G. Pavlovich, D. Goldstein, R. Little, J. England, and C. M. Peterson, “What is hemoglobin A1c? An analysis of glycated hemoglobins by electrospray ionization mass spectrometry,” *Clin. Chem.*, vol. 44, no. 9, pp. 1951–1958, Sep. 1998, doi: 10.1093/clinchem/44.9.1951.
- [164] C. D. Saudek and J. C. Brick, “The clinical use of hemoglobin A1c,” in *Journal of Diabetes Science and Technology*, 2009, vol. 3, no. 4, pp. 629–634, doi: 10.1177/193229680900300402.
- [165] I. C. Kuo *et al.*, “Glycated Hemoglobin and Outcomes in Patients with Advanced Diabetic Chronic Kidney Disease,” *Sci. Rep.*, vol. 6, Jan. 2016, doi: 10.1038/srep20028.
- [166] N. Alqahtani, W. A. G. Khan, M. H. Alhumaidi, and Y. A. A. R. Ahmed, “Use of Glycated Hemoglobin in the Diagnosis of Diabetes Mellitus and Pre-diabetes and Role of Fasting Plasma Glucose, Oral Glucose Tolerance Test,” *Int. J. Prev. Med.*, vol. 4, no. 9, p. 1025, 2013.
- [167] American Diabetes Association, “Standards of medical care in diabetes-2019,” *Diabetes Care*, vol. 42, no. Supplement 1, 2019, doi: 10.2337/dc19-Sint01.
- [168] American Diabetes Association, “Standards of medical care in diabetes-2010,” *Diabetes Care*, vol. 33, no. SUPPL. 1, pp. S11–S61, Jan. 2010, doi: 10.2337/dc10-S011.
- [169] R. Derr, E. Garrett, G. A. Stacy, and C. D. Saudek, “Is HbA1c affected by glycemic instability?,” *Diabetes Care*, vol. 26, no. 10, pp. 2728–2733, Oct. 2003, doi: 10.2337/diacare.26.10.2728.
- [170] L. C. Clark and C. Lyons, “Electrode systems for continuous monitoring in cardiovascular surgery,” *Ann. N. Y. Acad. Sci.*, vol. 102, no. 1, pp. 29–45, Dec. 2006, doi: 10.1111/j.1749-6632.1962.tb13623.x.
- [171] B. Solnica, J. W. Naskalski, and J. Sieradzki, “Analytical performance of glucometers used for routine glucose self-

- monitoring of diabetic patients,” *Clin. Chim. Acta*, vol. 331, no. 1–2, pp. 29–35, May 2003, doi: 10.1016/S0009-8981(03)00079-2.
- [172] D. C. Klonoff, “Continuous Glucose Monitoring;,” *Diabetes Care*, vol. 28, no. 5, pp. 1231–1239, 2005, doi: 10.2337/diacare.28.5.1231.
- [173] E. Kulcu, J. A. Tamada, G. Reach, R. O. Potts, and M. J. Lesho, “Physiological Differences Between Interstitial Glucose and Blood Glucose Measured in Human Subjects,” *Diabetes Care*, vol. 26, no. 8, pp. 2405–2409, 2003, doi: 10.2337/diacare.26.8.2405.
- [174] J. E. F. Ward, B. A. Stetson, and S. P. L. Mokshagundam, “Patient perspectives on self-monitoring of blood glucose: Perceived recommendations, behaviors and barriers in a clinic sample of adults with type 2 diabetes,” *J. Diabetes Metab. Disord.*, vol. 14, no. 1, p. 43, May 2015, doi: 10.1186/s40200-015-0172-z.
- [175] J. Yadav, A. Rani, V. Singh, and B. M. Murari, “Prospects and limitations of non-invasive blood glucose monitoring using near-infrared spectroscopy,” *Biomedical Signal Processing and Control*, vol. 18. Elsevier Ltd, pp. 214–227, 01-Apr-2015, doi: 10.1016/j.bspc.2015.01.005.
- [176] W. V. Gonzales, A. T. Mobashsher, and A. Abbosh, “The progress of glucose monitoring—A review of invasive to minimally and non-invasive techniques, devices and sensors,” *Sensors (Switzerland)*, vol. 19, no. 4. MDPI AG, 02-Feb-2019, doi: 10.3390/s19040800.
- [177] A. Nawaz, P. Øhlckers, S. Sælid, M. Jacobsen, and M. Nadeem Akram, “Review: Non-Invasive Continuous Blood Glucose Measurement Techniques,” *J. Bioinforma. Diabetes*, vol. 1, no. 3, pp. 1–27, Jun. 2016, doi: 10.14302/issn.2374-9431.jbd-15-647.
- [178] D. Bruen, C. Delaney, L. Florea, and D. Diamond, “Glucose sensing for diabetes monitoring: Recent developments,” *Sensors (Switzerland)*, vol. 17, no. 8. MDPI AG, 12-Aug-2017, doi: 10.3390/s17081866.
- [179] K. Maruo *et al.*, “New methodology to obtain a calibration model for noninvasive near-infrared blood glucose monitoring,” *Appl. Spectrosc.*, vol. 60, no. 4, pp. 441–448, Apr. 2006, doi: 10.1366/000370206776593780.

- [180] H. Von Lilienfeld-Toal, M. Weidenmüller, A. Xhelaj, and W. Mäntele, “A novel approach to non-invasive glucose measurement by mid-infrared spectroscopy: The combination of quantum cascade lasers (QCL) and photoacoustic detection,” in *Vibrational Spectroscopy*, 2005, vol. 38, no. 1–2, pp. 209–215, doi: 10.1016/j.vibspec.2005.02.025.
- [181] S. Liakat, K. A. Bors, L. Xu, C. M. Woods, J. Doyle, and C. F. Gmachl, “Noninvasive in vivo glucose sensing on human subjects using mid-infrared light,” *Biomed. Opt. Express*, vol. 5, no. 7, p. 2397, Jul. 2014, doi: 10.1364/boe.5.002397.
- [182] J. Kottmann, J. M. Rey, J. Luginbühl, E. Reichmann, and M. W. Sigrist, “Glucose sensing in human epidermis using mid-infrared photoacoustic detection,” *Biomed. Opt. Express*, vol. 3, no. 4, pp. 667–80, Apr. 2012, doi: 10.1364/BOE.3.000667.
- [183] M. A. Pleitez, T. Lieblein, A. Bauer, O. Hertzberg, H. Von Lilienfeld-Toal, and W. Mäntele, “In vivo noninvasive monitoring of glucose concentration in human epidermis by mid-infrared pulsed photoacoustic spectroscopy,” *Anal. Chem.*, vol. 85, no. 2, pp. 1013–1020, Jan. 2013, doi: 10.1021/ac302841f.
- [184] A. Bauer, O. Hertzberg, A. Küderle, D. Strobel, M. A. Pleitez, and W. Mäntele, “IR-spectroscopy of skin in vivo: Optimal skin sites and properties for non-invasive glucose measurement by photoacoustic and photothermal spectroscopy,” *J. Biophotonics*, vol. 11, no. 1, Jan. 2018, doi: 10.1002/jbio.201600261.
- [185] O. Amir *et al.*, “Continuous noninvasive glucose monitoring technology based on ‘occlusion spectroscopy,’” in *Journal of Diabetes Science and Technology*, 2007, vol. 1, no. 4, pp. 463–469, doi: 10.1177/193229680700100403.
- [186] A. M. K. Enejder *et al.*, “Raman spectroscopy for noninvasive glucose measurements,” *J. Biomed. Opt.*, vol. 10, no. 3, p. 031114, 2005, doi: 10.1117/1.1920212.
- [187] J. C. Pickup, F. Hussain, N. D. Evans, O. J. Rolinski, and D. J. S. Birch, “Fluorescence-based glucose sensors,” *Biosensors and Bioelectronics*, vol. 20, no. 12. Elsevier Ltd, pp. 2555–2565, 15-Jun-2005, doi: 10.1016/j.bios.2004.10.002.
- [188] D. C. Klonoff, “Overview of fluorescence glucose sensing: A

- technology with a bright future,” in *Journal of Diabetes Science and Technology*, 2012, vol. 6, no. 6, pp. 1242–1250, doi: 10.1177/193229681200600602.
- [189] B. H. Malik and G. L. Coté, “Real-time, closed-loop dual-wavelength optical polarimetry for glucose monitoring,” *J. Biomed. Opt.*, vol. 15, no. 1, p. 017002, 2010, doi: 10.1117/1.3290819.
- [190] G. Purvinis, B. D. Cameron, and D. M. Altrogge, “Noninvasive polarimetric-based glucose monitoring: An in vivo study,” *J. Diabetes Sci. Technol.*, vol. 5, no. 2, pp. 380–387, 2011, doi: 10.1177/193229681100500227.
- [191] H. Ullah, F. Hussain, and M. Ikram, “Optical coherence tomography for glucose monitoring in blood,” *Appl. Phys. B Lasers Opt.*, vol. 120, no. 2, pp. 355–366, Aug. 2015, doi: 10.1007/s00340-015-6144-7.
- [192] K. V. Larin, M. S. Eledrisi, M. Motamedi, and R. O. Esenaliev, “Noninvasive blood glucose monitoring with optical coherence tomography: A pilot study in human subjects,” *Diabetes Care*, vol. 25, no. 12, pp. 2263–2267, Dec. 2002, doi: 10.2337/diacare.25.12.2263.
- [193] O. K. Cho, Y. O. Kim, H. Mitsumaki, and K. Kuwa, “Noninvasive Measurement of Glucose by Metabolic Heat Conformation Method,” *Clin. Chem.*, vol. 50, no. 10, pp. 1894–1898, Oct. 2004, doi: 10.1373/clinchem.2004.036954.
- [194] F. Tang, X. Wang, D. Wang, and J. Li, “Non-Invasive Glucose Measurement by Use of Metabolic Heat Conformation Method,” *Sensors*, vol. 8, no. 5, pp. 3335–3344, May 2008, doi: 10.3390/s8053335.
- [195] S. I. Gusev *et al.*, “Application of terahertz pulsed spectroscopy for the development of non-invasive glucose measuring method,” in *Progress in Electromagnetics Research Symposium*, 2017, pp. 3229–3232, doi: 10.1109/PIERS.2017.8262313.
- [196] O. Cherkasova, M. Nazarov, and A. Shkurinov, “Noninvasive blood glucose monitoring in the terahertz frequency range,” *Opt. Quantum Electron.*, vol. 48, no. 3, pp. 1–12, Mar. 2016, doi: 10.1007/s11082-016-0490-5.

- [197] M. Hofmann, G. Fischer, R. Weigel, and D. Kissinger, "Microwave-based noninvasive concentration measurements for biomedical applications," *IEEE Trans. Microw. Theory Tech.*, vol. 61, no. 5, pp. 2195–2204, 2013, doi: 10.1109/TMTT.2013.2250516.
- [198] T. Yilmaz, A. Brizzi, R. Foster, M. Munoz, and Y. Hao, "A patch resonator for sensing blood glucose changes," in *2014 31th URSI General Assembly and Scientific Symposium, URSI GASS 2014*, 2014, doi: 10.1109/URSIGASS.2014.6930125.
- [199] R. Zhang, Z. Qu, H. Jin, S. Liu, Y. Luo, and Y. Zheng, "Noninvasive Glucose Measurement by Microwave Biosensor with Accuracy Enhancement," in *Proceedings - IEEE International Symposium on Circuits and Systems*, 2018, vol. 2018-May, doi: 10.1109/ISCAS.2018.8351711.
- [200] P. H. Siegel, Y. Lee, and V. Pikov, "Millimeter-wave non-invasive monitoring of glucose in anesthetized rats," in *International Conference on Infrared, Millimeter, and Terahertz Waves, IRMMW-THz*, 2014, doi: 10.1109/IRMMW-THz.2014.6956294.
- [201] P. H. Siegel, W. Dai, R. A. Klöner, M. Csete, and V. Pikov, "First millimeter-wave animal in vivo measurements of L-Glucose and D-Glucose: Further steps towards a non-invasive glucometer," in *International Conference on Infrared, Millimeter, and Terahertz Waves, IRMMW-THz*, 2016, vol. 2016-November, doi: 10.1109/IRMMW-THz.2016.7758908.
- [202] S. Saha *et al.*, "A Glucose Sensing System Based on Transmission Measurements at Millimetre Waves using Micro strip Patch Antennas," *Sci. Rep.*, vol. 7, no. 1, pp. 1–11, Dec. 2017, doi: 10.1038/s41598-017-06926-1.
- [203] S. Costanzo, "Non-invasive microwave sensors for biomedical applications: New design perspectives," *Radioengineering*, vol. 26, no. 2, pp. 406–410, Jun. 2017, doi: 10.13164/re.2017.0406.
- [204] N. K. Nikolova, "Microwave Biomedical Imaging," in *Wiley Encyclopedia of Electrical and Electronics Engineering*, John Wiley & Sons, Inc., 2014, pp. 1–22.
- [205] A. Rosen, M. A. Stuchly, and A. Vander Vorst, "Applications of RF/microwaves in medicine," *IEEE Trans. Microw. Theory Tech.*, vol. 50, no. 3, pp. 963–974, Mar. 2002, doi: 10.1109/22.989979.

- [206] R. L. Yadava, "RF/microwaves in bio-medical applications," in *Proceedings of the International Conference on Electromagnetic Interference and Compatibility*, 2003, vol. 2003-January, pp. 81–85, doi: 10.1109/ICEMIC.2003.1287768.
- [207] O. P. Gandhi, "Some basic properties of biological tissues for potential biomedical applications of millimeter waves," *J. Microw. Power*, vol. 18, no. 3, pp. 295–304, 1983, doi: 10.1080/16070658.1983.11689334.
- [208] O. Boric-Lubecke, Y. Nikawa, W. Snyder, J. Lin, and K. Mizuno, "Novel microwave and millimeter-wave biomedical applications," in *4th International Conference on Telecommunications in Modern Satellite, Cable and Broadcasting Services, ITELSIKS 1999 - Proceedings*, 1999, vol. 1, pp. 186–193, doi: 10.1109/TELSKS.1999.804724.
- [209] Z. D. Taylor *et al.*, "THz and mm-wave sensing of corneal tissue water content: In vivo sensing and imaging results," *IEEE Trans. Terahertz Sci. Technol.*, vol. 5, no. 2, pp. 184–196, Mar. 2015, doi: 10.1109/TTHZ.2015.2392628.
- [210] C. Yu, S. Fan, Y. Sun, and E. Pickwell-Macpherson, "The potential of terahertz imaging for cancer diagnosis: A review of investigations to date," *Quant. Imaging Med. Surg.*, vol. 2, no. 1, pp. 33–45, 2012, doi: 10.3978/j.issn.2223-4292.2012.01.04.
- [211] A. Y. Pawar, D. D. Sonawane, K. B. Erande, and D. V. Derle, "Terahertz technology and its applications," *Drug Invention Today*, vol. 5, no. 2. No longer published by Elsevier, pp. 157–163, 01-Jun-2013, doi: 10.1016/j.dit.2013.03.009.
- [212] L. Yu *et al.*, "The medical application of terahertz technology in non-invasive detection of cells and tissues: opportunities and challenges," *RSC Advances*, vol. 9, no. 17. Royal Society of Chemistry, pp. 9354–9363, 22-Mar-2019, doi: 10.1039/C8RA10605C.
- [213] P. H. Siegel, "Terahertz technology in biology and medicine," *IEEE Trans. Microw. Theory Tech.*, vol. 52, no. 10, pp. 2438–2447, Oct. 2004, doi: 10.1109/TMTT.2004.835916.
- [214] X. Yang *et al.*, "Biomedical Applications of Terahertz Spectroscopy and Imaging," *Trends in Biotechnology*, vol. 34, no. 10. Elsevier Ltd, pp. 810–824, 01-Oct-2016, doi:

- 10.1016/j.tibtech.2016.04.008.
- [215] A. G. Pakhomov, Y. Akyel, O. N. Pakhomova, B. E. Stuck, and M. R. Murphy, “Current State and Implications of Research on Biological Effects of Millimeter Waves: A Review of the Literature,” *Bioelectromagnetics*, vol. 19, no. 7, pp. 393–413, 1998, doi: 10.1002/(SICI)1521-186X(1998)19:7<393::AID-BEM1>3.0.CO;2-X.
- [216] R. Chandra, H. Zhou, I. Balasingham, and R. M. Narayanan, “On the Opportunities and Challenges in Microwave Medical Sensing and Imaging,” *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 7. IEEE Computer Society, pp. 1667–1682, 01-Jul-2015, doi: 10.1109/TBME.2015.2432137.
- [217] A. Winter, J. Laing, R. Paglione, and F. Sterzer, “Microwave Hyperthermia for Brain Tumors,” *Neurosurgery*, vol. 17, no. 3, pp. 387–399, Sep. 1985, doi: 10.1227/00006123-198509000-00001.
- [218] M. Salcman and G. M. Samaras, “Interstitial microwave hyperthermia for brain tumors,” *J. Neurooncol.*, vol. 1, no. 3, pp. 225–236, Sep. 1983, doi: 10.1007/BF00165607.
- [219] C. J. Simon, D. E. Dupuy, and W. W. Mayo-Smith, “Microwave ablation: Principles and applications,” in *Radiographics*, 2005, vol. 25, no. SPEC. ISS., doi: 10.1148/rg.25si055501.
- [220] C. L. Brace, “Microwave tissue ablation: Biophysics, technology, and applications,” *Critical Reviews in Biomedical Engineering*, vol. 38, no. 1. Begell House Inc., pp. 65–78, 2010, doi: 10.1615/CritRevBiomedEng.v38.i1.60.
- [221] J. R. Lajara Vizcaíno and J. Sebastián Pelegrí, *LabVIEW: Entorno gráfico de programación*, 2nd ed. MARCOMBO, S.A., 2011.
- [222] J. H. Scofield, “Frequency-domain description of a lock-in amplifier,” *Am. J. Phys.*, vol. 62, no. 2, pp. 129–133, Feb. 1994, doi: 10.1119/1.17629.
- [223] R. Burdett, “Amplitude Modulated Signals: The Lock-in Amplifier,” in *Handbook of Measuring System Design*, Chichester, UK: John Wiley & Sons, Ltd, 2005.
- [224] T. A. Lutz and S. C. Woods, “Overview of animal models of obesity,” *Curr. Protoc. Pharmacol.*, vol. 58, no. 1, pp. 5.61.1-5.61.18,

- 2012, doi: 10.1002/0471141755.ph0561s58.
- [225] M. L. Graham, J. L. Janecek, J. A. Kittredge, B. J. Hering, and H.-J. Schuurman, “The streptozotocin-induced diabetic nude mouse model: Differences between animals from different sources,” *Comp. Med.*, vol. 61, no. 4, pp. 356–360, 2011.
- [226] A. J. F. King, “The use of animal models in diabetes research,” *British Journal of Pharmacology*, vol. 166, no. 3. Wiley-Blackwell, pp. 877–894, Jun-2012, doi: 10.1111/j.1476-5381.2012.01911.x.
- [227] J. Shah, “Pharmacoeconomics of continuous drug administration using ALZET pumps,” *ALZET Spec. Deliv. Newsl.*, vol. 25, no. 1, pp. 5–6, 2008.
- [228] B. L. Furman, “Streptozotocin-Induced Diabetic Models in Mice and Rats,” *Curr. Protoc. Pharmacol.*, vol. 70, no. 1, pp. 5.47.1-5.47.20, Sep. 2015, doi: 10.1002/0471141755.ph0547s70.
- [229] R Core Team, “R: A Language and Environment for Statistical Computing.” R Foundation for Statistical Computing, Vienna, Austria, 2019.
- [230] J. O. Ramsay, H. Wickham, S. Graves, and G. Hooker, “fda: Functional Data Analysis.” 2020.
- [231] R. M. Balabin and S. V. Smirnov, “Variable selection in near-infrared spectroscopy: Benchmarking of feature selection methods on biodiesel data,” *Anal. Chim. Acta*, vol. 692, no. 1–2, pp. 63–72, Apr. 2011, doi: 10.1016/j.aca.2011.03.006.
- [232] N. C. Dingari, I. Barman, J. W. Kang, C.-R. Kong, R. R. Dasari, and M. S. Feld, “Wavelength selection-based nonlinear calibration for transcutaneous blood glucose sensing using Raman spectroscopy,” *J. Biomed. Opt.*, vol. 16, no. 8, p. 087009, 2011, doi: 10.1117/1.3611006.
- [233] Q. Zhang, Q. Li, and G. Zhang, “Rapid Determination of Leaf Water Content Using VIS/NIR Spectroscopy Analysis with Wavelength Selection,” *J. Spectrosc.*, vol. 27, no. 2, pp. 93–105, 2012, doi: 10.1155/2012/276795.
- [234] M. Vohland, M. Ludwig, S. Thiele-Bruhn, and B. Ludwig, “Determination of soil properties with visible to near- and mid-infrared spectroscopy: Effects of spectral variable selection,”

- Geoderma*, vol. 223–225, no. 1, pp. 88–96, Jul. 2014, doi: 10.1016/j.geoderma.2014.01.013.
- [235] A. G. Lalkhen and A. McCluskey, “Clinical tests: sensitivity and specificity,” *Contin. Educ. Anaesth. Crit. Care Pain*, vol. 8, no. 6, pp. 221–223, Dec. 2008, doi: 10.1093/BJACEACCP/MKN041.
- [236] K. Sørensen, “Receiver-operating characteristic curve analysis in diagnostic, prognostic and predictive biomarker research,” *Journal of Clinical Pathology*, vol. 62, no. 1, pp. 1–5, Jan-2009, doi: 10.1136/jcp.2008.061010.
- [237] P. J. Davis and P. Rabinowitz, *Methods of Numerical Integration*, 2nd ed. Courier Corporation, 2007.
- [238] D. M. Nathan, H. Turgeon, and S. Regan, “Relationship between glycated haemoglobin levels and mean glucose levels over time,” *Diabetologia*, vol. 50, no. 11, pp. 2239–2244, Nov. 2007, doi: 10.1007/s00125-007-0803-0.
- [239] Y. Tsuruo *et al.*, “Persistent median artery in the hand: A report with a brief review of the literature,” *Anatomical Science International*, vol. 81, no. 4, Springer, pp. 242–252, Dec-2006, doi: 10.1111/j.1447-073X.2006.00150.x.
- [240] B. Strauch and W. de Moura, “Arterial system of the fingers,” *J. Hand Surg. Am.*, vol. 15, no. 1, pp. 148–154, Jan. 1990, doi: 10.1016/S0363-5023(09)91123-6.
- [241] J. Miles, “R Squared, Adjusted R Squared,” in *Wiley StatsRef: Statistics Reference Online*, Chichester, UK: John Wiley & Sons, Ltd, 2014.
- [242] A. C. Cameron and F. A. G. Windmeijer, “R-squared measures for count data regression models with applications to health-care utilization,” *J. Bus. Econ. Stat.*, vol. 14, no. 2, pp. 209–220, 1996, doi: 10.1080/07350015.1996.10524648.
- [243] P. Sedgwick, “Pearson’s correlation coefficient,” *BMJ (Online)*, vol. 345, no. 7864. British Medical Journal Publishing Group, 07-Jul-2012, doi: 10.1136/bmj.e4483.
- [244] R. C. Gardner and R. W. J. Neufeld, “What the correlation coefficient really tells us about the individual,” *Can. J. Behav. Sci. / Rev. Can. des Sci. du Comport.*, vol. 45, no. 4, pp. 313–319, Oct. 2013,

doi: 10.1037/a0033342.

- [245] F. Dornuf *et al.*, “Classification of skin phenotypes caused by diabetes mellitus using complex scattering parameters in the millimeter-wave frequency range,” *Sci. Rep.*, vol. 7, no. 1, pp. 1–8, Dec. 2017, doi: 10.1038/s41598-017-06034-0.
- [246] L. Yang and Y. Hong, “Adaptive penalized splines for data smoothing,” *Comput. Stat. Data Anal.*, vol. 108, pp. 70–83, Apr. 2017, doi: 10.1016/j.csda.2016.10.022.
- [247] M. X. Rodríguez-Álvarez, D. J. Lee, T. Kneib, M. Durbán, and P. Eilers, “Fast smoothing parameter separation in multidimensional generalized P-splines: the SAP algorithm,” *Stat. Comput.*, vol. 25, no. 5, pp. 941–957, Sep. 2015, doi: 10.1007/s11222-014-9464-2.

8. APPENDIX

8.1. ETHICAL ISSUES

8.1.1. ANIMAL EXPERIMENTS

All the experimental procedures involved in the mice experiments were carried out according to European and Spanish laws and regulations (European convention ETS 1 2 3, about the use and protection of vertebrate mammals used in experimentation and other scientific purposes, Directive 2010/63/UE and Spanish Law 6/2013, and R.D. 53/2013 about the protection and use of animals in scientific research). Procedures were approved by the Animal Experimentation Ethical Committee of the CIEMAT according to all external and internal bio-safety and bio-ethics guidelines, and by Spanish competent authority with registered number PROEX 176/15.

8.1.2. PILOT CLINICAL STUDY

The clinical study protocol (Protocol code: FJD-ESPEC-DM-17-01) was approved by the Comité de Ética de la Investigación del Hospital Universitario Fundación Jiménez Díaz (CEI/CEIm-FJD). The clinical study was performed in accordance to the ethical principles regarding human experimentation established in the Declaration of Helsinki developed by the World Medical Association, the guidelines for Good Clinical Practice of the International Committee of Harmonization, and the Law 14/2017, of 3 July, on Biomedical Research. All the information related to the participants was processed according to the Organic Law 15/1999, of 13 December, on Protection of Personal Data.

INFORME DEL COMITE DE ÉTICA DE LA INVESTIGACION

Dra. Macarena Bonilla Porras, Vicepresidenta del COMITE DE ÉTICA DE LA INVESTIGACION DE LA FUNDACION JIMENEZ DIAZ

CERTIFICA:

Que en la reunión del CEIm-FJD que tuvo lugar el día 10 de abril de 2018 (acta nº 07/18) se evaluó el estudio referido y, ha decidido:


A P R O B A R

La propuesta para que se realice el estudio clínico titulado: **“Estudio de la correlación entre una medida no-invasiva de estados de hiperglucemia sostenida usando espectroscopia de ondas milimétricas y la Hb glicosilada en pacientes con Diabetes Mellitus tipo 1”**, Código Protocolo: FJD-ESPEC-DM-17-01. Versión del protocolo y la HIP: 2.1, 23 de Marzo de 2018.

Además, hace constar que:

1. En dicha reunión se cumplieron los requisitos establecidos en la legislación vigente –Real decreto 1090/2015 y Decreto 39/94 de la CAM– para que la decisión del citado CEIm sea válida.
2. El Estudio reúne las normas éticas estándar de nuestra Institución para la realización de este tipo de estudios.
3. Que se cumplen los preceptos éticos formulados en la Orden SAS 3470/2009 y la Declaración de Helsinki de la Asociación Médica mundial sobre principios éticos para las investigaciones médicas en seres humanos y en sus posteriores revisiones, así como aquellos exigidos por la normativa aplicable en función de las características del estudio.
4. El CEImFJD, tanto en su composición como en sus procedimientos, cumple con las normas de BPC (CPMP/CH/135/95) y con la legislación vigente que regula su funcionamiento, y que la composición del CEIm FJD es la indicada en el anexo I, teniendo en cuenta que en el caso de que algún miembro participe en el estudio o declare algún conflicto de interés no habrá participado en la evaluación ni en el dictamen.
5. Asimismo, hacemos constar no existe contraprestación económica para el centro y los investigadores.
6. Además, este comité recuerda la obligación de realizar el seguimiento del estudio de acuerdo a la legislación vigente

Lo que firmo en Madrid a 10 de abril de 2018


Dra. Macarena Bonilla Porras
Vicepresidenta CEImFJD

Conocido y conforme:


Alberto Montero Manso
Area Gestión Investigación

Vº Bº


Juan Antonio Alvaro de la Parra
Gerente FJD

ER_FIC 102/2017_FJD

**Anexo I
 COMPOSICIÓN DEL CEIm**

Presidente	Dr. Gabriel Herrero-Beaumont Cuenca. Médico Asistencial. Reumatología de FJD
Vicepresidente	Dra. Macarena Bonilla Porras. Farmacéutico de la FJD
Secretaria técnica	Dra. Lucía Llanos Jiménez. Farmacóloga de la FJD
Vocales	Dr. Javier Bécares Martínez. Farmacéutico de la FJD
	Dra. Miriam Blanco Rodríguez. Médico Asistencial. Pediatría de FJD
	Dr. Alfonso Cabello Ubeda. Médico Asistencial. Medicina Interna FJD
	Dña. Ana García Díaz, abogada
	D. Cesáreo Goicoechea Goicoechea. Lego no vinculado a la Institución
	Dr. Raúl Córdoba Mascuñano. Médico Asistencial. Hematología FJD
	Dra. Isabel Egocheaga Cabello. Médico de Atención Primaria. CS Isla de Oza
	Dña. María del Mar Jiménez del Castillo. DUE de la FJD
	Dra. Brezo Martínez-Amores Martínez. Médico Asistencial. Oncología del HRJC
	Dra. Dolores Martínez Pérez. Médico Asistencial. Cirugía Maxilofacial de FJD
	Dr. Germán Peces Barba. Médico Asistencial. Miembro Comité de Investigación. Neumología de la FJD
	Dr. Gonzalo Pizarro Sánchez. Médico Asistencial. Cardiología HUQM
	Dr. Francisco Javier Ruiz Hornillos. Miembro Comité de Investigación y Miembro del Comité de Ética Asistencial. Médico Asistencial. Alergología del HIE
	Dra. Olga Sánchez Pernaute (Secretaria Suplente). Médico Asistencial. Reumatología de la FJD

Anexo II

CENTROS E INVESTIGADORES PRINCIPALES Y COLABORADORES

Título: "Estudio de la correlación entre una medida no-invasiva de estados de hiperglucemia sostenida usando espectroscopía de ondas milimétricas y la Hb glicosilada en pacientes con Diabetes Mellitus tipo 1", Código Protocolo: FJD-ESPEC-DM-17-01. Versión del protocolo y la HIP: 2.1, 23 de Marzo de 2018

Fecha de actualización del anexo II: **10 de abril de 2018**

INVESTIGADOR PRINCIPAL	INVESTIGADORES COLABORADORES	CENTROS DE REALIZACIÓN
Eduarne Lecumberri Pascual	María José de la Cruz Carolina Suárez Lucía Llanos Jiménez	Fundación Jiménez Díaz Servicio de Endocrinología Unidad de Investigación Clínica

8.2. INFORMED CONSENTS

8.2.1. CONTROLS

HOJA DE INFORMACIÓN AL PARTICIPANTE (CONTROLES)

TÍTULO DEL ESTUDIO: Estudio de la correlación entre una medida no-invasiva de estados de hiperglucemia sostenida usando espectroscopia de ondas milimétricas y la Hb glicosilada en pacientes con Diabetes Mellitus tipo 1^o
Código Protocolo: FJD-ESPEC-DM-17-01, versión 2.1, 23 de Marzo de 2018

INVESTIGADORA PRINCIPAL: Edurne Lecumberri Pascual. Médico Adjunto Servicio de Endocrinología. Hospital Universitario Fundación Jiménez Díaz.

INTRODUCCIÓN:

Nos dirigimos a usted para informarle sobre un estudio de investigación en el que se le invita a participar. La intención de este escrito es que reciba la información correcta y suficiente para que pueda evaluar si quiere o no participar en este estudio. Para ello lea esta hoja informativa con atención y nosotros le aclararemos cualquier duda que le pueda surgir.

PARTICIPACIÓN VOLUNTARIA:

Debe saber que su participación en este estudio es voluntaria y que puede decidir no participar o cambiar su decisión y retirar el consentimiento en cualquier momento sin que por ello se altere la relación con su médico ni se produzca perjuicio alguno en su tratamiento.

DESCRIPCIÓN GENERAL DEL ESTUDIO:

Desde la Universidad Carlos III, en colaboración con el Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT) y con la Fundación Jiménez Díaz, se está desarrollando un nuevo método no invasivo (que no implique pinchazos) para la determinación del control del azúcar en sangre a través del análisis de un pliegue de piel en pacientes que tengan diabetes tipo 1. El tratamiento de esta enfermedad para evitar que aparezcan complicaciones a corto, medio y largo plazo requiere un estrecho control de los niveles de azúcar en sangre y un ajuste del mismo en función de cómo se encuentren. Para ello, es necesario que los pacientes se pinchen varias veces al día para determinar los niveles de azúcar, y que además acudan varias veces al año al médico para hacerse analítica que determine otros parámetros que dan idea del control de su enfermedad, como la Hemoglobina glicosilada (HbA1c).

Hasta este momento, se han realizado mediciones en animales, con buenos resultados. El objetivo de este proyecto es el desarrollo, validación y evaluación de esta nueva herramienta en seres humanos. Para ello, se necesita realizar la medición con la nueva herramienta tanto en pacientes con DM tipo 1 como en personas que como usted no tienen diabetes, y comparar los resultados con las mediciones habituales para el control de la diabetes. Por este motivo le invitamos a participar en el estudio. Si su médico cree que cumple los criterios para participar y usted está de acuerdo, tendrá que dar su consentimiento por escrito.

El estudio tiene una duración prevista de 6 meses, y participarán 30 personas, de las cuales 20 serán pacientes con DM tipo 1 y 10 serán personas sin diabetes. Las mediciones con la nueva herramienta se llevarán a cabo el día que usted acuda a su revisión habitual. Se llevarán a cabo 3 mediciones (al inicio, a los 3 y a los 6 meses). La medición consiste en colocar un pliegue de la piel espacio que hay entre el primer y el segundo dedo de la mano del paciente entre las dos sondas de medida del aparato. Para ello, el médico tomará con los dedos índice y pulgar de ambas manos un pliegue cutáneo en esa zona efectuando una pequeña tracción hacia afuera que permita la buena formación del pliegue. Una vez asegurado que ambos lados del pliegue son

paralelos se colocará el pliegue entre las sondas del equipo de medida, que habrán sido previamente separadas 70 mm. El técnico encargado de la toma de medidas procederá a cerrar el espacio entre sondas hasta que éstas entren en contacto con la piel, pinzando de esta forma el pliegue, sin que esto produzca el más mínimo dolor para el paciente. El médico mantendrá sujeto el pliegue de piel para asegurar su posición durante la realización de la medida. Tras aproximadamente 60 segundos, en los que el paciente deberá permanecer inmóvil (no notará absolutamente nada durante la medida), la caracterización se habrá completado. Tras ello, el técnico procederá a separar las sondas, liberando la piel del paciente.

Su participación en el estudio consiste en permitir que se le realicen estas mediciones y una analítica de sangre el día de la medición en dar su permiso para que se recoja información relevante sobre su enfermedad para este estudio. Las visitas y pruebas realizadas desde Endocrinología y Nutrición son las mismas que deben realizarse en su control habitual, es decir, no se hará ninguna prueba extraordinaria por el hecho de participar en el estudio salvo por la medición no invasiva.

BENEFICIOS Y RIESGOS DERIVADOS DE SU PARTICIPACIÓN EN EL ESTUDIO

Es probable que usted no se beneficie directamente de su participación en este estudio. Independientemente de su participación, usted recibirá la atención médica habitual. Este estudio puede contribuir al desarrollo de una nueva herramienta que en un futuro podría permitir un seguimiento de la enfermedad que evitara los pinchazos.

Usted no se va a someter a ningún riesgo por su participación en este estudio. La medición, como se le ha explicado, no conlleva riesgos. El pinchazo en una vena del brazo para hacerle la analítica puede causar algún efecto secundario como inflamación de la vena (flebitis), hematoma, trombo en la vena o inflamación del tejido bajo la piel, que suelen ser leves y resolverse sin incidencias.

CONFIDENCIALIDAD

El tratamiento, la comunicación y la cesión de los datos de carácter personal de todos los sujetos participantes se ajustará a lo dispuesto en la Ley Orgánica 15/1999, de 13 de diciembre de protección de datos de carácter personal. De acuerdo a lo que establece la legislación mencionada, usted puede ejercer los derechos de acceso, modificación, oposición y cancelación de datos, para lo cual deberá dirigirse a su médico del estudio. Los datos recogidos para el estudio estarán identificados mediante un código y solo su médico del estudio/colaboradores podrá relacionar dichos datos con usted y con su historia clínica. Por lo tanto, su identidad no será revelada a persona alguna salvo excepciones (si existe alguna situación especial por la que se necesitara conocer la identidad del sujeto para cumplir con algún requisito del estudio se debe explicar en este apartado), en caso de urgencia médica o requerimiento legal. Sólo se transmitirán a terceros y a otros países los datos recogidos para el estudio que en ningún caso contendrán información que le pueda identificar directamente, como nombre y apellidos, iniciales, dirección, nº de la seguridad social, etc. En el caso de que se produzca esta cesión, será para los mismos fines del estudio descrito y garantizando la confidencialidad como mínimo con el nivel de protección de la legislación vigente en nuestro país. El acceso a su información personal quedará restringido al médico del estudio/colaboradores, autoridades sanitarias, al Comité Ético de Investigación Clínica, cuando lo precisen para comprobar los datos y procedimientos del estudio, pero siempre manteniendo la confidencialidad de los mismos de acuerdo a la legislación vigente.

OTRA INFORMACIÓN RELEVANTE

Debe saber que puede ser excluido del estudio si el promotor o los investigadores del estudio lo consideran oportuno, ya sea por motivos de seguridad, por cualquier acontecimiento adverso que se produzca o porque consideren que no está cumpliendo con los procedimientos establecidos. En cualquiera de los casos, usted recibirá una explicación adecuada del motivo que ha ocasionado su retirada del estudio.

Al firmar la hoja de consentimiento adjunta, se compromete a cumplir con los procedimientos del estudio que se le han expuesto. Cuando acabe su participación recibirá el mejor tratamiento disponible y que su médico considere el más adecuado para su enfermedad.

Si tiene cualquier pregunta ahora o en cualquier momento del acerca del estudio, de su seguridad o de sus derechos, pregunte al investigador responsable del estudio o a los miembros de su equipo. La investigadora responsable del estudio es la Dra. Edurne Lecumberri, del Servicio de Endocrinología. En caso de preguntas o dudas acerca del estudio podrá ponerse en contacto con ella en el teléfono 915504846.

FORMULARIO DE CONSENTIMIENTO INFORMADO POR ESCRITO

TÍTULO DEL ESTUDIO: Estudio de la correlación entre una medida no-invasiva de estados de hiperglucemia sostenida usando espectroscopia de ondas milimétricas y la Hb glicosilada en pacientes con Diabetes Mellitus tipo 1*
 Código Protocolo: FJD-ESPEC-DM-17-01

INVESTIGADORA PRINCIPAL: Edurne Lecumberri Pascual. Médico Adjunto Servicio de Endocrinología. Hospital Universitario Fundación Jiménez Díaz.

Yo (Nombre y apellidos)

.....

- He leído la hoja de información que se me ha entregado.
- He podido hacer preguntas sobre el estudio.
- He recibido suficiente información sobre el estudio.
- He hablado con

(Nombre y apellidos del investigador)

Comprendo que mi participación es voluntaria.

Comprendo que puedo retirarme del estudio:

- Cuando quiera.
- Sin tener que dar explicaciones.
- Sin que esto repercuta en mis cuidados médicos.

Presto libremente mi conformidad para participar en el estudio.

En.....a..... de.....
 de.....

Firma del participante	Firma del investigador
Fecha: ___/___/___	Fecha: ___/___/___

8.2.2. PATIENTS

HOJA DE INFORMACIÓN AL PARTICIPANTE

TÍTULO DEL ESTUDIO: Estudio de la correlación entre una medida no-invasiva de estados de hiperglucemia sostenida usando espectroscopia de ondas milimétricas y la Hb glicosilada en pacientes con Diabetes Mellitus tipo 1"
Código Protocolo: FJD-ESPEC-DM-17-01, versión 2.1, 23 de Marzo de 2018

INVESTIGADORA PRINCIPAL: Edurne Lecumberri Pascual. Médico Adjunto Servicio de Endocrinología. Hospital Universitario Fundación Jiménez Díaz.

INTRODUCCIÓN:

Nos dirigimos a usted para informarle sobre un estudio de investigación en el que se le invita a participar. La intención de este escrito es que reciba la información correcta y suficiente para que pueda evaluar si quiere o no participar en este estudio. Para ello lea esta hoja informativa con atención y nosotros le aclararemos cualquier duda que le pueda surgir.

PARTICIPACIÓN VOLUNTARIA:

Debe saber que su participación en este estudio es voluntaria y que puede decidir no participar o cambiar su decisión y retirar el consentimiento en cualquier momento sin que por ello se altere la relación con su médico ni se produzca perjuicio alguno en su tratamiento.

DESCRIPCIÓN GENERAL DEL ESTUDIO:

Usted padece Diabetes tipo 1 (Dm tipo 1) y se encuentra en seguimiento por el Servicio de Endocrinología. Como sabe, el tratamiento de esta enfermedad para evitar que aparezcan complicaciones a corto, medio y largo plazo requiere un estrecho control de los niveles de azúcar en sangre y un ajuste del mismo en función de cómo se encuentren. Para ello, es necesario que los pacientes como usted se pinchen varias veces al día para determinar los niveles de azúcar, y que además acudan varias veces al año al médico para hacerse analítica que determine otros parámetros que dan idea del control de su enfermedad, como la Hemoglobina glicosilada (HbA1c).

Desde la Universidad Carlos III, en colaboración con el Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT) y con la Fundación Jiménez Díaz, se está desarrollando un nuevo método no invasivo (que no implique pinchazos) para la determinación del control del azúcar en sangre a través del análisis de un pliegue de piel en pacientes.

Hasta este momento, se han realizado mediciones en animales, con buenos resultados. El objetivo de este proyecto es el desarrollo, validación y evaluación de esta nueva herramienta en seres humanos. Para ello, se necesita realizar la medición con la nueva herramienta tanto en pacientes con DM tipo 1 como en personas sin diabetes, y comparar los resultados con las mediciones habituales para el control de la diabetes.

Si su médico cree que cumple los criterios para participar y usted está de acuerdo, tendrá que dar su consentimiento por escrito.

El estudio tiene una duración prevista de 6 meses, y participarán 30 personas, de las cuales 20 serán pacientes con Dm tipo 1 y 10 serán personas sin diabetes. Las mediciones con la nueva herramienta se llevarán a cabo el día que usted acuda a su revisión habitual. Se llevarán a cabo 3 mediciones (al inicio, a los 3 y a los 6 meses). La medición consiste en colocar un pliegue de la piel del espacio que hay entre el primer y el segundo dedo de la mano entre las dos sondas de medida del aparato. Para ello, el médico tomará con los dedos índice y pulgar de ambas manos un pliegue cutáneo en esa zona efectuando una pequeña tracción hacia afuera que permita la

buena formación del pliegue. Una vez asegurado que ambos lados del pliegue son paralelos se colocará el pliegue entre las sondas del equipo de medida, que habrán sido previamente separadas 70 mm. El técnico encargado de la toma de medidas procederá a cerrar el espacio entre sondas hasta que éstas entren en contacto con la piel, pinzando de esta forma el pliegue, sin que esto produzca el más mínimo dolor para el paciente. El médico mantendrá sujeto el pliegue de piel para asegurar su posición durante la realización de la medida. Tras aproximadamente 60 segundos, en los que el paciente deberá permanecer inmóvil (no notará absolutamente nada durante la medida), la caracterización se habrá completado. Tras ello, el técnico procederá a separar las sondas, liberando la piel del paciente.

Su participación en el estudio consiste en permitir que se le realicen estas mediciones y en dar su permiso para que se recoja información relevante sobre su enfermedad para este estudio. Las visitas y pruebas realizadas desde Endocrinología y Nutrición son las mismas que deben realizarse en el control de la diabetes normal, es decir, no se hará ninguna prueba extraordinaria por el hecho de participar en el estudio salvo por la medición no invasiva.

BENEFICIOS Y RIESGOS DERIVADOS DE SU PARTICIPACIÓN EN EL ESTUDIO

Es probable que usted no se beneficie directamente de su participación en este estudio. Independientemente de su participación, usted recibirá la atención médica habitual. Este estudio puede contribuir al desarrollo de una nueva herramienta que en un futuro podría permitir un seguimiento de la enfermedad que evitara los pinchazos.

Usted no se va a someter a ningún riesgo por su participación en este estudio. La medición, como se le ha explicado, no conlleva riesgos.

CONFIDENCIALIDAD

El tratamiento, la comunicación y la cesión de los datos de carácter personal de todos los sujetos participantes se ajustará a lo dispuesto en la Ley Orgánica 15/1999, de 13 de diciembre de protección de datos de carácter personal. De acuerdo a lo que establece la legislación mencionada, usted puede ejercer los derechos de acceso, modificación, oposición y cancelación de datos, para lo cual deberá dirigirse a su médico del estudio. Los datos recogidos para el estudio estarán identificados mediante un código y solo su médico del estudio/colaboradores podrá relacionar dichos datos con usted y con su historia clínica. Por lo tanto, su identidad no será revelada a persona alguna salvo excepciones (si existe alguna situación especial por la que se necesitara conocer la identidad del sujeto para cumplir con algún requisito del estudio se debe explicar en este apartado), en caso de urgencia médica o requerimiento legal. Sólo se transmitirán a terceros y a otros países los datos recogidos para el estudio que en ningún caso contendrán información que le pueda identificar directamente, como nombre y apellidos, iniciales, dirección, nº de la seguridad social, etc. En el caso de que se produzca esta cesión, será para los mismos fines del estudio descrito y garantizando la confidencialidad como mínimo con el nivel de protección de la legislación vigente en nuestro país. El acceso a su información personal quedará restringido al médico del estudio/colaboradores, autoridades sanitarias, al Comité Ético de Investigación Clínica, cuando lo precisen para comprobar los datos y procedimientos del estudio, pero siempre manteniendo la confidencialidad de los mismos de acuerdo a la legislación vigente.

OTRA INFORMACIÓN RELEVANTE

Debe saber que puede ser excluido del estudio si el promotor o los investigadores del

estudio lo consideran oportuno, ya sea por motivos de seguridad, por cualquier acontecimiento adverso que se produzca o porque consideren que no está cumpliendo con los procedimientos establecidos. En cualquiera de los casos, usted recibirá una explicación adecuada del motivo que ha ocasionado su retirada del estudio.

Al firmar la hoja de consentimiento adjunta, se compromete a cumplir con los procedimientos del estudio que se le han expuesto. Cuando acabe su participación recibirá el mejor tratamiento disponible y que su médico considere el más adecuado para su enfermedad.

Si tiene cualquier pregunta ahora o en cualquier momento del acerca del estudio, de su seguridad o de sus derechos, pregunte al investigador responsable del estudio o a los miembros de su equipo. La investigadora responsable del estudio es la Dra. Edurne Lecumberri, del Servicio de Endocrinología. En caso de preguntas o dudas acerca del estudio podrá ponerse en contacto con ella en el teléfono 915504846.

FORMULARIO DE CONSENTIMIENTO INFORMADO POR ESCRITO

TÍTULO DEL ESTUDIO: Estudio de la correlación entre una medida no-invasiva de estados de hiperglucemia sostenida usando espectroscopia de ondas milimétricas y la Hb glicosilada en pacientes con Diabetes Mellitus tipo 1"
Código Protocolo: FJD-ESPEC-DM-17-01

INVESTIGADORA PRINCIPAL: Edurne Lecumberri Pascual. Médico Adjunto Servicio de Endocrinología. Hospital Universitario Fundación Jiménez Díaz.

Yo (Nombre y apellidos)

.....

- He leído la hoja de información que se me ha entregado.
- He podido hacer preguntas sobre el estudio.
- He recibido suficiente información sobre el estudio.
- He hablado con

(Nombre y apellidos del investigador)

Comprendo que mi participación es voluntaria.

Comprendo que puedo retirarme del estudio:

- Cuando quiera.
- Sin tener que dar explicaciones.
- Sin que esto repercuta en mis cuidados médicos.

Presto libremente mi conformidad para participar en el estudio.

En,.....a..... de.....
de.....

Firma del participante	Firma del investigador
Fecha: ___/___/___	Fecha: ___/___/___