University of Strathclyde

Centre for Doctoral Training in Medical Devices and Health Technologies

Department of Biomedical Engineering

The Design and Development of a High-Throughput ATR-FTIR Serum Diagnostics Platform.

Duncan E. Finlayson

A thesis submitted to the Department of Biomedical Engineering, University of Strathclyde, in fulfilment of the requirements for the degree of Doctor of Engineering.

2020

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Signed: D. Finlay Date: 20th April 2021

Table of Contents

Acknowledgements

I was fortunate to be surrounded by great people whose expertise I drew on frequently and without whom this project would not have been possible.

First and foremost, a special thank you to my research supervisor Matthew J. Baker, for his support, guidance and granting this excellent opportunity. I would also like to thank the rest of the ClinSpec Dx team: Mark Hegarty, David Palmer, Benjamin Smith, and not least the lovely Holly J. Butler, for imparting invaluable advice, knowledge, and skills that helped me along the way. It was great to be part of such an exciting project with such wonderful people. Thanks to the rest of my friends and colleagues who helped me along the way: Julian, Alex, Loren, Sam, James, Kelly, Sayali, Lily, and especially my man Chris Rinaldi for being the most zen stress merchant I know.

My greatest appreciation to my lovely Mum, and wonderful Dad. You both fill me with inspiration every day. Mum, your kindness and support know no bounds and has encouraged me to keep going through thick and thin. Dad, your endless curiosity and jovial demeanour has coloured my mind and driven me to pursue my passions. Thank you from the bottom of my heart.

À mon amour Loriane. Merci de m'avoir soutenu et de m'avoir encouragé à avancer. Je peux toujours compter sur toi pour égayer mes journées avec de la bonne humeur, de la joie et des scones.

I dedicate this thesis to those who were lost along the way: granny Meg, auntie Ca, and uncle Billy.

Abstract

Attenuated Total Reflection Fourier Transform Infrared (ATR-FTIR) spectroscopy has demonstrable clinical ability and is well poised to address numerous clinical needs in virtue of operative ease, sample versatility, non-destructive testing, and good diagnostic prediction capabilities. Technical barriers stand in the way of clinical implementation, namely lack of high-throughput capabilities. Samples are individually deposited, prepared, analysed, then removed from a fixed Internal Reflection Element (IRE) amounting to a workflow bottleneck. IRE substrates are typically expensive and unamenable to manufacture and thus resigned to this workflow. We developed a low cost and disposable silicon (Si) IRE (SIRE) and universal infrared spectrometer accessory module enabling high-throughput analysis with high spectral quality and reproducibility. Spectral reproducibility was optimised through characterisation of intrinsic and extrinsic optical variance sources and subsequent mitigation. Si impurity variance had little impact on spectral reproducibility. Optimised batch sample preparation procedure and surface treatment diminished extrinsic variances. SIREs maintain comparable performance to a Diamond IRE in the detection and quantitation of archetypal carbohydrates, lipids, and proteins at clinically relevant concentrations demonstrating effectiveness across the biochemical 'fingerprint' region. A 300% increase in sample throughput over a conventional Diamond IRE was achieved. Sepsis, SIRS, and control patient samples were discriminated using spectra acquired from SIREs to demonstrate clinical suitability. A random forest classifier of sepsis (n=43), SIRS (n=59), and control (n=46) constructed from serum samples analysed on SIREs achieved high negative predictive value (93.55%), positive predictive value (76.08%), sensitivity (84.15%), specificity (90.43%), and accuracy (81.98%) for the detection of sepsis.

Project Outputs

Publications

1. Holly J. Butler, Paul M. Brennan, James M. Cameron, Duncan Finlayson, Mark G. Hegarty, Michael D. Jenkinson, David S. Palmer, Benjamin R. Smith & Matthew J. "**Development of high-throughput ATR-FTIR technology for rapid triage of brain cancer",** Baker. Nature Communications **10**(4501): pg. 1-9 (2019).

   DOI: https://doi.org/10.1038/s41467-019-12527-5

   Contributions: *Developed the platform and sample preparation protocols used for the study. Authored portions of the introduction (paragraph 6, 7, 8, & parts of 10), 'Si IRE evaluation' section of results, 'Si IRE production' and 'sample preparation' sections of the methods, and Figures 2, 3, and 4.*

2. Duncan Finlayson, Christopher Rinaldi, Matthew J. Baker. **"Is Infrared Spectroscopy Ready for the Clinic?",** Analytical Chemistry **91**(19): pg. 12117-12128 (2019)

   DOI: https://doi.org/10.1021/acs.analchem.9b02280

   Contributions: *Authored the introduction, 'Developing the translation', and 'IR spectroscopy of biofluids' sections. Edited the document and applied peer reviewer corrections.*

3. James M. Cameron, Holly J. Butler, David J. Anderson, Loren Christie, Lily Confield, Stuart Murray, Duncan Finlayson, Katie E. Spalding, Zanib Panni, Christopher Rinaldi, Alexandra Sala, Ashton G. Theakstone, Matthew J. Baker. **"Exploring pre-analytical factors for the optimisation of serum diagnostics: progressing the clinical utility of ATR-FTIR spectroscopy",** Vibrational Spectroscopy **109**(103092): pg. 1-13.

   DOI: https://doi.org/10.1016/j.vibspec.2020.103092

   Contributions: *Produced all work for section 3.7 'Sample drying'. Reviewed and edited prior to publication.*

Patents

1. Baker, M.J, Hegarty, M, Butler, H.J, Palmer, D & The University of Strathclyde 2018, *Infra-Red Spectroscopy System*, **WO2018/178669 A2**

2. Baker, M.J, Hegarty, M, Butler, H.J & The University of Strathclyde 2020, *sample container with integrated internal reflection element,* **WO2020/074918 A1**

Poster Presentations

1. Duncan E. Finlayson **"Investigation the Translation of ATR-FTIR to the Clinical Setting"**, CLIRCON17, 2017 Apr 02-05, Manchester.

2. Duncan E. Finlayson **"Design and Optimisation of Multi-Well Silicon Internal Reflection Elements"**, SPEC2018, 2018 June 10-15, Glasgow.

3. Duncan E. Finlayson **"Facilitating the Translation of ATR-FTIR to the Clinical Setting"**, BioMedEng18, 2018 Sep 06-07, London.

4. Duncan E. Finlayson **"Development of a High-Throughput ATR-FTIR System for Serum Diagnostics"**, BioMedEng19, 2019 Sep 05-06, London.

List of Abbreviations

A	Absorbance
ABS	Acrylonitrile Butadiene Styrene
AI	Artificial Intelligence
AI:AII	Amide I to Amide II ratio
ANOVA	Analysis of Variance
AOI	Angle of Incidence
ATR	Attenuated Total Reflectance
ATR-FTIR	Attenuated Total Reflectance Fourier Transform Infrared
AUC	Area Under the Curve
BaF_2	Barium Fluoride
BCG	Bromocresol Green
BCP	Bromocresol Purple
CAD	Computer Aided Design
CaF_2	Calcium Fluoride
cfDNA	Cell-free DNA
CFM	Cubic Feet per Minute
CO_2	Carbon Dioxide
COPD	Chronic Obstructive Pulmonary Disease
CRP	C-reactive Protein
CTHC	Constant Temperature Humidity Chambers
Cz	Czochralski
DF	Degrees of Freedom
Di	Diamond
DMLS	Direct Metal Laser Sintering
DNA	Deoxyribonucleic Acid
d_p	Penetration Depth
DSP	Double-Side Polished
Dstl	Defence Science & Technology Laboratory
DTGS	Deuterated Triglyceride Sulfate
EMSC	Extended Multiplicative Signal Correction
FDA	Food and Drug Administration
FFF	Fused Filament Fabrication
FTIR	Fourier Transform Infrared
FZ	Float-Zone
GCS	Glasgow Coma Score
Ge	Germanium
GUI	General User Interface
H_2O	Water
HCl	Hydrogen Chloride
HMDS	Hexamethyldisilazane
HPS	Human Pooled Serum
IL-10	Interleukin-10
IL-6	Interleukin-6
IL-8	Interleukin-8
IR	Infrared
IRE	Internal Reflection Element

IVD	In Vitro Diagnostic
IVD	In Vitro Diagnostic
IVDR	In Vitro Diagnostic Regulation
KBr	Potassium Bromide
KOH	Potassium Hydroxide
LBP	Lipopolysaccharide binding protein
LOD	Limit of Detection
LOQ	Limit of Quantification
LPCVD	Low Pressure Chemical Vapour Deposition
M	Mean
MCT	Mercury-Cadmium-Telluride
MDR	Medical Device Regulation
MHRA	Medicines & Healthcare products Regulatory Agency
Na	Sodium
NaCl	Sodium Chloride
NICE	National Institute for Health and Care Excellence
nm	Nanometre
NPV	Negative Predictive Value
O_2	Oxygen
OPD	Optical Path Difference
paCO2	partial pressure of Carbon Dioxide
PBS	Phosphate Buffered Saline
PC	Principal Component
PCA	Principal Component Analysis
PCT	Procalcitonin
PLA	Polylactic Acid
PLS	Partial Least Squares
PLS-DA	Partial Least Squares Discriminant Analysis
PLSR	Partial Least Squares Regression
POC	Poin-of-Care
PPV	Positive Predictive Value
PWM	Pulse Width Modulation
QCL	Quantum Cascade Laser
RF	Random Forest
RMSE	Root Mean Square Error
RMSEV	Root Mean Square Error of Validation
RSD%	Percentage Relative Standard Deviation
SD	Standard Deviation
SEM	Scanning Electron Microscope
Si	Silicon
Si1	Silicon Phonon 1
Si2	Silicon Phonon 2
SiC	Silicon Carbide
SIRS	Systemic Inflammatory Response Syndrome
SIRS	Systemic Inflammatory Response Syndrome
SMC	Scottish Microelectronics Centre
SMOTe	Synthetic Minority Oversampling Technique
SNR	Signal to Noise Ratio

SOFA	Sequential Organ Failure Assessment
SSP	Single Side Polished
sTREM-1	Soluble triggering receptor expressed on myeloid cells-1
T	Transmittance
TIR	Total Internal Reflection
TNF	Tumour Necrosis Factor
TNF-alpha	Tumour Necrosis Factor-alpha
UV	Ultraviolet
WCC	White Cell Count
ZnSe	Zinc Selenide
λ	Wavelength
λ^{-1}	Wavenumber

List of Figures

List of Tables

Chapter 1: Introduction

1.1. Introduction

Early and correct diagnosis leads to early intervention which leads to better patient outcomes [1]–[7]. Better patient outcomes lead to reduced economic burden on healthcare institutions [8]–[11]. These patently observable facts are the force driving the continuous advancement of medical diagnostic devices. In theory, patients who are diagnosed quicker are given the appropriate course of treatment earlier which increases the likelihood of survival, lowers treatment morbidity, and shortens time to recovery because the disease is detected at an earlier stage in its development. The longer a patient is sick, the greater the demand on resources and the higher the costs. This problem is exacerbated in vulnerable populations such as in low income areas and the elderly. An aging and expanding population, with the number of people aged over 65 expected to rise 71% by 2050 in developed countries [12], and an increasing prevalence of chronic disease is foreboding of the healthcare crisis that is to come. Access to rapid, low-cost, and accurate diagnostic platforms play a critical role in improving patient outcomes and alleviating economic strain on healthcare institutions. Another less obvious impact of improved accessibility to reliable diagnostic tests is the generation of reliable disease statistics which in turn guide policy making decisions such as awareness campaigns, resource allocations, and research priorities [13]. Improving diagnostic capabilities is therefore an important public health strategy.

Diagnosis is a process whereby a clinician must determine what is the causative conditions behind an array of signs and symptoms [14]. A clinical decision marks the beginning of a healthcare pathway aimed specifically at restoring the patient to health.

Diagnostic tests are one of the most powerful tools at the clinician's disposal to help guide this decision process and to monitor the effectiveness of therapy. There are numerous ways to improve diagnostic test capabilities. Accessibility to diagnostic tests, time to diagnosis, and diagnostic performance are critical. Generally, accessibility means reducing cost, but simple operation, clear results interpretation, ease of manufacture, and high sample throughput also help to improve accessibility to diagnostic tests [15]. There are numerous moving parts to improve time to diagnosis from the moment a test is ordered to the moment a clinical decision is made [14]. Diagnostic tests must therefore consider their place in clinical workflow and logistics. Often this necessitates the capacity to process high volumes of patient samples in short time frames while having minimal lab footprint. Conforming to existing clinical workflows is always preferable, for example by incorporating standard microscope slides into the design of the technology if necessary.

It is estimated that 70% of clinical decisions are influenced by laboratory tests [16], [17] and that everyone will experience at least one diagnostic error some of which lead to disastrous consequences [18], [19]. Failure to perform a differential diagnosis, such as distinguishing sepsis from Systemic Inflammatory Response Syndrome (SIRS) or post-surgical inflammation, is one of the root causes for poor outcomes and diagnostic tests are the best resource at the clinicians disposal to help through such situations [20]. From these findings, one could conclude that diagnostic performance has great consequences, and diagnostic tests should therefore be held to a lofty benchmark with exemplary specificity, sensitivity, and predictive values. Sepsis, for instance, is notoriously heterogenous with no known molecular signature to identify the condition [21]. Furthermore, the current diagnostic gold standard, blood cultures,

is time-consuming and frequently return false negative results particularly for neonatal patients [22]. Consequently, clinicians struggle to rule out sepsis when forming a differential diagnosis. This is critical as healthcare pathways must be initiated quickly (within 1 hour) and with confidence to avoid detracting from the real pathology, or exacerbating the global antibiotic resistance crisis through inappropriate administration of empirical antibiotics [23]. A rapid sepsis triage tool with good predictive capabilities would therefore be an invaluable resource to clinicians [22], [24]–[26].

1.2. Diagnostic Test Selection

Consideration should be given to whether *in vivo* or *in vitro* diagnostic testing is more desirable for a given application. *In vivo* has the advantage of capturing the physiological or molecular information of the patient in real time while *in vitro* diagnostics (IVDs) are easier to perform under controlled conditions and are repeatable [27]. Generally speaking, *in vitro* tests are less invasive than *in vivo* as they are conducted outside of the body although sampling method can vary from invasive to non-invasive (e.g., tissue biopsies versus liquid biopsies), and likewise for *in vivo* tests (e.g., Magnetic Resonance Imaging (MRI) or wearable sensors versus implantable sensors) [28]–[32]. Non-invasive *in vivo* devices generally monitor physiological signals (e.g., temperature, blood oxygenation, or electrical activity) or monitor secretions, such as sweat or saliva, that may lack pertinent biochemical signatures of disease present in other sample types (e.g., blood or serum) [29], [33]. Implantable *in*

vivo devices are useful for continuous monitoring of chronic conditions (e.g., blood glucose monitoring in diabetics), but need to be robust to resist degradation from foreign body responses (e.g., fibrous encapsulation) that may impair sensor functionality, and biocompatible to prevent adverse host reactions [34], [35]. Consequently, *in vivo* device development may require animal testing prior to clinical trials incurring ethical issues and stricter regulations [36]. For these reasons, it is generally less invasive, cheaper, and easier to standardise *in vitro* testing environments and parameters which in turn can enable high volume testing and widespread dissemination of tests at low cost [17], [27], [37]. This is useful when addressing prevalent conditions that require accessible and on demand testing such as sepsis [26], [38].

A diverse range of biosample types are available for *in vitro* monitoring which can be broadly categorised as tissues [39], cells [40], or biofluids [41], [42]. Tissues and cells can provide a more direct link to the site of pathological interest but may have limited availability, particularly where critical organs (e.g., brain tissue) are concerned [43], [44]. In oncological examinations, for instance, histology is considered the gold standard for determining the underlying pathology and directing personalised medicine [31], [45]. However, it is more likely that triage and periodic monitoring would comprise imaging and/or biofluid tests, such as the Prostate Specific Antigen (PSA) test for screening prostate cancer, as these tests are less invasive, quicker, and lack the subjectivity of tissue biopsies [46]. Liquid biopsies may even be on course to replace conventional tissue biopsies as a minimally invasive means of extracting neoplasm material from the blood and other biofluids [31], [43], [47]. For some systemic conditions biofluids, particularly blood, are an invaluable resource as

their molecular signatures reflect systemic physiological changes such as the change in pro- and anti-inflammatory cytokine/chemokine expression observed during sepsis progression [48], [49]. Acquisition of biofluid samples is generally less invasive, easier to obtain, storable, available in greater quantities, available for repeated testing, and the preparation before analysis less onerous in comparison to tissue samples [44], [50]. Biofluid samples are therefore the more compatible choice for low-cost and timely high-throughput testing [51].

An array of methods, such as immunochemistry [52], biosensing [38], [53], cultures [54], nucleic acid amplification tests (NAATs) [55], and spectroscopic techniques [56]–[59], are have been developed to perform *in vitro* biofluid diagnostic tests. In many cases, cultures represent the gold standard for determining infectious pathogens but, while highly specific in nature, are often low in clinical sensitivity (returning a false negative result in ~40% of patients who have a clinical diagnosis of sepsis), require large sample volumes, are labour intensive, and have notoriously lengthy lead times [60]–[63].

NAATs (e.g., polymerase chain reaction (PCR)) are also of interest for pathogen determination and are faster and often more clinically sensitive in comparison to culture tests, but have the potential for false positives due to the amplification of contaminating DNA or RNA, have difficulty obtaining quantitative results, and have complex sample preparation [64]–[66].

Immunoassays are inherently analytically specific since they utilise specific antibody interactions, but are notoriously expensive due to the use of antibody reagents, limit scope of molecular analysis to antibody-antigen affinity based

reactions, and in most cases restrict applicability to a single biomarker species [67], [68]. Multiplexed immunoassays do exist that enable detection of numerous biomarker targets, but are labour intensive and have proved difficult to implement as concerns have been raised over the reliability and reproducibility of these assays [69], [70].

Biosensors comprise a bioreceptor and a physicochemical transducer typically functioning through optical, thermal, piezoelectric, or electrochemical mechanisms [71]. Biosensing platforms are therefore highly customisable enabling ligand specific high-throughput screening and device miniaturisation, although impracticalities of sample preparation have largely remained a roadblock to clinical translation in infectious disease applications [72]–[75].

Mass spectrometry techniques enable high analytical specificity and sensitivity down to the femtomolar range, but compared to other techniques (e.g. vibrational spectroscopy) are slow due to rate limiting sample preparation steps, sample destructive, and are generally expensive in instrumentation and operation although progress has been made toward miniaturised low-cost systems [76]–[78].

Vibrational spectroscopic techniques, broadly categorised as Raman or Infrared (IR) spectroscopy, offer rapid, easy to operate, non-invasive, non-destructive, and reproducible testing requiring no reagents, low sample volumes, and minimal sample preparation [79]–[81]. High dimensional data of the patient sample is acquired comprising spectral signatures unique to biochemistry under interrogation enabling high chemical specificity [82]. Raman spectroscopy utilises expensive lasers, is susceptible to intense fluorescence background noise and suffers a fundamental limitation of weak signal produced from Raman scattering requiring long acquisition

times to compensate [83], [84]. IR spectrometers do not suffer this limitation and are lower in cost because lasers are not required, but perform poorly when interrogating aqueous environments often necessitating the drying of biological specimens prior to analysis [85]. Both techniques interrogate samples using a single point of analysis, however the rate limiting factor of Raman is inherent to the principles of its operation. Thus, IR spectroscopy is a good candidate for rapid, low-cost testing provided it is possible to establish reproducible sample preparation protocols to enable high-throughput sample processing [86], [87].

The pathophysiology under examination needs to be well defined for biomarker specific approaches to be clinically effective as the chemical shifts observed must correlate well with the presence and progression of disease. By contrast, vibrational spectroscopic analysis of biofluids offer a method for observing a global change in biomolecular content reflected by the body's systemic response to disease [88]. In the context of sepsis, this approach likely holds a significant advantage over biomarker specific approaches which have thus far not proven effective in reliably detecting this condition [89]. Early sepsis management is critical to improving patient outcomes, so a test would must return a result in less than 3 hours and ideally less than 1 hour [26], [90]. Ideally, the test would be inexpensive and high throughput as clinicians may wish to order high volumes of precautionary tests in order to rule out sepsis from a differential diagnosis. Furthermore, the test should be minimally invasive and require only small amounts of sample to enable neonatal testing where sepsis is the cause of significant mortality and morbidity [91], [92]. Preferably, the test should minimise the prevalence of false negatives, exhibiting high sensitivity and NPV (each >98%) for the detection of sepsis, in order to avoid unintentional cessation of empirical

antibiotics which may jeopardize patient wellbeing [26]. Good specificity and PPV (>85%) is also desirable to limit false positives as inappropriate administration of antibiotics can compound the global antimicrobial resistance crisis and has unknown impact on the well-being of neonates [22], [26]. For the reasons presented in this section, it is believed that an IR spectroscopic biofluid IVD with high-throughput capabilities would be well poised to meet these requirements by virtue of minimal sample preparation, low cost, reproducibility, rapidity, and ability to observe systemic biochemical changes.

1.3. IR Spectroscopic Biofluid Diagnostics

Biofluid diagnostics are a high value diagnostic resource because biofluid samples, such as blood, saliva, and urine, are easy to collect, minimally invasive, contain a wealth of biochemical information, and can be collected inexpensively. Biofluids are perhaps the most effective means to carry out objective, cost-effective, and rapid diagnosis and triaging of patients [93]. Novel biofluid diagnostic techniques are therefore attractive solutions for improving diagnostic capabilities. However, standardisation of the pre-analytical factors of biofluid spectroscopic lab tests need to be established as inconsistencies in the pre-analytical phase is a significant source of error [94]. A selection of pre-analytical considerations are the methods of collection (e.g., sample volume, draw site, patient fasting requirements, transportation), storage (e.g., temperature, number of freeze-thaw cycles), processing (e.g., addition of anticoagulants, dilution, centrifugation time, droplet deposition), and measurement

parameters (e.g., operators, spectrometer configuration/manufacturer, substrates) [95]. Most of these factors are dependent on the biofluid type under investigation (e.g. cerebrospinal fluid, saliva, serum, etc.) [51]. Hence careful consideration should be given to sample type selection.

Biomolecules, or any other molecule for that matter, interact with light in ways unique to the biomolecule under investigation. Molecular bonds continuously vibrate at the zero-point energy and most biomolecules will absorb specific frequencies of infrared (IR) light unique to their composition causing their bonds to vibrate at a greater energy level. The study of these interactions between IR light and matter is called IR spectroscopy. Barring symmetric diatomic molecules (e.g., oxygen), all biomolecules absorb mid-IR light at specific frequencies, hence it can be inferred that almost all the biomolecules in a biofluid sample will interact with a broadband beam of incident IR light. Thus, an analysis can be carried out that interrogates the entire biochemical profile of a patient sample. The resulting dataset is a spectrum accounting for the extent to which each frequency of mid-IR light was attenuated by the sample. It is a complex dataset as it is representative of a complex biochemistry and can be difficult to interpret through direct inspection of the spectrum. However, this is the perfect scenario to employ multivariate or machine learning techniques that can automatically recognise patterns indicative of sample class. The underlying principles of IR spectroscopy and machine learning are detailed in the next chapter.

Many widely used biofluidic tests and devices focus on the analysis of a single target biomarker, although rarely do diagnostic procedures rely on just one test. Instead, a combination of screening and diagnostic tests often influence clinical decisions [96]. Univariate analysis of a biomarker can be appropriate when the

pathological state is well characterised, for example the monitoring of blood glucose levels in diabetics, but can be inadequate when investigating ambiguous differential diagnoses, syndromic conditions, or comorbidities. In these cases, an array of biomarkers, such as lipid panels as predictors of coronary heart disease [97], in combination with patient history and medical imaging may paint a more detailed pictured, but numerous results can be difficult for an individual to digest. Systems that use scores or classification criteria have been developed that combine screening and diagnostic test observations to identify and categorise pathological states such as the Glasgow Coma Score (GCS), SIRS criteria and the Sequential Organ Failure Assessment (SOFA) score [98], [99]. These are effective tools, but incur an element of assessor subjectivity and may not account for patient heterogeneity [100], [101]. Machine learning is a rapidly growing field well poised to address these limitations of current diagnostic modalities. Algorithms are trained against highly dimensional datasets to recognise patterns in the data and classify accordingly [102], [103].

Machine learned algorithms are capable of learning and adapting as they are exposed to more data. The utility of this in the context of diagnostic devices is clear; the human biological system is enormously complex and varies considerably within and between patient groups due to intrinsic and extrinsic factors such as age, environment, and lifestyle. Machine learning can examine many biological components simultaneously whilst also adapting to resist heterogeneity within sample data [104]. Some of the difficulties in implementing machine learning approaches to clinical diagnostics lie in collecting the large amounts of data required to construct predictive models, validation of models according to clinical gold standards, and the objective comparison of algorithms across studies [88], [105], [106]. IR spectroscopic

techniques are useful in this regard as they offer a way to quickly generate high dimensionality data by probing the entire biochemical profile of a sample simultaneously. Machine learning is a subset of Artificial Intelligence (A.I.), a discipline gaining traction within the broader healthcare community with applications in imaging, genetics, electrodiagnostics, monitoring, disability and/or mental health evaluation, mass screening, and more [107]. Scotland is emerging as a leader with innovations like Current Health's remote patient monitor [108], MyWay Digital Health's patient management platform [109], and MindMate's Alzheimer support app [110]. A.I., particularly machine learning, is applicable to a broad range of healthcare applications that many feel will transform the healthcare landscape [111].

IR spectroscopy is an analytical technique used extensively in the food [112], [113], pharmaceutical [114], [115], petrochemical [116], environmental [117], [118], and forensics industries [119], [120]. While there has not been widespread implementation of IR spectroscopy in the clinical laboratory, there exists a wealth of clinical research [121]–[129], and IR spectroscopy is an excellent technique for the analysis of biological materials [56]. It is easy to operate, low cost, non-destructive, label-free, and requires low sample volumes. A biochemical fingerprint of the sample is collected theoretically providing a snapshot of the subject's health condition. One obstacle in IR spectroscopy is the difficulty in analysing samples with high water content. Biological samples almost always have a high water content and must be dried prior to analysis [86]. Another limitation is the potential for spectral distortions, such as resonant Mie scattering or other dispersion artefacts, that occur when structures of similar magnitudes to the wavelength of IR light, such as cells, are present in the sample [130]. This can scatter IR light and lower signal quality. These particular issues

can be avoided through selection of appropriate sample type and IR mode [56]. Whole blood, for instance, is full of erythrocytes that can interfere with the spectral contribution from other biomolecular species in the sample [128]. Whole blood or plasma samples also require anticoagulants to prevent clotting during storage and it has been shown that selection of different anticoagulants is represented in the IR spectra [95]. This means that different anticoagulation procedures across healthcare centres may impact the IR analysis. Serum is an ideal sample type to avoid the extra pre-processing or anticoagulants required for the analysis of whole blood or plasma as clotting factors (e.g., fibrinogen) and cells (e.g., erythrocytes) are removed through centrifugation after a whole blood sample has been allowed to clot [131]. However, whilst serum collection can be reproducible, it comes with the caveat of an extra coagulation step increasing processing time [132], [133]. Furthermore, precautions still need to be taken to minimise pre-analytical variance. Blood collection method should standardised and avoid venous stasis, haemolysis, and draw site error (e.g., away from drip arm or site of infection) [134], [135]. Serum should be analysed fresh if possible as FTIR spectroscopy has been found to be sensitive enough to distinguish fresh from a single freeze-thaw cycle [95]. Despite lacking cellular components, serum retains a wealth of biochemical information, is easily derived from whole blood, accessible, can be obtained reproducibly, and is a common sample type in routine clinical laboratory diagnostics [136]. Serum is therefore often favoured for infrared spectroscopic analysis [85], [136].

Factors limiting the translation of IR spectroscopy to the clinical laboratory are variability induced during preanalytical stages and through the requirement to dry samples [137], confinement to a single point of analysis [86], and perhaps reluctance

form spectrometer manufacturers to dedicate instrument design for clinical applications [88]. Use of serum over whole blood may minimise the risk of resonant scattering, but samples still need to be dried. Drying can often result in gelation and cracking patterns, and a phenomenon known as the 'coffee ring' effect whereby dried films are chemically and physically heterogeneously distributed across the sample film [137]. Consequently, composition can vary greatly across the film presenting a risk of undermining reproducibility and sensitivity of analyses. Efforts to minimise the coffee ring effect include the analysis of minute volumes (<1µl), dilution of the sample, or spraying sample onto the sample site surface [138]–[140]. Measurement of the bulk sample using a particularly sensitive mode or IR spectroscopy called Attenuated Total Reflectance Fourier Transform Infrared (ATR-FTIR) spectroscopy has been shown to minimise the impact of biofluid drying patterns without need for dilution effectively providing an average spectrum of the dried droplet [88], [106], [141]. Furthermore, ATR-FTIR is highly favoured in the research community for the analysis of serum as samples can be placed directly onto the ATR sampling site for quick measurement without sophisticated preparation protocols eliminating spectral contributions from a slide substrate while maintaining high reproducibility [56], [106], [121], [128], [142]–[145]. ATR is therefore the FTIR spectroscopy mode best suited for reproducible and sensitive clinical testing [87]. However, in any FTIR mode multiple sample scans are necessary to reduce artifacts and variations in the spectra due to environmental conditions and sample heterogeneity [84]. ATR-FTIR is restricted by a fixed point of analysis as the optical substrates required to perform ATR are expensive, fixed, high refractive index crystals (e.g. Diamond) that are not compatible with the design of a high-throughput system (e.g. microtiter plates) [86]. Samples must be laboriously

dried individually and removed before subsequent sample analysis culminating in a workflow bottleneck and destroying the sample droplets in the process [138]. Hence, an alternative, low cost, and readily interchangeable ATR crystal is required to enable batch processing and, by extension, rapid analysis capabilities.

1.4. Translation of Infrared Spectroscopy to the Clinic

IR spectroscopic techniques have demonstrable utility in biofluids [93], [137], cytological [146], [147], histological [148], and microbiological [149] studies. Research articles have promised excellent disease detection capabilities over a wide variety of diseases and across numerous biofluid sample types such as whole blood [150], serum [151], plasma [152], saliva [153], sputum [154], amniotic fluid [155], cerebrospinal fluid [156], urine [157], bile [123], and even tears [158]. The evidence that IR spectroscopy is a powerful diagnostic tool with respect to predictive and investigative capabilities is clear, so the obvious question is: why is this technique not used in the clinical laboratory? The answer is that demonstrating potential as a powerful predictive tool does not necessarily demonstrate clinical suitability. New diagnostic techniques arise from a wide variety of origins, but broadly speaking they are propelled either by ‘technical push’ where existing technologies are adapted to enter the clinical market or ‘clinical pull’ where clinical problems call for novel technical solutions [159]. Academic or commercial sector researchers and entrepreneurs seeking to translate their work into the clinical setting generally fall within the former category. Consequently, their technologies may not be suitable for

the clinical environment for a variety of reasons such as usability, lab footprint, incompatibility with automation, cost, complexity, and length of procedure. All these factors could disrupt clinical workflows, a ramification which cannot be underestimated. Continued end user input and support from clinicians, laboratory technicians, or any other clinical operators, is indispensable for resolving issues during the development stages and preparing the most market ready product before initiating clinical trials. Infrared spectroscopy is one such 'technical push' field that has remained on the periphery of the healthcare industry to date. There are, however, several examples of start-up companies closing in on this goal and IR spectroscopic techniques are widely used in the pharmaceuticals industry [114].

The projects presented here were carried out in partnership with ClinSpec Diagnostics, a Strathclyde University spin-out company developing spectroscopic brain cancer diagnostics using artificial intelligence and ATR-FTIR. Other notable IR spectroscopy diagnostic companies are: Biotech Resources, Cireca Theranostics, and Glyconics. Biotech Resources developed a spectroscopic system to deliver malaria screening to low-income areas earning international acclaim [129]. Cireca Theranostics developed 'spectral histopathology' protocols for objective, reagent free, machine learning based analyses of tissue sections for use in situations where histopathologists are unable to establish a consensus on how to classify a sample [160], [161]. Glyconics, a Cambridge based IR spectroscopy diagnostics company, partnered with Spectrolytic, a German IR spectrometer supplier and manufacturer to develop a handheld Point-of-Care (POC) infrared spectroscopy device specifically for use in the clinic. This partnership presents a good case study for 'technical push' clinical translation. Spectrometer manufacturers are generally unwilling to adapt their products

to the clinical environment because it is not a market they are established in and therefore view expansion into this market as a significant risk. Confirming existence of a consumer base through examination of an unmet clinical need, proof-of-concept studies, and financial projections through health economic analysis is a good way to reassure manufacturers that tailoring their instruments toward clinical applications will help the growth of their company. Glyconics are in the early stages of a clinical trial to validate claims that the hand-held device can be used to differentiate between chronic obstructive pulmonary disease (COPD) from other disease states and monitor the conditions progression [162], [163].

Above all else, translation of technology into the clinical marketplace requires there to be a clearly defined clinical need. There is no shortage of unmet clinical needs, but appropriately fitting a solution is difficult. Most entrepreneurial ventures of POC diagnostics fail in the development stage because of a fundamental disconnect between the technical specifications of the core technology and the clinical needs [164]. Some of the common technical oversights, such as complexity or time to result, have already been mentioned, but from the clinical view it is equally egregious not to consider the specific healthcare pathways the diagnostic fits into. Health economic analysis is a wise strategy to determine if a technology is suited to a clinical application. It needs to be determined if a medical device, therapy, or diagnostic technique can improve patient outcomes whilst being economically feasible.

Health economic analysis predicts the cost and effect on public health of a specific course of action and compares it to all other alternative courses of action. It is an indispensable tool in clarifying that a diagnostic device, therapy, or any healthcare pathway is both effective in its purpose and feasible to implement [165], [166].

Convincing important stakeholders, such as healthcare providers, patient groups, clinicians, and investors, is made substantially easier through evidence provided by a health economic analysis. Such an analysis will also help define the end market strategy for the product.

Manufacturers must obtain regulatory clearance before an IVD can be delivered to the clinical market. To operate in the European Union (EU) manufacturers must comply with the In Vitro Device Regulation (IVDR) (formerly IVD Directive) and in the United States manufacturers must comply with the Food and Drugs Administration (FDA) title 21 of the Code of Federal Regulations (21 CFR) [167], [168]. Both the IVDR and FDA have classification systems for IVDs which dictates the degree of regulatory control necessary and the regulatory pathways that must be followed. Although the IVDR and FDA have a different system of IVD classification, the principle is the same in the sense that IVDs are classified according to intended use and associated risk. In order of increasing risk, the IVDR classifies IVDs as class A, B, C, or D, and the FDA classifies IVDs as class I, II, or III. All IVDR compliant manufacturers must establish a ISO13485 certified Quality Management System (QMS), perform post market surveillance, and prove compliance to a notified body prior to CE marking a product [169]. All 21 CFR compliant manufacturers must establish ‘general controls’ in alignment with 21 CFR part 807 (e.g., design controls and suitable labelling of products) and class II or III devices must establish ‘special controls’ (e.g., 21 CFR part 820 compliant QMS and post market surveillance) [37].

A critical requirement for IVD regulatory approval is the validation of IVD safety and efficacy through clinical trials [170]. Clinical trials assess the performance and reproducibility of the IVD on appropriate study populations and under the

conditions of intended use [171]. A trial protocol needs to be established and peer reviewed during set-up of a clinical trial detailing the aims, scientific rational, procedures and methodology, patient eligibility criteria and recruitment plan, and data analysis strategy of the trial [172]. In the UK, trial protocols must receive ethical approval and authorisation from the Medicines and Healthcare Products Regulatory Agency (MHRA) ensuring the trial has been planned in alignment with Good Clinical Practice (GCP) standards [173]. An ethics committee will determine if the likely benefits of the test will outweigh any possible risks and if steps have been taken to mitigate risks for the patient including: the provision of informed consent, confidentiality, and appropriate consideration for vulnerable populations such as the elderly and disabled [174].

1.4.1. Previous and Parallel Work

Two significant bodies of work predated the projects presented within this thesis. The first was a series of studies demonstrating the ability to discriminate between brain cancer and non-cancer patient through spectroscopic analysis with high sensitivity and specificity using only blood serum samples and an ATR-FTIR spectrometer [121], [122]. ATR-FTIR is a particularly sensitive technique that holds numerous advantages over other IR spectroscopy modes (see chapter 2 section 2.2.1.3). The first study achieved a sensitivity and specificity of 93.75% and 96.53% for cancer vs non-cancer classification respectively from 97 patient samples [121]. The second study expanded the patient dataset to 433 patients and achieved sensitivity and specificity of 91.5% and 83.0% respectively and also classified patient samples according to a greater number of brain cancer subgroups [122].

The second significant piece of work was a health economic analysis informed by an expert clinical focus group that demonstrated that an ATR-FTIR spectroscopy serum based brain cancer test could improve patient outcomes in terms of Quality Adjusted Life Years (QALYs) in an economically beneficial way by reducing diagnostic delays in the current clinical pathway [175]. The findings of this health economic evaluation were that such a test has a strong probability of being cost-effective if the price of the test is less than £40 pounds. The development of the brain cancer clinical test ran parallel to the projects described herein and is still under development for eventual commercialisation by ClinSpec Dx. The hardware developed throughout the projects described within this thesis is the core technology that enabled a prospective clinical validation study for the rapid triage of brain cancer patients [176]. Brain cancer diagnostics are not the focal point of this thesis but will be referred to. This thesis will instead present an ATR-FTIR serum based sepsis test to demonstrate the viability of this technology as a general clinical diagnostics platform. The technical requirements of the core ATR-FTIR technology developed and clinical need for a sepsis serum test are detailed below.

1.4.2. An ATR-FTIR Serum Diagnostics Platform

ATR-FTIR spectroscopy is a specific form of infrared spectroscopy able to achieve high quality spectral data. Operation is simple: samples are placed on top of a crystal of high refractive index, typically diamond, called an Internal Reflection Element (IRE), and is interrogated by an IR beam directed through the IRE. The IRE is fixed within a top-plate so that the IR beam can be delivered at a specific angle such that Total Internal Reflection (TIR) occurs in the IRE. The reason for this is because

TIR produces a phenomenon known as an evanescent wave which protrudes beyond the IRE surface and into the sample (Figure 1.1) [112]. This is the only portion of the IR beam to interact with the sample during ATR-FTIR spectroscopy. Detailed descriptions on how IR spectroscopy operates and how ATR compares to other FTIR spectroscopy modes is discussed in the section 2.2 of the next chapter.

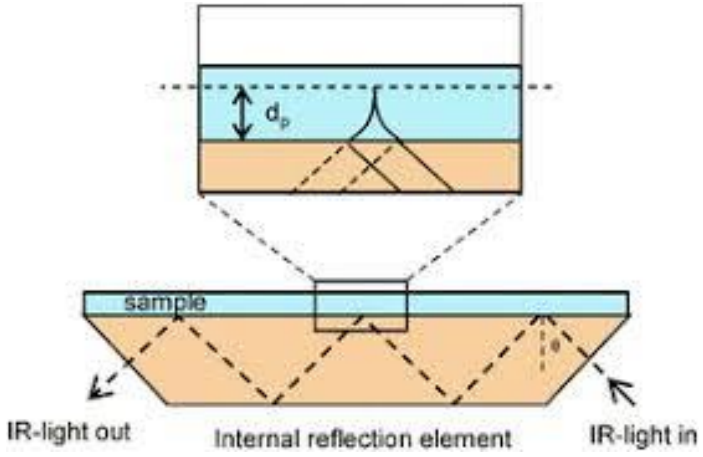


Figure 1.1. *Schematic of the IR beam trajectory during ATR-FTIR spectroscopy. Image credit:* [177]

This fixed IRE configuration presents the inherent design flaw of traditional ATR-FTIR spectrometers. The user is limited to analysing one sample at a time and is forced to clear away a sample, destroying it, before preparing the next sample. FTIR spectroscopy is otherwise a non-destructive technique. A significant benefit of ATR-FTIR spectroscopy is that it is reagent free and does not require extensive sample preparation usually only requiring samples to be dried upon the IRE [56]. Despite this, a fixed IRE creates an inherent process bottleneck that converts even simple sample preparation into a laborious and gruelling task. It is common practice to require biological sample triplicates for reliable analysis, which compounds this bottleneck issue. Typically for biofluid ATR-FTIR analysis sample preparation would take this form: sample collection, separation (e.g., centrifugation for serum analysis), sample

deposition onto the IRE (*ca.* ~3µl per well generally), sample dehydration, and finally the sample scan. Crucially, the sample dehydration step can take anywhere from 15-120 minutes. It should be noted that each of these steps may incur unwanted variance and should therefore be standardised as far as possible [137], [138]. This is a crucial factor in the development of a reproducible and reliable diagnostic tool.

The solution presented here looks to remove the fixed IRE and replace it with an interchangeable and disposable IRE slide capable of carrying multiple samples at once. This bears with it a number of technical difficulties. Firstly, IREs are typically made from expensive materials, such as diamond, or germanium, which are not amenable to scaled production. Scalable and cheap production being an important characteristic of a clinical ATR-FTIR platform in view of the health economic analysis discussed in section 1.3.1. Silicon is a promising alternative IRE substrate as cheap production of silicon-based devices is well demonstrated and its manufacturability is well understood. Silicon is used extensively in the semiconductor and microelectronics industry. High grade silicon is therefore a ubiquitous low-cost substance and silicon foundries are present everywhere. However, the optical properties of conventional silicon IREs do not favour IR biospectroscopic applications [178]. Silicon unfortunately absorbs light in biologically relevant portions of the IR light spectrum (wavenumbers below 1500cm^{-1}) meaning pertinent sample information may be lost. This also somewhat undermines the purpose of a machine learning spectral approach if the technique is not able to interrogate all components of the sample simultaneously. Although this problem cannot be eliminated it can be mitigated. Limiting the length of silicon material that the beam must travel through, also simply called the beam pathlength, reduces the amount of IR light lost to the silicon lattice. The geometry of

the IRE can be designed such that the beam only undergoes a single internal reflection as opposed to multiple internal reflections (Figure 1.2) [179]. A single bounce IR system has the drawback of reduced sensitivity because the IR beam has less opportunities to interact with the sample. Whether or not this is an acceptable compromise or if this limitation can be circumvented is one of the core aspects under investigation throughout this thesis.

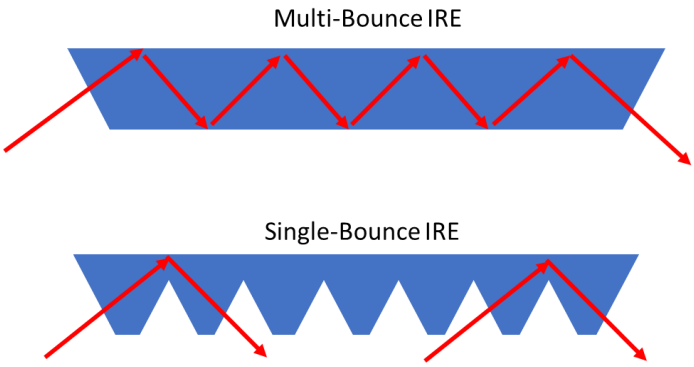


Figure 1.2. *Comparison of multi-bounce and single bounce IRE beam trajectories.*

The interchangeable IRE slide designed in this project will host multiple optically active 'single bounce' grooved areas across the length of the slide and has been dubbed the Single Internal Reflection Element (SIRE). The envisioned clinical workflow is simple (Figure 1.3). A clinician is deliberating on a differential diagnosis between sepsis and SIRS or would like to rule out sepsis and requests a SIRE sepsis test. Routine blood sample collection techniques are used to acquire patient samples [180]. Collection should be done using Rapid Serum Tubes (RTS) to minimise clotting time (~5min) and ideally a transilluminator instead of a tourniquet to eliminate venous stasis associated errors [181]–[183]. Separation should be done immediately with centrifugation parameters sufficient to avoid sedimentation and haemolysis (e.g. 1100-1500g for 10 minutes) [134]. To save time, a background scan of the SIRE may be

acquired while the patient sample is being centrifuged. A small volume, usually 3µl, of serum is then deposited onto each sampling area of the SIRE. The SIREs are then moved to a controlled heating environment and allowed to dry for a set amount of time. The samples are then ready for analysis, after which the sample slides may be stored for analysis at a later date if required. Results are returned to the user in an easily interpretable manner. The SIRE slides could easily be equipped with traceability mechanisms, such as a barcode, to help ensure patient data is not lost. How the SIRE was designed, developed, fabricated, and tested is one of the central discussions of this thesis. Since conventional IR spectrometers are not built to interface SIREs, a custom-built accessory unit with a SIRE docking port was also developed. This too will be presented.

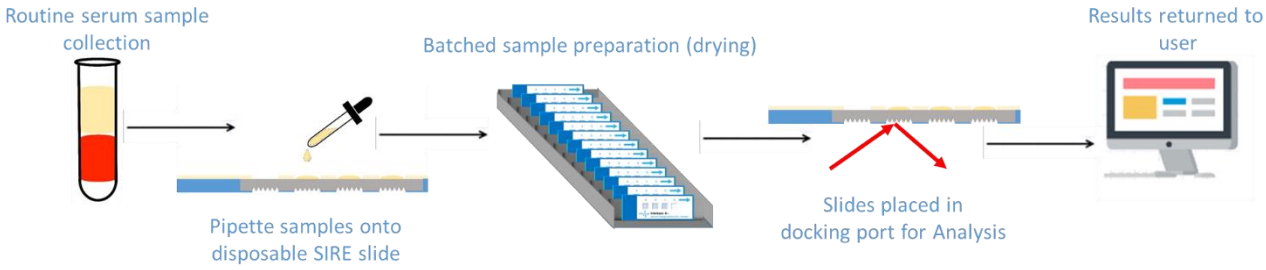


Figure 1.3. *Clinical process flowchart of a SIRE serum test from test ordered to results returned.*

1.5. Sepsis: An Unmet Clinical Need

Sepsis is a syndromic condition characterised by the extreme flux of both anti-inflammatory and proinflammatory cytokines evoked by a blood borne infection. The latter is the only characteristic that clearly differentiates sepsis from Systemic Inflammatory Response Syndrome (SIRS). Like sepsis, SIRS is described as an

inflammatory cytokine storm, only it can be caused by a variety non-infectious insults such as pancreatitis [184], trauma [185], leukaemia [186], invasive surgery, and burns [187]. One of the difficulties in distinguishing between SIRS and sepsis is that any one of these conditions will invariably affect the microbiome which could lead to infection like symptoms. Sepsis has a mortality rate of *ca*. ~30% which worsens as the condition progresses into severe sepsis (20-50%), and septic shock (40-80%) [188], [189]. In the USA, there is an incidence of 3 sepsis cases per 1000 people and accounts for 20% (210,000) of all annual in-hospital deaths [190], [191]. Sepsis is one of the most common neonatal complications and may lead to lifelong disabilities and is the biggest killer of children under five [192]–[194]. It is estimated that the condition costs $24 billion in the USA annually [195]. In the UK the annual cost of sepsis is estimated to be between £7.42-£10.2 billion and that effective sepsis management could save £170 million pounds annually [196]. The former figures account for 3.76-5.17% of the UK's total £197.4 billion healthcare expenditure [197], [198]. The total deaths due to sepsis in England are estimated to be approximately 36,900 out of 123,000 cases each year [199]. However, an article published in the Lancet in January 2020 reports that global sepsis deaths have been underestimated with new estimates of 48 million sepsis cases and 11 million sepsis deaths globally each year [200]. This accounts for almost 20% of all deaths worldwide each year. Consequently, financial estimates cited previously may also be underestimated especially considering the reporters express concern on the lack of clarity over incidence, prevalence, and care pathways of sepsis [197]. The overwhelming consensus is that early detection and effective triaging of sepsis will help alleviate all these alarming statistics [200]–[202]. If the spectroscopic test developed within this project could be marketed at £40 per test, a simple calculation

forecasts that the cost to healthcare to test every sepsis incidence in the UK would be £1.476 million (£40 x 36,900) annually, a fraction of the £170 million that the National Institute for Health and Care Excellence (NICE) estimate could be saved by promoting effective sepsis management annually. Even at a unit cost of £100 per test (maximum £3.69 million annually) it is plausible that the test could be cost effective and still represent a sizable market potential. Clearly this is a rudimentary assessment as realistically tests would not be utilised in 100% of cases. An in-depth health economic assessment would be required for accurate figures.

The clinical definitions of sepsis have been changed numerous times and there does not seem to be a satisfactory resolution to this debate due in large part to the highly heterogenous clinical presentation [203], [204]. Clinicians will make a judgement on whether or not a patient is presenting with sepsis based on physical examinations, vital signs and microbiological tests, but this is largely non-specific [205]. Generally, clinicians will make a diagnosis if two of the four SIRS criteria below are observed along with a documented or suspected infection [206]:

1. Temperature >38°C or <36°C
2. Heart rate >90 beats per minute.
3. Respiratory rate > 20 breaths per minute or PaCO2 < 32mmHg.
4. White blood cells count > 12,000/cu mm, or 10% immature (band) forms.

It remains evident that there is currently no clear way of performing a differential diagnosis of SIRS against sepsis. This is important because the clinical response is different depending on the diagnosis [207]. In particular, aggressive fluid resuscitation within the first six hours and antibiotic administration within the first

hour of suspected sepsis is crucial in lowering mortality and morbidity [190], [208]. The current standard of care is the administration of a ‘resuscitation bundle’ within six hours of suspected sepsis aimed at delivering early goal directed therapy (Table 1.1) [209]. Although improvements to patient outcomes following introduction of the resuscitation bundle were good (34.4% relative risk reduction for mortality), uptake of the bundle was low with a reported 14% of 1232 patients across 18 centres in the UK receiving the bundle [210]. To address this the ‘sepsis six’ pathway was introduced as an initial resuscitation bundle designed to offer basic intervention within the first hour of suspected sepsis [211]. The sepsis six bundle comprises three diagnostic tests and three interventions (Table 1.1) [212]. In one hospital, administration of appropriate care within the first hour improved form 29% to 63% of sepsis cases using this simplified approach [212]. However, non-specific markers are still relied upon to triage patients toward these healthcare pathways as there is currently no known molecular signature to identify sepsis [21]. An effective test must return a result in less than 3 hours to halt inappropriate antibiotic treatments and should return a result in less than 1 hour. However, there are few currently proposed molecular diagnostic tests capable of returning a result in less than 4 hours [22], [26]. A rapid, objective, inexpensive, and reliable spectroscopic serum test could therefore further improve uptake of resuscitation bundles.

Table 1.1. *Steps undertaken for sepsis resuscitation bundle and sepsis six pathway.*

Resuscitation bundle:	**Sepsis six pathway (all to be completed within 1h):**
1. Measure lactate within first 6h	1. Administer high-flow oxygen
2. Obtain blood cultures (before antibiotic administration)	2. Obtain blood cultures (before antibiotic administration)
3. Administer broad spectrum antibiotics within 3h	3. Administer empiric IV antibiotics
4. Administer IV fluids	4. Measure serum lactate
5. Achieve Central Venous Pressure (CAP) > 8 mmHg within 6h	5. Administer IV fluids
6. Achieve Mean Arterial Pressure (MAP) > 65mmHg within 6h	6. Commence urine output measurements
7. Achieve central venous oxygen saturation ($ScvO_2$) > 70% within 6h	

To clarify, laboratory biomarker tests are routinely used in the screening of sepsis, but none possess the sensitivity and/or specificity to reliably help the clinician arrive at a diagnosis [213]. However, it is recognised that biomarker tests do improve the classification of sepsis [203]. Lactate tests are the most common laboratory test ordered [189]. Sepsis is a highly metabolic condition; therefore, lactate levels rise sharply and serve as a strong indicator of disease progression. Lactate tests have high sensitivity, but low specificity for sepsis predictions. Of course, anti- and pro-inflammatory serum cytokines, such as procalcitonin [214], c-reactive protein [215], [216], Tumour Necrosis Factor (TNF) [217], [218] and Interleukin-6 [219], and many more may also be requested [189], [213]. Unfortunately, none of these have emerged as a definitive diagnostic tool although it remains to be seen how well a combination of these markers through multivariate analysis or machine learning can predict sepsis and monitor sepsis progression.

The gold standard diagnostic test for sepsis are blood cultures. Blood cultures can take up to five days to return a result [26], require large blood volumes, and determine the causative infectious agent in only 20-40% of sepsis cases [62]. Clearly this is an ineffectual tool to combat sepsis especially in the neonate where drawing large volumes of blood should be avoided. However, it is the only way to confirm the presence of infection and clinicians must know the specific causative pathogen as well as infection source in order to effectively treat the patient [220]. The inability to quickly determine infection and causative pathogen caused the definition of sepsis to include 'suspected infection' giving clinicians more freedom to act. Clinicians also struggle to 'rule out' sepsis when forming a differential diagnosis and cannot risk delaying sepsis therapy which may lead to inappropriate therapies being administered, or detracting from the real underlying pathology [205]. This further highlights the need for quick and specific testing to enable triaging of patients with suspected sepsis.

It is envisioned that an objective, rapid, and low-cost spectroscopic SIRE sepsis test could provide a sensitive and specific triaging tool by predicting if a patient has sepsis or not which may be particularly useful for scenarios where the clinician is seeking to rule out sepsis. A definitive diagnosis would require knowledge of the causative agent and infection site, whereas the proposed test would only inform on the patients' status as someone with sepsis or not. A positive result will initiate the sepsis six bundle and a negative result will help clinicians decide to terminate administration of broad-spectrum antibiotics and prevent resuscitation if not required. For example, a patient may present at the emergency department with several days of vomiting, diarrhoea, fever (>38◦C), a fast heartbeat (>90bpm), and an acute abdomen [221]. Depending on the patient history, a clinician may reasonably determine the patient has

pancreatitis, but equally cannot discount the presence of an infection and possibility of sepsis. In compliance with sepsis six pathway, the clinician must administer broad spectrum antibiotics without yet determining a diagnosis. In this scenario, the clinician may order amylase, glucose, or lipid panels to investigate a pancreatitis diagnosis, but these tests do not prove the absence of sepsis [222]. The critical factor is therefore to confirm if sepsis is present or not and rule it out to pursue other healthcare pathways, or indeed initiate the sepsis six pathway. A SIRE sepsis test would be ordered, and a positive/negative prediction made on whether the patient has sepsis, SIRS, or neither condition.

Since false negatives are associated with greater risk to the patient than false positives, it has been identified from the literature that a NPV and sensitivity in excess of 98%, and PPV and specificity in excess of 85% would be desirable to reliably identify patients with and without sepsis [22], [24]–[26]. The test developed within this project will be measured against these predictive values and compared to predictive values achieved with common laboratory biomarker sepsis tests. It is hoped that the low cost and speed of the test encourage clinicians to err on the side of caution and order a test even in situations where the presence of sepsis is in doubt promoting early detection and improving patient outcomes [190], [191]. To be effective, this test will likely need to be carried out within the emergency department. This is to achieve a low result turnaround time which must be less than 3 hours to enable clinicians to focus on the real pathology, halt antibiotic treatments early, and prevent aggressive resuscitation and would preferably be less than 1 hour for early detection or to prevent antibiotic treatment entirely [26]. Tests would likely be ordered by the emergency

department where most sepsis patients are received [211]. Routine serum sample collection techniques would be employed to facilitate this.

Given the prevalence of sepsis, quick onset, lack of detectability, and economic burden; the requirement for high-throughput, low-cost, rapid, and reliable blood testing methods is clear. ATR-FTIR spectroscopy, using the SIREs described previously, is well poised to fulfil these requirements at the point of care. Moreover, the low sample volume required to perform ATR-FTIR is especially beneficial given the pronounced clinical need in neonatal sepsis diagnostics. The complex and heterogenous biochemistry make sepsis an ideal candidate to demonstrate a machine learning based spectroscopic approach. Most importantly, the result could be delivered quickly and clearly made possible by simple sample preparation requirements and envisaged compatibility with existing clinical workflows. Furthermore, this technology could be present at the point-of-care instead of in a centralised microbiological centre. Crucially, it is probable that this test could deliver results to the clinician within 60 minutes of symptom presentation thereby fulfilling a notorious unmet clinical need [205]. The diagnostic capabilities of SIREs with respect to sepsis detection will be demonstrated in this thesis.

1.6. Aims and Objectives

The central discussion of this thesis is: can infrared spectroscopy be used to process a high capacity of samples in a clinically useful way? The fundamental aim of this project is to answer this through creation of a clinically useful ATR-FTIR serum

diagnostics platform for the triaging of patients with suspected sepsis. Chapters three through six will demonstrate each aspect of this aim.

Lack of high sample throughput capabilities has thus far been a major barrier for the translation of ATR-FTIR to the clinical setting. Chapter three will present the design and development of a SIRE and spectrometer accessory unit that together enable rapid, low cost, and high-throughput ATR-FTIR analysis. Design rationale will be led by analytical performance, usability characteristics, cost, and manufacturing practicality.

SIRE spectrum reproducibility will be optimised in chapter four. The SIRE and accessory unit will be characterised and evaluated with respect to a conventional ATR-FTIR spectrometer set-up with a high sensitivity diamond IRE. This is to determine if the high-throughput ATR-FTIR system maintains the analytical qualities expected of an ATR-FTIR spectrometer. There will be a focus on determining how reliably SIREs can analyse serum samples. Optimal pre-analytical serum sample preparation procedures for the SIRE slide will also be established.

SIRE analytical performance in terms of analyte detection and quantitation will be characterised in chapter five. Diagnostic tests typically quantify a biomolecule concentration within a sample and compare this value to a reference range. To assess diagnostic capabilities of SIREs, it is appropriate to determine quantitative abilities of clinical biomarkers at clinically relevant concentrations. Quantitative analysis of biomolecules within the three of the four major biomolecule groups (carbohydrates, lipids, and proteins) will be demonstrated using SIREs. Quantitative capabilities and limits of detection have been compared to that of a diamond IRE and to clinically

accepted methods characterised within the literature. Biomarkers routinely used in the diagnosis and prognosis of sepsis and SIRS were purposefully selected to verify that spectral changes could be linked to changes within the sample of relevant biomolecules.

Finally, to support claims that SIREs are clinically useful it is imperative to demonstrate that an unmet clinical need can be fulfilled. Chapter six demonstrates that SIREs are capable of aiding in the triage of sepsis through classification of sepsis, SIRS, and control patient samples from a cohort of 148 patient samples. Predictive capabilities are compared against competing sepsis biomarker tests as well as idealistic NPV (>98%), PPV (>85%), sensitivity (>98%), and specificity (>85%) as identified in the literature [22], [26].

1.7. References

[1] M. R. Hoenig, C. N. Aroney, and I. A. Scott, "Early invasive versus conservative strategies for unstable angina and non-ST elevation myocardial infarction in the stent era," *Cochrane Database Syst. Rev.*, no. 3, 2010, doi: 10.1002/14651858.cd004815.pub3.

[2] D. P. Becker, J. D. Miller, J. D. Ward, R. P. Greenberg, H. F. Young, and R. Sakalas, "The outcome from severe head injury with early diagnosis and intensive management," *J. Neurosurg.*, vol. 47, no. 4, pp. 491–502, 1977, doi: 10.3171/jns.1977.47.4.0491.

[3] P. M. Farrell *et al.*, "Evidence on improved outcomes with early diagnosis of cystic fibrosis through neonatal screening: Enough is enough!," *J. Pediatr.*, vol. 147, no. 3 SUPPL., 2005, doi: 10.1016/j.jpeds.2005.08.012.

[4] B. Heidari, "Rheumatoid arthritis: Early diagnosis and treatment outcomes," *Casp. J. Intern. Med.*, vol. 2, no. 1, pp. 161–170, 2011.

[5] F. L. Pagan, "Improving outcomes through early diagnosis of Parkinson's disease.," *Am. J. Manag. Care*, vol. 18, no. 7 Suppl, pp. 176–182, 2012.

[6] S. S. Kim, J. H. Kim, and I. J. Kim, "Current challenges in diabetic nephropathy: Early diagnosis and ways to improve Outcomes," *Endocrinol. Metab.*, vol. 31,

no. 2, pp. 245–253, 2016, doi: 10.3803/EnM.2016.31.2.245.

[7] J. Lavelle, M. Chilutti, and R. Grundmeier, “Delayed Antimicrobial Therapy Increases Mortality and Organ Dysfunction in Pediatric Sepsis,” *Crit Care Med*, vol. 42, no. 11, pp. 2409–2417, 2015, doi: 10.1097/CCM.0000000000000509.Delayed.

[8] R. Viti, A. Marcellusi, P. Sciattella, and O. Viapiana, “Economic evaluation of spondyloarthritis: economic impact of diagnostic delay in Italy,” *Clin. Outcomes Res.*, vol. 10, pp. 45–51, 2018.

[9] T. R. Gildea, S. D. Byfield, D. S. Wilson, and C. C. Quinn, “A retrospective analysis of delays in the diagnosis of lung cancer and associated costs,” *Clin. Outcomes Res.*, vol. 9, pp. 261–269, 2017.

[10] T. R. Einarson, B. G. Bereza, A. Acs, and R. Jensen, “Systematic literature review of the health economic implications of early detection by screening populations at risk for type 2 diabetes,” *Curr. Med. Res. Opin.*, vol. 33, no. 2, pp. 331–358, 2017, doi: 10.1080/03007995.2016.1257977.

[11] The Adademy of Medical Sciences and Cancer Research UK, “Accelerating the translation of early detection and diagnosis research in cancer,” 2018.

[12] Work, “Global Health and Aging,” 2011. Accessed: Apr. 25, 2019. [Online]. Available: https://www.who.int/ageing/publications/global_health.pdf.

[13] A. Jutel, “Sociology of diagnosis: A preliminary review,” *Sociol. Heal. Illn.*, vol. 31, no. 2, pp. 278–299, 2009, doi: 10.1111/j.1467-9566.2008.01152.x.

[14] E. P. Balogh, B. T. Miller, and J. R. Ball, *Improving diagnosis in health care*. 2016.

[15] D. Kuupiel, V. Bawontuo, and T. Mashamba-Thompson, “Improving the Accessibility and Efficiency of Point-of-Care Diagnostics Services in Low- and Middle-Income Countries: Lean and Agile Supply Chain Management,” *Diagnostics*, vol. 7, no. 4, p. 58, 2017, doi: 10.3390/diagnostics7040058.

[16] T. Badrick, “Evidence-Based Laboratory Medicine,” *Clin Biochem Rev*, vol. 34, no. August, pp. 43–46, 2013.

[17] P. Wang, “Developing In Vitro Diagnostics for Commercialization and Clinical Implementation,” in *Intellectual Property-Regulator*, vol. 1, no. 3, 2019, pp. 1–6.

[18] D. E. Newman-Toker and M. A. Makary, “Measuring Diagnostic Errors in Primary Care,” *JAMA Intern. Med.*, vol. 173, no. 6, p. 425, 2013, doi: 10.1001/jamainternmed.2013.225.

[19] H. Singh, G. D. Schiff, M. L. Graber, I. Onakpoya, and M. J. Thompson, “The global burden of diagnostic errors in primary care,” *BMJ Qual. Saf.*, vol. 26, no. 6, pp. 484–494, 2017, doi: 10.1136/bmjqs-2016-005401.

[20] D. W. S. Jennifer R. Schneider, Dave D. Chadee, Akio Mori, Jeanne Romero-Severson, “Types and Origins of Diagnostic Errors in Primary Care Settings,”

JAMA Intern. Med., vol. 23, no. 1, pp. 1–7, 2008, doi: 10.1038/jid.2014.371.

[21] T. Evans, "Diagnosis and Management of Sepsis," *Clin. Med. (Northfield. Il).*, vol. 18, no. 2, pp. 146–149, 2018.

[22] P. C. Ng and H. S. Lam, "Biomarkers for Late-Onset Neonatal Sepsis: Cytokines and Beyond," *Clin. Perinatol.*, vol. 37, no. 3, pp. 599–610, 2010, doi: 10.1016/j.clp.2010.05.005.

[23] I. S. Pradipta *et al.*, "Antibiotic Resistance in Sepsis Patients : Evaluation and Recommendation of Antibiotic Use," vol. 5, no. 6, 2013, doi: 10.4103/1947-2714.114165.

[24] L. Srinivasan and M. C. Harris, "New technologies for the rapid diagnosis of neonatal sepsis," *Curr. Opin. Pediatr.*, vol. 24, no. 2, pp. 165–171, 2012, doi: 10.1097/MOP.0b013e3283504df3.

[25] T. G. Connell, M. Rele, D. Cowley, J. P. Buttery, and N. Curtis, "How reliable is a negative blood culture result? Volume of blood submitted for culture in routine practice in a children's hospital," *Pediatrics*, vol. 119, no. 5, pp. 891–896, 2007, doi: 10.1542/peds.2006-0440.

[26] M. Sinha, J. Jupe, H. Mack, T. P. Coleman, S. M. Lawrence, and I. Fraley, "Emerging Technologies for Molecular Diagnosis of Sepsis," *Clin. Microbiol. Rev.*, vol. 31, no. 2, pp. 1–26, 2018.

[27] M. Omidi *et al.*, "Characterization of biomaterials," in *Biomaterials for Oral and Dental Tissue Engineering*, no. September, Elsevier Ltd, 2017, pp. 97–115.

[28] K. D. Daniel *et al.*, "Implantable Diagnostic Device for Cancer Monitoring," *Biosens. Bioelectron.*, vol. 24, no. 11, pp. 3252–3257, 2009, doi: 10.1016/j.bios.2009.04.010.Implantable.

[29] K. Guk *et al.*, "Evolution of Wearable Devices with Real-Time Disease Monitoring for Personalized Healthcare," *Nanomaterials*, vol. 9, no. 819, pp. 1–23, 2019.

[30] D. C. Klonoff, "The Benefits of Implanted Glucose Sensors," *J. Diabetes Sci. Technol.*, vol. 1, no. 6, pp. 797–800, 2007.

[31] M. Ilié and P. Hofman, "Pros: Can tissue biopsy be replaced by liquid biopsy?," *Transl. lung cancer Res.*, vol. 5, no. 4, pp. 420–423, 2016, doi: 10.21037/tlcr.2016.08.06.

[32] A. R. Parikh *et al.*, "Liquid versus tissue biopsy for detecting acquired resistance and tumor heterogeneity in gastrointestinal cancers," *Nat. Med.*, vol. 25, no. September, pp. 1415–1421, 2019, doi: 10.1038/s41591-019-0561-9.

[33] A. E. G. Cass and S. Sharma, "Microneedle Enzyme Sensor Arrays for Continuous In Vivo Monitoring," in *Methods in Enzymology*, 1st ed., vol. 589, Elsevier Inc., 2017, pp. 413–427.

[34] V. G. Gavalas, J. Wang, and L. G. Bachas, "Chapter 18 Membranes for the development of biosensors," in *Membrane Science and Technology*, vol. 8, no.

C, Elsevier, 2003, pp. 379–392.

[35] B. D. Malhotra and M. A. Ali, “Chapter 7: Nanostructured Biomaterials for In Vivo Biosensors,” in *Nanomaterials for Biosensors*, 2018, pp. 183–219.

[36] H. E. Koschwanez and W. M. Reichert, “In vitro, in vivo and post explantation testing of glucose-detecting biosensors: Current methods and recommendations,” *Biomaterials*, vol. 28, no. 25, pp. 3687–3703, 2007, doi: 10.1016/j.biomaterials.2007.03.034.

[37] K. Achyuthan and D. Whitten, “Design Considerations for High Throughput Screening and In Vitro Diagnostic Assays,” *Comb. Chem. High Throughput Screen.*, vol. 10, no. 6, pp. 399–412, 2007, doi: 10.2174/138620707781996475.

[38] K. M. Lei, P. I. Mak, M. K. Law, and R. P. Martins, “CMOS biosensors for: In vitro diagnosis-transducing mechanisms and applications,” *Lab Chip*, vol. 16, no. 19, pp. 3664–3681, 2016, doi: 10.1039/c6lc01002d.

[39] H. Fabian, N. A. N. Thi, M. Eiden, P. Lasch, J. Schmitt, and D. Naumann, “Diagnosing benign and malignant lesions in breast tissue sections by using IR-microspectroscopy,” *Biochim. Biophys. Acta - Biomembr.*, vol. 1758, no. 7, pp. 874–882, 2006, doi: 10.1016/j.bbamem.2006.05.015.

[40] M. Al-Abbadi, “Basics of cytology,” *Avicenna J. Med.*, vol. 1, no. 1, p. 18, 2011, doi: 10.4103/2231-0770.83719.

[41] T. Tuuminen, “Urine as a specimen to diagnose infections in twenty-first century: Focus on analytical accuracy,” *Front. Immunol.*, vol. 3, no. MAR, pp. 1–6, 2012, doi: 10.3389/fimmu.2012.00045.

[42] S. Mondello, U. Muller, A. Jeromin, J. Streeter, R. L. Hayes, and K. K. Wang, “Blood-based diagnostics of traumatic brain injuries,” *Expert Rev. Mol. Diagn.*, vol. 11, no. 1, pp. 65–78, 2011, doi: 10.1586/erm.10.104.

[43] D. Fernández-Lázaro, J. L. G. Hernández, A. C. García, A. C. Martínez, J. Mielgo-Ayuso, and J. J. Cruz-Hernández, “Liquid biopsy as novel tool in precision medicine: Origins, properties, identification and clinical perspective of cancer’s biomarkers,” *Diagnostics*, vol. 10, no. 4, pp. 1–17, 2020, doi: 10.3390/diagnostics10040215.

[44] D. Marx *et al.*, “Proteomics and Metabolomics for AKI Diagnosis,” *Semin. Nephrol.*, vol. 38, no. 1, pp. 63–87, 2018, doi: 10.1016/j.semnephrol.2017.09.007.

[45] H. Amalou and B. J. Wood, “Biopsy and personalized medicine,” *Nature Reviews Gastroenterology and Hepatology*, vol. 9, no. 11. Nature Publishing Group, p. 683, Oct. 02, 2012, doi: 10.1038/nrgastro.2012.100-c1.

[46] E. Kovac *et al.*, “Association of Baseline Prostate-Specific Antigen Level With Long-term Diagnosis of Clinically Significant Prostate Cancer Among Patients Aged 55 to 60 Years: A Secondary Analysis of a Cohort in the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Sc,” *JAMA Netw. open*, vol. 3, no. 1, pp. 1–11, 2020, doi: 10.1001/jamanetworkopen.2019.19284.

[47] G. De Rubis, S. Rajeev Krishnan, and M. Bebawy, “Liquid Biopsies in Cancer Diagnosis, Monitoring, and Prognosis,” *Trends Pharmacol. Sci.*, vol. 40, no. 3, pp. 172–186, 2019, doi: 10.1016/j.tips.2019.01.006.

[48] S. L. Fan, M. John Lee, and D. G. Remick, “Diagnosing Sepsis - The Role of Laboratory Medicine,” *Physiol. Behav.*, vol. 460, no. 1, pp. 203–210, 2016, doi: 10.1016/j.physbeh.2017.03.040.

[49] M. Meleti, D. Cassi, P. Vescovi, G. Setti, T. A. Pertinhez, and M. E. Pezzi, “Salivary biomarkers for diagnosis of systemic diseases and malignant tumors. A systematic review,” *Med. Oral Patol. Oral Cir. Bucal*, vol. 25, no. 2, pp. 299–310, 2020, doi: 10.4317/medoral.

[50] A. Rohman, A. Windarsih, E. Lukitaningsih, M. Rafi, K. Betania, and N. A. Fadzillah, “The use of FTIR and Raman spectroscopy in combination with chemometrics for analysis of biomolecules in biomedical fluids: A review,” *Biomed. Spectrosc. Imaging*, vol. 8, no. 3–4, pp. 55–71, 2020, doi: 10.3233/bsi-200189.

[51] L. B. Leal, M. S. Nogueira, R. A. Canevari, and L. F. C. S. Carvalho, “Vibration spectroscopy and body biofluids: Literature review for clinical applications,” *Photodiagnosis Photodyn. Ther.*, vol. 24, pp. 237–244, 2018, doi: 10.1016/j.pdpdt.2018.09.008.

[52] S. K. Vashist and J. H. T. Luong, “Immunoassays: An overview,” in *Handbook of Immunoassay Technologies: Approaches, Performances, and Applications*, Elsevier, 2018, pp. 1–18.

[53] P. Mehrotra, “Biosensors and their applications - A review,” *J. Oral Biol. Craniofacial Res.*, vol. 6, no. 2, pp. 153–159, May 2016, doi: 10.1016/j.jobcr.2015.12.002.

[54] C. Giuliano, C. R. Patel, and P. B. Kale-Pradhan, “A guide to bacterial culture identification and results interpretation,” *Pharm. Ther.*, vol. 44, no. 4, pp. 192–200, 2019.

[55] M. Fakruddin *et al.*, “Nucleic acid amplification: Alternative methods of polymerase chain reaction,” *J. Pharm. Bioallied Sci.*, vol. 5, no. 4, pp. 245–252, 2013, doi: 10.4103/0975-7406.120066.

[56] M. J. Baker *et al.*, “Using Fourier transform IR spectroscopy to analyze biological materials,” *Nat. Protoc.*, vol. 9, no. 8, pp. 1771–1791, 2014, doi: 10.1038/nprot.2014.110.

[57] H. J. Butler *et al.*, “Using Raman spectroscopy to characterize biological materials,” *Nat. Protoc.*, vol. 11, no. 4, pp. 664–687, 2016, doi: 10.1038/nprot.2016.036.

[58] S. Zhang, G. A. Nagana Gowda, T. Ye, and D. Raftery, “Advances in NMR-based biofluid analysis and metabolite profiling,” *Analyst*, vol. 135, no. 7, pp. 1490–1498, 2010, doi: 10.1039/c000091d.

[59] J. Heinemann, B. Noon, D. Willems, K. Budeski, and B. Bothner, “Analysis of

Raw Biofluids by Mass Spectrometry Using Microfluidic Diffusion-Based Separation," *Anal Methods*, vol. 9, no. 3, pp. 385–392, 2017, doi: 10.1039/C6AY02827F.Analysis.

[60] K. B. Laupland and L. Valiquette, "The changing culture of the microbiology laboratory.," *Can. J. Infect. Dis. Med. Microbiol.*, vol. 24, no. 3, pp. 125–8, 2013, doi: 10.1155/2013/101630.

[61] S. Obaro, "Updating the diagnosis of bacterial meningitis," *Lancet Infect. Dis.*, vol. 19, no. 11, pp. 1160–1161, 2019, doi: 10.1016/S1473-3099(19)30549-3.

[62] N. Schaub, R. Frei, and C. Mueller, "Addressing unmet clinical needs in the early diagnosis of sepsis," *Swiss Med. Wkly.*, vol. 141, no. July, pp. 1–4, 2011, doi: 10.4414/smw.2011.13244.

[63] L. Ljungström *et al.*, "Clinical evaluation of commercial nucleic acid amplification tests in patients with suspected sepsis," *BMC Infect. Dis.*, vol. 15, no. 199, pp. 1–10, 2015, doi: 10.1186/s12879-015-0938-4.

[64] M. Vaneechoutte and J. Van Eldere, "The possibilities and limitations of nucleic acid amplification technology in diagnostic microbiology," *J. Med. Microbiol.*, vol. 46, no. 3, pp. 188–194, 1997, doi: 10.1099/00222615-46-3-188.

[65] R. Hans and N. Marwaha, "Nucleic acid testing-benefits and constraints," *Asian J. Transfus. Sci.*, vol. 8, no. 1, pp. 2–3, 2014, doi: 10.4103/0973-6247.126679.

[66] M. M. Rönn, L. Mc Grath-Lone, B. Davies, J. D. Wilson, and H. Ward, "Evaluation of the performance of nucleic acid amplification tests (NAATs) in detection of chlamydia and gonorrhoea infection in vaginal specimens relative to patient infection status: A systematic review," *BMJ Open*, vol. 9, no. 1, pp. 1–9, 2019, doi: 10.1136/bmjopen-2018-022510.

[67] C. Nistor and J. Emnéus, "Chapter 9 Immunoassay: potentials and limitations," in *Comprehensive Analytical Chemistry*, vol. 44, 2005, pp. 375–427.

[68] S. B. Nimse, M. D. Sonawane, K. S. Song, and T. Kim, "Biomarker detection technologies and future directions," *Analyst*, vol. 141, no. 3, pp. 740–755, 2016, doi: 10.1039/c5an01790d.

[69] J. A. Bastarache, T. Koyama, N. L. Wickersham, D. B. Mitchell, R. L. Mernaugh, and L. B. Ware, "Accuracy and Reproducibility of a Multiplex Immunoassay Platform : A Validation Study," *J. Immunol. Methods*, vol. 367, no. 1–2, pp. 33–39, 2012, doi: 10.1016/j.jim.2011.01.005.Accuracy.

[70] A. A. Ellington, I. J. Kullo, K. R. Bailey, and G. G. Klee, "Measurement and quality control issues in multiplex protein assays: A case study," *Clin. Chem.*, vol. 55, no. 6, pp. 1092–1099, 2009, doi: 10.1373/clinchem.2008.120717.

[71] S. N. Sawant, "Development of Biosensors from Biopolymer Composites," in *Biopolymer Composites in Electronics*, Elsevier Inc., 2017, pp. 353–383.

[72] A. C. Carpenter, I. T. Paulsen, and T. C. Williams, "Blueprints for biosensors: Design, limitations, and applications," *Genes (Basel).*, vol. 9, no. 375, pp. 1–28, 2018, doi: 10.3390/genes9080375.

[73] M. L. Sin, K. E. Mach, P. K. Wong, and J. C. Liao, “Advances and challenges in biosensor-based diagnosis of infectious diseases,” *Expert Rev. Mol. Diagn.*, vol. 14, no. 2, pp. 225–244, 2014, doi: 10.1586/14737159.2014.888313.

[74] M. Ritzi-Lehnert, “Development of chip-compatible sample preparation for diagnosis of infectious diseases,” *Expert Rev. Mol. Diagn.*, vol. 12, no. 2, pp. 189–206, 2012, doi: 10.1586/erm.11.98.

[75] R. Lal, “Integrated biosensors: Promises and problems,” *Bioelectrochemistry Bioenerg.*, vol. 27, no. 2, pp. 121–139, 1992, doi: 10.1016/0302-4598(92)87037-U.

[76] S. Banerjee, “Empowering Clinical Diagnostics with Mass Spectrometry,” *ACS Omega*, vol. 5, no. 5, pp. 2041–2048, 2020, doi: 10.1021/acsomega.9b03764.

[77] D. J. Swiner, S. Jackson, B. J. Burris, and A. K. Badu-Tawiah, “Applications of mass spectrometry for clinical diagnostics: The influence of turnaround time,” *Anal. Chem.*, vol. 92, no. 1, pp. 183–202, 2020, doi: 10.1021/acs.analchem.9b04901.

[78] M. Hermes *et al.*, “Mid-IR hyperspectral imaging for label-free histopathology and cytology,” *J. Opt.*, vol. 20, no. 2, p. 023002, Jan. 2018, doi: 10.1088/2040-8986/AAA36B.

[79] O. J. Old *et al.*, “Vibrational spectroscopy for cancer diagnostics,” *Anal. Methods*, vol. 6, no. 12, pp. 3901–3917, 2014, doi: 10.1039/c3ay42235f.

[80] K. Kong, C. Kendall, N. Stone, and I. Notingher, “Raman spectroscopy for medical diagnostics - From in-vitro biofluid assays to in-vivo cancer detection,” *Adv. Drug Deliv. Rev.*, vol. 89, pp. 121–134, 2015, doi: 10.1016/j.addr.2015.03.009.

[81] B. R. Wood *et al.*, “Application of Vibrational Spectroscopy and Imaging to Point-of-Care Medicine : A Review,” vol. 72, pp. 52–84, 2018, doi: 10.1177/0003702818791939.

[82] V. Balan *et al.*, “Vibrational spectroscopy fingerprinting in medicine: From molecular to clinical practice,” *Materials (Basel).*, vol. 12, no. 18, pp. 1–40, 2019, doi: 10.3390/ma12182884.

[83] K. Eberhardt, C. Stiebing, C. Matthaüs, M. Schmitt, and J. Popp, “Advantages and limitations of Raman spectroscopy for molecular diagnostics: An update,” *Expert Rev. Mol. Diagn.*, vol. 15, no. 6, pp. 773–787, 2015, doi: 10.1586/14737159.2015.1036744.

[84] F. Faghihzadeh, N. M. Anaya, L. A. Schifman, and V. Oyanedel-Craver, “Fourier transform infrared spectroscopy to assess molecular-level changes in microorganisms exposed to nanoparticles,” *Nanotechnol. Environ. Eng.*, vol. 1, no. 1, pp. 1–16, 2016, doi: 10.1007/s41204-016-0001-8.

[85] D. Rohleder *et al.*, “Comparison of mid-infrared and Raman spectroscopy in the quantitative analysis of serum,” *J. Biomed. Opt.*, vol. 10, no. 3, pp. 1–10, 2005, doi: 10.1117/1.1911847.

[86] C. Hughes *et al.*, “Assessing the challenges of Fourier transform infrared spectroscopic analysis of blood serum,” *J. Biophotonics*, vol. 7, no. 3–4, pp. 180–188, 2014, doi: 10.1002/jbio.201300167.

[87] K. M. Dorling and M. J. Baker, “Highlighting attenuated total reflection Fourier transform infrared spectroscopy for rapid serum analysis,” *Trends in Biotechnology*, vol. 31, no. 6. pp. 327–328, 2013, doi: 10.1016/j.tibtech.2013.03.010.

[88] M. J. Baker *et al.*, “Clinical applications of Infrared and Raman spectroscopy : state of play and future challenges,” *Analyst*, vol. 143, pp. 1735–1757, 2018, doi: 10.1039/c7an01871a.

[89] C. Pierrakos, D. Velissaris, M. Bisdorff, J. C. Marshall, and J. L. Vincent, “Biomarkers of sepsis: Time for a reappraisal,” *Crit. Care*, vol. 24, no. 287, pp. 1–15, 2020, doi: 10.1186/s13054-020-02993-5.

[90] S. B. Armen *et al.*, “Improving Outcomes in Patients With Sepsis,” *Am. J. Med. Qual.*, vol. 31, no. 1, pp. 56–63, 2016, doi: 10.1177/1062860614551042.

[91] A. L. Shane, P. J. Sánchez, and B. J. Stoll, “Neonatal sepsis,” *Lancet*, vol. 390, pp. 1–11, 2017, doi: 10.1016/S0140-6736(17)31002-4.

[92] F. Kim, R. A. Polin, and T. A. Hooven, “Neonatal sepsis,” *BMJ*, vol. 371, pp. 1–4, 2020, doi: 10.1136/bmj.m3672.

[93] A. L. Mitchell, K. B. Gajjar, G. Theophilou, F. L. Martin, and P. L. Martin-hirsch, “Vibrational spectroscopy of biofluids for disease screening or diagnosis : translation from the laboratory to a clinical setting,” *J. Biophotonics*, vol. 7, no. 3–4, pp. 153–165, 2014, doi: 10.1002/jbio.201400018.

[94] E. Schleicher, “The clinical chemistry laboratory: Current status, problems and diagnostic prospects,” *Anal. Bioanal. Chem.*, vol. 384, no. 1, pp. 124–131, 2006, doi: 10.1007/s00216-005-0185-4.

[95] L. Lovergne, P. Bouzy, V. Untereiner, and R. Garnotel, “Biofluid infrared spectro-diagnostics: pre-analytical considerations for clinical applications,” *Faraday Discuss.*, vol. 187, pp. 521–537, 2016, doi: 10.1039/c5fd00184f.

[96] U. P. Rohr *et al.*, “The value of in vitro diagnostic testing in medical practice: A status report,” *PLoS One*, vol. 11, no. 3, pp. 1–16, 2016, doi: 10.1371/journal.pone.0149856.

[97] L. Kumar and A. L. Das, “Assessment of Serum Lipid Profile in Patients of Coronary Artery Disease: A Case-Control Study,” *Int. J. Contemp. Med. Res. [IJCMR]*, vol. 5, no. 5, pp. 59–62, 2018, doi: 10.21276/ijcmr.2018.5.5.42.

[98] G. L. Sternbach, “The Glasgow Coma Scale,” *J. Emerg. Med.*, vol. 19, no. 1, pp. 67–71, 2000, doi: 10.1016/S0736-4679(00)00182-7.

[99] A. E. Jones, S. Trzeciak, and J. A. Kline, “The Sequential Organ Failure Assessment score for predicting outcome in patients with severe sepsis and evidence of hypoperfusion at the time of emergency department presentation,” *Crit. Care Med.*, vol. 37, no. 5, pp. 1649–1654, 2009, doi:

10.1097/CCM.0b013e31819def97.

[100] P. V Dong and O. L. Cremer, “Limitations of the use of the Glasgow Coma Scale in intensive care patients with non-neurological primary disease: a search for alternatives,” *Crit. Care*, vol. 15, no. Suppl 1, p. P506, 2011, doi: 10.1186/cc9421.

[101] S. Lambden, P. F. Laterre, M. M. Levy, and B. Francois, “The SOFA score - Development, utility and challenges of accurate assessment in clinical trials,” *Crit. Care*, vol. 23, no. 1, pp. 1–9, 2019, doi: 10.1186/s13054-019-2663-7.

[102] M. A. Myszczynska *et al.*, “Applications of machine learning to diagnosis and treatment of neurodegenerative diseases,” *Nat. Rev. Neurol.*, vol. 16, no. 8, pp. 440–456, 2020, doi: 10.1038/s41582-020-0377-8.

[103] D. Bone, S. L. Bishop, M. P. Black, M. S. Goodwin, C. Lord, and S. S. Narayanan, “Use of machine learning to improve autism screening and diagnostic instruments: effectiveness, efficiency, and multi-instrument fusion,” *J. Child Psychol. Psychiatry Allied Discip.*, vol. 57, no. 8, pp. 927–937, 2016, doi: 10.1111/jcpp.12559.

[104] D. P. Lewis, T. Jebara, and W. S. Noble, “Support vector machine learning from heterogeneous data: An empirical analysis using protein sequence and structure,” *Bioinformatics*, vol. 22, no. 22, pp. 2753–2760, 2006, doi: 10.1093/bioinformatics/btl475.

[105] C. J. Kelly, A. Karthikesalingam, M. Suleyman, G. Corrado, and D. King, “Key challenges for delivering clinical impact with artificial intelligence,” *BMC Med.*, vol. 17, no. 195, pp. 1–9, 2019, doi: 10.1186/s12916-019-1426-2.

[106] M. J. Baker *et al.*, “Developing and understanding biofluid vibrational spectroscopy: A critical review,” *Chem. Soc. Rev.*, vol. 45, no. 7, pp. 1803–1818, 2016, doi: 10.1039/c5cs00585j.

[107] F. Jiang *et al.*, “Artificial intelligence in healthcare: Past, present and future,” *Stroke Vasc. Neurol.*, vol. 2, no. 4, pp. 230–243, 2017, doi: 10.1136/svn-2017-000101.

[108] “A Full-Service Remote Healthcare Platform - Current Health.” https://currenthealth.com/ (accessed Oct. 09, 2020).

[109] N. T. Conway, B. Allardice, D. J. Wake, and S. G. Cunningham, “User Experiences of an Electronic Personal Health Record for Diabetes,” *J. Diabetes Sci. Technol.*, vol. 13, no. 4, pp. 744–750, 2019, doi: 10.1177/1932296818818837.

[110] C. McGoldrick, S. Crawford, and J. J. Evans, “MindMate: A single case experimental design study of a reminder system for people with dementia,” *Neuropsychol. Rehabil.*, vol. 0, no. 0, pp. 1–21, 2019, doi: 10.1080/09602011.2019.1653936.

[111] T. Davenport and R. Kalakota, “The potential for artificial intelligence in healthcare,” *Futur. Healthc. J.*, vol. 6, no. 2, pp. 94–102, 2019.

[112] D. W. Sun, *Infrared Spectroscopy for Food Quality Analysis and Control*. Elsevier Inc., 2009.

[113] A. Hashimoto and T. Kameoka, "Applications of infrared spectroscopy to biochemical, food, and agricultural processes," *Appl. Spectrosc. Rev.*, vol. 43, no. 5, pp. 416–451, 2008, doi: 10.1080/05704920802108131.

[114] G. N. Kalinkova, "Infrared spectroscopy in pharmacy," *Vib. Spectrosc.*, vol. 19, no. 2, pp. 307–320, 1999, doi: 10.1016/S0924-2031(99)00017-X.

[115] J. Luypaert, D. L. Massart, and Y. Vander Heyden, "Near-infrared spectroscopy applications in pharmaceutical analysis," *Talanta*, vol. 72, no. 3, pp. 865–883, 2007, doi: 10.1016/j.talanta.2006.12.023.

[116] P. A. Pantoja, J. López-Gejo, C. A. O. do Nascimento, and G. A. C. Le Roux, "Application of Near-Infrared Spectroscopy to the Characterization of Petroleum," *Anal. Charact. Methods Crude Oil. Relat. Prod.*, pp. 221–243, 2017, doi: 10.1002/9781119286325.ch8.

[117] C. W. Brown, P. F. Lynch, and M. Ahmadjian, "Applications of Infrared Spectroscopy in Petroleum Analysis and Oil Spill Identification," *Appl. Spectrosc. Rev.*, vol. 9, no. 1, pp. 223–248, 1975, doi: 10.1080/05704927508081491.

[118] T. Visser, "Infrared Spectroscopy in Environmental Analysis," *Encycl. Anal. Chem.*, vol. 10, pp. 1–21, 2006, doi: 10.1002/9780470027318.a0832.

[119] A. Takamura, K. Watanabe, T. Akutsu, and T. Ozawa, "Soft and Robust Identification of Body Fluid Using Fourier Transform Infrared Spectroscopy and Chemometric Strategies for Forensic Analysis," *Sci. Rep.*, vol. 8, no. 1, pp. 1–10, 2018, doi: 10.1038/s41598-018-26873-9.

[120] A. V. Ewing and S. G. Kazarian, "Infrared spectroscopy and spectroscopic imaging in forensic science," *Analyst*, vol. 142, no. 2, pp. 257–272, 2017, doi: 10.1039/c6an02244h.

[121] J. R. Hands *et al.*, "Attenuated Total Reflection Fourier Transform Infrared (ATR-FTIR) spectral discrimination of brain tumour severity from serum samples," *J. Biophotonics*, vol. 7, no. 3–4, pp. 189–199, 2014, doi: 10.1002/jbio.201300149.

[122] J. R. Hands *et al.*, "Brain tumour differentiation : rapid stratified serum diagnostics via attenuated total reflection Fourier-transform infrared spectroscopy," *J. Neurooncol.*, vol. 127, no. 3, pp. 463–472, 2016, doi: 10.1007/s11060-016-2060-x.

[123] V. Untereiner *et al.*, "Bile analysis using high-throughput FTIR spectroscopy for the diagnosis of malignant biliary strictures : a pilot study in 57 patients," *J. Biophotonics*, vol. 7, no. 3–4, pp. 241–253, 2014, doi: 10.1002/jbio.201300166.

[124] M. J. Baker, E. Gazi, M. D. Brown, J. H. Shanks, P. Gardner, and N. W. Clarke, "FTIR-based spectroscopic analysis in the identification of clinically aggressive prostate cancer," *Br. J. Cancer*, vol. 99, no. 11, pp. 1859–1866, 2008, doi:

10.1038/sj.bjc.6604753.

[125] B. R. Wood, M. A. Quinn, F. R. Burden, and D. Mcnaughton, “An Investigation into FTIR Spectroscopy as a Biodiagnostic Tool for Cervical Cancer,” *Biospectroscopy*, vol. 2, pp. 143–153, 1996.

[126] A. Travo, C. Paya, G. Déléris, J. Colin, and I. Forfar, “Potential of FTIR spectroscopy for analysis of tears for diagnosis purposes,” *Anal. Bioanal. Chem.*, vol. 406, pp. 2367–2376, 2014, doi: 10.1007/s00216-013-7607-5.

[127] J. Titus, E. Viennois, D. Merlin, and A. G. U. Perera, “Minimally invasive screening for colitis using attenuated total internal reflectance fourier transform infrared spectroscopy,” *J. Biophotonics*, vol. 10, pp. 465–472, 2017, doi: 10.1002/jbio.201600041.

[128] K. Gajjar *et al.*, “Fourier-transform infrared spectroscopy coupled with a classification machine for the analysis of blood plasma or serum: A novel diagnostic approach for ovarian cancer,” *Analyst*, vol. 138, no. 14, pp. 3917–3926, 2013, doi: 10.1039/c3an36654e.

[129] A. Khoshmanesh, M. W. A. Dixon, S. Kenny, L. Tilley, D. Mcnaughton, and B. R. Wood, “Detection and Quantification of Early-Stage Malaria Parasites in Laboratory Infected Erythrocytes by Attenuated Total Reflectance Infrared Spectroscopy and Multivariate Analysis,” *Anal. Chem.*, vol. 86, no. 9, pp. 4379–4386, 2014.

[130] P. Bassan *et al.*, “Reflection contributions to the dispersion artefact in FTIR spectra of single biological cells,” *Analyst*, vol. 134, no. 6, pp. 1171–1175, 2009, doi: 10.1039/b821349f.

[131] M. K. Tuck *et al.*, “Standard Operating Procedures for Serum and Plasma Collection: Early Detection Research Network Consensus Statement Standard Operating Procedure Integration Working Group,” *J Proteome Res*, vol. 8, no. 1, pp. 113–117, 2009, doi: 10.1089/bio.2015.0059.

[132] J. Guo, N. Schupf, R. P. Mayeux, and Y. Gu, “Reproducibility of serum cytokines in an elderly population,” *Immun. Ageing*, vol. 17, no. 29, pp. 1–10, 2020, doi: 10.1186/s12979-020-00201-0.

[133] A. Harper *et al.*, “Reproducibility of Serum Potassium Values in Serum From Blood Samples Stored for Increasing Times Prior to Centrifugation and Analysis,” *J. Clin. Lab. Anal.*, vol. 30, no. 3, pp. 244–247, 2016, doi: 10.1002/jcla.21843.

[134] J. H. Livesey, M. J. Ellis, and M. J. Evans, “Pre-analytical requirements.,” *Clin. Biochem. Rev.*, vol. 29, no. Suppl i, pp. S11-15, 2008, [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/18852848%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2556574.

[135] G. Lima-Oliveira, G. Lippi, G. L. Salvagno, G. Picheth, and G. C. Guidi, “Laboratory Diagnostics and Quality of Blood Collection,” *J. Med. Biochem.*, vol. 34, no. 3, pp. 288–294, 2015, doi: 10.2478/jomb-2014-0043.

[136] H. J. Byrne, F. Bonnier, J. McIntyre, and D. R. Parachalil, “Quantitative analysis of human blood serum using vibrational spectroscopy,” *Clin. Spectrosc.*, vol. 2, no. March, p. 100004, 2020, doi: 10.1016/j.clispe.2020.100004.

[137] J. M. Cameron, H. J. Butler, D. S. Palmer, and M. J. Baker, “Biofluid spectroscopic disease diagnostics: A review on the processes and spectral impact of drying,” *J. Biophotonics*, vol. 11, pp. 1–12, 2018, doi: 10.1002/jbio.201700299.

[138] L. Lovergne, G. Clemens, V. Untereiner, R. A. Lukaszweski, G. D. Sockalingum, and M. J. Baker, “Investigating Optimum Sample Preparation for Infrared Spectroscopic Serum Diagnostics,” *Anal. Methods*, vol. 7, no. 17, pp. 6921–7382, 2015, doi: 10.1039/C5AY00502G.

[139] J. Ollesch, H. M. Heise, and T. Behrens, “FTIR spectroscopy of biofluids revisited: an automated approach to spectral biomarker identification,” *Analyst*, vol. 138, pp. 4092–4102, 2013, doi: 10.1039/c3an00337j.

[140] E. Diessel, P. Kamphaus, K. Grothe, R. Kurte, U. Damm, and H. M. Heise, “Nanoliter serum sample analysis by mid-infrared spectroscopy for minimally invasive blood-glucose monitoring,” *Appl. Spectrosc.*, vol. 59, no. 4, pp. 442–451, 2005, doi: 10.1366/0003702053641360.

[141] E. Diessel, S. Willmann, P. Kamphaus, R. Kurte, U. Damm, and H. M. Heise, “Glucose quantification in dried-down nanoliter samples using mid-infrared attenuated total reflection spectroscopy,” *Appl. Spectrosc.*, vol. 58, no. 4, pp. 442–450, 2004, doi: 10.1366/000370204773580293.

[142] J. R. Hands *et al.*, “Investigating the rapid diagnosis of gliomas from serum samples using infrared spectroscopy and cytokine and angiogenesis factors,” *Anal. Bioanal. Chem.*, vol. 405, no. 23, pp. 7347–7355, 2013, doi: 10.1007/s00216-013-7163-z.

[143] M. Khanmohammadi, A. B. Garmarudi, M. Ramin, and K. Ghasemi, “Diagnosis of renal failure by infrared spectrometric analysis of human serum samples and soft independent modeling of class analogy,” *Microchem. J.*, vol. 106, pp. 67–72, 2013, doi: 10.1016/j.microc.2012.05.006.

[144] H. Ghimire, M. Venkataramani, Z. Bian, Y. Liu, and A. G. U. Perera, “ATR-FTIR spectral discrimination between normal and tumorous mouse models of lymphoma and melanoma from serum samples,” *Sci. Rep.*, vol. 7, no. 16993, pp. 1–9, 2017, doi: 10.1038/s41598-017-17027-4.

[145] V. E. Sitnikova, M. A. Kotkova, T. N. Nosenko, T. N. Kotkova, D. M. Martynova, and M. V. Uspenskaya, “Breast cancer detection by ATR-FTIR spectroscopy of blood serum and multivariate data-analysis,” *Talanta*, vol. 214, no. 120857, pp. 1–8, 2020, doi: 10.1016/j.talanta.2020.120857.

[146] B. Bird, M. J. Romeo, M. Diem, K. Bedrossian, N. Laver, and S. Naber, “Cytology by infrared micro-spectroscopy: Automatic distinction of cell types in urinary cytology,” *Vib. Spectrosc.*, vol. 48, no. 1, pp. 101–106, 2008, doi:

10.1016/j.vibspec.2008.03.006.

[147] J. M. Schubert *et al.*, “Spectral Cytopathology of Cervical Samples: Detecting Cellular Abnormalities in Cytologically Normal Cells,” *Lab. Invest.*, vol. 90, no. 7, p. 1068, Jul. 2010, doi: 10.1038/LABINVEST.2010.72.

[148] D. Mayerich, M. J. Walsh, A. Kadjacsy-Balla, P. S. Ray, S. M. Hewitt, and R. Bhargava, “Stain-less staining for computed histopathology.,” *Technology*, vol. 3, no. 1, pp. 27–31, Mar. 2015, doi: 10.1142/S2339547815200010.

[149] D. Naumann, D. Helm, and H. Labischinski, “Microbiological characterizations by FT-IR spectroscopy,” *Nature*, vol. 351, no. 6321, pp. 81–82, May 1991, doi: 10.1038/351081a0.

[150] Y. C. Shen, A. G. Davies, E. H. Linfield, P. F. Taday, D. D. Arnone, and T. S. Elsey, “Determination of Glucose Concentration in Whole Blood using FTIR Spectroscopy,” *J. Biol. Phys.*, vol. 29, pp. 129–133, 2003.

[151] F. Elmi, A. Fayyaz, M. Mitra, H. Alinezhad, and N. Nikbakhsh, “Application of FT-IR spectroscopy on breast cancer serum analysis,” *Spectrochim. Acta - Mol. Biomol. Spectrosc.*, vol. 187, no. A, pp. 87–91, 2017, doi: 10.1016/j.saa.2017.06.021.

[152] E. Barlev, U. Zelig, O. Bar, C. Segev, and S. Mordechai, “A novel method for screening colorectal cancer by infrared spectroscopy of peripheral blood mononuclear cells and plasma,” *J. Gastroenterol.*, vol. 51, no. 3, pp. 214–221, 2016, doi: 10.1007/s00535-015-1095-7.

[153] D. A. Scott, D. E. Renaud, S. Krishnasamy, P. Meriç, N. Buduneli, and Ş. Çetinkalp, “Diabetes-related molecular signatures in infrared spectra of human saliva,” *Diabet. Metab. Syndr.*, vol. 2, no. 48, pp. 1–9, 2010.

[154] P. D. Lewis *et al.*, “Evaluation of FTIR Spectroscopy as a diagnostic tool for lung cancer using sputum,” *BMC Cancer*, vol. 10, no. 640, pp. 1–10, 2010, doi: 10.1186/1471-2407-10-640.

[155] K. Liu, T. C. Dembinski, and H. H. Mantsch, “Rapid determination of fetal lung maturity from infrared spectra of amniotic fluid,” *Am J Obs. Gynecol*, vol. 178, no. 2, pp. 234–241, 1998.

[156] D. Yonar, L. Ocek, B. I. Tiftikcioglu, Y. Zorlu, and F. Severcan, “Relapsing-Remitting Multiple Sclerosis diagnosis from cerebrospinal fluids via Fourier transform infrared spectroscopy coupled with multivariate analysis,” *Sci. Rep.*, vol. 8, no. 1025, pp. 1–13, 2018, doi: 10.1038/s41598-018-19303-3.

[157] K. V Oliver, A. Vilasi, A. Maréchal, S. H. Moochhala, R. J. Unwin, and P. R. Rich, “Infrared vibrational spectroscopy : a rapid and novel diagnostic and monitoring tool for cystinuria,” *Nat. Publ. Gr.*, vol. 6, no. 34737, pp. 1–7, 2016, doi: 10.1038/srep34737.

[158] A. Travo, C. Paya, G. Déléris, J. Colin, B. Mortemousque, and I. Forfar, “Potential of FTIR spectroscopy for analysis of tears for diagnosis purposes,” *Anal. Bioanal. Chem.*, vol. 406, no. 9–10, pp. 2367–2376, 2014, doi:

10.1007/s00216-013-7607-5.

[159] B. C. Wilson, M. Jermyn, and F. Leblond, “Challenges and opportunities in clinical translation of biomedical optical spectroscopy and imaging,” *J. Biomed. Opt.*, vol. 23, no. 3, pp. 1–13, 2018, doi: 10.1117/1.JBO.23.3.030901.

[160] M. Diem, A. Ergin, S. Remiszewski, and X. Mu, “A Reagent-Free , High Resolution Lung Cancer Diagnostic Method Based on Phenotypic Infrared Spectral Imaging,” *J. Thorac. Oncol.*, vol. 11S, no. 11, p. S288, 2016, doi: 10.1016/j.jtho.2016.09.066.

[161] O. Theisen, K. Gerwert, S. Remiszewski, C. M. Thompson, and M. Diem, “A method for the comparison of multi-platform spectral histopathology (SHP) data sets,” *Analyst*, vol. 140, pp. 2465–2472, 2015, doi: 10.1039/c4an01879f.

[162] S. C. Whiteman, Y. Yang, J. M. Jones, and M. A. Spiteri, “FTIR spectroscopic analysis of sputum : Preliminary findings on a potential novel diagnostic marker for COPD,” *Ther. Adv. Respir. Dis.*, vol. 2, no. 1, pp. 23–31, 2008, doi: 10.1177/1753465807087972.

[163] N. Patel *et al.*, “FTIR spectroscopic profiling of COPD: Identification of distinct spectral signatures and correlation to COPD status,” *Thorax*, vol. 65, no. Suppl 4, pp. A124–A125, 2010, doi: 10.1136/thx.2010.150987.12.

[164] B. H. Weigl *et al.*, “The Value of Clinical Needs Assessments for Point-of-Care Diagnostics,” *Point Care.*, vol. 23, no. 1, pp. 1–7, 2012, doi: 10.1038/jid.2014.371.

[165] Institute for Clinical and Economical Review, “Guide to Understanding Health Technology Assessment (HTA),” 2018.

[166] A. Dang, N. Likhar, and U. Alok, “Importance of Economic Evaluation in Health Care: An Indian Perspective,” *Value Heal. Reg. Issues*, vol. 9, no. 6, pp. 78–83, 2016, doi: 10.1016/j.vhri.2015.11.005.

[167] European Commision, “Regulation (EU) 2017/746 of the European Parliament and of the council of 5 April 2017 on in vitro diagnostic medical devices and repealing Directive 98/79/EC and Commission Decision 2010/227/EU,” *Official Journal of the European Union*. 2017.

[168] Food and Drugs Administration, “Overview of IVD Regulation.” https://www.fda.gov/medical-devices/ivd-regulatory-assistance/overview-ivd-regulation (accessed Nov. 07, 2020).

[169] M. Reinikainen and M. Suppo, “BSI: Explaining IVD classification issues.” pp. 1–24.

[170] Medical Device Innovation Consortium, “Developing Clinical Evidence for Regulatory and Coverage Assessments in In Vitro Diagnostics (IVDs).” pp. 1–45, 2019.

[171] L. Ferrante Di Ruffano, C. J. Hyde, K. J. McCaffery, P. M. M. Bossuyt, and J. J. Deeks, “Assessing the value of diagnostic tests: A framework for designing and evaluating trials,” *Br. Med. J.*, vol. 344, pp. 18–22, 2012, doi:

10.1136/bmj.e686.

[172] S. Ria *et al.*, “Applying for regulatory approval of a clinical trial of a medical device in the UK – a practical guide,” *Br. Dent. J.*, vol. 225, no. 11, pp. 1033–1036, 2018, doi: 10.1038/sj.bdj.2018.1033.

[173] World Health Organization., *Handbook for good clinical research practice (GCP): guidance for implementation*. 2002.

[174] A. W and M. Al-Sayed, “Human subjects in clinical trials: Ethical considerations and concerns,” *J. Transl. Sci.*, vol. 4, no. 6, pp. 1–5, 2018, doi: 10.15761/jts.1000239.

[175] E. Gray *et al.*, “Health economic evaluation of a serum-based blood test for brain tumour diagnosis: Exploration of two clinical scenarios,” *BMJ Open*, vol. 8, no. 5, pp. 1–9, 2018, doi: 10.1136/bmjopen-2017-017593.

[176] H. J. Butler *et al.*, “Development of high-throughput ATR-FTIR technology for rapid triage of brain cancer,” *Nat. Commun.*, vol. 10, no. 1, pp. 1–9, 2019, doi: 10.1038/s41467-019-12527-5.

[177] L. Vaccari and F. D’Amico, “Infrared Spectroscopy Basic Concepts on Theory and Instrumentation.” https://www.ceric-eric.eu/wp-content/uploads/2018/07/PaGES4_EVENT1_Vaccari-D-Amico.pdf.

[178] M. Koç and E. Karabudak, “History of spectroscopy and modern micromachined disposable Si ATR-IR spectroscopy,” *Appl. Spectrosc. Rev.*, vol. 53, no. 5, pp. 420–438, 2018, doi: 10.1080/05704928.2017.1366341.

[179] B. Vasilev, H. Schumacher, U. Ku, and J. W. Bartha, “Applications of Microstructured Silicon Wafers as Internal Reflection Elements in Attenuated Total Reflection Fourier Transform Infrared Spectroscopy,” *Appl. Spectrosc.*, vol. 64, no. 9, pp. 1022–1027, 2010.

[180] Clinical and Laboratory Standards Institute, *Collection of Diagnostic Venous Blood Specimens - 7th Edition*. 2017.

[181] W. Y. Ng and C. P. Yeo, “Thrombin-accelerated quick clotting serum tubes: An evaluation with 22 common biochemical analytes,” *Adv. Hematol.*, vol. 2013, pp. 1–8, 2013, doi: 10.1155/2013/769479.

[182] G. Lima-Oliveira *et al.*, “Elimination of the venous stasis error for routine coagulation testing by transillumination,” *Clin. Chim. Acta*, vol. 412, no. 15–16, pp. 1482–1484, 2011, doi: 10.1016/j.cca.2011.04.008.

[183] G. Lima-Oliveira *et al.*, “Transillumination: A new tool to eliminate the impact of venous stasis during the procedure for the collection of diagnostic blood specimens for routine haematological testing,” *Int. J. Lab. Hematol.*, vol. 33, no. 5, pp. 457–462, 2011, doi: 10.1111/j.1751-553X.2011.01305.x.

[184] P. K. Garg, “Chapter 2C: Pathophysiology of systemic inflammatory response syndrome and multiorgan dysfunction syndrome in acute pancreatitis,” in *Pancreatitis*, 2017, pp. 29–37.

[185] J. M. Lord *et al.*, “The systemic immune response to trauma: an overview of pathophysiology and treatment,” *Physiol. Behav.*, vol. 176, no. 1, pp. 139–148, 2018, doi: 10.1016/S0140-6736(14)60687-5.The.

[186] J. Dunnick, B. Herman, and J. A. Rose, *Pediatric emergencies presenting to urgent care centers*. Elsevier Inc., 2017.

[187] R. K. Chakraborty and B. Burns, *Systemic Inflammatory Response Syndrome*. StatPearls Publishing, 2020.

[188] G. S. Martin, “Sepsis, severe sepsis and septic shock: changes in incidence, pathogens and outcomes,” *Expert Rev Anti Infect Ther*, vol. 6, no. 10, pp. 701–706, 2012, doi: 10.1586/eri.12.50.Sepsis.

[189] S. Y. Cho and J. H. Choi, “Biomarkers of Sepsis,” *Crit. Rev. Clin. Lab. Sci.*, vol. 50, no. 1, pp. 23–36, 2013, doi: 10.3947/ic.2014.46.1.1.

[190] R. L. Gauer, W. Army, F. Bragg, and N. Carolina, “Early Recognition and Management of Sepsis in Adults,” *Fisrt Six Hours*, vol. 88, no. July, pp. 44–52, 2013.

[191] H. Il Kim and S. Park, “Sepsis: Early recognition and optimized treatment,” *Tuberc. Respir. Dis. (Seoul).*, vol. 82, no. 1, pp. 6–14, 2018, doi: 10.4046/trd.2018.0041.

[192] K. A. Simonsen, A. L. Anderson-Berry, S. F. Delair, and H. Dele Davies, “Early-onset neonatal sepsis,” *Clin. Microbiol. Rev.*, vol. 27, no. 1, pp. 21–47, 2014, doi: 10.1128/CMR.00031-13.

[193] J. H. Pek *et al.*, “Neurocognitive impairment after neonatal sepsis: Protocol for a systematic review and meta-analysis,” *BMJ Open*, vol. 10, no. 6, pp. 1–5, 2020, doi: 10.1136/bmjopen-2020-038816.

[194] J. S. T. Fung, S. Akech, N. Kissoon, M. O. Wiens, M. English, and J. Mark Ansermino, “Determining predictors of sepsis at triage among children under 5 years of age in resource-limited settings: A modified Delphi process,” *PLoS One*, vol. 14, no. 1, pp. 1–14, 2019, doi: 10.1371/journal.pone.0211274.

[195] K. N. Iskander *et al.*, “Sepsis: Multiple abnormalities, heterogeneous responses, and evolving understanding,” *Physiol. Rev.*, vol. 93, no. 3, pp. 1247–1288, 2013, doi: 10.1152/physrev.00037.2012.

[196] NICE, *Sepsis: recognition, assessment and early management*, no. 51. 2016.

[197] N. Hex, J. Retzler, C. Bartlett, and M. Arber, “The Cost of Sepsis Care in the UK,” *York Heal. Econ. Consort.*, no. February, pp. 1–46, 2017.

[198] J. Cooper, “Healthcare expenditure, UK Health Accounts - Office for National Statistics,” *UK Government Health Accounts*, 2019. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthcaresystem/bulletins/ukhealthaccounts/2017 (accessed Oct. 10, 2020).

[199] National Institute for Care Excellence, “Putting NICE guidance into practice. Resource impact report: Sepsis: the recognition, diagnosis and early

management," 2016.

[200] K. E. Rudd *et al.*, "Global, regional, and national sepsis incidence and mortality, 1990–2017: analysis for the Global Burden of Disease Study," *Lancet*, vol. 395, no. 10219, pp. 200–211, 2020, doi: 10.1016/S0140-6736(19)32989-7.

[201] S. K. Nachimuthu and P. J. Haug, "Early detection of sepsis in the emergency department using Dynamic Bayesian Networks.," *AMIA Annu. Symp. Proc.*, vol. v.2012, pp. 653–662, 2012.

[202] P. Worapratya and P. Wuthisuthimethawee, "Septic shock in the ER: Diagnostic and management challenges," *Open Access Emerg. Med.*, vol. 11, pp. 77–86, 2019, doi: 10.2147/OAEM.S166086.

[203] A. Leligdowicz and M. A. Matthay, "Heterogeneity in sepsis: New biological evidence with clinical applications," *Crit. Care*, vol. 23, no. 1, 2019, doi: 10.1186/s13054-019-2372-2.

[204] N. De Prost, K. Razazi, and C. Brun-Buisson, "Unrevealing culture-negative severe sepsis," *Crit. Care*, vol. 17, no. 5, p. 1, 2013, doi: 10.1186/cc13022.

[205] T. E. Sweeney, O. Liesenfeld, and L. May, "Diagnosis of bacterial sepsis: why are tests for bacteremia not sufficient?," *Expert Rev. Mol. Diagn.*, vol. 19, no. 11, pp. 959–962, 2019, doi: 10.1080/14737159.2019.1660644.

[206] F. Gül, M. K. Arslantaş, İ. Cinel, and A. Kumar, "Changing definitions of sepsis," *Turk Anesteziyoloji ve Reanimasyon Dern. Derg.*, vol. 45, no. 3, pp. 129–138, 2017, doi: 10.5152/TJAR.2017.93753.

[207] A. C. Heffner, J. M. Horton, M. R. Marchick, and A. E. Jones, "Etiology of Illness in Patients with Severe Sepsis Admitted to the Hospital from the Emergency Department," *Clin. Infect. Dis.*, vol. 6, no. 50, pp. 814–820, 2010, doi: 10.1161/CIRCULATIONAHA.110.956839.

[208] J. Wise, "Sepsis should be treated within one hour, says NICE," *BMJ*, vol. 356, p. j1257, 2017, doi: 10.1136/bmj.j1257.

[209] H. B. Nguyen, W. Sen Kuan, M. Batech, P. Shrikhande, M. Mahadevan, and C.-H. Li, "Outcome effectiveness of the severe sepsis resuscitation bundle with addition of lactate clearance as a bundle item: a multi-national evaluation," *Crit. Care*, vol. 15, no. R229, pp. 1–10, 2011, doi: 10.1186/cc10469.

[210] R. Daniels, T. Nutbeam, G. McNamara, and C. Galvin, "The sepsis six and the severe sepsis resuscitation bundle: A prospective observational cohort study," *Emerg. Med. J.*, vol. 28, no. 6, pp. 507–512, 2011, doi: 10.1136/emj.2010.095067.

[211] The UK sepsis trust, "A Toolkit for Emergency Departments," 2014.

[212] P. Kumar, M. Jordan, J. Caesar, and S. Miller, "Improving the management of sepsis in a district general hospital by implementing the 'Sepsis Six' recommendations," *BMJ Qual. Improv. Reports*, vol. 4, no. 1, pp. 1–6, 2015, doi: 10.1136/bmjquality.u207871.w4032.

[213] C. Pierrakos and J. L. Vincent, “Sepsis biomarkers: A review,” *Crit. Care*, vol. 14, no. 1, pp. 1–18, 2010, doi: 10.1186/cc8872.

[214] B. M. Tang, G. D. Eslick, J. C. Craig, and A. S. McLean, “Accuracy of procalcitonin for sepsis diagnosis in critically ill patients: systematic review and meta-analysis,” *Lancet Infectious Diseases*, vol. 7, no. 3. Lancet Infect Dis, pp. 210–217, Mar. 2007, doi: 10.1016/S1473-3099(07)70052-X.

[215] P. Póvoa *et al.*, “C-reactive protein as an indicator of sepsis,” *Intensive Care Med.*, vol. 24, no. 10, pp. 1052–1056, 1998, doi: 10.1007/s001340050715.

[216] B. Clyne and J. S. Olshaker, “The C-reactive protein,” *J. Emerg. Med.*, vol. 17, no. 6, pp. 1019–1025, 1999, doi: 10.1016/S0736-4679(99)00135-3.

[217] T. Calandra *et al.*, “Prognostic values of tumor necrosis factor/cachectin, interleukin-1, interferon-α, and interferon-γ in the serum of patients with septic shock,” *J. Infect. Dis.*, vol. 161, no. 5, pp. 982–987, 1990, doi: 10.1093/infdis/161.5.982.

[218] R. F *et al.*, “High tumor necrosis factor serum level is associated with increased survival in patients with abdominal septic shock: A prospective study in 59 patients,” *Surgery*, vol. 120, no. 5, pp. 801–807, 1996.

[219] R. T. Patel, K. I. Deen, D. Youngs, J. Warwick, and M. R. B. Keighley, “Interleukin 6 is a prognostic indicator of outcome in severe intra-abdominal sepsis,” *Br. J. Surg.*, vol. 81, no. 9, pp. 1306–1308, 1994, doi: 10.1002/bjs.1800810914.

[220] R. S. N. Panday, E. M. J. Lammers, N. Alam, and P. W. B. Nanayakkara, “An overview of positive cultures and clinical outcomes in septic patients: A sub-analysis of the Prehospital Antibiotics Against Sepsis (PHANTASi) trial,” *Crit. Care*, vol. 23, no. 1, pp. 1–9, 2019, doi: 10.1186/s13054-019-2431-8.

[221] NHS, “Acute pancreatitis - Symptoms,” 2018. https://www.nhs.uk/conditions/acute-pancreatitis/symptoms/ (accessed Oct. 11, 2020).

[222] R. Mitchell, M. Byrne, and J. Baille, “Pancreatitis,” *Lancet*, vol. 361, pp. 1447–1455, 2003, doi: 10.1136/bmj.296.6628.1065-b.

Chapter 2: Experimental Principles and Methodology

2.1. Vibrational Spectroscopy

Spectroscopy is defined as the study of the resultant spectra when matter interacts with or produces electromagnetic radiation (e.g., infrared light). Vibrational spectroscopy in particular is the non-destructive interrogation of molecular bonding environments for qualitative and quantitative analysis. It is separated into two broad categories: Raman and Infrared (IR) spectroscopy. These techniques utilise different mechanisms to elicit different spectral responses, but both rely on the principle that all molecules continuously vibrate at frequencies unique to the composition of the molecule. IR spectroscopy detects light absorption due to changes in dipole moment and Raman spectroscopy detects shifts in incident beam frequency due to inelastic scattering of light following bond polarization [1]. Generally, the techniques are considered complementary because some molecular vibrations are IR active while others are Raman active. IR spectroscopy is an absorbance based technique and is the technique under development and investigation within the scope of the studies presented throughout.

2.1.1. Absorbance

When a beam of light transmits through an absorbing medium the emerging beam is of lower intensity than the incident beam. The pathlength of the beam through the material, l (length of material the beam needs to travel through), the molar absorptivity of the material, ε, and the concentration of the absorbing material, c, all increase absorbance of the beam's energy into the medium and thus reduces beam

intensity. The Beer-Lambert law (Equation 2.1) relates these properties to the attenuation of the propagating beam [2].

$$A = \varepsilon c l$$

Equation 2.1. *Beer-Lambert law.*

If the intensity of the incident beam, I_0, and intensity of transmitted beam, I, are known, the absorbance, A, and similarly the transmittance, T, of the beam can be calculated (Equation 2.2). Figure 2.1 shows a graphical representation of the Beer-Lambert law.

$$A = \log\left(\frac{I_0}{I}\right) = log\left(\frac{1}{T}\right)$$

Equation 2.2. *Intensity, absorbance, and transmittance relationship.*

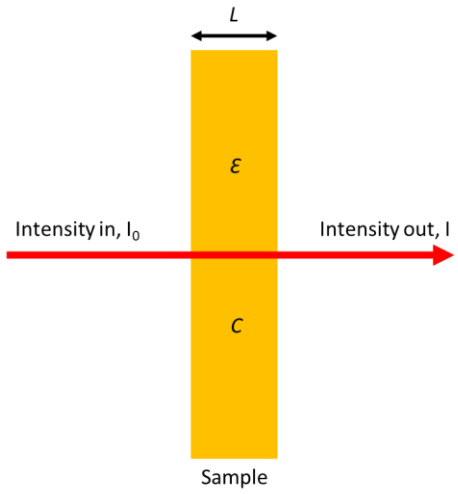


Figure 2.1. *Illustration of the Beer-Lambert law.*

Absorbance spectroscopy techniques operate by first acquiring a reference scan (also called the background scan) using the exact same experimental set-up, but with a blank sample cartridge, then acquiring the sample scan. This captures all the excess spectral information gained from the environment, such as from molecules in the air,

which is then removed from the sample scan. In this way the spectral profile of the sample can be isolated from any environmental factors. The reference scan can be considered the incident beam, I_0, and the sample can be considered the transmitted beam, I, of Equation 2.3., hence:

$$A = -log\left(\frac{sample}{reference}\right)$$

Equation 2.3. *Absorbance, sample spectrum, and reference spectrum relationship.*

These relationships are widely used to ascertain the concentration of an unknown material in a solution by probing the sample with a single wavelength of light [2]. Spectroscopic approaches scan a sample across a range of wavelengths. The output is therefore a distribution of absorbance values across a range of discrete wavelength intervals. Vibrational spectroscopy is slightly different in that it is common practice to represent data using wavenumbers, λ^{-1} (cm^{-1}), which is the reciprocal of wavelength λ (nm). This is because wavenumber is directly proportional to frequency according to the wave equation (Equation 2.4). Figure 2.2 shows a simple IR absorbance spectrum of water along with the bond vibrations responsible for each absorbance peak. Water is capable of three different vibrational modes because it is a nonsymmetric triatomic molecule (H_2O). The absorption peaks are often referred to as bands because they do not cover a singular point, but rather are spread over a small range of wavenumbers. This is because absorption bands are typically a combination of smaller peaks superimposed on each other to form a larger absorption band [3].

$$v = c\lambda^{-1}$$

Equation 2.4. *Wave equation. v = frequency, c = speed of light, λ^{-1} = wavenumber.*

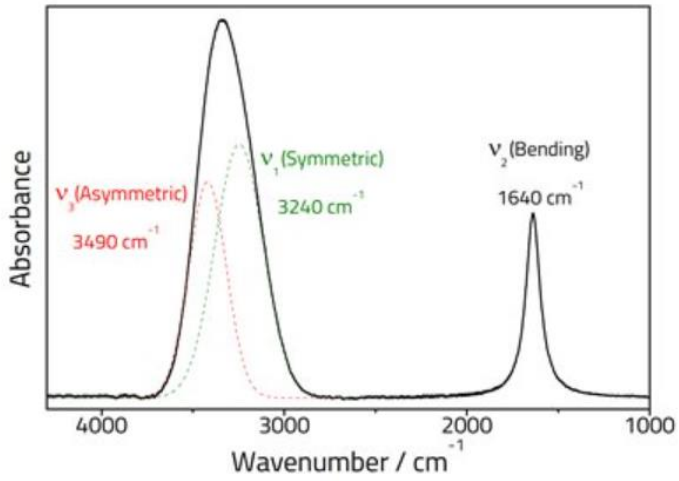


Figure 2.2. *Simple IR absorbance spectrum of water. Asymmetric and symmetric absorption bands superimpose to form a broader peak at roughly 3300cm^{-1}. Image credit:* [4]

2.1.2. Molecular Vibrations

2.1.2.1. Degrees of Freedom and Dipole Moments

The atoms in a molecule are continuously moving relative to each other such that the centre of mass of the molecule remains fixed. All molecules can move in three basic ways: translation, rotation, and vibration. There are three translational directions along the x, y, and z axis and three rotational directions around the x, y, and z axis in three-dimensional space. All linear molecules, for example carbon dioxide (CO_2), have a rotational movement around one axis which does not result in relative motion between atoms to each other and thus cannot be observed. Vibrational movements, called vibrational modes, are a little more complicated and describe how each atom within a molecule moves relative to other atoms in the molecule. There are six normal vibrational modes: stretching, asymmetric stretching, wagging, twisting, scissoring, and rocking. These can be further categorised into stretching, ν, in-plane bending, δ,

out-of-plane bending, γ, vibrational modes (Figure 2.3). The summation of the possible ways in which a molecule can move accounts for the molecules degrees of freedom (DF). The degrees of freedom in a molecule is 3N where N is the number of atoms in the molecule. Three degrees of freedom are translational (x, y, z) and three rotational (xy, xz, yz) which are subtracted from the total degrees of freedom to find the number of vibrational modes. From this the number of vibrational modes can be calculated as 3N-6, or 3N-5 for linear molecules because one of the rotational modes in linear molecules cannot be observed. For example, earlier water was shown to have three vibrational modes (Figure 2.2) because it is non-symmetric and contains only three atoms.

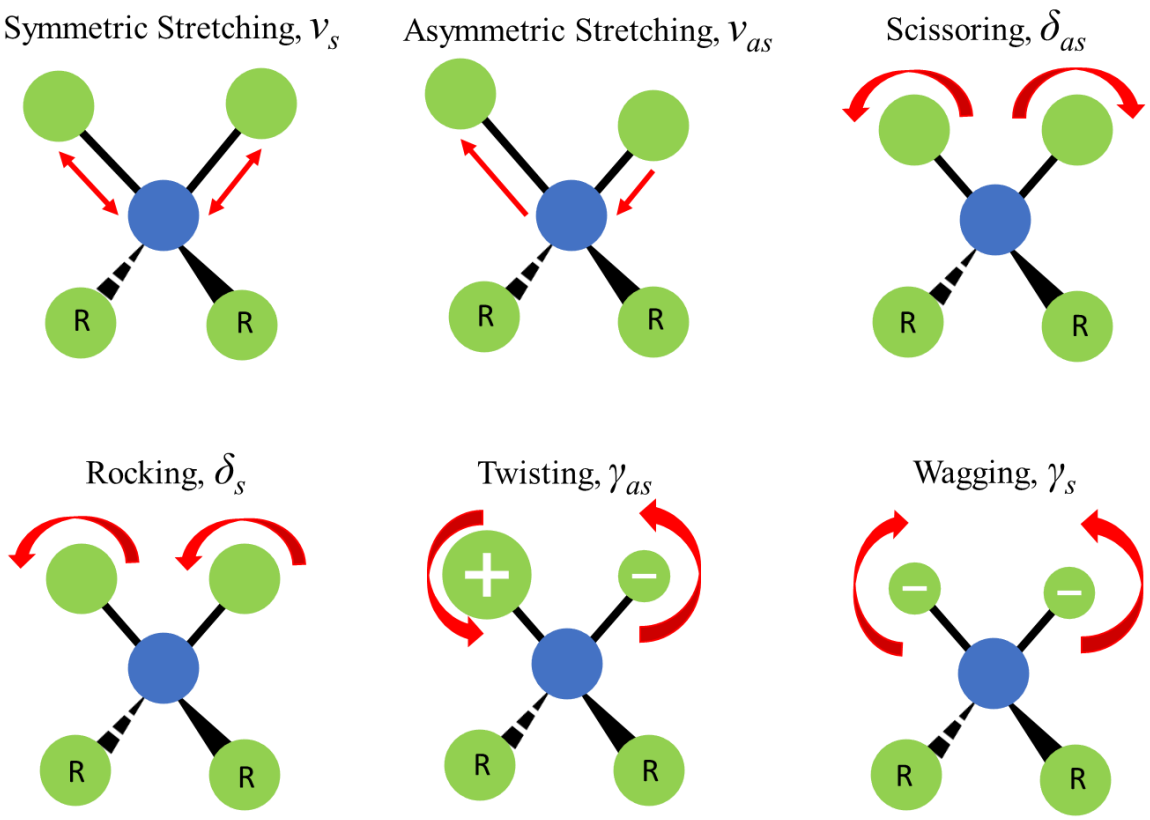


Figure 2.3. *Illustrations of the movements observed in the six normal vibrational modes of molecular bonds. R indicates an unspecified number of atoms in a chain. +ve and -ve symbols represent movement out of and into the page, respectively. S = symmetric and as = asymmetric.*

A Dipole moment exists in a shared bond where the atoms in the bond possess a different electronegative charge. Heteronuclear molecules, such as hydrogen chloride (HCl), exhibit dipole moments and are thus polar, but homonuclear molecules, such as oxygen (O_2), do not exhibit dipole moments and are thus non-polar. When molecules vibrate the dipole moment fluctuates which creates an electric field that interacts with the electric field of light photons. If the frequency of light corresponds to the frequency of vibration, energy from the photon is absorbed into the bond and the bond vibrates with a greater amplitude. IR absorption can only occur if there is a net change in the dipole moment of the molecule during the vibration [5]. For instance, CO_2 has polar C=O bonds, but because the molecule is symmetrical the dipoles cancel each other out during the symmetric stretching vibrational mode leaving a net dipole moment change of zero. If the bonds vibrate in the symmetric stretch vibrational mode, then the electric fields generated by each bond also cancel out and there is no net change in dipole moment. However, a net change in dipole moment will occur if the molecule vibrates in the asymmetric stretch and scissor bend vibrational modes. In these modes an electric field is generated which can interact with light photons. Frequency matches generally occur in the IR frequency range of electromagnetic radiation. Hence, the symmetric stretch mode of CO_2 is said to be IR inactive because it cannot absorb IR radiation, while asymmetric scissor bending modes are said to be IR active because they can absorb IR radiation.

2.1.2.2. Vibrational Frequency and Energy

Molecular vibrational modes are analogous to the mechanical behaviour of two masses adjoined by a spring. Consider a simple diatomic molecule. One atom is fixed to a wall and the other is oscillating continuously with equilibrium bond length, r

(Figure 2.4). The bond (spring) is continuously stretching and contracting in a motion described as a harmonic oscillator due to the attractive forces of the atomic charges and the repulsive forces of the atomic nuclei.

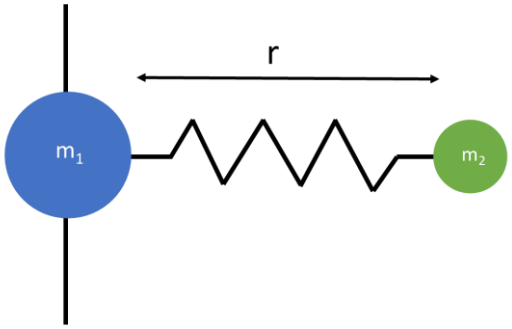


Figure 2.4. *Schematic of a mass-spring model of a molecular bond. m_1 and m_2 represent the mass of two different atoms.*

A relationship between the potential energy of the spring and the frequency can be derived by borrowing relationships from classical physics to arrive at an approximation of bond behaviour. Hooke's law (Equation 2.5) provides a relationship between the force applied to a spring and the distance the spring is stretched. Newton's second law (Equation 2.6) provides a relationship between the mass of an object and acceleration. Since acceleration is a time dependant variable it follows that frequency of oscillation can be derived from the system. Reducing the differential equations (Equation 2.7) gives a relationship between frequency of oscillation, ν, mass of the atoms, m, and the spring constant (bond strength), k (Equation 2.8) [6], [7].

$$F = -kx \qquad F = ma = m\frac{d^2x}{dt^2}$$

Equation 2.5. *Hooke's Law* ***Equation 2.6.*** *Newton's 2nd law*

$$-kx = m\frac{d^2x}{dt^2}$$

Equation 2.7. *Second order linear differential equation*

$$v = \frac{1}{2\pi}\sqrt{\frac{k}{m}}$$

Equation 2.8. *Frequency of oscillation*

Variable assignments: *F = force, k = spring constant, x = distance, m = mass, a = acceleration, v = frequency.*

The harmonic oscillator can be presented graphically as a parabola of energy varying as a function of bond length (Figure 2.5). The equilibrium bond length, r, between the atoms will always remain the same, however the amplitude of the oscillation can change. The Heisenberg uncertainty principle states that if the location of a particle is known its momentum cannot be known and vice versa [1]. Thus, atoms are in constant motion since stopping would imply that its position can be known which violates the Heisenberg uncertainty principle. Until acted upon, the bond will remain oscillating in its ground state which is the lowest energy state and smallest bond length the bond can exist at. The ground state is also known as zero-point energy, E_0. Here the analogy of the mechanical spring and the reality of the molecular bond depart. Energy levels of a bond are quantised meaning bond energy can only assume discrete magnitudes of energy whereas a spring can oscillate through a continuous range of energy levels. Photons can excite bonds to transition from the ground state to higher energy levels, but the photon must be of equal energy to the energy difference otherwise the photon will simply transmit through the molecule. When a quantum of

incident photon is of equal energy to the bond excitation energy the photon energy will be absorbed which will evoke a transition from the ground state, E_0, to the first energy level, E_1, also known as the fundamental transition [8]. A photon of double energy will excite the bond energy to the second overtone, E_2, and triple energy to the third overtone, E_3, and so on.

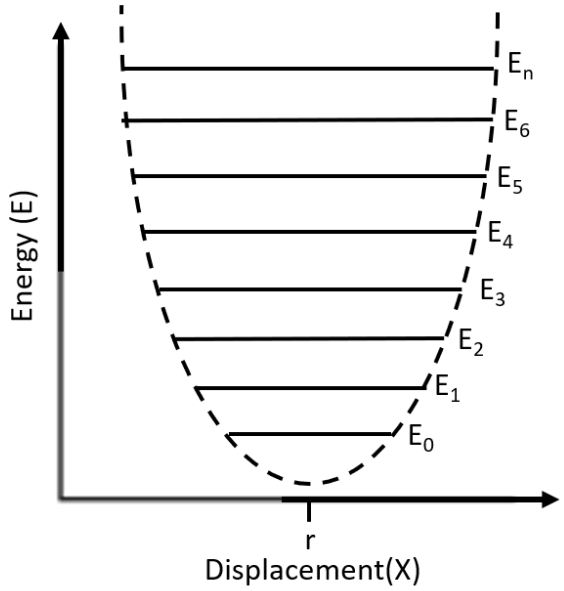


Figure 2.5. *The harmonic oscillator where bond displacement, X, is a function of energy, E. r indicates where the equilibrium bond length lies. Bonds can only assume discrete energy levels as indicated by the solid lines.*

There is a problem with the harmonic oscillator model in that it fails to consider what happens when too much energy is put into the system. The spring or bond would be stretched too far and break. Springs snap and bonds dissociate. An anharmonic oscillator is more appropriate to represent this (Figure 2.6). Again, the average intermolecular distance remains unchanged regardless of bond energy, but the oscillations are increasingly asymmetric as the bond energy increases. If enough energy is put into the system, the bond will break. Another effect is that the energy levels are not regularly separated like in the harmonic oscillation model. In contrast to

the harmonic oscillator model, the impact of this is that photon energy levels required to evoke the fundamental transition and successive overtones are not equally spaced.

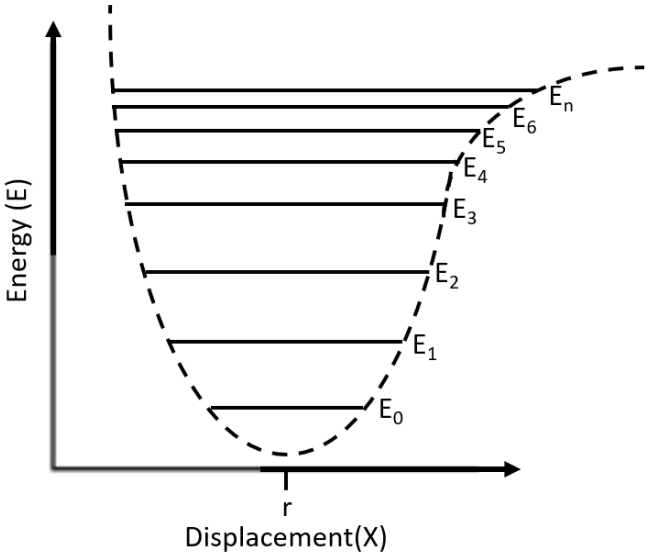


Figure 2.6. *The anharmonic oscillator where increasing bond energy levels tend towards bond dissociation. Energy levels the bond can assume are not spaced at even energy intervals. r represents the equilibrium bond length at every energy level.*

The absorption of photon energy is the observable effect exploited in IR spectroscopy. Photon energy and frequency are directly proportional according to Planck's relationship (Equation 2.9). This means precise frequencies of light are capable of evoking energy level transitions in molecular bonds. Such frequencies of light are said to be resonant with the molecular bond. Resonant frequency is defined as the frequency at which an input stimulus will elicit the greatest response, in this case excitation of a vibrational mode. Since frequency is inversely proportional to wavelength (Equation 2.10), it follows that wavelength is also inversely proportional to energy.

$$E = hv$$

Equation 2.9. *Planck's relationship*

$$v = \frac{c}{\lambda}$$

Equation 2.10. *Frequency wavelength relationship*

Variable Assignments: *E = energy, h = Planck's constant, v = frequency, c = speed of light, λ = wavelength.*

The IR spectrum is generally split into three different regions: far-IR (10-400cm^{-1}), mid-IR (400-4000cm^{-1}), and near-IR (4000-12500cm^{-1}). Typically, it is the mid-IR region that is used in biochemical analysis because fundamental transitions tend to occur in the mid-IR wavenumber range. Overtones are generally observed in the near-IR region and rotational energies are observed in the microwave region of the light spectrum [9].

2.2. Fourier Transform Infrared Spectroscopy

There are two primary methods used to perform IR spectroscopy: dispersive IR spectroscopy, and Fourier Transform IR (FTIR). Dispersive spectroscopy is an older technique and has been largely superseded by FTIR spectroscopy. The former scans the sample through individual IR frequencies while the latter scans over a wide range of frequencies simultaneously and then uses mathematical transforms to interpret the data. FTIR spectroscopy holds a few advantages over dispersion IR. The two main advantages are known as the Fellgett's advantage and Jacquinot's throughput advantage [10]. Fellgett's advantage states that multiplexed measurements, such as in FTIR spectroscopy, will observe an improvement to signal to noise ratio in comparison to direct measurements. The Jacquinot throughput advantage is due to the

improvement in beam throughput of FTIR spectrometers compared to dispersion spectrometers. Dispersion spectrometers require slits to produce monochromatic light which reduces light intensity whereas FTIR spectrometers direct polychromatic light through a circular aperture called a Jacquinot stop.

Fourier transformation is a mathematical process for transforming time domain data to the frequency domain and vice-versa [11]. It is a powerful tool used to decompose a signal into the frequencies that comprise the signal. The significance of this in the context of IR spectroscopy is that for a beam of known bandwidth and discretised frequencies, all the frequencies of light before and after the beam passes through a sample can be known and frequencies of light lost to the sample can be inferred. Hence, the absorbance spectrum of the sample can be determined. Equation 2.11 shows the mathematical expression of a Fourier transformation of a function in the time domain, f(t), to the frequency domain, F(ω):

$$F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-i\omega t}\, dt$$

Equation 2.11. *Fourier transform.*

2.2.1. FTIR Spectrometer Instrumentation and Modes

2.2.1.1. The Michelson Interferometer

FTIR spectrometers deliver the beam to the sample as waveforms known as interference fringes [12]. Two beams of matching frequency will interact with one another to varying degrees depending on their relative phase. Beams in phase with each other (θ = 0°, and θ = 360°) have all constituent wavelengths in sync and will

combine in constructive interference to produce a beam of doubled intensity. Out of phase beams ($\theta = 180°$) have all constituent wavelengths out of sync and will destructively interfere and extinguish each other. Phases in between will boost and reduce intensity to varying degrees through elimination and amplification of different wavelengths. The reason for generating an interference pattern is because an uninterfered (planar) wave contains all unconvoluted frequencies from a broadband IR source, whereas an interference fringe contains a distribution of all frequencies because every wavelength of light produces a characteristic interference pattern [12], [13].

In FTIR spectroscopy, interference patterns are made possible by the Michelson interferometer; the salient optical component of an FTIR spectrometer. The Michelson interferometer consists of a fixed mirror, a linearly translatable mirror, and a beam splitter (Figure 2.7). Light enters the interferometer from an IR light source, usually a globar light source made form a Silicon Carbide rod heated to 1000-1650°C. 50% of the light is transmitted through the non-absorbing beam splitter to the fixed mirror, and the remaining 50% is reflected to the moveable mirror. The moveable mirror is continuously moving back and forth which has the effect of extending and reducing the beam pathlength of the moveable mirror relative to the fixed mirror. When the two beams recombine at the beam splitter, the beams interact resulting in interference. The phase difference is dependent on the position of the moveable mirror relative to fixed mirror which is called the Optical Path Difference (OPD). OPD can be adjusted to modulate the spectral resolution of the scan. The interference pattern passes through the sample and onto a detector, typically a Mercury-Cadmium-Telluride (MCT) or Deuterated Triglyceride Sulfate (DTGS) detector. Measurement

of the interference pattern is called an interferogram where intensity is plotted as a function of OPD. A Fourier transformation is performed on the interferogram to produce a spectrum of absorbance (or transmittance) against wavenumber (Figure 2.8). The large peak in the middle of the interferogram is called the 'centre burst' and is where all wavelengths constructively interfere since OPD = 0.

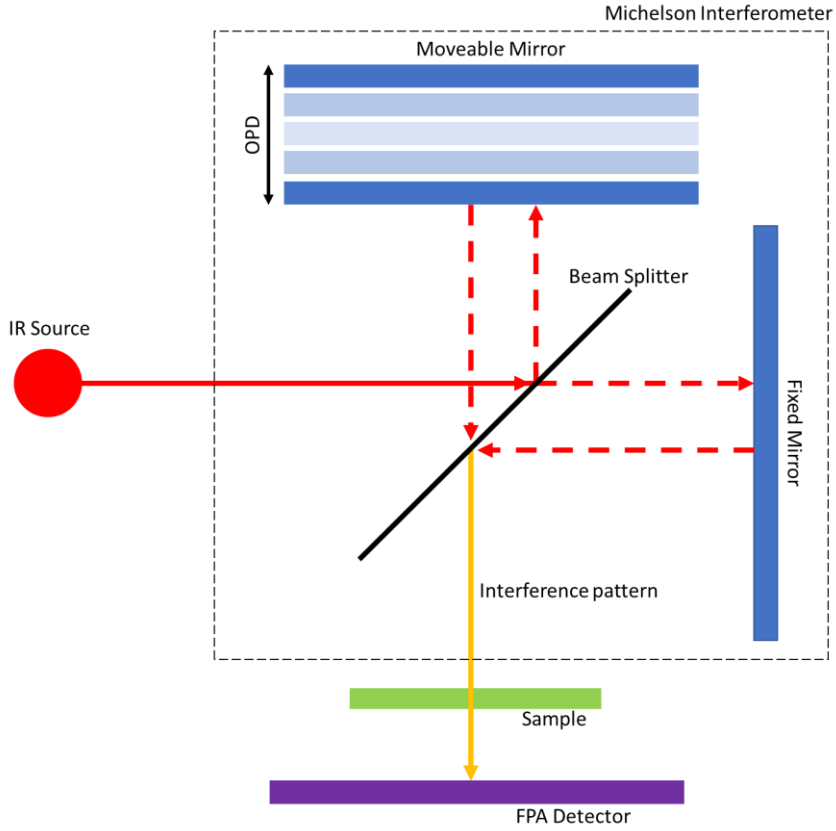


Figure 2.7. *Schematic of an IR spectrometer set-up including the Michelson Interferometer.*

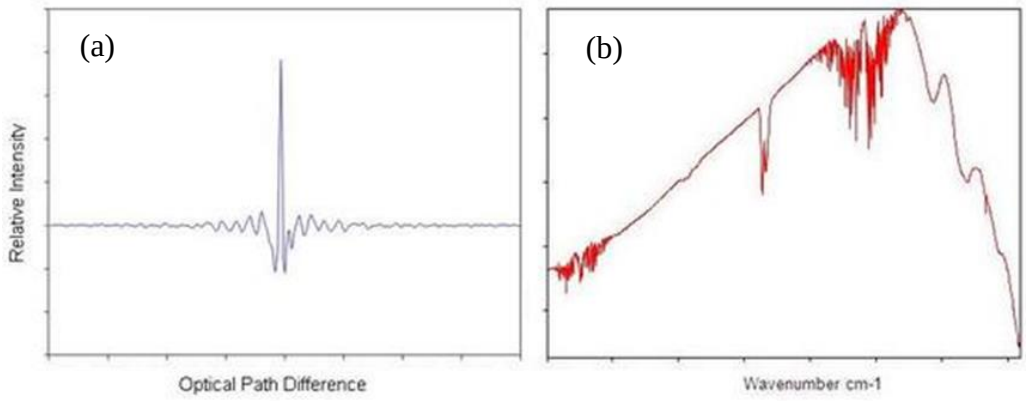


Figure 2.8. *Graphs of the IR signal detected by an IR spectrometer before and after Fourier transformation (a) an interferogram as recorded by the FPA detector. (b) an IR energy spectrum following Fourier transformation. Image credit:* [14].

2.2.1.2. Atmospheric Correction

A reference scan known as the background spectrum must be acquired prior to acquisition of sample spectra. This is achieved by simply scanning the sample compartment without a sample in it. Background spectra will contain spectral information pertaining to the humidity in the room, material of the optical components in the spectrometer, air composition and anything else that may interact with the IR beam. Essentially, the background spectrum corrects for environmental and experimental set-up factors unrelated to the sample under investigation. Following a sample scan, the sample interferogram is divided by the background interferogram and Fourier transformation is carried out to obtain a sample only spectrum (see Equation 2.3. from section 2.1.1). Shown previously in Figure 2.8.(b) is a typical background spectrum.

2.2.1.3. FTIR Spectrometer Sampling Modes

FTIR spectrometers may be configured in numerous ways to facilitate the analysis of different sample types. The three most common FTIR spectrometer modes are transmission, transflectance, and internal reflectance (such as Attenuated Total Reflectance (ATR)) mode. Most FTIR spectrometers are usually capable of employing all three modes through use of interchangeable sample compartment accessory units. Each have been presented as simplified illustrations in Figure 2.9.

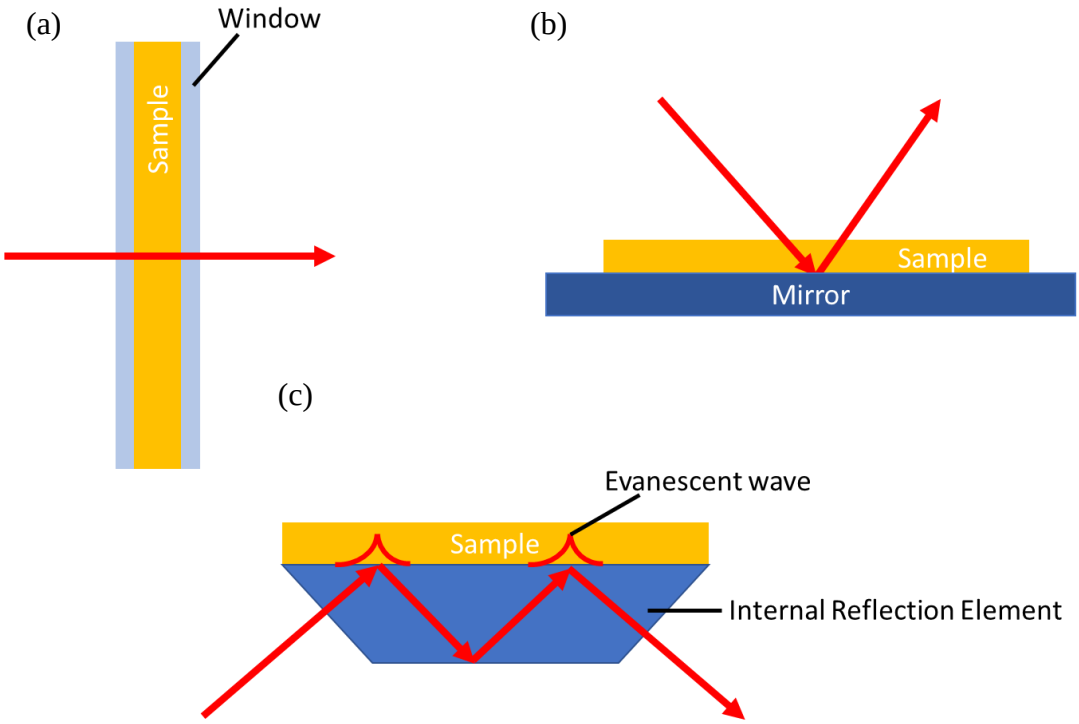


Figure 2.9. *Schematics of the three common FTIR spectrometer sampling modes: (a) transmission mode, (b) transflectance mode, and (c) attenuated total reflectance mode.*

Transmission mode quite simply transmits an IR beam directly through the sample; however, the sample usually needs to be supported by an IR transparent substrate such as sodium chloride (NaCl), barium fluoride (BaF_2) or calcium fluoride (CaF_2) windows. When using transflectance mode the sample is mounted upon an IR

reflective substrate and the beam is transmitted through the sample, reflected, then transmitted back out of the sample approximately doubling the beam pathlength through the sample in comparison to transmittance mode [15]. Sample preparation can present issues when using both techniques. Solid samples sometimes need to be pressed into KBr pellets which can be inconsistent and incur variance in the spectra. Aqueous samples may dissolve NaCl windows or contain air bubbles within sample windows. Conversely, sample preparation is straightforward when using ATR mode and does not experience any of these issues. Liquids can be deposited directly onto the ATR crystal as can solids although pressure may need to be applied to create good contact between the sample and the ATR crystal [16].

ATR mode takes a different approach to transmission and transflectance modes in the sense that the IR beam does not directly pass through the sample instead making use of the evanescent wave phenomena observed during Total Internal Reflection (TIR). TIR occurs when light from travelling through a medium reaches the boundary of another medium at such an angle that instead of exiting the medium the light reflects inwards (Figure 2.10). This will occur if the light incidence is below the critical angle, θ_c, which is determined by the refractive indices of the respective media (Equation 2.12). At the point of TIR an evanescent wave is produced that protrudes beyond the boundary of the substrate and into the sample, decaying exponentially to a penetration depth, d_p. The sample attenuates the evanescent wave by absorbing energy at resonant frequencies. Since the evanescent wave is the only component of the beam to interact with the sample, the effective pathlength in ATR is the penetration depth. The substrate is a prism called the Internal Reflection Element (IRE) and is made from a highly refractive material such as germanium (Ge), silicon carbide (SiC), diamond (Di), or

zinc selenide (ZnSe). Each material holds advantages and disadvantages and are generally selected based on application. For instance, Di is resistant to chemical attack so would be a good choice for highly corrosive substances and Ge has a high refractive index so is good for analysis of thin layers due to the low penetration depth. Penetration depth is dependent on the angle of incidence, θ_i, wavelength, λ, and refractive indices of the IRE, n_1, and of the sample, n_2 (Equation 2.13). Since refractive index of a material is dependent on wavelength, the critical angle and penetration depth will vary across the mid-IR spectrum [17]. The refractive index is a complex with both real and imaginary components where the real component holds when there is no absorption, and the imaginary component holds within an absorption band and is directly related to the extinction coefficient of the Beer-Lambert law [18]. Thus, in ATR anomalous dispersion of the refractive index will occur because the refractive index of the sample changes rapidly in the vicinity of an absorbance maxima [19]. This leads to shifts in peak position and band shape in comparison to the pure transmission spectrum of a substance and requires correction if ATR spectra is to be directly compared to transmission mode spectra [20].

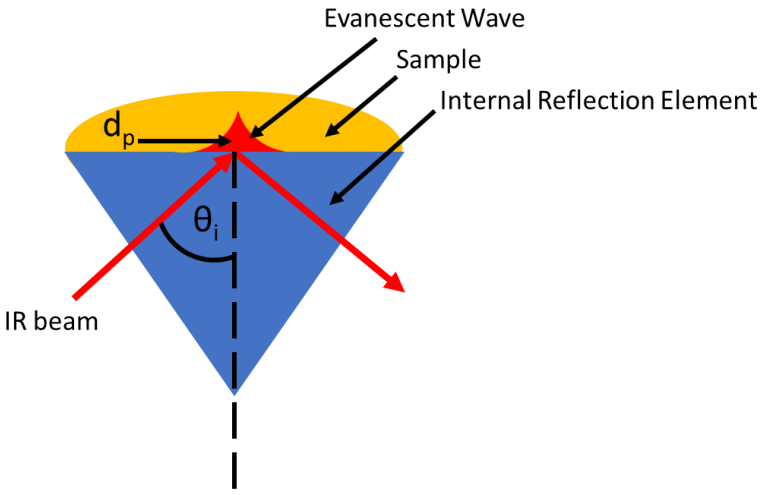


Figure 2.10. *Illustration of attenuated total reflection displaying total internal reflection and protrusion of an evanescent wave at a penetration depth, d_p, into a sample.*

$$\theta_c = \sin^{-1}\left(\frac{n_2}{n_1}\right)$$

Equation 2.12. *Critical angle formula. TIR will occur when $\theta_i < \theta_c$.*

$$d_p = \frac{\lambda}{2\pi n_1\sqrt{\sin^2\theta_i - \left(\frac{n_1}{n_2}\right)^2}}$$

Equation 2.13. *Evanescent wave penetration depth formula.*

Variable Assignments: *λ = wavelength, n_1 = refractive index of IRE, n_2 = refractive index of sample, θ_i = angle of incidence.*

IRE prisms may be designed to permit either single or multiple internal reflections. Multiple internal reflections allow for a single beam to probe a sample with multiple evanescent waves, thus increasing spectral absorbance and signal to noise ratio. However, extending beam pathlength through the IRE increases signal lost to the IREs own absorptivity. The extent to which this is a problem depends on the IRE material. For instance, Di is relatively IR transparent in the 900-1800cm^{-1} wavenumber region and is therefore the preferred choice for biological sample analysis by many spectroscopists [16]. It is, of course, prohibitively expensive, and not amenable to

high-throughput sample analysis systems. There does not exist a high-throughput system for ATR-FTIR spectroscopy. This is thought to be a major roadblock between IR spectroscopy and clinical translation [21], [22]. Impracticality of high-throughput ATR is owed to the high-cost and low manufacturability of IRE materials and the need to reliably clean the IRE between samples [23].

As discussed in chapter 1, ATR holds numerous advantages over transmission and transflectance modes. Generally, ATR-FTIR spectroscopy is more flexible, quicker, and reproducible with a higher signal-to-noise ratio [16], [17], [24], [25]. Broadly speaking, this is because there are fewer physical effects requiring correction or special sample preparation precautions to be concerned about. In transmission and transflection mode the IR beam must transmit through the bulk sample medium. Therefore, sample thickness and uniformity must be controlled with respect to the samples molar extinction coefficient [26]. ATR only requires that the sample forms intimate contact with the IRE and preferably with a minimum sample thickness no lower than the penetration depth of the evanescent wave across the IRE surface (~3µm) [16], [17], [21]. Thus, the difficulty experienced using transmission and transflectance modes to ensure repeatable sample thickness is mitigated [27]. This essentially reduces sample preparation requirements and allows better control over experimental parameters [28].

Transmission spectra are susceptible to a number of artefacts such as light scattering [29], thin film interference patterns [30], refraction and dispersion [31] that require computational correction. The IR windows used to confine samples in transmission mode can cause dispersion and refraction of the polychromatic IR light which results in a wavelength dependant focal length (i.e. chromatic aberration) [31].

Thin film interference patterns may arise from the need to sandwich a thin film of sample between thick IR window substrate can create multiple reflections within the film leading to optical effects such as interference fringes, oscillatory backgrounds, and peak shifts [30], [32]. Scattering is where light is deflected away from its trajectory preventing light reaching the IR detector which may be erroneously measured as absorbance into the sample medium. This can be a major problem for the study of non-homogenously distributed biological materials in transmission and transflection modes [21], [33], [34]. For instance, resonant Mie scattering occurs when an absorbing particle of a similar magnitude to the incident light (e.g. 7-8μm erythrocytes) resulting in broad oscillations across the baseline and distortion of position and intensity of absorbance bands [35]. This effect arises from the dependence of Mie scattering on refractive index combined with a wavelength dependent variation of the refractive index across particle absorption resonance [36]. The resulting anomalous dispersion artefact thus originates from the mixing of the real (reflective) and imaginary (absorptive) components of the refractive index [34].

Anomalous dispersion of the refractive index and/or scattering effects may arise for a number of reasons [16], [19], [37]. In transflection mode the resonant Mie scattering contributions to anomalous dispersion can be further complicated by reflection artefacts [38]. Moreover, in transflection mode the use of low-emissivity slides may influence spectra through an electric-field standing wave artefact potentially leading to spectral differences based solely on sample thickness and not biomolecular content [39]. By comparison, all of the above effects are either greatly reduced or non-existent in ATR FTIR spectroscopy as the sample is interrogated by an evanescent wave confined at the interface [17], [21], [24], [25], [33], [40].

2.2.2. Infrared Spectroscopy of Biofluid Samples

Investigation of biological materials in the field of IR spectroscopy has been extensive [29], [41]–[43] leading to a detailed understanding of the spectral responses observed from complex biological samples [44], [45]. Vast amounts of information exist pinpointing what absorbance peaks will be observed for specific molecular bonds. Since molecular vibrational frequencies are unique to the molecular bond, peak assignments can be inferred for a sample of unknown molecular composition. Peaks on a spectrum from a biological sample can be roughly identified according to the four major biomolecule categories: proteins, carbohydrates, lipids, and nucleic acids due to the functional groups that are characteristic of each category [46]. For instance, carbonyl (C=O) groups tend to absorb around 1740cm^{-1} through symmetric stretching vibrations (C=Oν_s) so the presence of a peak here would suggest high lipid content [7]. Similarly, a sample with proteins will present prominent peaks at 1650cm^{-1} and 1550cm^{-1} due to vibrations of the amide functional groups (O=C-N) found in proteins. Consequently, these peaks are named the amide I (1650cm^{-1}) and amide II (1550cm^{-1}) bands. Figure 2.11 shows a typical biological sample spectrum with absorbance peaks annotated with the major biomolecule categories, molecular bonds and vibrational modes associated to the peaks. The region between 900-1800cm^{-1} is known as the 'fingerprint' region as this region is particularly rich with absorbance peaks of biological samples and is therefore highly useful in discriminant analysis hence the name fingerprint.

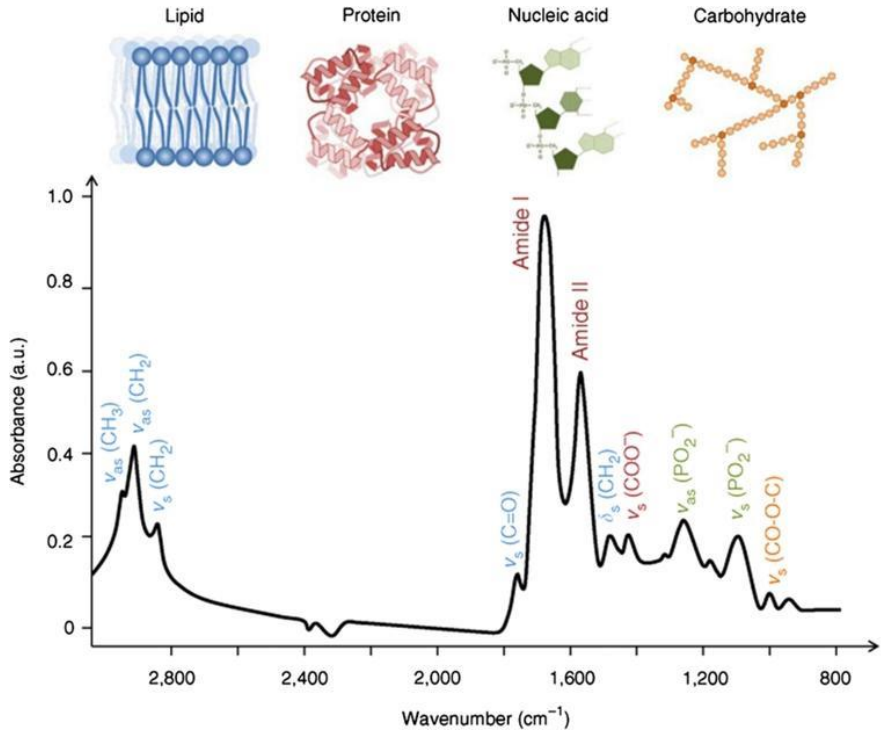


Figure 2.11. *Typical biological sample spectrum with significant peaks annotated and the major biological classes these peaks are associated with. Image credit:* [16].

Biofluids are of particular interest clinically because they are generally easy to acquire and rich in information. A major difficulty when using IR spectroscopy for the analysis of biofluid samples is that fresh samples tend to have a high water content. Water is a strong absorber of IR light and will obscure other absorbance peaks present in the spectrum [23]. In transmission mode the IR beam must travel through the entire sample thickness and in transflection mode the effective pathlength is approximately twice that of transmission leading the spectra to be dominated by water absorptions. The sample thickness could be reduced to 5-8μm to limit pathlength, but neither transmission nor transflectance mode are sensitive enough to detect absorbance bands of interest at such limited pathlengths [23]. ATR-FTIR spectroscopy is useful for analysis of samples with high water content due to the limited effective pathlength of the evanescent wave [17], [47], [48]. This means only a thin layer of sample in

proximity to the IRE is interrogated limiting interference from bulk water absorption [23], [49]. Nevertheless, dehydrating samples prior to analysis is usually the most desirable approach to eliminate the water signal in order to ascertain as much information as possible on biomolecules of interest as spectral contributions from water are still apparent in the fingerprint region of aqueous biofluid samples and could lead to false interpretation [50]. Furthermore, water gives rise to anomalous dispersion effects which distort the OH band and compromise quantitative analysis in ATR mode [37]. 'Digital drying' through computational correction of aqueous sample spectra may emerge as a quicker approach, but molecular components that exist in low concentrations may not be detected at the surface until the droplet is fully dried [51]. This is due to the Vroman effect whereby there is a delay in high molecular weight proteins (e.g. Kininogen) attaching to the IRE surface while low molecular weight proteins (e.g. Albumin) attach quickly from the time the droplet has been deposited [52]. Clearly this is pertinent to the ATR spectroscopy study of serum especially since constituents that are in proximity to the interface are preferentially sampled [23]. However, drying samples may give rise to other issues. Cracked drying profiles can affect the contact efficiency of the sample with the IRE, and the 'coffee-ring' phenomenon can give rise to non-homogenous biomolecule distribution across the biofilm [53], [54]. The coffee-ring effect describes a phenomenon whereby molecules are distributed in concentric rings dependent on their molecular weight following droplet drying [52]. While not as detrimental to ATR measurements compared to transmission or transflectance mode because ATR acquires a spectrum averaged across the ATR crystal and is less susceptible to scattering [21], [55], drying patterns still present a risk of variation across sample measurements since it may not be possible to

maintain a minimum sample thickness greater than the evanescent wave penetration depth (~3μm) across the sampling area [17]. Well defined and controlled sample drying conditions are therefore an essential step of sample preparation and will be critical in reducing variance across clinical centres.

2.3. Spectral Pre-Processing

Generally, biological analysis of IR spectra is only interested in investigating differences in sample biochemistry, and nothing else. Before analysis it is usually desirable to perform signal processing steps to reduce variability and improve subsequent analysis. The techniques available to do this are called pre-processing techniques because they are steps that precede processing through data analysis such as generation of classification algorithms, or construction of concentration calibration curves.

Some aberrations, such as spectral artefacts, cannot be accounted for by using a background scan [56]. Differences in samples, the environment, and the instrument between scans can accumulate variance and lead to misguided conclusions when, for instance, comparing results from different studies. A few examples are differences in patient lifestyle, pipetting technique, storage conditions, sample age, room temperature, substrate, acquisition parameters, instrument model, scattering, instrument age, and instrument calibration [22], [57], [58]. A plethora of pre-processing techniques are available to reduce the influence of such confounding factors and can be broadly categorised as either filtering techniques, or model-based

techniques [59]. Filtering techniques perform transformations to remove undesired variations and arrive at ‘better’ spectra. Model-based techniques fit the data to a mathematical model. Some broad pre-processing categories are data smoothing, normalisation, feature selection, baseline correction, quality testing, and class balancing.

2.3.1. Baseline Correction

The baseline can deviate between spectra for various reasons such as light scattering, light throughput, or black body radiation [60]. Rubberband baseline correction is a common method that calculates a convex polygonal line whose vertices touch the troughs of the absorbance spectrum but does not cross spectrum line (Figure 2.12) [61], [62]. This is called the convex hull and is removed to reposition the spectrum at a new level baseline.

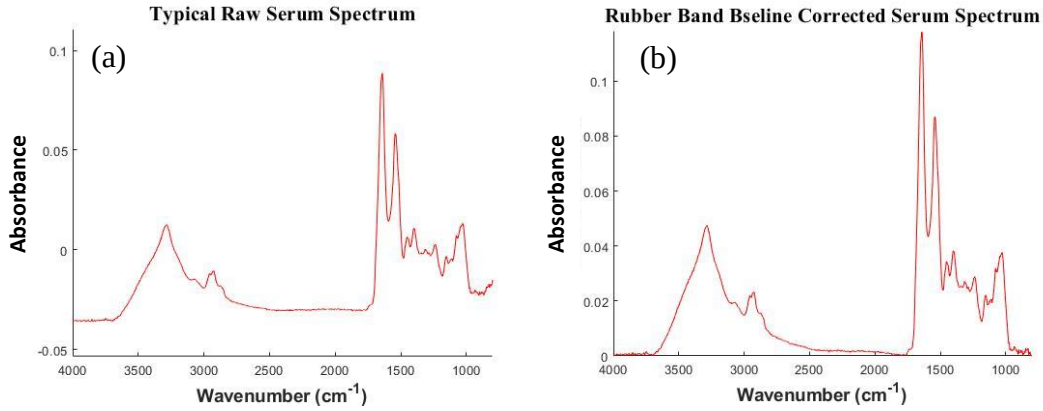


Figure 2.12. *IR spectrum of human pooled serum before and after baseline correction (a) raw IR (b) IR spectrum after rubber band baseline correction.*

2.3.2. Normalisation

Normalisation is commonly used to eliminate differences in sample thickness, light throughput, or environmental conditions. It is a process that scales all spectra within a reference frame. Three simple normalisation methods are min-max normalisation, amide I normalisation, and vector normalisation [59], [63]–[65]. Min-max normalisation finds the minimum and maximum intensity value across the dataset then subtracts the minimum value from every datapoint and divides this value by the range (Equation 2.14). This sets the minimum value to zero and scales the entire dataset between zero and the maximum. Amide I normalisation divides every value by the amide I intensity since this is generally the largest value in the dataset (Equation 2.15). This sets the value of the amide I peak to one and all other datapoints are a ratio of this value. As an alternative, vector normalisation is performed by calculating the vector length, N, by finding the square root of the sum of all square values in the dataset. Each of the raw data variables are then divided by the vector length. All spectra are thus scaled to each other by letting the sum of squared standard deviation equal one (Equation 2.16). The entire wavenumber range is normalised in vector normalisation and min-max normalisation. In amide I normalisation the amide 1 peak will always be the same and features at other wavenumbers may be made clearer. There are therefore different situations each approach is more appropriate for. For instance, vector normalisation is the most widely used since it scales the whole wavenumber region, preventing exaggeration of features across the spectrum. Conversely, if variation is expected to occur at another place in the spectrum from the amide I peak it may make sense to normalise around the amide I peak so that variations elsewhere in the spectra are more easily observed. Min-max normalisation guarantees all features

will have the same scale but does not handle outliers well as outliers will dictate the bounds between which the data is normalized [65].

$$S = \frac{(S_r - S_{min})}{(S_{max} - S_{min})}$$

Equation 2.14. *Min-max normalisation equation.*

$$S = \frac{S_r}{S_{AmideI}}$$

Equation 2.15. *Amide I normalisation equation.*

$$N = \sqrt{S_1^2 + S_2^2 + \ldots + S_n^2} \qquad S = \frac{S_r}{N}$$

Equation 2.16. *Vector length and vector normalisation equations*

Variable Assignments: *S = normalised spectrum, S_r = Raw spectrum, S_{min} = minimum value of entire dataset, S_{max} = maximum value of entire dataset, S_{amideI} = average value at amide I peak, N = the vector length.*

Objects with a similar size magnitude to the wavelength of mid-IR light (2.5-25µm) within the sample under investigation, such as cellular structures, will result in light scattering that will distort the absorbance spectrum [58], [59]. This resonant Mie scattering typically presents as a large curved band (Figure 2.13) [36], [66]. As mentioned previously, other forms of scattering may arise, for example from uneven surfaces that can redirect light such that the IR light does not reach the detector. These artefacts can be removed by Extended Multiplicative Signal Correction (EMSC) which effectively baseline corrects and normalises the data [38]. EMSC is a model-based technique aimed at reducing the multiplicative (amplitude) and additive (offset) variations between spectra such as variations caused by differences in sample pathlength. Raw spectra are fitted to a reference spectrum, relying on the fact that overall shape of IR spectra of biological samples is generally consistent [67].

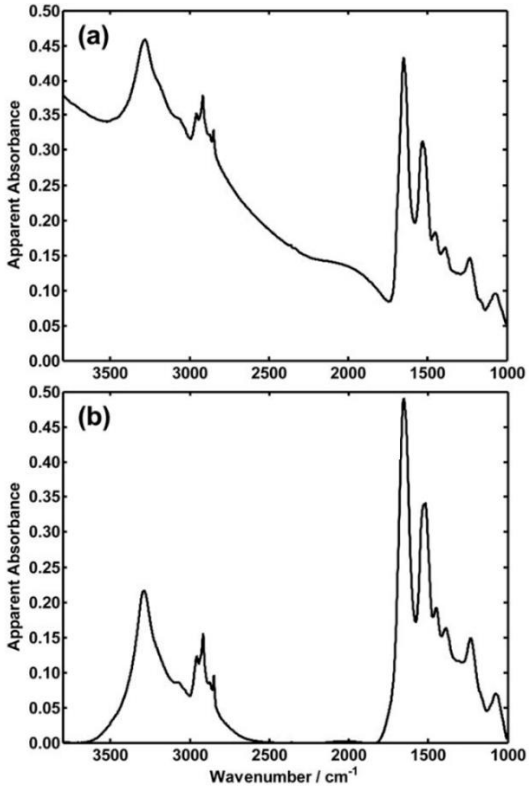


Figure 2.13. *FTIR image spectrum of prostate tissue before and after pre-processing by EMSC (a) raw spectrum where resonant Mie scattering is evident (b) corrected spectrum. Adapted from:* [29]

2.3.3. Feature Selection

Feature selection simply aims to focus in on data that are of importance in analysis. A single IR spectrum is a complicated dataset that typically includes approximately 3600 datapoints (400-4000cm^{-1} wavenumber range at 1cm^{-1} resolution). For a 100 patient sample set with three biological sample repetitions and three instrument repetitions there will be a total of 3.24 million datapoints. If the analysis is performed using advanced computational techniques such as machine learning, it could take days to compute all the variables, many of which will not add value to the analysis. Many spectroscopists therefore decide to cut the wavenumber

range between 900-1800cm^{-1} to capture only the fingerprint region of biological samples, or they might choose the reduce the dataset further by cutting more precise wavenumber ranges if the peaks of interest are already known prior to analysis. Another technique for reducing the dataset is by data binning where absorbance intensities over a given wavenumber interval, called a binning window, are averaged together thereby preserving the wavenumber range, but reducing the resolution. For instance, data binning a spectrum range of 400-4000cm^{-1} with a binning window of 4cm^{-1} will reduce the dataset from 3600 datapoints to 900 datapoints.

2.4. Data Analysis Techniques

A single serum sample may contain many thousands of biomolecule species and as a result the IR spectrum is complex. Analysis of such data may necessitate anything from simple univariate peak intensity analysis to advanced multivariate statistical approaches.

2.4.1. Univariate Analysis

Percentage Relative Standard Deviation (RSD%) and Signal-to-Noise Ratio (SNR) are two widely used metrics to assess spectral quality. There are many ways to calculate SNR, but the most common method used in FTIR spectroscopy is to take the maximum absorbance intensity (e.g., amide I peak) and divide it by the standard deviation of a portion of the baseline to represent noise (Equation 2.17) [53], [59],

[68]–[70]. Figure 2.14 shows the spectral regions that are typically used to represent the signal and noise portions of the SNR equation. RSD% is calculated by dividing the standard deviation of a dataset by the mean of the dataset and multiplying the result by 100 (Equation 2.18). This metric is used to express the precision and repeatability of spectral data gathering regimes [71]. The lower the RSD% the less dispersed the measurements of a dataset are relative to the mean.

$$SNR = \frac{x_s}{\sigma_n}$$

Equation 2.17. *Signal to Noise Ratio Equation*

$$RSD\% = \frac{\sigma}{\bar{x}} x\ 100$$

Equation 2.18. *Relative Standard Deviation Equation*

Variable Assignments: *x_s = signal at selected peak, σ_n = standard deviation of selected noise region, σ = standard deviation, $\bar{x}$ = mean.*

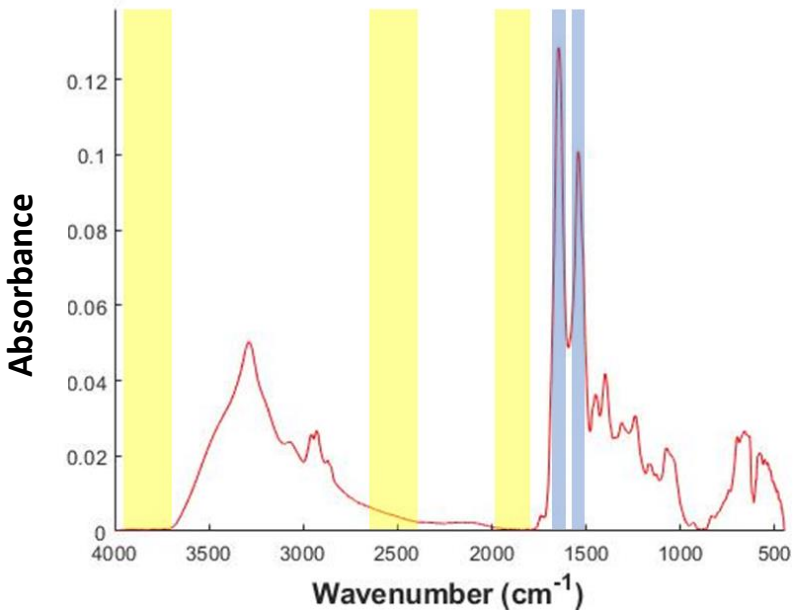


Figure 2.14. *Typical biological IR spectrum with areas commonly used to represent regions of noise (yellow) and to represent signal (blue) in SNR equations.*

The Limit of Detection (LOD) and Limit of Quantification (LOQ) are defined as the lowest concentration of a measurand that can be reliably detected at and quantified at by an analytical procedure [72]. Both are specific to the molecule under investigation but can nevertheless serve as an impartial comparator of the performance characteristics of techniques [73]. LOD and LOQ are commonly determined through calibration curves using Equations 2.19 and 2.20 below [74]:

$$LOD = {}^{3S_a}/_{b}$$

Equation 2.19. *Limit of Detection equation.*

$$LOQ = {}^{10S_a}/_{b}$$

Equation 2.20. *Limit of Quantification equation.*

Variable Assignments: S_a *= The standard deviation of the response and* b *= The gradient of the regression line.*

2.4.2. Multivariate Analysis and Machine Learning

Big data sets almost always need to be reduced to make sense of them. Feature selection techniques described earlier and feature extraction techniques both aim to reduce the dimensionality of a dataset to make interpretation easier. In contrast to feature selection techniques where portions of the data are deemed superfluous and discarded, feature extraction techniques aim to convert large, complicated datasets into smaller, interpretable datasets whilst utilising all the available information [75]. Given the diverse nature of IR absorbance spectra, it is often appropriate to employ such multivariate analysis approaches [22]. Principal Component Analysis (PCA) is a dimension reducing technique and the most common multivariate technique used in

the field of IR spectroscopy as it provides a robust method to reduce the data into a more digestible visualization [57]. Similarly, Partial Least Squares (PLS) is also a data reduction technique that helps identify where the variance lies in a dataset [76]. The primary difference is PCA does not require knowledge of the sample classes for analysis whereas PLS does. This is because PCA is an unsupervised technique and PLS is a supervised technique. Supervised techniques are dependent on the labelling of the datasets (e.g. cancer vs non-cancer) while unsupervised techniques are not [77]. By themselves both PCA and PLS are exploratory techniques used to visualise data but can be extended for regression (e.g., PLSR) or discriminant analysis (e.g., PLS-DA) [57], [76].

Data reduction techniques can be used to develop classification algorithms. This is especially useful for disease diagnostics as models can be trained to classify patient samples into disease or non-disease groups [46], [61]. Training algorithms to recognise patterns in data is referred to as machine learning. There is no consensus on which multivariate approach is best to build FTIR spectral classification models with [22], but two widely used machine learning techniques for IR spectroscopic clinical studies are PLS-DA and Random Forest (RF) [78]–[85]. How to perform PCA, PLS, and RF methods will be described here.

2.4.2.1. Principal Component Analysis

Principal Component Analysis (PCA) is a quick, intuitive, unbiased, and reliable exploratory technique frequently used to identify clustering patterns, outliers, and trends during the initial assessment of complex spectral data [57], [61]. To

distinguish variance in a spectral dataset the correlation between each spectrum at each specific wavenumber would need to be observed. The dimensionality of spectral datasets is too large to make this possible, thus the dimensionality must be reduced. PCA achieves this by finding linear combinations, called Principal Components (PCs), which are weighted sums derived from the initial variables (i.e., absorbencies at each wavenumber) [86]. In the context of spectroscopy, these PCs are not related to specific spectral bands but instead relate to a collection of bands each with a different variance, or weighting, contribution to the PC.

In PCA, the data is restructured along axes that best describes the variance in correlation to the wavenumbers. A rigid rotation of the original axes is taken such that there is maximum variation across the first new axis (PC1) [87]. The second axis (PC2) is the direction of maximum residual variance orthogonal to the first axis and every subsequent axis is the direction of greatest residual variance orthogonal to the last [88]. Consequently, PCs are uncorrelated and successive PCs explain less variance than the last. To help visualise a PCA transformation, Figure 2.15 shows a simple 2D scatterplot and where a subsequent PCA would transform the data to.

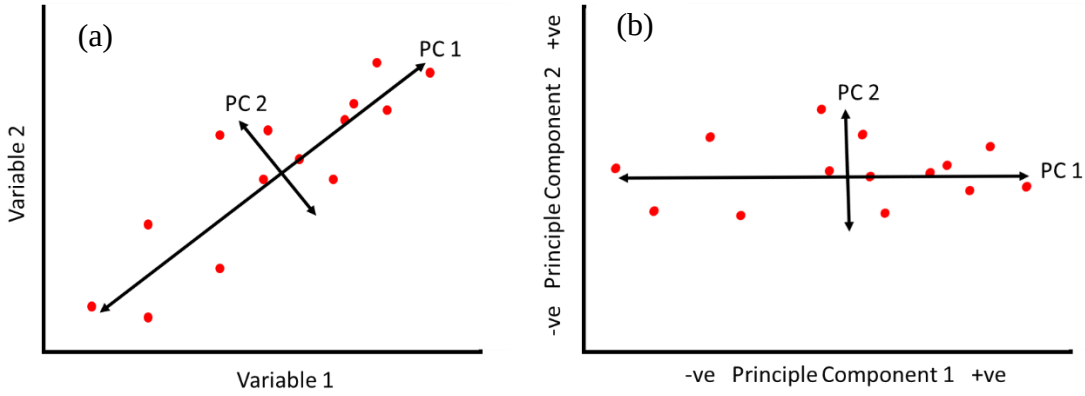


Figure 2.15.*(a) Simple two variable scatter plot with PC 1 and 2 annotated indicating the directions in the data where the first and second greatest amount of variance lies, respectively. (b) the same data readjusted to resemble the dataset after PCA transformation indicating where the greatest differences in the dataset lie.*

All sample spectra collected can be represented as a data matrix where X represents all observations (absorbencies at each wavenumber) and Y represents all dependent variables (disease class, or sample concentrations, etc.) (Figure 2.16).

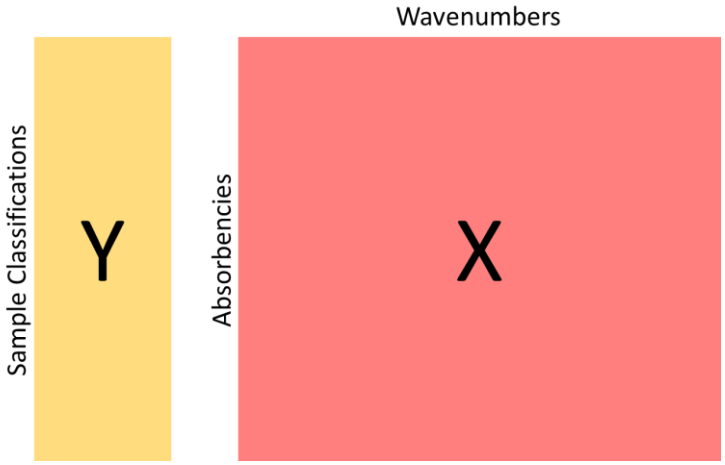


Figure 2.16. *Large IR spectral datasets can be separated into the dependant variable matrix Y and the data matrix X.*

The focus is on reducing the dimensionality of the data matrix, X. X is a *n x p* matrix where *n* is the number of samples and *p* is the number of dimensions (i.e., wavenumbers). First, the data matrix is usually mean-centred so that all absorbance

values are expressed as a deviation from their mean since PCs are only concerned with the variation in the data [87]. The mean of each column in X is calculated and subtracted from each value in the column setting the mean of each column to zero and shifting the centre of X to the coordinate system origin [86]. If large differences in variance of the initial variables does not reflect relative importance (i.e., some variables are expected to be dispersed across a broader range than others) then the data matrix could be standardized since PCA will be sensitive to this [89], [90]. However, this should only be done in exceptional circumstances so will not be discussed further [91]. The covariance matrix, S, is then found using the data matrix, X, its transpose, X^T, and the number of samples, *n* [89].

Variance is a measure of the spread of data within a dataset, but only applies to one dimension. Covariance is a measure of how two dimensions vary with respect to each other. More specifically, it is a measure of linearity between two dimensions whereby large positive covariance indicates well correlated dimensions, large negative covariance indicates inverse correlation, and zero covariance indicates uncorrelated dimensions [92]. The covariance matrix captures the covariance between all possible pairs of dimensions in the original dataset and is fundamental to many multivariate analysis techniques [26], [93].

Variance is extracted from the original dataset through orthogonal linear transformation of the intercorrelated variables to a new uncorrelated set of variables (PC scores) ranked in decreasing order of explained variance [76]. The covariance matrix produced from the original data is decomposed into a score matrix of scaling coefficients (eigenvalues), and a loadings matrix (PCs, also called eigenvectors) [57]. An eigenvector is a linear view of directions the original data is dispersed across in

multidimensional space. Eigenvalues are coefficients of the eigenvectors and are an exact measure of how much variance is explained by the linear combinations defined by the corresponding eigenvector [94]. Thus, PCs can be ranked according to the amount of variance explained.

Since PCs are uncorrelated, the PCs that account for the least amount of variance in the dataset may be discarded. This is how the dimensionality of the dataset is reduced. Ideally, PCs representing the least amount of variance would only contain information related to spectral noise or any other unwanted experimental signal, although this is highly dependent on the study [86]. The number of PCs kept can be determined either by setting a threshold for the total variance captured (e.g. >90%) [90], using a scree plot [95], or by cross-validation based on mean square error of spectrum reconstruction [61], [96].

The last step in PCA is to recast the original data onto the PCs that have been derived. The variance within the dataset can be visualised from either the sample scatter plots, or wavenumber loadings plots (Figure 2.17) [57]. The scores matrix represents variance in the sample direction of the original data and is used to deduce relationships between samples, with clustering inferring sample homogeneity and separation inferring sample heterogeneity. The loadings matrix represents variance in the wavenumber direction and is used to associate distinct spectral markers to the relationships observed in the PC scores plot.

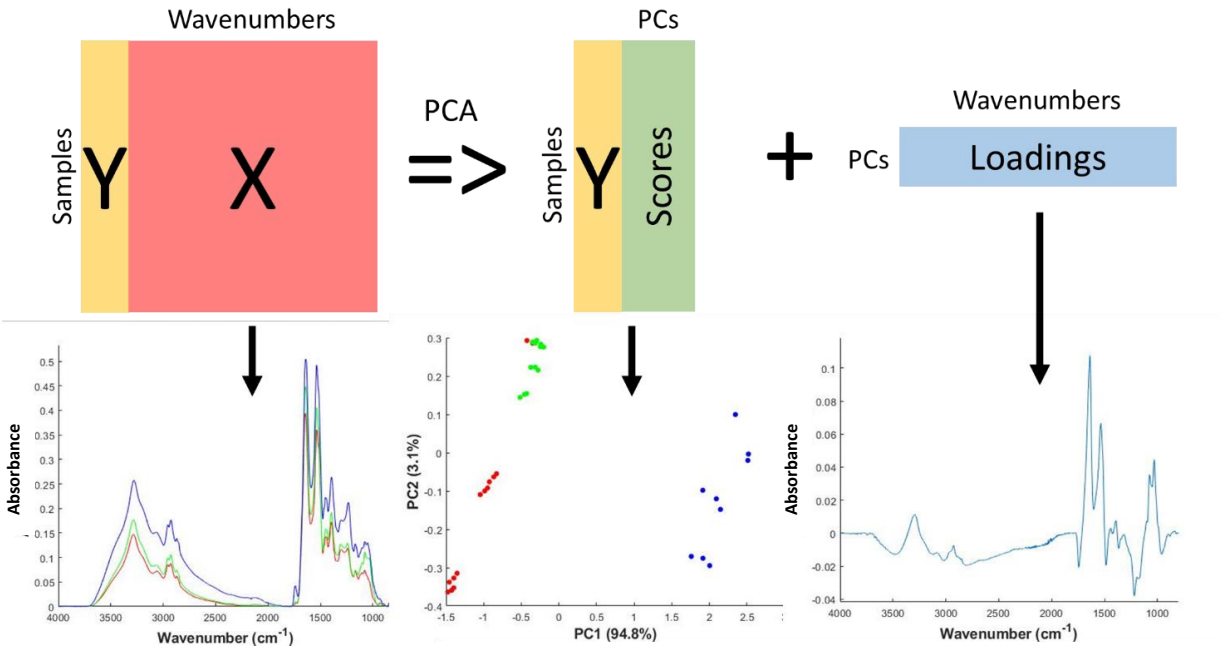


Figure 2.17. *PCA transformation inputs and outputs. PC scores of different PC directions can be plotted against one another to observe if groups can be well discriminated. Loading plots of individual PCs indicate the degree to which specific wavenumbers contribute variance in the PC.*

PCA is unbiased due to its unsupervised nature as no information pertaining to data class is used to construct the linear combinations. Thus, resultant scores and loadings represent true variance within the dataset. However, PCA only identifies total dataset variance and does not discriminate within-group from between-group variance [97]. Consequently, class separation can only be elucidated if within-class variance is less than between-class variance. This is to say that misleading class separation can appear in a PCA scores plot if variance is introduce through sample preparation issues, inappropriate data processing, or experimental bias [85].

2.4.2.2. Partial Least Squares Regression

Partial Least Squares (PLS) is similar to PCA in that wavenumbers are used to construct linear combinations, except in PLS both the data matrix and the class matrix

are decomposed [88]. PLS is therefore a supervised method as opposed to unsupervised. It is an iterative process where latent structures, so called factors, are extracted one at a time [85]. Factors, equivalent to PCs, are extracted by finding a linear combination of predictors that will have maximal covariance with a linear combination of the response [97]. Thus, PLS models attempt to identify factors which contribute systematic majority of variance in predictors versus associated response, enabling regression analysis [98]. A main advantage of PLS is that it can handle noisy data with many collinear, or even missing, variables and is able to model numerous response variables simultaneously [99], [100].

PLS is not a single technique but encapsulates a suite of closely related techniques [85], [100]–[104]. Initially, both observation and response matrices are typically mean centred. The original variables of each matrix are transformed into new latent variables via linear orthogonal transformations capturing maximum variance in the data matrix, X, and maximum variance in the response variables, Y [76]. The latent structures (latent variable scores and loadings) are then decomposed such that the latent variable X scores share maximal covariance with the Y latent variable scores [85]. This process is repeated iteratively whereby at each iteration the linear combination (i.e., factor, or principal component) of maximal covariance is extracted. Doing so deflates the X and Y matrices by subtracting the rank-one principal component [101]. The residuals can then be used in the next iteration with this process continuing for as many components as is desired [104]. Ultimately, a linear relationship, so called PLS Regression (PLSR), is formed whereby the latent variables of X serve as predictors of the Y latent variables which can be used to estimate original unknown Y response variables [85], [103]. When Y matrix constituents are categorical

response variables, (e.g., disease states), a discriminant model can be derived from the PLSR method [97], [105]. In this case, class membership must be coded in a class membership matrix, Y, as a numerical representation of response categories [101]. This is called PLS Discriminant Analysis (PLS-DA).

PLS models have a tendency to over-fit to data with the ability to create class separations even with random input data [88]. Model accuracy and robustness must therefore be validated. This can be achieved by partitioning the sample set into a subset used to build models (training set) and a subset reserved purely to test the model's predictive capabilities against (test set) [85]. Use of a test and training set is the standard across many machine learning methods. In many cases it may not be possible to commit portions of the data to the test set due to sample size limitations. To address this, resampling procedures known as cross-validation are commonly used. There are multiple ways to implement cross-validation regimes, such as the leave-one-out, leave-n-out, or Monte Carlo cross-validation methods [85], [106]. Essentially a predefined number of models are constructed in parallel each using different combinations of samples within the training and test sets enabling all available sample data to be used both for training and testing predictive models [98], [100]. Prediction results across all model repetitions can then be averaged to give a single estimate of model performance. In PLSR, the average and standard deviation of the Root Mean Square Error or the Validation set (RMSEV) and r^2 values are used to evaluate model performance. Cross-validation is the standard approach for PLSR analysis, and widely used across many machine learning and multivariate techniques [100]. Cross-validation is also used to determine the optimum number of components to be used by PLS models [103]. This is achieved by selecting the minimum number of components beyond which the

combined models' predictive capabilities no longer improve. In particular, the number of components which minimise average RMSEV is typically selected [98].

2.4.2.3. Random Forest

Random Forest is a supervised machine learning method and is possibly the most widely used classification tool in chemometrics [107]. Unlike PCA and PLS, Random Forest (RF) does not decompose the data into PCs. Instead, the method relies on an ensemble of decision trees [108]. A decision tree is a simple concept that splits the data into groups based on a series of 'yes/no' questions about features of the data. Each decision in the tree is called a node, and the final node at the bottom of the decision tree is called a leaf node where a prediction is ultimately made about the sample's classification (Figure 2.18). Ideally, at each node the data splits into subsets of data that are as different to each other as possible. As with PLS, models are constructed using a training set, and how well the decision tree operates is evaluated against a test set. If a new sample from the test set is incorrectly classified by the decision tree, one reason could be because the model was too good at recognising the features of the training data. This is called overfitting because the decision tree has been fitted too closely to the training dataset. This is a common problem for classic decision trees [109].

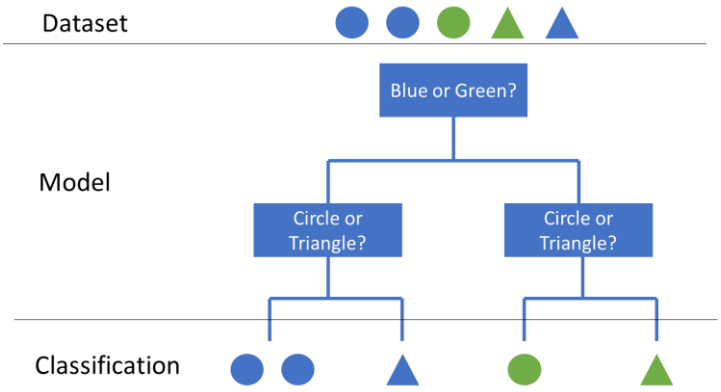


Figure 2.18. *Simple decision tree model used to classify components of an input dataset.*

RF overcomes overfitting issues by generating numerous trees (hence 'forest') where the features of interest at each node are selected from random feature subsets and the data used to train the trees are selected at random from the training set (hence 'random') [108], [109]. Each tree 'votes' on what the classification of a sample is and the classification with the most votes is assigned to the sample (Figure 2.19). The method feeds samples that were not used in the construction of the decision trees into each tree to test the robustness of the system.

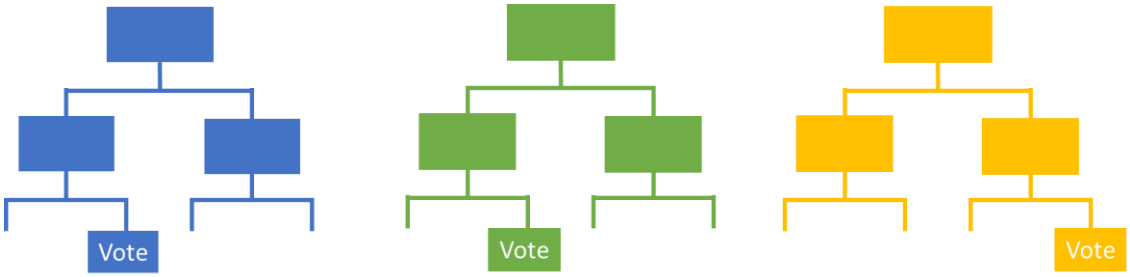


Figure 2.19. *RF creates numerous decision trees that vote on an input sample's classification to determine what is most likely to be the correct classification.*

Decision trees determine how to split the features at each node using the Gini index [110]. The Gini index measures the impurity of features with respect to class in terms of the quality of data splits at the parent node [111]. Gini impurity is a measure of the probability that a node has incorrectly classified a feature and can be calculated

rather simply using the fraction of samples split into each class at each parent node (Equation 2.21) [109], [111], [112]. Some features will incur lower Gini impurity than others and is consequently deemed more important for classifying data correctly. Each time a feature is used to make a split at a node the two children nodes will have a Gini impurity value lower than the parent node [109]. The extent to which a specific feature decreases Gini impurity across the depth of a decision tree can be calculated as the sum of Gini index reduction at every node where the feature was used to form a split [113]. Thus, the relative importance of a features with respect to their ability correctly classify samples can be estimated. Relative importance of features (e.g., wavenumbers) can then be presented graphically to give a deeper insight into the attributes responsible for classification (Figure 2.20).

$$I_G(n) = \sum_{i=1}^{j} P_i(1 - P_i)$$

Equation 2.21. *Gini impurity equation.*

Variable assignment: *I_G = Gini impurity, n = node number, P_i = relative frequency of class, i, in node, n, j = number of classes.*

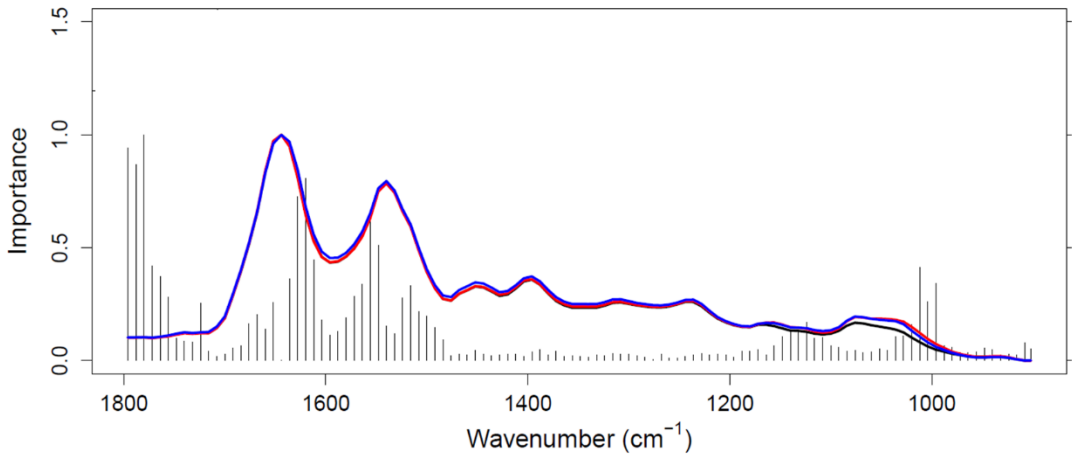


Figure 2.20. *Gini importance plot indicating which wavenumbers of the IR spectrum were most important for classifying data in a RF classifier.*

2.4.2.4. Class Balancing

In situations where there is an unequal amount of classes within a study sample set, the dataset is said to be imbalanced. This can cause problems when using classification algorithms because the algorithm will mainly focus on learning patterns associated with the majority class and will tend to underperform when trying to recognise the minority class [114]. Thus, classes need to be balanced to optimise classifier model performance. One approach is to under-sample the majority class by subtracting majority class samples from the training dataset, but this is an unfortunate waste of valuable data [115]. Conversely, one could over-sample the minority class by duplicating samples within the training dataset, but this could lead to overfitting of the classification model to specific minority class sample inputs [116]. Synthetic Minority Oversampling Technique (SMOTe) is a common pre-processing technique designed to reduce the class imbalance in a dataset by creating new synthetic minority class samples from the original sample set [117]. SMOTe selects a minority class sample and locates a predefined amount, usually 5, of the samples randomly chosen k-nearest minority neighbours in feature space and connects line segments to each feature vector (i.e., sample). The difference between a chosen feature vector and a number (depending on how much over-sampling is needed) of randomly selected nearest neighbours is found and multiplied by a random number between 0 and 1. This is then added to the chosen feature vector creating a new point randomly positioned along the line segments adjoining features. Hence, a new synthetic sample is randomly interpolated from a combination of nearest minority neighbours in feature space [115],

[118]. Figure 2.21 shows an illustration of samples in feature space and where synthetic samples may be generated using SMOTe.

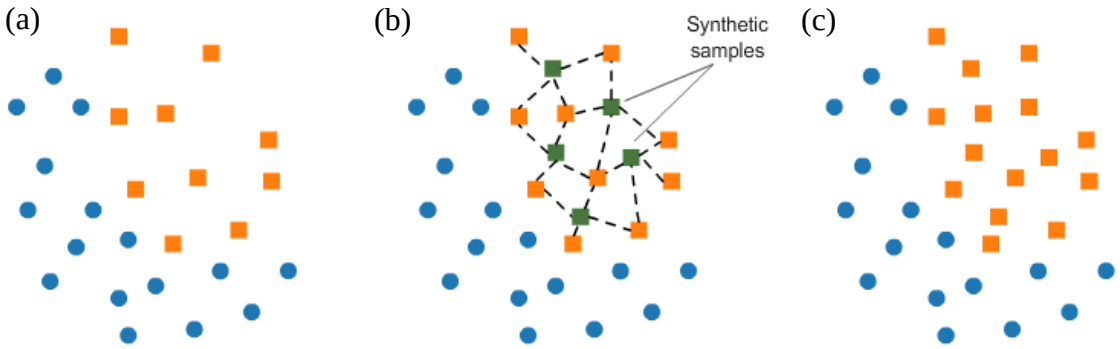


Figure 2.21. *Illustration of SMOTe. (a) raw sample set in feature space. (b) lines drawn between nearest neighbours of minority class in feature space where new synthetic samples are created. (c) Balanced dataset. Image credit:* [119]

2.5. Fabrication and Prototyping Techniques

Effective design often necessitates rigorous preliminary testing using bespoke parts and prototypes before finalising a working product. The viability of scaled production should always be at the forefront of decisions. The later in the product design and development cycle a problem is identified, the more expensive it can become to address the issue [120]. If possible, the processes and designs used in development should closely resemble the designs and fabrication processes of the final product(s). However, there are industries where this may not be possible (e.g., microelectronics, high-performance optics, heavy machinery) and design verification may need to be carried out in alternative ways [121], [122]. A combination of manufacturing processes is usually required to produce multicomponent devices. The

tooling costs to set-up large scale manufacturing processes can be prohibitive and are unwise to invest in if designs have not been validated through prototyping. Techniques employed within this project for the fabrication of prototype models and finalised designs are described here.

2.5.1. 3D Printing

3D printing is an excellent low-cost method to rapidly produce designs for testing before moving to scaled manufacturing techniques such as injection moulding [123]. There are numerous 3D printing methods available each with relative advantages and disadvantages [124]. A common method is Fused Filament Fabrication (FFF) [125], [126]. Thermoplastic filament, usually Polylactic Acid (PLA) or Acrylonitrile Butadiene Styrene (ABS), is extruded through a heated circular die. The plastic enters the extruder cold, is liquefied using heat, and rapidly cools upon exiting the die with the help of fans. As the plastic is cooling, the nozzle moves across a build following a predefined pattern in a 2D plane. The patterns are created using Computer Aided Design (CAD) files. The filament is continually pushed through during printing to give a constant flow until the 2D pattern is complete. The build plate then moves down (z-direction relative to the print head) by the width of the extruded plastic. Another 2D layer is then printed on top and the heat of the plastic exiting the extruder fuses the layers together. The printer repeatedly adds 2D layers on top of each other in this manner until the 3D CAD model is complete. 3D printing techniques are referred to as additive manufacturing because of how material is successively added in layers to build a complete shape. Figure 2.22 shows a schematic of FFF in practice:

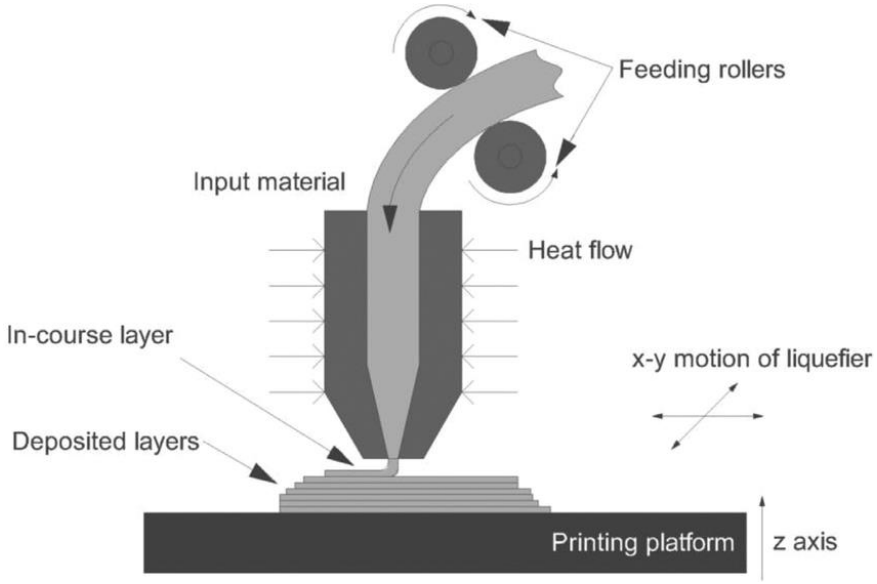


Figure 2.22. *Schematic of an FFF 3D printer. Filament is extruded through a heated nozzle while a X-Y translational printer head dispenses filament in a pre-defined pattern. Image credit:* [127]

Selective Laser Sintering (SLS) is another 3D printing method that holds some similarities to FFF, but the raw material is provided as a powder instead of a filament meaning a more diverse range of materials, such as metals or ceramics, may be used for fabrication [128]. A laser is directed at a bed of powder following the successive 2D planes of the 3D model (Figure 2.23). The laser is only powerful enough to sinter the powder without inducing melting. Once each 2D pattern is completed, the powder bed moves down and fresh powder is laid on top of the build. FFF and SLS are two of the technique used within this project, but there are many other additive manufacturing techniques, such as Stereolithography (SLA), Laminated Object Manufacturing (LOM), or Laser Engineered Net Shaping (LENS), that maybe more suitable rapid prototyping approaches depending on the requirements [129].

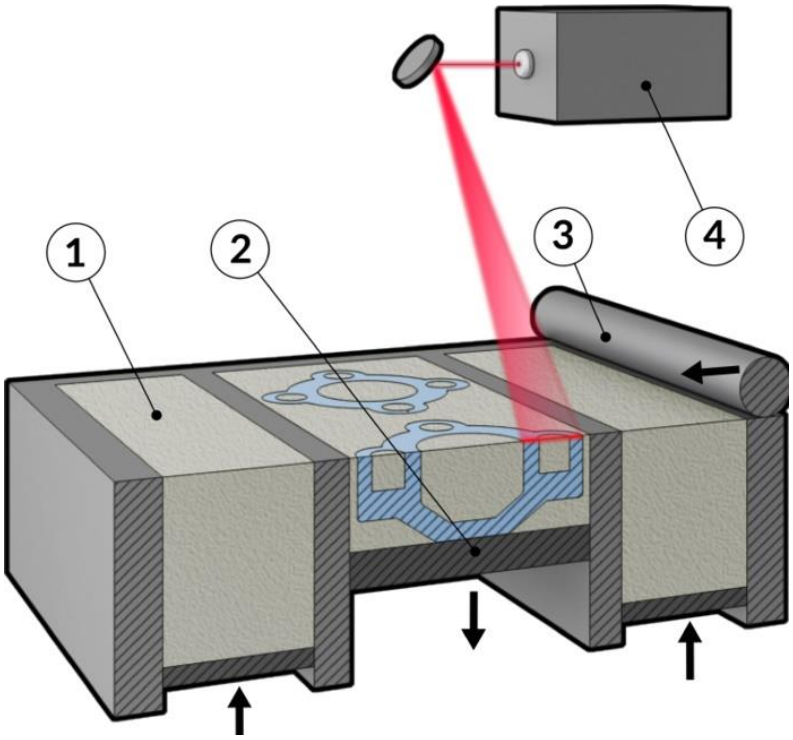


Figure 2.23. *Schematic of an DMLS 3D printer. A laser is directed at raw material in a pre-defined pattern. Raw material is repeatedly applied on top of the structure until the print is complete. (1) powder metal, (2) powder bed build plate, (3) powder roller, (4) laser source. Image credit:* [130]

2.5.2. Injection Moulding

Injection moulding is a simple yet effective high capacity production method capable of forming complex shapes from a diverse range of materials [131], [132]. Thermoplastics are the most common material used, but metals and glasses may be processed through injection moulding too [133]. Raw material in the form of pellets is fed through a hopper into a heated chamber and melted. The molten material is then injected at high pressure into a mould the shape of the final product (Figure 2.24). Depending on the feature geometry and process it may be difficult to fabricate small features (e.g., <1mm in size) due to the likelihood of trapping small air pockets which will weaken the part [134]. This may require refinement of the injection moulding

process, or more sophisticated machinery capable of producing micron or submicron scale features [135].

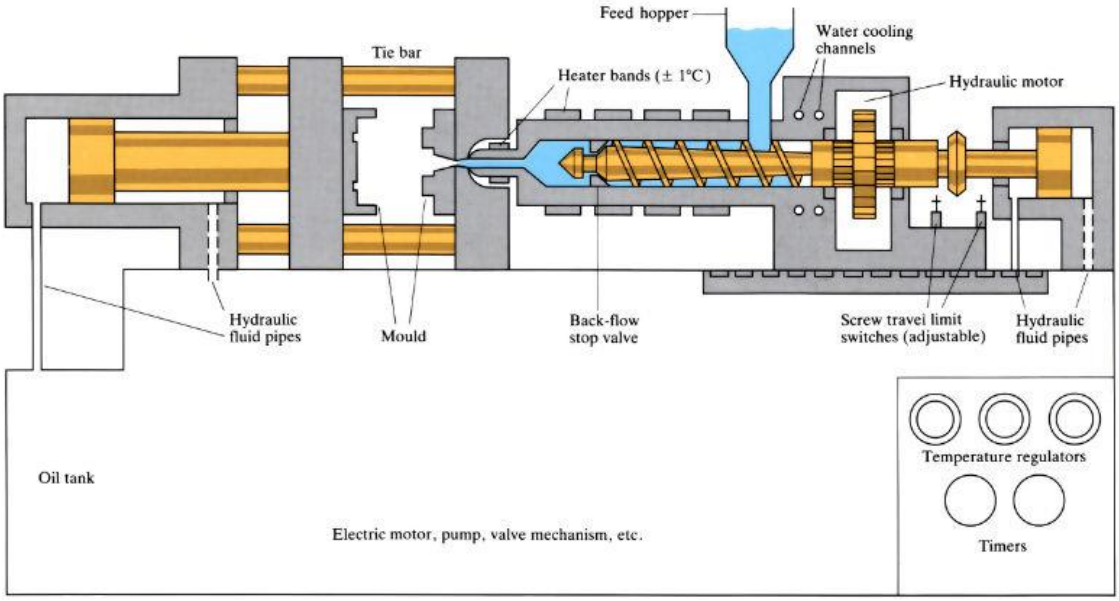


Figure 2.24. *Schematic of an injection moulding machine. Raw material is melted in a screw barrel and injected into the mould cavity at high pressure. Image credit:* [136]

The process incurs a low overhead cost, but high initial set-up cost [137]. The tools required to injection mould components are bespoke and expensive, but once this has been accounted for the operational costs and raw materials are comparatively low due to the high manufacturing output and low cost of materials.

2.5.3. Silicon Microfabrication

Microfabrication encapsulates a suite of techniques aimed at fabricating structures on the micron scale. Silicon is often central to the processes due to the unique crystallographic properties and abundance of the material. Silicon etching specifically is of huge importance to the semiconductor and microelectronics industry.

Conceptually, silicon etching is quite straight forward; silicon planes are selectively removed using etchant chemicals to create well defined geometries. The underlying mechanisms are complex, but well characterised in the literature [138]–[142]. Silicon has a crystal lattice structure and each crystal plane has different relative properties. A system known as Miller indices are used to denote planes within a crystal structure (Figure 2.25). It is important to know that silicon planes with different miller indices possess different characteristics. Miller indices of crystal planes are denoted with plain and curly brackets (i.e., '( )', '{ }') and the vector directions perpendicular to these planes are denoted with square or pointed brackets (i.e., '[ ]', '< >') . The former is always given when selecting a Si wafer because it indicates the crystal plane of the major exposed faces of the wafer.

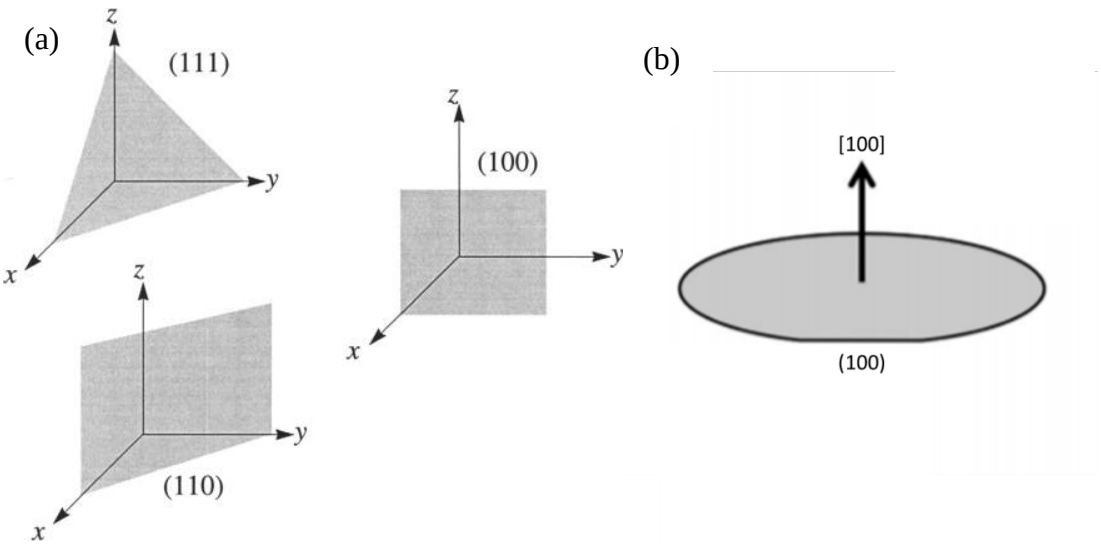


Figure 2.25. *Illustration of Miller indices. (a) the three most commonly used Miller indices, (111), (110), and (100), and what they look like in 3D space. (b) a diagram of a Silicon wafer and the how Miller indices correctly annotated for a (100) orientated Silicon wafer. Image credit:* [143], [144]

Different crystal planes intercept the Si crystal lattice in different ways which impacts the surface energy of the exposed faces [145]. For instance, Si(100) has a higher surface energy than Si(111), thus Si(100) is more vulnerable to chemical attack

than Si(111). This anisotropic property allows material to be selectively removed from a Si wafer. Various techniques have been developed to do this, the most common of which is chemical wet etching by potassium hydroxide (KOH). The KOH rapidly erodes specific planes of the Si crystalline structure, most notably the (100) and (110) directions, while having almost zero effect on the (111) crystal plane direction (etch rate ratio approximately $E_{\{100\}}:E_{\{110\}}:E_{\{111\}} = 400:600:1$ [146]).

In the present studies, an anisotropic KOH etch was used to make v-shaped groove patterns on a silicon substrate. Figure 2.26 presents a flow diagram detailing each step of this process. First, an Oxide layer was thermally grown on the surface of the silicon wafers using the wet oxidation approach. A silicon nitride layer was applied using Low Pressure Chemical Vapour Deposition (LPCVD) until a layer of approximately 2000Å was formed. Wafers were then placed in a plasma asher to dehydrate the surface. It is essential that during the subsequent steps there is good adhesion between the photoresist layer and the wafer. To help this, Hexamethyldisilazane (HMDS) is applied via the vapour prime process. The HMDS coating promotes photoresist wettability by bonding surface silanol groups and thereby increasing the hydrophobicity of the silicon's surface and removing polar molecules (e.g. water) which would otherwise occupy space the photoresist could adhere [147]. Photoresist was applied through the spin coating technique which provides a more uniform spread of photoresist across the wafer in comparison to spray coating. Improved uniformity helps avoid random errors from occurring during subsequent processes. The wafer was then aligned under a custom-built Chromium photomask (Compugraphics, England) taking care to align perfectly with the flat of the silicon wafer as miniscule deviations can result in large defects arising during the KOH etch

stage. The first patterning stages began when photoresist was exposed to UV light directed through the photomask. UV exposure weakens crosslinks in the photoresist which can then be selectively cleaned away using developer solution to reveal the underlying silicon nitride hard mask. The expose silicon nitride was then etched using a plasma etch method to expose the underlying silicon oxide. The remaining photoresist was then exposed by once again using a plasma asher. The exposed silicon dioxide was etched away using hydrofluoric acid, the wafer was rinsed, and the silicon was finally ready for the anisotropic KOH etch. The wafers are submerged in 33% KOH solution for 80°C for 150 minutes. Minor alterations to any of these three parameters can have large implications for the final product. To finish the process, the remaining nitride mask and oxide layers are stripped away in the same manner as before, the wafers are thoroughly rinsed and a final photoresist layer is applied before the wafers are diced into the individual 33x6mm chips using a diamond tipped saw. This final photoresist layer prevents debris from the dicing process and dust from shipping getting in the grooves before the slides can be used. SIREs are stored within protective packaging for as long as possible at this point as this is the most reliable way to ensure no contamination occurs upon leaving the cleanroom. When it is time to use the SIREs the protective photoresist layer must be removed by placing the slides into acetone and leaving in an ultrasonic bath for five minutes. Slides are then rapidly transferred to Isopropyl Alcohol (IPA), then rinsed with deionized water before drying with Nitrogen gas.

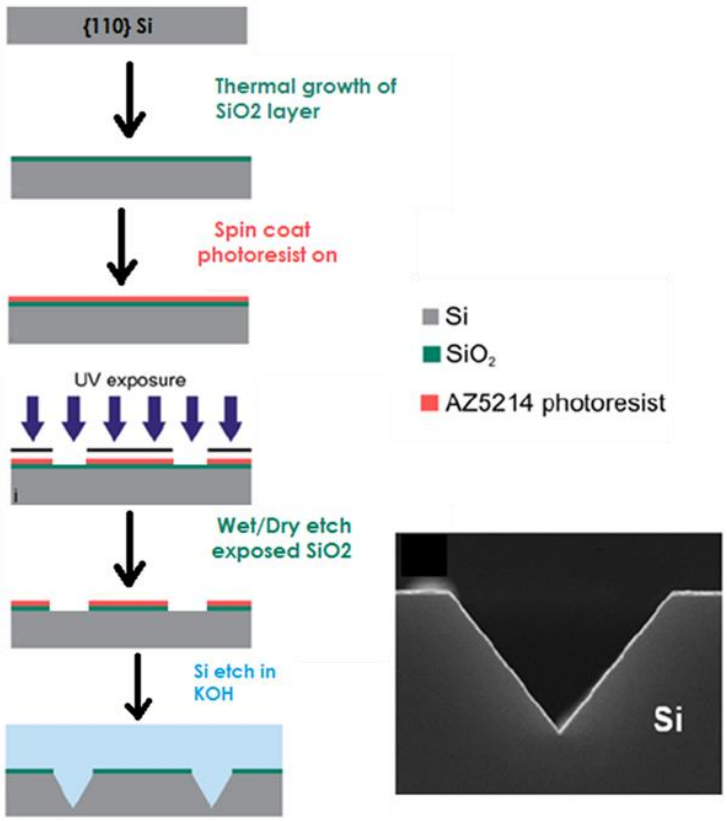


Figure 2.26. *Flowchart diagram of the anisotropic wet etching process used to fabricate the optically active areas of the SIREs (modified from Smith et. al.* [148]*).*

Etch quality was assessed using a Scanning Electron Microscope (SEM) (TESCAN VEGA3, TESCAN) and reflectance spectrometer (Nanospec 3000, Nanometrics Inc.). All silicon microfabrication, metrology, and imaging were carried out using the cleanroom facilities at the Edinburgh University Scottish Microelectronics Centre (SMC).

The parameters of all microfabrication procedures, such as etchant concentration, temperature and etch duration, of microfabrication process are tightly controlled and slight adjustments can have a big impact on the final product. Furthermore, these procedures are conducted in highly controlled environments called ‘cleanrooms’ which require specialised personnel. Miniscule particles can have

devastating effects on the fabrication process. For instance, a human hair could destroy micron scale features across the entire length of the hair. The already high cost of the specialised equipment, reagents, and materials required for microfabrication is therefore compounded by high quality control and design costs. Consequently, microfabrication techniques are often not suitable for early stage development. Fortunately, microfabrication is a highly scalable technique where costs are minimised upon high volume production. This is owed to the development of high capacity equipment and widespread establishment of large microfabrication foundries used to produce microelectronics and semiconductors.

2.6. References

[1] P. Larkin, "Introduction: Infrared and Raman Spectroscopy," in *Infrared and Raman Spectroscopy*, 2011, pp. 1–5.

[2] D. F. Swinehart, "The Beer-Lambert Law," *J. Chem. Educ.*, vol. 39, no. 7, pp. 333–335, 1962, doi: 10.1021/ed039p333.

[3] R. Bacani, *Gel Characterization: From Molecules to Nanostructure to Macroproperties*. Elsevier Inc., 2019.

[4] Specac, "Molecular Vibration Explained | Animated Guides," 2019. https://www.specac.com/en/news/calendar/2019/03/molecular-vibrations (accessed Jun. 13, 2020).

[5] N. B. Colthup, L. H. Daly, and S. E. Wiberley, "Chapter 1 - Vibrational and Rotational Spectra," in *Introduction to Infrared and Raman Spectroscopy (Second Edition)*, 1975, pp. 1–68.

[6] A. T. Fromhold, "Quantum mechanics," in *Encyclopedia of Physical Science and Technology (Third Edition)*, 2014, pp. 109–110.

[7] C. Berthomieu and R. Hienerwadel, "Fourier transform infrared (FTIR) spectroscopy," *Photosynthesis Research*, vol. 101, no. 2–3. pp. 157–170, 2009, doi: 10.1007/s11120-009-9439-x.

[8] D. N. Sathyanarayana, *Vibrational Spectroscopy: Theory and Applications*. 2004.

[9] D. I. Ellis and R. Goodacre, “Metabolic fingerprinting in disease diagnosis: Biomedical applications of infrared and Raman spectroscopy,” *Analyst*, vol. 131, no. 8, pp. 875–885, 2006, doi: 10.1039/b602376m.

[10] H. Susi and D. M. Byler, “Resolution-Enhanced Fourier Transform Infrared Spectroscopy of Enzymes,” *Methods Enzymol.*, vol. 130, no. 13, pp. 290–311, 1985.

[11] F. E. Harris, “Chapter 13 - Integral transforms,” in *Mathematics for Physical Science and Engineering*, 2014, pp. 453–486.

[12] K. P. Möllmann and M. Vollmer, “Fourier transform infrared spectroscopy in physics laboratory courses,” *Eur. J. Phys.*, vol. 34, no. 6, 2013, doi: 10.1088/0143-0807/34/6/S123.

[13] N. Jaggi, “Fourier Transform Infrared Spectroscopy,” in *Handbook of Applied Solid State Spectroscopy*, 2006, pp. 411–450.

[14] N. Birkner and Q. Wang, “How an FTIR Spectrometer Operates,” 2020. https://chem.libretexts.org/Bookshelves/Physical_and_Theoretical_Chemistry_Textbook_Maps/Supplemental_Modules_(Physical_and_Theoretical_Chemistry)/Spectroscopy/Vibrational_Spectroscopy/Infrared_Spectroscopy/How_an_FTIR_Spectrometer_Operates (accessed Jun. 13, 2020).

[15] S. I. Amma, J. Luo, C. G. Pantano, and S. H. Kim, “Specular reflectance (SR) and attenuated total reflectance (ATR) infrared (IR) spectroscopy of transparent flat glass surfaces: A case study for soda lime float glass,” *J. Non. Cryst. Solids*, vol. 428, pp. 189–196, 2015, doi: 10.1016/j.jnoncrysol.2015.08.015.

[16] M. J. Baker *et al.*, “Using Fourier transform IR spectroscopy to analyze biological materials,” *Nat. Protoc.*, vol. 9, no. 8, pp. 1771–1791, 2014, doi: 10.1038/nprot.2014.110.

[17] S. G. Kazarian and K. L. A. Chan, “ATR-FTIR spectroscopic imaging: recent advances and applications to biological systems,” *Analyst*, vol. 138, no. 7, pp. 1940–1951, 2013, doi: 10.1039/C3AN36865C.

[18] P. J. Larkin, “Instrumentation and Sampling Methods,” in *Infrared and Raman Spectroscopy*, 2011, pp. 27–54.

[19] M. Miljković, B. Bird, and M. Diem, “Line shape distortion effects in infrared spectroscopy,” *Analyst*, vol. 137, no. 17, pp. 3954–3964, 2012, doi: 10.1039/c2an35582e.

[20] M. Romeo and M. Diem, “Correction of dispersive line shape artifact observed in diffuse reflection infrared spectroscopy and absorption/reflection (transflection) infrared micro-spectroscopy,” *Vib. Spectrosc.*, vol. 38, no. 1, pp. 129–132, 2005, doi: 10.1016/j.vibspec.2005.04.003.

[21] C. Hughes *et al.*, “Assessing the challenges of Fourier transform infrared spectroscopic analysis of blood serum,” *J. Biophotonics*, vol. 7, no. 3–4, pp. 180–188, 2014, doi: 10.1002/jbio.201300167.

[22] M. J. Baker *et al.*, “Developing and understanding biofluid vibrational

spectroscopy: A critical review," *Chem. Soc. Rev.*, vol. 45, no. 7, pp. 1803–1818, 2016, doi: 10.1039/c5cs00585j.

[23] H. Fabian, P. Lasch, and D. Naumann, "Analysis of biofluids in aqueous environment based on mid-infrared spectroscopy," *J. Biomed. Opt.*, vol. 10, no. 3, pp. 1–10, 2005, doi: 10.1117/1.1917844.

[24] F. L. Martin *et al.*, "Distinguishing cell types or populations based on the computational analysis of their infrared spectra," *Nat. Protoc.*, vol. 5, no. 11, pp. 1748–1760, 2010, doi: 10.1038/nprot.2010.133.

[25] K. M. Dorling and M. J. Baker, "Highlighting attenuated total reflection Fourier transform infrared spectroscopy for rapid serum analysis," *Trends in Biotechnology*, vol. 31, no. 6. pp. 327–328, 2013, doi: 10.1016/j.tibtech.2013.03.010.

[26] D. Ami, P. Mereghetti, and S. Maria, "Multivariate Analysis for Fourier Transform Infrared Spectra of Complex Biological Systems and Processes," in *Multivariate Analysis in Management, Engineering and the Sciences*, 2013, pp. 189–220.

[27] J. Grdadolnik, "ATR-FTIR spectroscopy: Its advantages and limitations," *Acta Chim. Slov.*, vol. 49, no. 3, pp. 631–642, 2002.

[28] H. J. Gulley-stahl, J. A. Haas, K. A. Schmidt, A. P. Evan, and A. J. Sommer, "Attenuated Total Internal Reflection Infrared Spectroscopy (ATR-FTIR): A Quantitative Approach for Kidney Stones Analysis," *Appl. Spectrosc.*, vol. 63, no. 7, pp. 759–766, 2009, doi: 10.1366/000370209788701044.Attenuated.

[29] P. Bassan *et al.*, "FTIR microscopy of biological cells and tissue: Data analysis using resonant Mie scattering (RMieS) EMSC algorithm," *Analyst*, vol. 137, no. 6, pp. 1370–1377, 2012, doi: 10.1039/c2an16088a.

[30] S. W. King and M. Milosevic, "A method to extract absorption coefficient of thin films from transmission spectra of the films on thick substrates," *J. Appl. Phys.*, vol. 111, no. 7, pp. 1–9, 2012, doi: 10.1063/1.3700178.

[31] K. L. A. Chan and S. G. Kazarian, "Correcting the effect of refraction and dispersion of light in FT-IR spectroscopic imaging in transmission through thick infrared windows," *Anal. Chem.*, vol. 85, no. 2, pp. 1029–1036, 2013, doi: 10.1021/ac302846d.

[32] T. Konevskikh, A. Ponossov, R. Blümel, R. Lukacs, and A. Kohler, "Fringes in FTIR spectroscopy revisited: Understanding and modelling fringes in infrared spectroscopy of thin films," *Analyst*, vol. 140, no. 12, pp. 3969–3980, 2015, doi: 10.1039/c4an02343a.

[33] S. Hariri *et al.*, "FTIR bio-spectroscopy scattering correction using natural biological characteristics of different cell lines," *Analyst*, vol. 144, no. 19, pp. 5810–5828, 2019, doi: 10.1039/c9an00811j.

[34] P. Bassan *et al.*, "Reflection contributions to the dispersion artefact in FTIR spectra of single biological cells," *Analyst*, vol. 134, no. 6, pp. 1171–1175,

2009, doi: 10.1039/b821349f.

[35] S. Barzegari Banadkoki, F. Torkamani Azar, and F. Hosseini Shirazi, "Estimation and Reduction of Resonant Mie Scattering (RMieS) From IR Spectra of Biological Cells by Optimization Algorithm," *J. Med. Biol. Eng.*, vol. 39, no. 3, pp. 431–441, 2018, doi: 10.1007/s40846-018-0423-9.

[36] P. Bassan, H. J. Byrne, F. Bonnier, J. Lee, and P. Gardner, "Resonant Mie scattering in infrared spectroscopy of biological materials – understanding the 'dispersion artefact,'" *Analyst*, vol. 134, no. 1, pp. 1586–1593, 2009, doi: 10.1039/b904808a.

[37] M. Hancer, R. P. Sperline, and J. D. Miller, "Anomalous dispersion effects in the IR-ATR spectroscopy of water," *Appl. Spectrosc.*, vol. 54, no. 1, pp. 138–143, 2000, doi: 10.1366/0003702001948222.

[38] P. Bassan *et al.*, "Resonant Mie Scattering (RMieS) correction of infrared spectra from highly scattering biological samples," *Analyst*, vol. 135, no. 5, pp. 268–277, 2010, doi: 10.1039/b921056c.

[39] P. Bassan *et al.*, "The inherent problem of transflection-mode infrared spectroscopic microscopy and the ramifications for biomedical single point and imaging applications," *Analyst*, vol. 138, no. 1, pp. 144–157, 2013, doi: 10.1039/c2an36090j.

[40] A. L. Mitchell, K. B. Gajjar, G. Theophilou, F. L. Martin, and P. L. Martin-hirsch, "Vibrational spectroscopy of biofluids for disease screening or diagnosis : translation from the laboratory to a clinical setting," *J. Biophotonics*, vol. 7, no. 3–4, pp. 153–165, 2014, doi: 10.1002/jbio.201400018.

[41] F. Lyng, E. Gazi, and P. Gardner, "Preparation of Tissues and Cells for Infrared and Raman Spectroscopy and Imaging," *Biomed. Appl. Synchrotron Infrared Microspectrosc.*, pp. 145–189, 2011, doi: 10.1515/9783110276138.7.

[42] S. Sabbatini, C. Conti, G. Orilisi, and E. Giorgini, "Infrared spectroscopy as a new tool for studying single living cells: Is there a niche?," *Biomed. Spectrosc. Imaging*, vol. 6, no. 3–4, pp. 85–99, 2017, doi: 10.3233/bsi-170171.

[43] G. Sankari *et al.*, "Analysis of serum immunoglobulins using Fourier transform infrared spectral measurements," *Biol. Med.*, vol. 2, no. 3, pp. 42–48, 2010, doi: 10.4172/0974-8369.1000066.

[44] V. Zohdi, D. R. Whelan, B. R. Wood, J. T. Pearson, K. R. Bambery, and M. J. Black, "Importance of tissue preparation methods in FTIR micro-spectroscopical analysis of biological tissues: 'Traps for new users,'" *PLoS One*, vol. 10, no. 2, pp. 1–11, 2015, doi: 10.1371/journal.pone.0116491.

[45] Z. Movasaghi, S. Rehman, and I. U. Rehman, "Fourier transform infrared (FTIR) spectroscopy of biological tissues," *Appl. Spectrosc. Rev.*, vol. 43, no. 5, pp. 134–179, 2008, doi: 10.1080/05704928.2016.1230863.

[46] M. C. D. Santos, Y. M. Nascimento, J. M. G. Araújo, and K. M. G. Lima, "ATR-FTIR spectroscopy coupled with multivariate analysis techniques for the

identification of DENV-3 in different concentrations in blood and serum: A new approach," *RSC Adv.*, vol. 7, no. 41, pp. 25640–25649, 2017, doi: 10.1039/c7ra03361c.

[47] H. Martens, S. W. Bruun, I. Adt, G. D. Sockalingum, and A. Kohler, "Pre-processing in biochemometrics: Correction for path-length and temperature effects of water in FTIR bio-spectroscopy by EMSC," *J. Chemom.*, vol. 20, no. 8–10, pp. 402–417, 2007, doi: 10.1002/cem.1015.

[48] A. Barth, "Infrared spectroscopy of proteins," *Biochim. Biophys. Acta*, vol. 1767, no. 9, pp. 1073–1101, 2007, doi: 10.1016/j.bbabio.2007.06.004.

[49] S. J. Parikh, K. W. Goyne, A. J. Margenot, F. N. D. Mukome, and F. J. Calderón, "Soil chemical insights provided through vibrational spectroscopy," in *Advances in Agronomy*, vol. 126, Elsevier, 2014, pp. 1–148.

[50] F. Bonnier, M. J. Baker, and H. J. Byrne, "Vibrational spectroscopic analysis of body fluids: Avoiding molecular contamination using centrifugal filtration," *Anal. Methods*, vol. 6, no. 14, pp. 5155–5160, 2014, doi: 10.1039/c4ay00891j.

[51] A. Sala *et al.*, "Rapid analysis of disease state in liquid human serum combining infrared spectroscopy and 'digital drying,'" *J. Biophotonics*, vol. 13, no. 9, pp. 1–10, 2020, doi: 10.1002/jbio.202000118.

[52] J. M. Cameron, H. J. Butler, D. S. Palmer, and M. J. Baker, "Biofluid spectroscopic disease diagnostics: A review on the processes and spectral impact of drying," *J. Biophotonics*, vol. 11, pp. 1–12, 2018, doi: 10.1002/jbio.201700299.

[53] L. Lovergne, G. Clemens, V. Untereiner, R. A. Lukaszweski, G. D. Sockalingum, and M. J. Baker, "Investigating Optimum Sample Preparation for Infrared Spectroscopic Serum Diagnostics," *Anal. Methods*, vol. 7, no. 17, pp. 6921–7382, 2015, doi: 10.1039/C5AY00502G.

[54] L. Lovergne, P. Bouzy, V. Untereiner, and R. Garnotel, "Biofluid infrared spectro-diagnostics: pre-analytical considerations for clinical applications," *Faraday Discuss.*, vol. 187, pp. 521–537, 2016, doi: 10.1039/c5fd00184f.

[55] M. J. Baker *et al.*, "Clinical applications of Infrared and Raman spectroscopy : state of play and future challenges," *Analyst*, vol. 143, pp. 1735–1757, 2018, doi: 10.1039/c7an01871a.

[56] Å. Rinnan, "Pre-processing in vibrational spectroscopy-when, why and how," *Anal. Methods*, vol. 6, no. 18, pp. 7124–7129, 2014, doi: 10.1039/c3ay42270d.

[57] C. L. M. Morais, K. M. G. Lima, M. Singh, and F. L. Martin, "Tutorial: multivariate classification for vibrational spectroscopy in biological samples," *Nat. Protoc.*, vol. 15, no. 7, pp. 2143–2162, 2020, doi: 10.1038/s41596-020-0322-8.

[58] H. J. Butler, B. R. Smith, R. Fritzsch, P. Radhakrishnan, D. S. Palmer, and M. J. Baker, "Optimised spectral pre-processing for discrimination of biofluids via ATR-FTIR spectroscopy," *Analyst*, vol. 143, no. 24, pp. 6121–6134, 2018, doi:

10.1039/c8an01384e.

[59] R. Gautam, S. Vanga, F. Ariese, and S. Umapathy, “Review of multidimensional data processing approaches for Raman and infrared spectroscopy,” *EPJ Tech. Instrum.*, vol. 2, no. 8, pp. 1–38, 2015, doi: 10.1140/epjti/s40485-015-0018-6.

[60] F. Qian, Y. Wu, and P. Hao, “A fully automated algorithm of baseline correction based on wavelet feature points and segment interpolation,” *Opt. Laser Technol.*, vol. 96, pp. 202–207, 2017, doi: 10.1016/j.optlastec.2017.05.021.

[61] J. Trevisan, P. P. Angelov, P. L. Carmichael, A. D. Scott, and F. L. Martin, “Extracting biological information with computational analysis of Fourier-transform infrared (FTIR) biospectroscopy datasets: Current practices to future perspectives,” *Analyst*, vol. 137, no. 14, pp. 3202–3215, 2012, doi: 10.1039/c2an16300d.

[62] S. Giguere, T. Boucher, C. J. Carey, S. Mahadevan, and M. D. Dyar, “A Fully Customized Baseline Removal Framework for Spectroscopic Applications,” *Appl. Spectrosc.*, vol. 71, no. 7, pp. 1457–1470, 2017, doi: 10.1177/0003702817695624.

[63] B. J. Lee *et al.*, “Discrimination and prediction of the origin of Chinese and Korean soybeans using Fourier transform infrared spectrometry (FT-IR) with multivariate statistical analysis,” *PLoS One*, vol. 13, no. 4, pp. 1–16, 2018, doi: 10.1371/journal.pone.0196315.

[64] O. A. Akanbi, I. S. Amiri, E. Fazeldehkordi, O. A. Akanbi, I. S. Amiri, and E. Fazeldehkordi, “Chapter 4 – Feature Extraction,” in *A Machine-Learning Approach to Phishing Detection and Defense*, 2015, pp. 45–54.

[65] J. Han, M. Kamber, and J. Pei, “Data Preprocessing,” in *Data Mining*, 2012, pp. 83–124.

[66] K. R. Bambery, B. R. Wood, and D. McNaughton, “Resonant Mie scattering (RMieS) correction applied to FTIR images of biological tissue samples,” *Analyst*, vol. 137, no. 1, pp. 126–132, 2012, doi: 10.1039/c1an15628d.

[67] N. K. Afseth and A. Kohler, “Extended multiplicative signal correction in vibrational spectroscopy, a tutorial,” *Chemom. Intell. Lab. Syst.*, vol. 117, pp. 92–99, 2012, doi: 10.1016/j.chemolab.2012.03.004.

[68] A. Subramanian and L. Rodriguez-saona, “Chapter 7 - Fourier Transform Infrared (FTIR) Spectroscopy,” in *Infrared Spectroscopy for Food Quality Analysis and Control*, 2009, pp. 146–173.

[69] X. Zhang *et al.*, “Profiling serologic biomarkers in cirrhotic patients via high-throughput Fourier transform infrared spectroscopy: Toward a new diagnostic tool of hepatocellular carcinoma,” *Transl. Res.*, vol. 162, no. 5, pp. 279–286, 2013, doi: 10.1016/j.trsl.2013.07.007.

[70] P. Lasch, “Spectral pre-processing for biomedical vibrational spectroscopy and microspectroscopic imaging,” *Chemom. Intell. Lab. Syst.*, vol. 117, pp. 100–

114, 2012, doi: 10.1016/j.chemolab.2012.03.011.

[71] B. Hofko, M. Z. Alavi, H. Grothe, D. Jones, and J. Harvey, “Repeatability and sensitivity of FTIR ATR spectral analysis methods for bituminous binders,” *Mater. Struct.*, vol. 50, no. 187, pp. 1–15, 2017, doi: 10.1617/s11527-017-1059-x.

[72] D. A. Armbruster and T. Pry, “Limit of Blank, Limit of Detection and Limit of Quantification,” *Clin Biochem Rev*, vol. 29, no. I, pp. 49–52, 2008.

[73] H. Senyuva, “Analytical Methods: Overview of Methods of Analysis for Chemical Hazards,” in *Encyclopedia of Food Safety*, vol. 1, Elsevier Ltd., 2014, pp. 152–158.

[74] A. Shrivastava and V. Gupta, “Methods for the determination of limit of detection and limit of quantitation of the analytical methods,” *Chronicles Young Sci.*, vol. 2, no. 1, p. 21, 2011, doi: 10.4103/2229-5186.79345.

[75] E. A. Abdel Maksoud, S. Barakat, and M. Elmogy, “Medical Images Analysis Based on Multilabel Classification,” in *Machine Learning in Bio-Signal Analysis and Diagnostic Imaging*, Elsevier Inc., 2019, pp. 209–245.

[76] E. K. Kemsley, “Discriminant analysis of high-dimensional data: A comparison of principal components analysis and partial least squares data reduction methods,” *Chemom. Intell. Lab. Syst.*, vol. 33, no. 1, pp. 47–61, 1996, doi: 10.1016/0169-7439(95)00090-9.

[77] R. Goodacre, “Explanatory analysis of spectroscopic data using machine learning of simple, interpretable rules,” *Vib. Spectrosc.*, vol. 32, pp. 33–45, 2003, doi: 10.1016/S0924-2031(03)00045-6.

[78] Y. El Khoury *et al.*, “Serum-based differentiation between multiple sclerosis and amyotrophic lateral sclerosis by Random Forest classification of FTIR spectra,” *Analyst*, vol. 144, no. 15, pp. 4647–4652, 2019, doi: 10.1039/c9an00754g.

[79] B. R. Smith *et al.*, “Combining random forest and 2D correlation analysis to identify serum spectral signatures for neuro-oncology,” *Analyst*, vol. 141, no. 12, pp. 3668–3678, 2016, doi: 10.1039/c5an02452h.

[80] C. L. Song, M. Z. Vardaki, R. D. Goldin, and S. G. Kazarian, “Fourier transform infrared spectroscopic imaging of colon tissues: evaluating the significance of amide I and C–H stretching bands in diagnostic applications with machine learning,” *Anal. Bioanal. Chem.*, vol. 411, no. 26, pp. 6969–6981, 2019, doi: 10.1007/s00216-019-02069-6.

[81] H. Chen, G. Tang, W. Ai, and K. Cai, “Use of Random Forest in FTIR Analysis of LDL Cholesterol and Tri-Glycerides for Hyperlipidemia,” pp. 5–8, 2015, doi: 10.1002/btpr.2161.

[82] F. Großerueschkamp *et al.*, “Marker-free automated histopathological annotation of lung tumour subtypes by FTIR imaging,” *Analyst*, vol. 140, no. 7, pp. 2114–2120, 2015, doi: 10.1039/c4an01978d.

[83] G. O. Ildiz, S. Bayari, A. Karadag, and E. Kaygisiz, “Fourier Transform Infrared Spectroscopy Based Complementary Diagnosis Tool for Autism Spectrum Disorder in Children and Adolescents,” *Molecules*, vol. 25, pp. 2079–2091, 2020.

[84] S. Roy, D. Perez-Guaita, S. Bowden, P. Heraud, and B. R. Wood, “Spectroscopy goes viral: Diagnosis of hepatitis B and C virus infection from human sera using ATR-FTIR spectroscopy,” *Clin. Spectrosc.*, vol. 1, no. 100001, pp. 1–14, 2019, doi: 10.1016/j.clispe.2020.100001.

[85] B. Worley and R. Powers, “Multivariate Analysis in Metabolomics,” *Curr. Metabolomics*, vol. 1, no. 1, pp. 92–107, 2013, doi: 10.2174/2213235x11301010092.

[86] G. Gauglitz and T. Vo-Dinh, “Chapter 22 - Optical Spectroscopy,” in *Handbook of Spectroscopy*, vol. 1, 2003, pp. 441–468.

[87] I. A. Cowe and J. W. McNicol, “The Use of Principal Components in the Analysis of Near-Infrared Spectra.,” *Appl. Spectrosc.*, vol. 39, no. 2, pp. 257–266, 1985, doi: 10.1366/0003702854248944.

[88] J. G. Kelly *et al.*, “Biospectroscopy to metabolically profile biomolecular structure: A multistage approach linking computational analysis with biomarkers,” *J. Proteome Res.*, vol. 10, no. 4, pp. 1437–1448, 2011, doi: 10.1021/pr101067u.

[89] T. Korenius, J. Laurikkala, and M. Juhola, “On principal component analysis, cosine and Euclidean measures in information retrieval,” *Inf. Sci. (Ny).*, vol. 177, no. 22, pp. 4893–4905, 2007, doi: 10.1016/j.ins.2007.05.027.

[90] I. Jolliffe, “Principal components analysis,” *Aten. Primaria*, vol. 12, no. 6, pp. 333–338, 1993, doi: 10.4135/9781446251119.n72.

[91] M. G. Borgognone, J. Bussi, and G. Hough, “Principal component analysis in sensory analysis: Covariance or correlation matrix?,” *Food Qual. Prefer.*, vol. 12, no. 5–7, pp. 323–326, 2001, doi: 10.1016/S0950-3293(01)00017-9.

[92] D. S. Sivia and J. Skilling, *Data Analysis: A Bayesian Tutorial (Second Edition)*, vol. 2. 2006.

[93] W. B. Wu and H. Xiao, “Covariance Matrix Estimation in Time Series,” in *Handbook of Statistics*, vol. 30, Elsevier B.V., 2012, pp. 187–209.

[94] I. T. Jollife and J. Cadima, “Principal component analysis: A review and recent developments,” *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.*, vol. 374, no. 20150202, pp. 1–16, 2016, doi: 10.1098/rsta.2015.0202.

[95] R. B. Cattell, “The Scree Test for the number of factors,” *Multivariate Behav. Res.*, vol. 1, no. 2, pp. 245–276, 1966, doi: 10.1207/s15327906mbr0102.

[96] S. Wold, “Cross-Validatory Estimation of the Number of Components in Factor and Principal Components Models,” *Technometrics*, vol. 20, no. 4, pp. 397–405, 1978.

[97] L. Wang and B. Mizaikoff, “Application of multivariate data-analysis techniques to biomedical diagnostics based on mid-infrared spectroscopy,” *Anal. Bioanal. Chem.*, vol. 391, no. 5, pp. 1641–1654, 2008, doi: 10.1007/s00216-008-1989-9.

[98] H. J. Byrne, F. Bonnier, J. McIntyre, and D. R. Parachalil, “Quantitative analysis of human blood serum using vibrational spectroscopy,” *Clin. Spectrosc.*, vol. 2, no. March, p. 100004, 2020, doi: 10.1016/j.clispe.2020.100004.

[99] D. Cozzolino, “The Sample , the Spectra and the Maths — The Critical Pillars in the Development of Robust and Sound Applications of Vibrational Spectroscopy,” *Molecules*, vol. 25, no. 3674, pp. 1–16, 2020.

[100] S. Wold, M. Sjöström, and L. Eriksson, “PLS-regression: A basic tool of chemometrics,” *Chemom. Intell. Lab. Syst.*, vol. 58, no. 2, pp. 109–130, 2001, doi: 10.1016/S0169-7439(01)00155-1.

[101] R. Rosipal, “Nonlinear partial least squares: An overview,” in *Chemoinformatics and Advanced Machine Learning Perspectives: Complex Computational Methods and Collaborative Techniques*, 2010, pp. 169–189.

[102] Y. Liu and W. Rayens, “PLS and dimension reduction for classification,” *Comput. Stat.*, vol. 22, no. 2, pp. 189–208, 2007, doi: 10.1007/s00180-007-0039-y.

[103] A.-L. Boulesteix, “PLS Dimension Reduction for Classification with Microarray Data,” *Stat. Appl. Genet. Mol. Biol.*, vol. 3, no. 1, pp. 1–30, 2004, doi: 10.2202/1544-6115.1075.

[104] P. Geladi and B. R. Kowalski, “Partial Least-Squares Regression: A Tutorial,” *Anal. Chim. Acta*, vol. 185, pp. 1–17, 1986.

[105] M. Barker and W. Rayens, “Partial least squares for discrimination,” *J. Chemom.*, vol. 17, no. 3, pp. 166–173, 2003, doi: 10.1002/cem.785.

[106] Q. S. Xu, Y. Z. Liang, and Y. P. Du, “Monte Carlo cross-validation for selecting a model and estimating the prediction error in multivariate calibration,” *J. Chemom.*, vol. 18, no. 2, pp. 112–120, 2004, doi: 10.1002/cem.858.

[107] S. Lee, H. Choi, K. Cha, and H. Chung, “Random forest as a potential multivariate method for near-infrared (NIR) spectroscopic analysis of complex mixture samples: Gasoline and naphtha,” *Microchem. J.*, vol. 110, pp. 739–748, 2013, doi: 10.1016/j.microc.2013.08.007.

[108] T. M. Oshiro, P. S. Perez, and J. A. Baranauskas, “How many trees in a random forest?,” in *Machine Learning and Data Mining in Pattern Recognition*, 2012, vol. MLDM, pp. 154–168, doi: 10.1007/978-3-642-31537-4_13.

[109] Y. Qi, “Chapter 11 - Random Forest for Bioinformatics,” in *Ensemble Machine Learning: Methods and Applications*, 2012, pp. 307–323.

[110] F. A. Farris, “The Gini Index and Measures of Inequality,” *Am. Math. Mon.*, vol. 117, no. 10, pp. 851–864, 2010, doi: 10.4169/000298910X523344.

[111] M. Pal, “Random forest classifier for remote sensing classification,” *Int. J. Remote Sens.*, vol. 26, no. 1, pp. 217–222, 2005, doi: 10.1080/01431160412331269698.

[112] S. Nembrini, I. R. König, and M. N. Wright, “The revival of the Gini importance?,” *Bioinformatics*, vol. 34, no. 21, pp. 3711–3718, 2018, doi: 10.1093/bioinformatics/bty373.

[113] K. J. Archer and R. V. Kimes, “Empirical characterization of random forest variable importance measures,” *Comput. Stat. Data Anal.*, vol. 52, no. 4, pp. 2249–2260, 2008, doi: 10.1016/j.csda.2007.08.015.

[114] X. Guo, Y. Yin, C. Dong, G. Yang, and G. Zhou, “On the class imbalance problem,” in *4th International Conference on Natural Computation, ICNC*, 2008, vol. 4, pp. 192–201, doi: 10.1109/ICNC.2008.871.

[115] X. Y. Liu, J. Wu, and Z. H. Zhou, “Exploratory undersampling for class-imbalance learning,” *IEEE Trans. Syst. Man, Cybern. Part B Cybern.*, vol. 39, no. 2, pp. 539–550, 2009, doi: 10.1109/TSMCB.2008.2007853.

[116] A. Abraham and S. M. A. Elrahman, “A Review of Class Imbalance Problem,” *J. Netw. Innov. Comput.*, vol. 1, pp. 332–340, 2013.

[117] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: Synthetic Minority Over-Sampling Technique,” *J. Artif. Intell. Res.*, vol. 16, no. Sept. 28, pp. 321–357, 2002, doi: 10.1613/jair.953.

[118] W. Juanjuan, X. Mantao, W. Hui, and Z. Jiwu, “Classification of imbalanced data by using the SMOTE algorithm and locally linear embedding,” in *2006 8th international Conference on Signal Processing*, 2006, pp. 1–4, doi: 10.1109/ICOSP.2006.345752.

[119] R. Alencar, “Resampling strategies for imbalanced datasets,” *Kaggle*, 2018. https://www.kaggle.com/rafjaa/resampling-strategies-for-imbalanced-datasets (accessed Feb. 07, 2021).

[120] G. Boothroyd, “Product design for manufacture and assembly,” in *Computer-Aided Design*, vol. 26, no. 7, 1994, pp. 505–520.

[121] H. Kaeslin, “Chapter 1 - Introduction to Microelectronics,” in *Top-Down Digital VLSI Design*, 2015, pp. 1–40.

[122] C. K. Chua, S. H. Teh, and R. K. L. Gay, “Rapid Prototyping Versus Virtual Prototyping in Product Design and Manufacturing,” *Int. J. Adv. Manuf. Technol.*, vol. 15, pp. 597–603, 1999.

[123] T. Rayna and L. Striukova, “From rapid prototyping to home fabrication: How 3D printing is changing business model innovation,” *Technol. Forecast. Soc. Change*, vol. 102, pp. 214–224, 2016, doi: 10.1016/j.techfore.2015.07.023.

[124] P. Shah, R. Racasan, and P. Bills, “Comparison of different additive manufacturing methods using computed tomography,” *Case Stud. Nondestruct. Test. Eval.*, vol. 6, no. B, pp. 69–78, 2016, doi: 10.1016/j.csndt.2016.05.008.

[125] S.-H. Ahn, M. Montero, D. Odell, S. Roundy, and P. K. Wright, “Anisotropic material properties of fused deposition modeling ABS,” *Rapid Prototyp. J.*, vol. 8, no. 4, pp. 248–257, 2002, doi: 10.1108/13552540210441166.

[126] A. Bellini and S. Güçeri, “Mechanical characterization of parts fabricated using fused deposition modeling,” *Rapid Prototyp. J.*, vol. 9, no. 4, pp. 252–264, 2003, doi: 10.1108/13552540310489631.

[127] R. Jerez-Mesa, J. A. Travieso-Rodriguez, X. Corbella, R. Busqué, and G. Gomez-Gras, “Finite element analysis of the thermal behavior of a RepRap 3D printer liquefier,” *Mechatronics*, vol. 36, pp. 119–126, 2016, doi: 10.1016/j.mechatronics.2016.04.007.

[128] J. P. Kruth, P. Mercelis, J. Van Vaerenbergh, L. Froyen, and M. Rombouts, “Binding mechanisms in selective laser sintering and selective laser melting,” *Rapid Prototyp. J.*, vol. 11, no. 1, pp. 26–36, 2005, doi: 10.1108/13552540510573365.

[129] K. S. Prakash, T. Nancharaih, and V. V. S. Rao, “Additive Manufacturing Techniques in Manufacturing -An Overview,” *Mater. Today Proc.*, vol. 5, no. 2, pp. 3873–3882, 2018, doi: 10.1016/j.matpr.2017.11.642.

[130] ManufacturingGuide, “Direct Laser Metal Sintering, DLMS.” https://www.manufacturingguide.com/en/direct-laser-metal-sintering-dmls.

[131] W. G. Frizelle, “Injection Molding Technology,” in *Applied Plastics Engineering Handbook: Processing, Materials, and Applications: Second Edition*, Second Edi., Elsevier Inc., 2017, pp. 191–202.

[132] M. T. MacLean-Blevins, “Process selection—which plastics process to use?,” in *Designing Successful Products with Plastics*, 2018, pp. 51–77.

[133] V. Goodship, “Injection Molding of Thermoplastics,” in *Design and Manufacture of Plastic Components for Multifunctionality*, Elsevier Inc., 2016, pp. 103–170.

[134] R. Surace, G. Trotta, V. Bellantone, and I. Fassi, “The Micro Injection Moulding Process for Polymeric Components Manufacturing,” in *New Technologies - Trends, Innovations and Research*, 2012, pp. 65–90.

[135] U. M. Attia, S. Marson, and J. R. Alcock, “Micro-injection moulding of polymer microfluidic devices,” *Microfluid. Nanofluidics*, vol. 7, no. 1, pp. 1–28, 2009, doi: 10.1007/s10404-009-0421-x.

[136] The Open University, “Injection moulding,” *OpenLearn*, 2017. https://www.open.edu/openlearn/science-maths-technology/engineering-technology/manupedia/injection-moulding (accessed Nov. 13, 2020).

[137] S. Ebnesajjad, “Injection Molding,” in *Fluoroplastics (Second Edition)*, vol. 2, 2015, pp. 236–281.

[138] L. Csepregi, “Micromechanics: A Silicon Microfabrication Technology,” *Microelectron. Eng.*, vol. 3, no. 1–4, pp. 221–234, 1985.

[139] O. Tabata, R. Asahi, H. Funabashi, K. Shimaoka, and S. Sugiyama, “Anisotropic etching of silicon in TMAH solutions,” *Sensors Actuators A.*, vol. 34, no. 1, pp. 51–57, 1992, doi: 10.1016/0924-4247(92)80139-T.

[140] W. H. Smart and K. Subramanian, “The Use of Silicon Microfabrication Technology in Painless Blood Glucose Monitoring,” *Diabetes Technol. Ther.*, vol. 2, no. 4, pp. 549–559, 2000.

[141] K. E. Bean, “Anisotropic Etching of Silicon,” *IEEE Trans. Electron Devices*, vol. 25, no. 10, pp. 1185–1193, 1978.

[142] A. Noori, S. Upadhyaya, and P. R. Selvanganapathy, “Materials and Microfabrication Processes for Microfluidic Devices,” in *Materials and Microfabrication Processes for Microfluidic Devices*, 2008, pp. 35–92.

[143] M. L. Reed and G. K. Fedder, “Photolithographic Microfabrication,” in *Handbook of Sensors and Actuators*, 1998, pp. 13–61.

[144] M. Hopcroft, “What is the Young’s Modulus of Silicon?,” *J. Microelectromechanical Syst.*, vol. 19, no. 2, pp. 229–238, 2010.

[145] M. A. Gosálvez, I. Zubel, and E. Viinikka, “Wet Etching of Silicon,” in *Handbook of Silicon Based MEMS Materials and Technologies: Second Edition*, 2015, pp. 470–502.

[146] N.-T. Nguyen, “Fabrication technologies,” in *Micromixers (Second Edition)*, 2012, pp. 113–161.

[147] M. C. B. A. Michielsen, V. B. Marriott, J. J. Ponjee, H. van der Wel, F. J. Touwslager, and J. A. H. M. Moonen, “Priming of Silicon Substrates with Trimethysilyl Containing Compounds,” *Microelectron. Eng.*, vol. 11, pp. 475–480, 1990.

[148] C. L. C. Smith, N. Stenger, A. Kristensen, N. A. Mortensen, and S. I. Bozhevolnyi, “Gap and channeled plasmons in tapered grooves: A review,” *Nanoscale*, vol. 7, no. 21, pp. 9355–9386, 2015, doi: 10.1039/c5nr01282a.

[149] L. Miloudi *et al.*, “ATR-IR coupled to partial least squares regression (PLSR) for monitoring an encapsulated active molecule in complex semi-solid formulations,” *Analyst*, vol. 143, no. 10, pp. 2377–2389, 2018, doi: 10.1039/c8an00547h.

[150] B. R. Smith, M. J. Baker, and D. S. Palmer, “PRFFECT: A versatile tool for spectroscopists,” *Chemom. Intell. Lab. Syst.*, vol. 172, no. October 2017, pp. 33–42, 2018, doi: 10.1016/j.chemolab.2017.10.024.

Chapter 3: Design and Fabrication of a High-Throughput ATR-FTIR Platform

3.1. Introduction

ATR-FTIR has irrefutable potential as a powerful IVD device yet still eludes the clinical laboratory [1]–[3]. One of the main stumbling blocks for the translation of ATR-FTIR spectroscopy to the clinic is thought to be a single point of analysis which prohibits high-throughput clinical workflow [4], [5]. ATR-FTIR does not require expensive reagents or skilled operators, but samples must be individually dried upon the IRE, analysed, then removed before another sample may be processed. This is a time-limiting factor, compounded by the customary requirement to perform technical and sample replicates for accuracy, inherently prohibiting the instalment of a batched processing system for high-throughput sample analysis. The IRE is the crux of the issue, as current mainstream designs have highly expensive fixed components. There are two features which must be developed to achieve high-throughput clinical ATR-FTIR spectroscopy:

1. Low-cost and readily interchangeable IREs capable of interrogating biological samples.
2. An accessory module for interfacing these interchangeable IREs to an FTIR spectrometer platform.

For clinical applications, a specific spectrometer model should be used as it will be difficult to validate diagnostic tests if different spectrometers that acquire spectra with varying degrees of spectral quality are used. Regulatory approval would therefore be easier to obtain on a dedicated spectrometer make and model. However, for non-clinical applications users may benefit from the flexibility offered by a

universal accessory module capable of interfacing any spectrometer platform. The IRE is the salient component of an ATR-FTIR spectrometer, differentiating it from alternative FTIR modes and is the site at which a sample is interrogated by the IR beam. It is made from a material with high refractive index and excellent transmissivity of the IR spectrum. Di and Ge among several other materials adequately fit this profile, but usually these materials cannot be manufactured at the low costs conducive to a high-throughput ATR workflow. Si is widely avoided due to its high absorptivity of IR light in the spectral 'fingerprint' region critical for biological sample analysis [6]. Silicon substrates used in transflection and transmission FTIR modes bypass this limitation by having small pathlengths through the silicon crystal. Emulating a short IR pathlength through a low-cost silicon IRE therefore presents a promising option to resolve issues currently inhibiting high-throughput ATR. Micromachined single-bounce silicon IREs that reduce signal loss via Si lattice absorption have previously been demonstrated [7]. However, the question remains what the optimal silicon crystal thickness and geometry would be for biological sample analysis. Cheaper fabrication and greater material availability of silicon permits more elaborate IRE design, enabling multiple optically active surfaces to be etched into a single unit for sample replicates. Presented are the design decisions, development pathway, and experimentation of a low-cost disposable SIRE sample slide. The development of a custom ATR accessory module for semi-automated indexing of sample slides across the IR aperture of FTIR spectrometers is also described. Feasibility of scaled production of sample slides and potential fabrication issues is discussed.

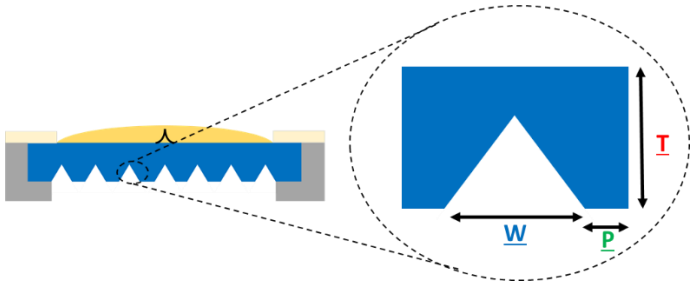
3.2. Materials and Methods

3.2.1. Silicon Internal Reflection Element Design Experimentation

The SIRE sample slide is designed to host four optically active regions: three for sample triplicates and one for a background scan. Background scans serve to automatically remove IR signal of the environment from sample scans. Sample and background sites are positioned at 9mm intervals to match the pitch of multi-channel pipettes aiding future integration into high-throughput workflows. Optically active areas comprise a 5x5mm arrays of v-shaped grooves aligned parallel to the incident beam allowing the beam to enter through numerous points then exit after internal reflection and avoid multiple internal reflections. The reader is referred to chapter 2 section 2.5.3 for particulars on how v-grooves were etched into the silicon.

To investigate the effects of silicon thickness on signal extinction, three different thicknesses of silicon wafers with the following specifications were acquired in duplicate: 150mm diameter, P-type, Double-Side Polished (DSP) Czochralski grown (Cz) crystals (Semefab, Glenrothes). Ideally, undoped Si wafers would be used, but because dopants are desirable in the majority of Si wafer applications it is difficult to obtain low quantities of undoped Si wafers at reasonable cost and in reliable supply. However, even at typical dopant concentrations between approximately 10^{13} and 10^{16} atoms per centimetre cubed, Si wafer purity is still greater than 99.9999% [8]. Thus, absorbances due to dopants are sometimes considered negligible [9]. For practical reasons, P-type Si was selected for all Si wafers.

36 different SIRE designs were fabricated with eight replicates of each design. Three design parameters were investigated: wafer thickness, v-groove width, and v-groove pitch (Figure 3.1). Table 3.1 presents the dimensions each parameter was tested at. Cz refers to the method of silicon crystal growth technique. The Czochralski (Cz) technique is by far the more common, and cheaper, of the two predominant silicon fabrication techniques. The other approach is known as the Float-Zone (FZ) method which produces extremely high purity silicon wafers, but is limited in both its production capacity and maximum six-inch wafer size [10]. The ramifications of these limitations faced by FZ wafers is a far higher cost per unit SIRE in comparison to Cz wafers and far lower availability. For this reason, a compromise has been made and Cz grown wafers have been used in all SIRE devices.



Design parameter	Dimension (µm)			
SIRE thickness		380	525	675
Groove pitch		25	50	100
Groove width	100	150	200	250

Figure 3.4. *Cross-sectional schematic of a SIRE optically active area with design parameters under investigation notated. These are: SIRE thickness, T, groove width, W, groove pitch, P.*

Table 3.1. *Dimension levels the investigated design parameters were tested at.*

SIRE slides feature four optically active areas, three of which are available for the analysis of a sample. 3µl of Human Pooled Serum (HPS) (Sigma-Aldrich, USA) was pipetted into the three sample areas of each SIRE design (Table 3.1) and air dried for 1 hour. Slides were analysed manually on a Spectrum2 IR spectrometer

(PerkinElmer, USA) and a Quest ATR accessory (Specac, UK) attachment with specular reflectance puck. Background scans were acquired from the fourth empty area of the SIRE slide. Spectra were acquired from the 4000-450cm^{-1} range using 16 co-added scans and a resolution set to 4cm^{-1} with 1cm^{-1} data spacing. Quest ATR internal mirror angles were calibrated to optimise transmitted energy throughput before each experiment (see section 3.2.3 for details). A Matlab (MathWorks, USA) based General User Interface (GUI) developed in-house was used to process spectra.

Each sample area was scanned across all 288 SIREs (36 designs, 8 replicates). A total of 864 spectra were collected over the course of the experiment. Spectra collected from each individual design were averaged to reduce experimental biases. All spectra were then baseline corrected, outlier spectra removed, and vector normalised. An outlier removal step took out any spectra that had more than 5% of data points out with two standard deviations from the mean of all 864 spectra (Figure 3.2). 117 spectra were removed by this method and were removed before vector normalisation as outliers can influence the vector normalisation step. It is believed this high number of outliers is due to the lack of an established sample preparation protocol and automated slide indexing unit to precisely position slides at the time of experimentation resulting in signs of scattering artefacts present in some spectra. Signal-to-Noise Ratio (SNR), Amide I absorbance, and Amide II absorbance was used to measure and compare SIRE designs. Averaged and Relative Standard Deviation (RSD%) values were used in each case. SNR was calculated using in the standard way (chapter 2, section 2.4.1). The noise component was taken as the standard deviation between baseline wavenumbers (1800-1850cm^{-1}) and the signal component was taken as the average amide I band (1625-1675cm^{-1}) [11], [12].

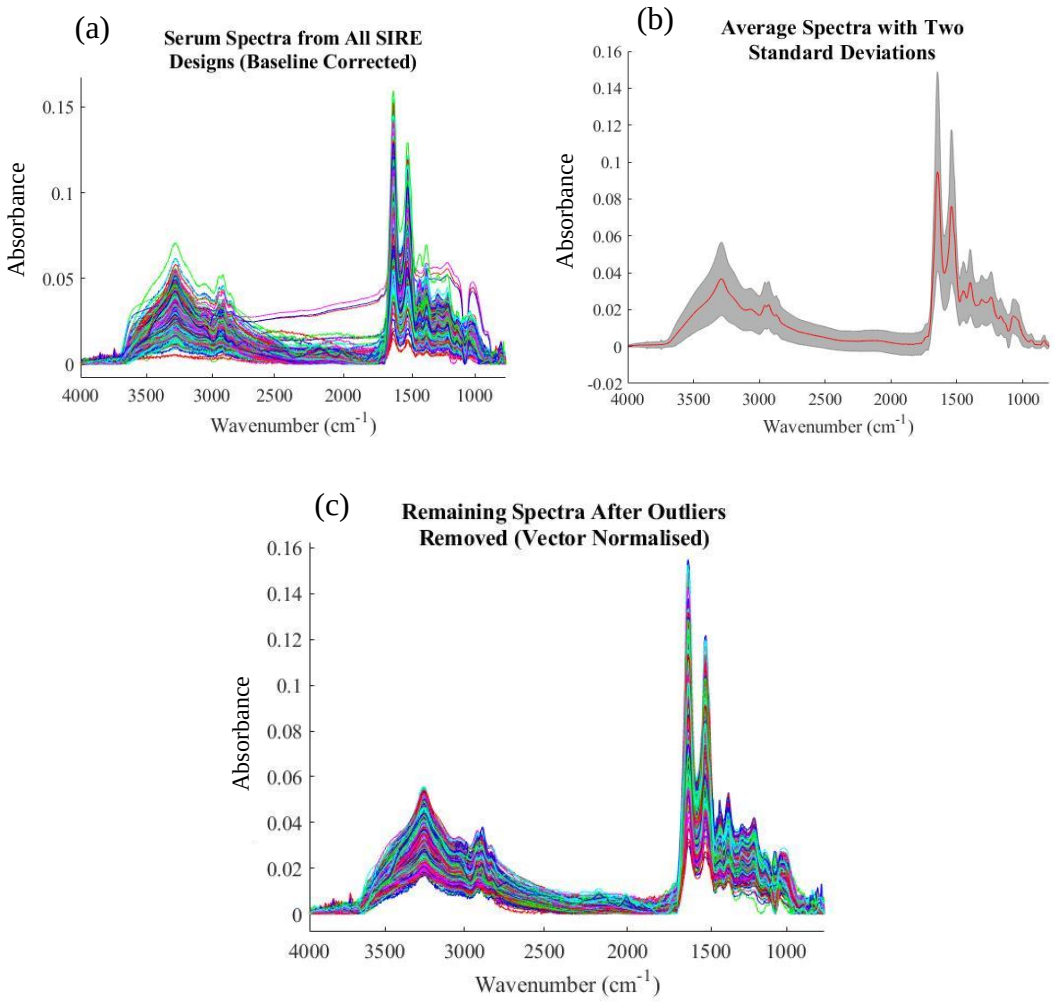


Figure 3.2. *(a) All spectra acquired across every SIRE Design. Spectra have been rubber band baseline corrected. Several spectra seem to be impacted by scattering artefacts likely due to the lack of established sample preparation protocol. (b) Average of all spectra collected across every SIRE design (red) with the two standard deviations from the mean indicated by the shaded region (grey). (c) Remaining SIRE design study spectra after outliers have been removed. These spectra have been vector normalised and are in the format subsequent analysis was carried out on.*

To explore the mechanism by which increasing silicon thickness diminished serum spectra, spectra of the silicon lattice itself were obtained. To do this a Gold mirror was placed at the specular reflectance puck aperture so that a background scan of the environment could be acquired and then the SIREs themselves were scanned in

the same way a sample would be scanned. Average spectra of the three available thicknesses (375µm, 525µm, and 675µm) were scanned for comparison.

3.2.2. SIRE Sample Slide Components and Assembly

The sample slide assembly comprises the SIRE, a Polylactic Acid (PLA) receptacle, and an adhesive label with graphics. Cleaned SIREs are placed in the receptacle and fixed in place with the adhesive label using RH07 adhesive (UPM Raflatac UK, Scarborough). SIREs are inset in a PLA receptacle for the following reasons:

i. Contribute structural integrity to the slide.
ii. Ease of handling for integration into the clinical laboratory (same size as standard microscope slide).
iii. Avoid any direct contact with the SIRE thereby reducing risk of sample contamination.
iv. Partition adjacent samples and prevent leakage of samples from one compartment to the next.
v. Provide an option for future traceability mechanisms (e.g., barcode, radio frequency tag, etc).

A rough schematic of the fully assembled SIRE sample slide is presented in Figure 3.3. PLA receptacles were fabricated using a Fused Filament Fabrication (FFF) 3D printer (Ultimaker, Netherlands). The receptacle dimensions conform to standard microscope slides (75x25x1mm) with an off-centre lipped inner cavity to host the SIRE. The optically active areas of SIREs are positioned groove side down into the

cavity. Consequently, SIREs are raised by approximately 0.6mm from the IR beam aperture when placed flat on it. The impact of this clearance above the aperture was tested by comparing the beam energy throughput with and without the receptacle and also to a diamond IRE. The label fixes the SIRE in place and indicates to the user the background scan area (position '0') and three sampling areas (positions '1', '2', and '3'). Numerous other sample slide designs were conceptualised and prototyped before arriving at the design described above. A collection of alternative and previous designs has been presented in Appendix I (chapter 8). Fabrication tolerances were calculated from the dimensions of ten SIREs and ten PLA receptacles. Measurements were taken using a pair of digital callipers (RS components, UK). Simulations were carried out to determine if injection moulding could be used for reliable scaled production of the plastic receptacle component (simulations constructed using MoldFlow by Autodesk). Two thermoplastics were selected for simulations, Polypropylene (PP) and Acrylonitrile Butadiene Styrene (ABS), because they are widely available, inexpensive, and commonly used for injection moulding. Material and process parameters were selected from generic material pre-sets available within MoldFlow. For ABS, mould temperature was set to 60°C, melt temperature 230°C, and maximum injection pressure to 180MPa. For PP, mould temperature was set to 50°C, melt temperature 220°C, and maximum injection pressure to 180MPa. Injection time was set automatically by MoldFlow. Clamp opening time was set to 5 seconds for both materials. A variety of injector pin configurations were attempted to determine which provided the greatest chance of manufacturing a high-quality part.

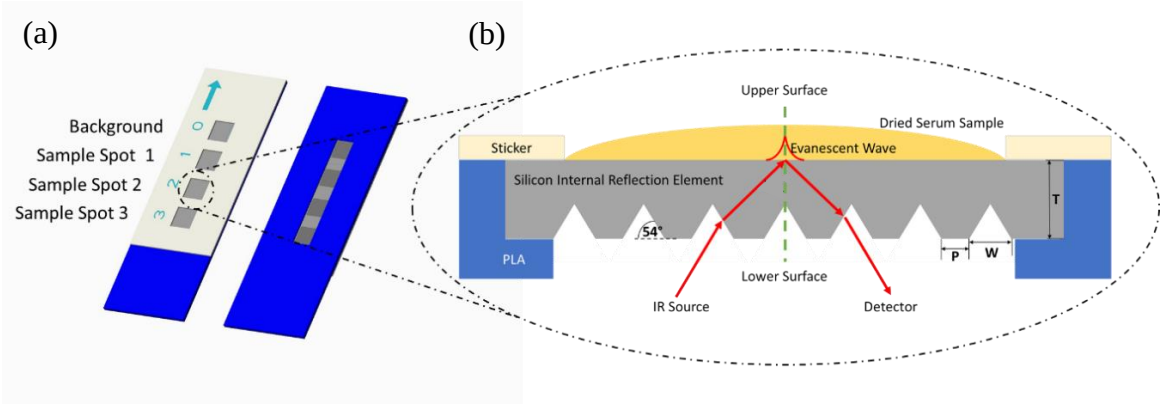


Figure 3.3. *(a) Top and bottom image of the fully assembled SIRE sample slide. Slides are inserted onto the accessory unit from position zero first where a background scan is acquired then the slide is indexed and scanned sequentially across sample containing positions 1, 2, and 3. (b) Annotated cross-sectional schematic of one optically active area of the SIRE sample slide.*

3.2.3. Custom ATR Accessory Unit Design

The salient function of the accessory module was to facilitate the use of SIREs on an IR spectrometer. It is desirable that the accessory module is compatible with a range, ideally all, IR spectrometer platforms. An ATR accessory (Figure 3.4) contains a system of highly engineered optics that deliver the beam to the IRE which is embedded into a top-plate under normal circumstances. Several accessory module prototypes were constructed throughout the design process. Many conditions that had to be met were only revealed through these prototypes ultimately crafting a very different solution to the design that was initially envisioned. Previous models have been presented in Appendix II (chapter 8).

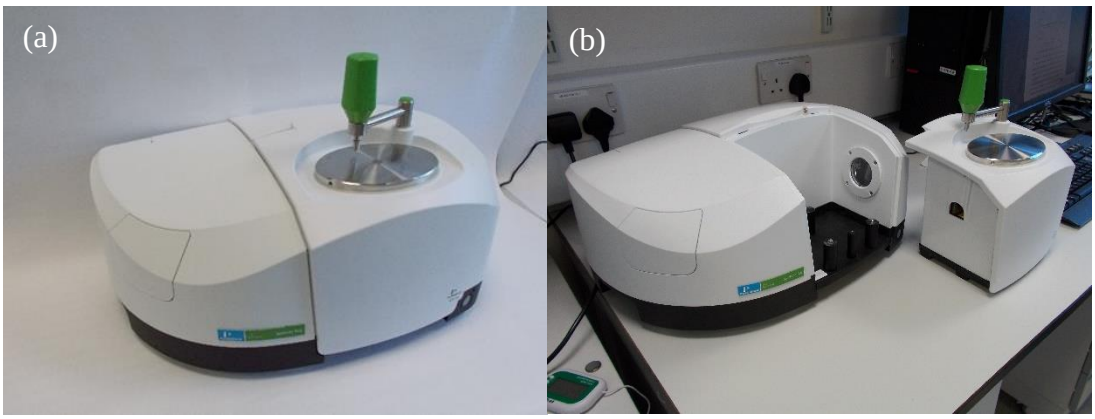


***Figure 3.4**. Image of a Spectrum2; (a) with accessory module attached, and (b) with accessory module detached. The metallic oval on top is known as the top-plate.*

A notable feature of the Spectrum2 is the off-centre focal point (Figure 3.5). Normally focal points are centred within the sample presenting the difficult engineering problem that optics must be adjustable to enable optimisation of beam throughput when interchanging from the Spectrum2 to other spectrometer platforms. Early investigations determined optimal Angle of Incidence (AOI) of beam to the SIRE to be ~30° for the analysis of serum. However, this was determined using a Cary 660 Agilent Technologies spectrometer so may not be applicable to the atypical Spectrum2 beam configuration and AOIs less than 30° could not be tested (for more details see Appendix II, chapter 8).

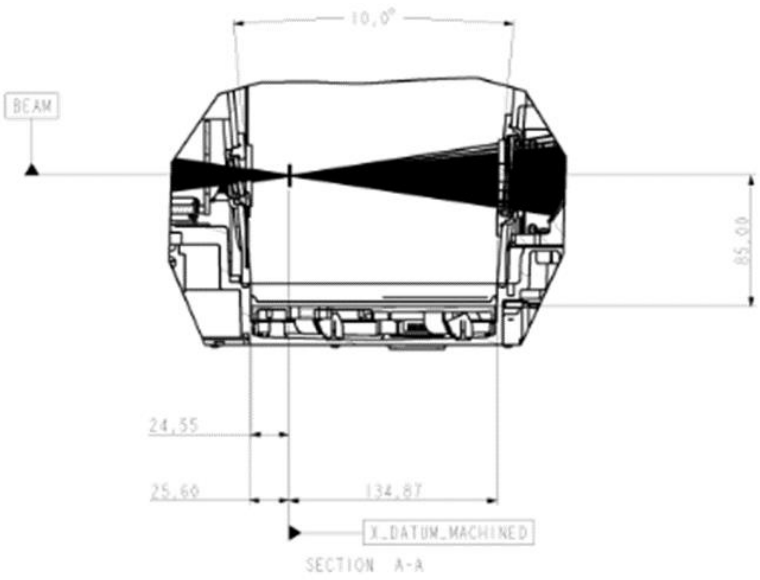


Figure 3.5. *Schematic of the beam trajectory (solid black) through the sample compartment of the Spectrum2. PerkinElmer spectrometers feature a focal point which is offset left from centre of the sample compartment.*

A universal ATR accessory device called the Quest ATR by Specac was introduced to circumvent the complication of constructing a system of optics capable of coping with the different focal points set by manufacturers. The Quest ATR unit features four adjustable gold mirrors and can be used on any spectrometer platform (Figure 3.6.(a)). An interchangeable top-plate puck system allows the user to quickly swap IRE substrates and gives the option to install a specular reflectance puck. The specular reflectance puck is simply a metal puck with a 3.8mm hole in the centre to enable an IR beam to be delivered to external (e.g., a gold mirror) or internal (i.e., an IRE) reflectance substrates. The hole acts as an aperture fixing the beam spot size to 3.8mm and ensuring the amount of light reaching the SIRE is consistent. For all analysis in this chapter, a SIRE sample slide guide was 3D printed and slides were indexed manually across the specular reflectance puck because more sophisticated automated slide indexers were not yet available (Figure 3.6.(b)). An image of the specular reflectance puck has been presented in Figure 3.6.(c), the SIRE is positioned

on top on the aperture at the centre of the specular reflectance puck. An IR absorbing black putty is situated on the beam-facing side of the puck around the aperture to prevent stray beams reflecting into the system and causing variations in light throughput. The precise Angle of Incidence (AOI) of the Quest ATR is not known, but the nominal AOI of the IR beam to the sample surface is specified as 45° [13]. However, the energy throughput of the Quest ATR can be optimised by removing the front cover plate then adjusting the internal Quest ATR unit mirrors using a 2.5mm ball driver while simultaneously monitoring the spectrometer's beam energy throughput until the highest energy throughput is achieved (Figure 3.6.(d)). Presumably, the AOI tends towards 30° when beam throughput has been optimised in accordance with preliminary investigations (Appendix II, chapter 8). Since the beam originates from a globar light source and is not collimated, the light will arrive at the SIRE outside the chief beam AOI (approximately ±10°).

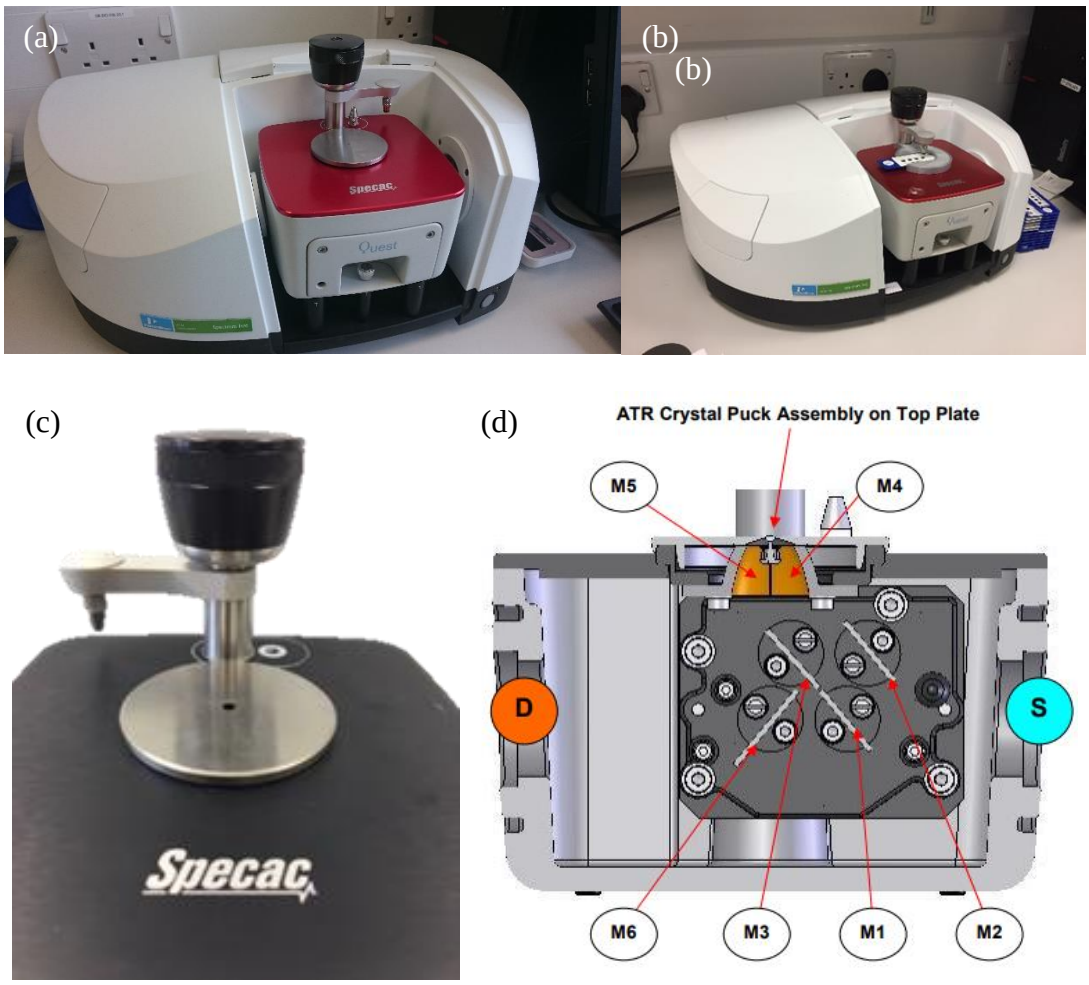


Figure 3.6. *(a) The Specac Quest ATR accessory module installed within the Spectrum2. The circular silver disk on top of the Quest is an interchangeable IRE crystal puck. (b) a SIRE slide guide for manual indexing along with a SIRE slide have been placed on top of the specular reflectance puck. (c) A close up image of the specular reflectance puck. The SIRE is placed on the aperture for analysis. (d) Open view of the internal mirror arrangement of the Quest ATR. S and D represent source and detector, respectively. Mirrors M1, M2, M3, and M6 are all adjustable, flat, highly reflective gold mirrors and M4 and M5 are fixed aspheric gold-coated focussing mirrors. Taken from:* [13]

A custom built automated slide indexer was built specifically to interface the Quest ATR unit. The IR beam was delivered to the SIRE sample slide through the aperture of the specular reflectance puck. The slide indexer moved the SIRE sample slide sequentially through all four optically active areas for analysis. PLA prototypes revealed that plastic components were too flexible and vibrations during operation

would negatively affect performance (Appendix II, chapter 8). The final design was constructed with improved support mechanisms and sturdier materials to minimise vibrations and rotational movements of the slide. In brief, the final design utilises a stepper motor and lead screw to drive a carriage in which the SIRE sample slide resides. A microcontroller and motor driver circuit, programmed to actuate the stepper motor upon user input via push button, can precisely position the slide based on lead screw turn count. Positions are predefined in the microcontroller program such that slides will be positioned above the specular reflectance puck aperture from slide position '0', '1', '2', and '3' before returning to '0' again, the home position. Red LEDs indicate to the user which position the slide is in. Two slide shafts eliminate rotational movement of the carriage around the drive shaft. The carriage itself surrounds the sample slide at the sides and on top minimising vibrational movement. Openings have been made at the top of the carriage to allow easy removal of sample slides and to not disturb serum samples deposited in the sample areas. A more detailed description of the indexer unit construction is given below.

A linear translation mechanism was mounted upon a metal baseplate (Figure 3.7). The baseplate is a 142mm x 144mm x 3.8mm plate with a 62mm off-centre diameter hole to cover the area around and sit flush with the Quest ATR specular reflectance puck. A 40mm x 40mm square was also cut out of the rear-facing baseplate edge to accommodate the Quest ATRs clamping arm. The baseplate was fastened using screws fitted into two M3x0.5mm holes tapped through the baseplate and into the surface of the Quest ATR at the source (left) and detector (right) facing sides. The linear translator comprised a slide carriage (35mm x 80mm x 25mm), flat 1.85V RS Pro hybrid stepper motor (RS components, UK), 116mm M4x0.7mm lead screw (Igus

UK, Northampton), flanged lead screw nut (13mm M4x0.7mm thread), 18mm cylindrical drive coupling (3.5mm inner diam. (ID), 7mm outer diam. (OD)), ball bearing (4mm ID, 8mm OD, RS components, UK), and two 120mm x 4mm cylindrical slide shafts (The Metals Warehouse, Nottingham) all supported by two mounting blocks fitted at the left and right side of the baseplate using four M3x0.5mm studs. The baseplate, slide carriage, and mounting blocks (left: 20mm x 25mm x 7mm, right: 32mm x 40mm 24mm) are all custom components milled from 6063-T6 anodised aluminium as well as the lead screw nut and drive coupling which was machined from CZ121 brass (Bar Knight Precision Engineers, Glasgow). The slide shafts were inserted through two 4.2mm through holes drilled at the top and bottom of the slide carriage. The lead screw meshes with the lead screw nut which is fixed within a central 6mm through hole of the slide carriage via two M2x0.4mm screws. Both left and right mounting blocks featured a central through hole (4.2mm and 8.5mm diameter respectively) to host the bearing, stepper motor shaft, drive coupling, and lead screw. The lead screw was connected using two M4 nuts to the ball bearing that was push fit within an 8mm blind hole concentric to the 4.2mm through hole of the left mounting block. Two M3 bolts were used to mount the stepper motor onto the right-side mounting block and the motor shaft was coupled to the lead screw via two M3x0.5mm grub screws through the drive coupling. A flat stepper motor was selected to fit within the limited space of the sample compartment. M4x0.5 thread was cut to 3mm depth either slide of the slide shafts which were then fixed to corresponding holes tapped into the mounting blocks above and below the lead screw. The stepper motor is controlled using an Arduino Pro Mini 328 microcontroller (Arduino, Italy) and a DRV8825 motor driver (Pololu, USA) which users can operate using a EOZ push

button (RS components, UK). Users can see which position the SIRE slide is in, but to reduce risk of confusion Kingbright red LEDs (RS components, UK) indicate to the user which SIRE slide position is located above the Quest ATR aperture. For full electronic schematics, see Appendix III (chapter 8).

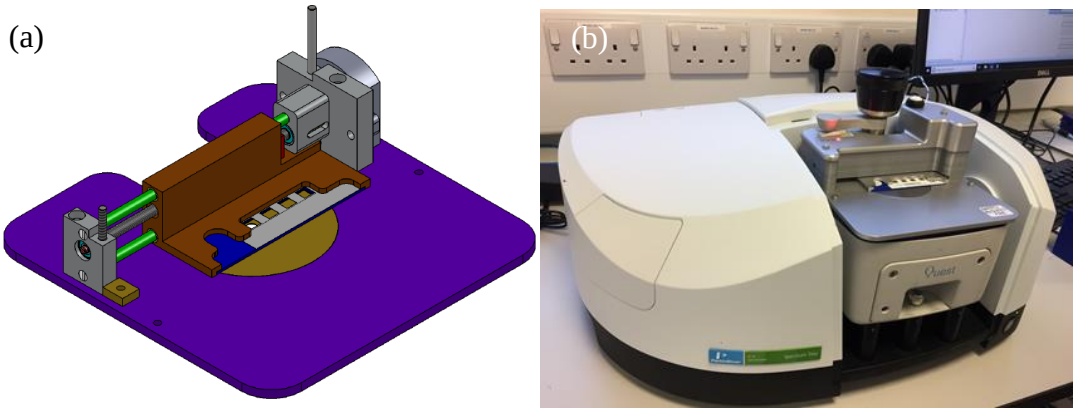


Figure 3.7. *The custom built slide indexing unit intended to be complementary to the Quest ATR universal accessory module (a) as a CAD model showing an assembly of a selection of core components: baseplate (purple), specular reflectance puck (light brown), slide carriage (dark brown), slide shafts (green), mounting blocks, lead screw, and stepper motor (grey). (b) slide indexing unit mounted within the Spectrum2 on top of the Quest ATR unit with a SIRE sample slide inserted into the slide carriage. A 3D printed PLA cover has been added which provides secure housing for the red LEDs and push button.*

3.3. Results

3.3.1. Silicon Internal Reflection Element Design Optimisation

There are many potential sources for error in the data such as instrumental error, sample preparation error, and sample error that should be considered when

evaluating SIRE design reliability [14]. Instrumental error of the sample slides is very low (RSD<4%) may arise from either the SIREs or the spectrometer itself and has been estimated by calculating the average RSD% of the raw silicon spectra acquired from the three optically active regions upon individual sample slides. The RSD% of the Amide I/II ratio across individual sample slides was calculated to examine the error contribution by the sample itself since this is not a metric that could be affected by instrumentation, but instead should be an inherent property of the sample (i.e., Amide I/II ratio is purely influenced by the sample, not by SIRE design provided sample thickness remains constant [15], [16]). The average RSD for Amide I/II, and by extension the sample error, is less than 3%. It is difficult to categorize the sample preparation error contribution, but an observation made during the course of the experiment was that small differences in the uniformity of serum droplet spreading would have a major impact on spectral quality by introducing noise and distorting the baseline of spectra. Sample preparation error will be further explored in chapter 4 section 4.3.2.1 to investigate optimal sample preparation techniques for use with the SIRE sample slides.

The average and RSD% values for signal to noise ratio (SNR) of every SIRE design has been presented in Table 3.2. To fully clarify which SIRE design produces the highest quality of signal and least inter-slide variability in the context of serum analysis; average signal intensity and RSD% values of the Amide I and Amide II bands has been presented as high, reliable absorption at these bands is indicative of good information capture (Table 3.3, and Table 3.4).

Table 3.2. *(top) Average signal to noise ratios achieved from every SIRE design. Greener values indicate a higher signal to noise ratio while redder values indicate a lower signal to noise ratio. (bottom) Relative Standard Deviation of every SIRE design. Greener values indicate lower RSD values, while redder values indicate higher RSD values.*

Average Signal to Noise Ratio

		Width (µm)			
Thickness (µm)	Pitch (µm)	100	150	200	250
380	25	152.9	271.8	164.8	136.7
	50	186.0	233.3	185.2	178.7
	100	174.7	181.6	166.8	225.5
525	25	214.3	148.4	117.4	152.8
	50	160.6	137.7	191.9	146.3
	100	220.9	159.7	146.2	202.4
675	25	131.5	140.9	139.3	158.5
	50	165.7	177.5	151.3	151.5
	100	133.9	124.3	86.1	88.6

Relative Standard Deviation (%) of Signal to Noise Ratio

		Width (µm)			
Thickness (µm)	Pitch (µm)	100	150	200	250
380	25	21.0	24.3	25.7	13.0
	50	30.4	18.9	22.5	30.4
	100	60.1	48.5	20.3	41.9
525	25	49.7	34.2	20.1	39.5
	50	43.7	27.3	73.0	29.2
	100	38.8	46.3	28.8	45.5
675	25	50.2	46.5	31.5	38.3
	50	34.0	50.6	48.0	46.6
	100	57.7	40.6	37.8	44.7

Table 3.3. *(top) Average amide I absorbance achieved from every SIRE design. Greener values indicate a higher amide I absorbance while redder values indicate a lower amide I absorbance. (bottom) Relative Standard Deviation of amide I absorbance achieved from every SIRE design. Greener values indicate lower RSD values, while redder values indicate higher RSD values.*

Average Amide I Absorbance

		Width (µm)			
Thickness (µm)	Pitch (µm)	100	150	200	250
380	25	0.084	0.085	0.083	0.082
	50	0.083	0.089	0.087	0.086
	100	0.076	0.079	0.082	0.085
525	25	0.086	0.080	0.081	0.074
	50	0.081	0.083	0.079	0.084
	100	0.077	0.077	0.081	0.087
675	25	0.078	0.079	0.076	0.079
	50	0.080	0.079	0.080	0.080
	100	0.065	0.071	0.068	0.072

Relative Standard Deviation (%) of Amide I Absorbance

		Width (µm)			
Thickness (µm)	Pitch (µm)	100	150	200	250
380	25	4.2	3.2	5.0	2.1
	50	5.9	3.1	5.5	4.1
	100	6.6	4.8	4.7	4.0
525	25	6.0	9.0	6.3	5.6
	50	4.5	4.9	11.6	4.5
	100	12.0	4.1	2.8	3.7
675	25	6.9	7.9	5.0	6.8
	50	8.7	8.8	6.4	4.4
	100	8.8	10.4	7.4	7.2

Table 3.4. *(top) Average amide II absorbance achieved from every SIRE design. Greener values indicate a higher amide II absorbance while redder values indicate a lower amide II absorbance. (bottom) Relative Standard Deviation of amide II absorbance achieved from every SIRE design. Greener values indicate lower RSD values, while redder values indicate higher RSD values.*

Average Amide II Absorbance

		Width (μm)			
Thickness (μm)	Pitch (μm)	100	150	200	250
380	25	0.066	0.067	0.066	0.065
	50	0.067	0.070	0.067	0.067
	100	0.062	0.064	0.065	0.067
525	25	0.067	0.063	0.064	0.060
	50	0.065	0.066	0.063	0.066
	100	0.064	0.062	0.065	0.068
675	25	0.063	0.064	0.062	0.064
	50	0.064	0.064	0.064	0.065
	100	0.057	0.059	0.057	0.059

Relative Standard Deviation (%) of Amide II Absorbance

		Width (μm)			
Thickness (μm)	Pitch (μm)	100	150	200	250
380	25	3.5	2.7	4.2	1.5
	50	4.1	2.7	3.2	3.4
	100	5.8	3.8	3.3	3.3
525	25	4.7	6.5	5.0	6.1
	50	4.2	3.0	8.5	3.4
	100	8.8	4.0	3.5	3.1
675	25	6.1	6.5	3.9	4.7
	50	7.7	8.0	6.1	3.6
	100	7.1	7.4	6.1	6.3

It is hard to decipher a clear advantage for spectral quality of one design strategy over another in terms of average SNR. The data indicates thinner SIREs have an improved SNR over thicker SIRE designs. However, average SNR did not seem to be greatly affected by changes in the width of the beam coupling grooves, W, or groove

pitch, P. RSD% of SNR is considerably improved in thinner SIREs with smaller pitch. In general, a groove width of 200µm appears to give better SNR RSD% values, but the best recorded RSD% design SIRE parameters are 380µm thickness, 250µm width, and 25µm pitch. Both Amide 1 and Amide II band average and RSD% of signal intensity is improved in thinner SIREs. Wider groove sizes and smaller pitch also appear to improve both average signal intensity and RSD%. Overall, RSD% is maximised in 380µm thick SIREs with groove widths of 250µm, and 25µm pitch. Amide I and amide II average signal intensity is also good with these design parameters.

The absorption of light by a sample is directly proportional to the beam's pathlength through the sample as stated by Beer-Lamberts law [17]. This same rule applies to the IRE itself and can be observed very clearly by the difference in baselines of the raw spectra of different silicon thicknesses presented in Figure 3.8.(a). Similar to how interference from water signal can be limited by shortening sample pathlength, the peaks from 500-1500cm^{-1} wavenumbers are also diminished as the thickness decreases due to decreased interaction with the silicon lattice (819cm^{-1}, 896cm^{-1}, and 962cm^{-1}), or impurities such as interstitial oxygen (1107cm^{-1}) or substitutional carbon (605cm^{-1}). This is more easily observed after the spectra have been rubber band baseline corrected (Figure 3.8.(b)). A change in coupling angle as SIRE thickness varies could also be responsible for the observed change in baseline. However, the groove etch angle is fixed at 54.7° across all SIRE designs meaning that the angle of the coupling surface relative to the beam should remain fixed. It is perhaps more likely, then, that attenuation of the beam due to Si thickness is responsible for baseline differences in accordance with the Beer-Lambert law [9], [18]. Nevertheless, it is

evident that less beam energy is transmitted to the detector as SIRE thickness increases. Artefacts observed in the form of a ripple in the baseline of the spectra is exacerbated as thickness increases. This is most noticeable in the rippled baseline of the 675µm thick SIRE spectra. The source of the ripple artefact could be interference fringes arising from unintentional multiple internal reflections occurring as a result of the greater IRE thickness, although further experimentation would be required to confirm this [19], [20]. The theoretical upper and lower limits for IR attenuation by SIRE thickness could not be investigated within this experiment due to practical limitations. 380µm thick silicon wafers are about as thin as readily available six-inch silicon wafers can be, and 675µm thick wafers are the maximum thickness of silicon wafers that can fit in standard industrial six-inch silicon wafer carriers. It is expected that beyond a thickness of 675µm, silicon IREs will completely extinguish any signal passing through [9]. In contrast, using SIREs that are thinner than 380µm may improve SNR, but would likely be very difficult to handle and too brittle to analyse samples that require pressure application.

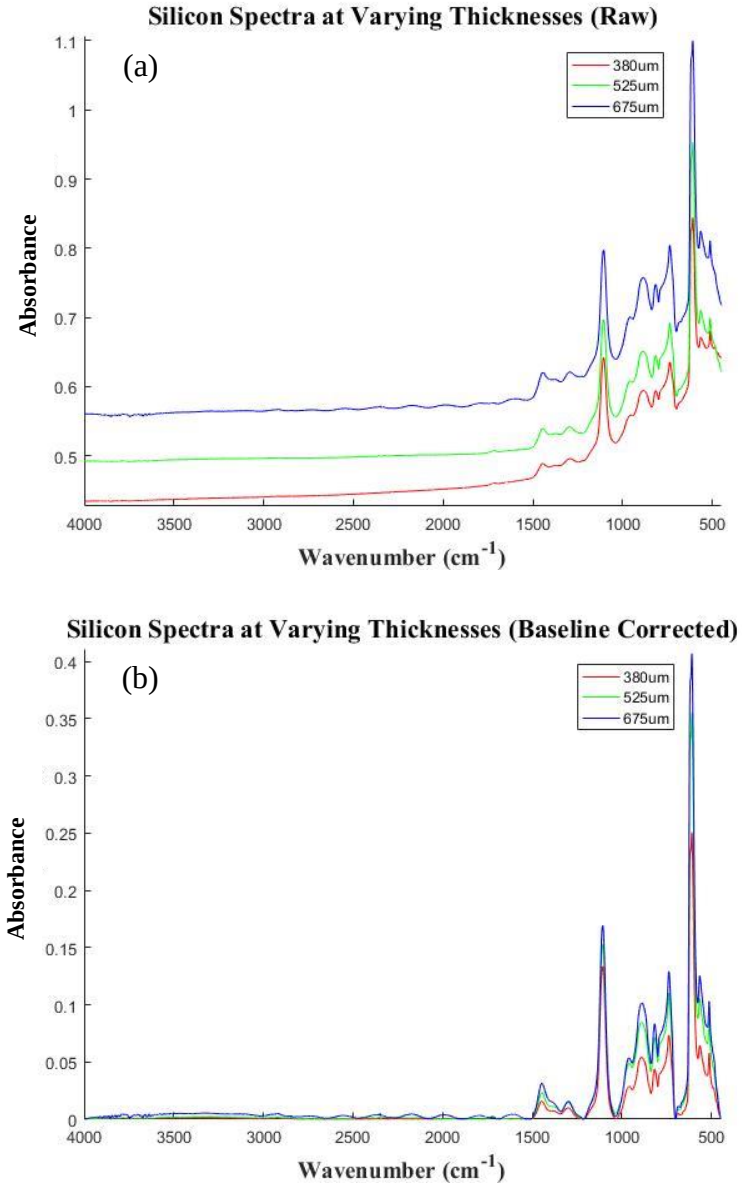


Figure 3.8. *Spectra of silicon SIREs at every thickness tested. Notable peaks are assigned as follows: interstitial oxygen – 1107cm^{-1}, substitutional carbon – 607cm^{-1}, silicon lattice phonons - 819cm^{-1}, 896cm^{-1}, and 962cm^{-1}. (a) Raw spectra of the silicon reveal a greater level of attenuation of the IR beam when thickness increases as indicated by the rising baseline. (b) Rubber band corrected spectra clearly displays the exacerbated effects of absorbance bands as SIRE thickness increases. Also noteworthy is the noisy 'rippled' baseline artefact present in the baseline of 675µm thick SIRE samples.*

Clinically significant information is located at wavenumbers above 900cm^{-1} of serum spectra [6], [21]. Clearly by analysing Figure 3.8 it can be observed that there are several absorption peaks that may prove problematic. It may be the case that SIREs

are not suitable for diagnostic tests where univariate analysis of carbohydrates [6], and perhaps to some extent nucleic acids, are crucial biomolecular indicators, such as diabetes [22]. Fortunately, multivariate and machine learning approaches are well suited to deal with this limitation by not relying heavily on information of one specific point of the spectrum. Fabricating SIREs from Floatzone (FZ) wafers is an obvious solution to mitigating interference from impurities since they lack the interstitial oxygen absorption peaks at 1107cm^{-1} [9], but their high costs (£29.50 per Cz compared to *ca*. ~£100 per FZ wafer) and lack of availability are unfortunately prohibitive within the scope of this prototyping stage. However, this may not prohibit FZ wafers from future implementation as costs could still lie within financial constraints defined by the brain tumour project introduced in chapter 1 section 1.4.1, which dictates a test cost of no more than £40, or any other serum diagnostic applications not yet evaluated through health economic analysis [23]. The unit cost of Cz based SIRE sample slides is roughly £5.90, and unit cost FZ based SIRE sample slides is estimated to be £8.30. A brief breakdown of the sample slide finances is presented in Table 3.5. A batch is equivalent to 25 wafers which equates to 1500 SIRE slides. A manufacturing yield of 85% was assumed per batch (i.e., a batch of 1500 slides will yield 1275 useable devices). Not included is the cost of a one-off £906 purchase of a photomask which can be reused indefinitely. Quoted values are based on expenditures for the present study and all would decrease upon scaled production therefore from a manufacturing perspective it is feasible tests can be delivered at a cost of less than £40. However, it should be noted that there are many other factors that will increase overall costs such as packaging, accessory unit manufacturing, distribution, commercialisation, quality assurance, training, regulatory clearance, market surveillance, maintenance, and more

[24]–[26]. Although the costs of these additional factors have not been fully evaluated in this early stage of development it is likely they would be significant. The consequence of this is the diagnostic performance must be of higher quality to compensate for higher cost. For instance, health economic assessment has shown if the SIRE brain cancer test cost rises from £40 to £100 it will likely still be cost effective in the primary care setting so long as clinical specificity is greater than 90% [23].

Table 3.5. *Breakdown of expenses for a single and for a batch SIRE sample slides.*

	Unit Cost (£)	Batch Cost (£)
Cz Wafer	29.50	737.5
FZ wafer	100	2,500
Microfabrication	N/A	6,602
PLA receptacle	0.5	637.5
Label	0.5	637.5
Sample Slide (Cz)	**6.90**	**8,589.5**
Sample Slide (FZ)	**8.30**	**10,352**

Figure 3.9 highlights the difference between spectra of serum collected using SIREs and diamond IREs. The spectra of silicon have been overlaid to illustrate that the absorption bands observed are missing from SIRE based serum spectra when compared to diamond IRE based serum spectra. The fundamental differences between these two IREs for serum analysis will be explored further in chapter 4 section 4.3.3.

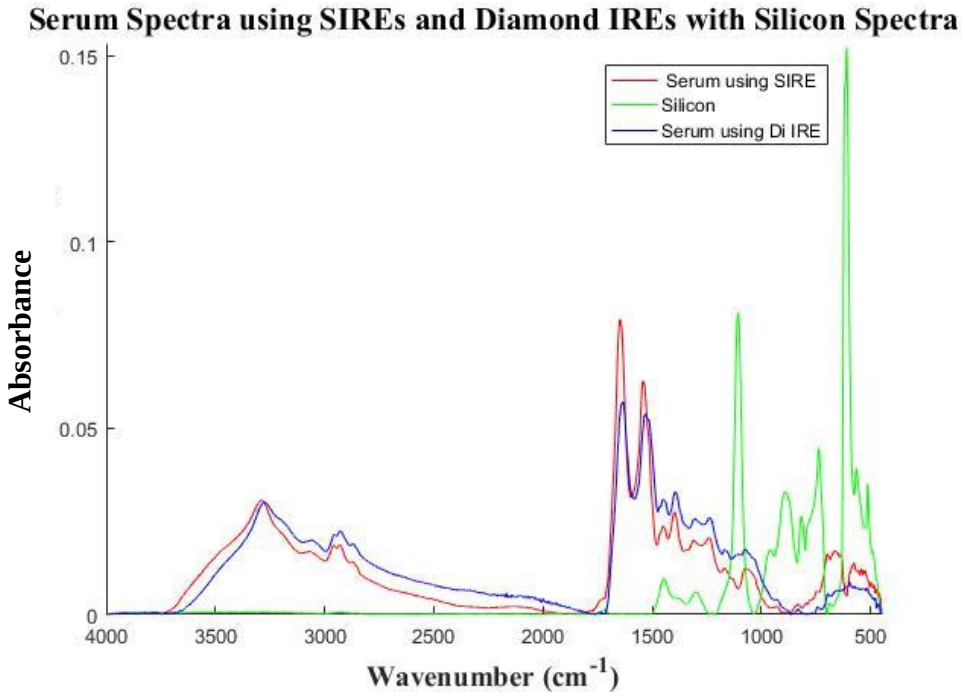


Figure 3.9. *Serum spectra collected using both a SIRE and a Diamond IRE. Spectra of silicon has been overlaid to show clearly that there is a dip in signal at 1107cm^{-1} in the serum spectra collected using a SIRE due to attenuation of the IR beam by the SIRE.*

3.3.2. Silicon Internal Reflection Element Fabrication Complications

Silicon microfabrication encompasses a huge variety of highly precise techniques, so small operator errors can translate into gross device defects. The nature of these defects range anywhere from small isolated imperfections to flaws spanning an entire silicon wafer.

3.3.2.1. Hard Mask Undercutting

Silicon etching proceeds by chemically attacking the silicon lattice in a directionally dependent matter as outlined in chapter 2, section 2.5.3. No crystal plane in silicon is impervious to KOH, but the degree of resistance to chemical attack differs greatly. For silicon planes <100>, <110> and <111> in 33% KOH at 80°C the etch

rate ratio is roughly 400:600:1 [27]. However, etch ratios differ widely in the literature therefore etch times should be established in house, using the literature as a guideline [28], [29]. This is in part because etch ratio is dependent on KOH concentration, temperature, etching geometries, and solvent flow (e.g., stirring) [30]. The effect of unintentional etching of the <111> direction is grooves are about 4-6µm bigger in both length and breadth than intended (Figure 3.10). This results in silicon material etched away from underneath the silicon nitride hardmask, also referred to as undercutting.

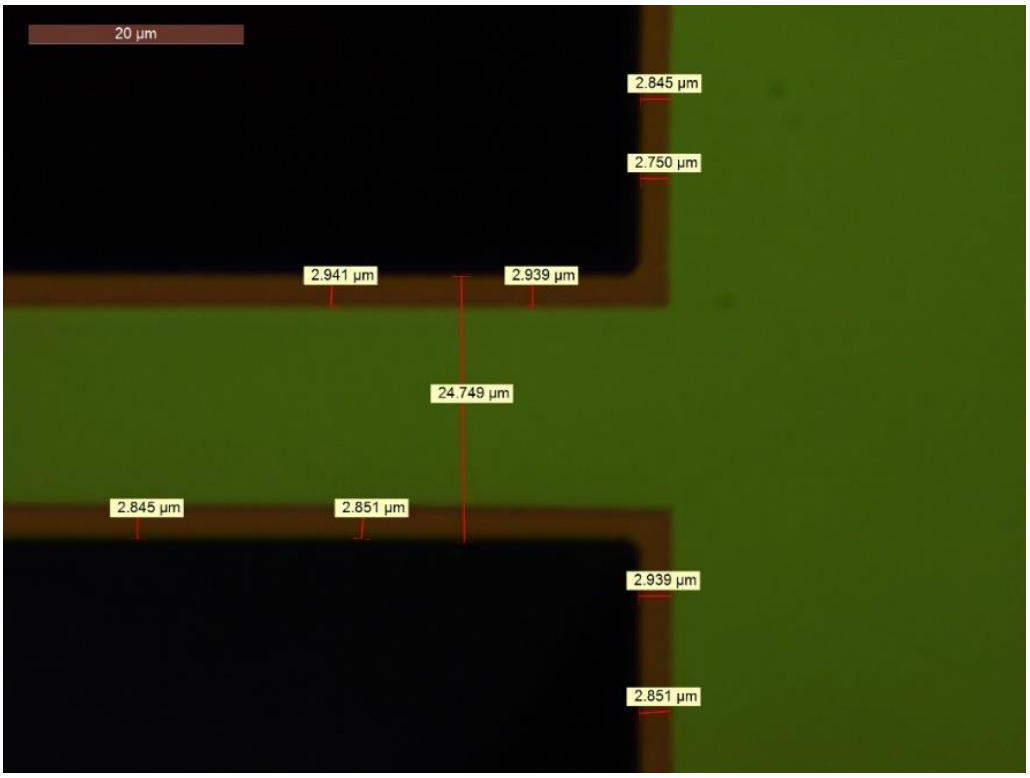


Figure 3.10. *Microscope image of SIRE grooves after KOH etch. Blacked out areas are deep grooves and everything else is a Nitride hard mask. Due to thin film reflection effects, light scattered from Nitride lying atop silicon appears green, and Nitride suspended in air appears brown* [31]*. Measuring the overhanging Nitrile indicates the grooves are ca. 6µm wider and longer than intended. Images acquired using a reflectance spectrometer.*

In this application 2-3µm is a tolerable degree of undercutting at each etch face. However, extensive undercutting may cause structural failure of the groove walls and

result in debris lodged within the grooves (Figure 3.11). Silicon-on-silicon debris is very difficult to remove by means of mechanical cleaning due to static charge [32] pinning the debris in place and chemical cleaning regimes are not an option because any attempt to remove the silicon particles will in turn damage the underlying silicon surface. Defects of this type can ruin an entire batch of silicon wafers which amounts to thousands of pounds worth of damages and huge time delays.

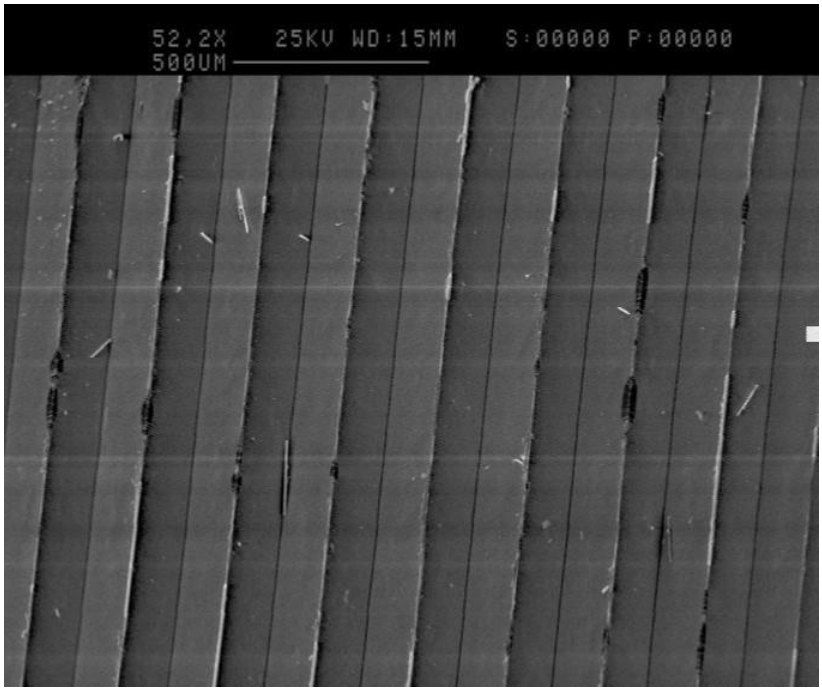


Figure 3.11. *SEM image of SIRE grooves. The peaks of the grooves have been underetched to the point of collapse, lodging silicon debris in the troughs.*

Mask undercutting is also an issue when manifesting as a result of photomask misalignment with etches being much wider and deeper than the designer intended (Figure 3.12). Undercutting is highly sensitive to small errors in photomask alignment and defects typically uniformly affect all devices printed upon a wafer [33]. Highly skilled operators ensure the effect of photomask misalignment are kept to minimum, but perfect alignment with the crystal plane is near impossible. In this application, coarse sidewalls were also a concern as this could scatter IR light (Figure 3.13). This

was either a consequence of incorrect orientation of the etching planes forming etch steps along the <111> plane [34], [35], or by using low KOH concentrations (<30%) and not stirring the solution during the etching process [36], [37]. In the latter scenario the mechanism for sidewall damage is the entrapment of hydrogen bubbles upon the silicon surface that randomly and inadvertently mask small areas lowering the uniformity of the etch pattern.

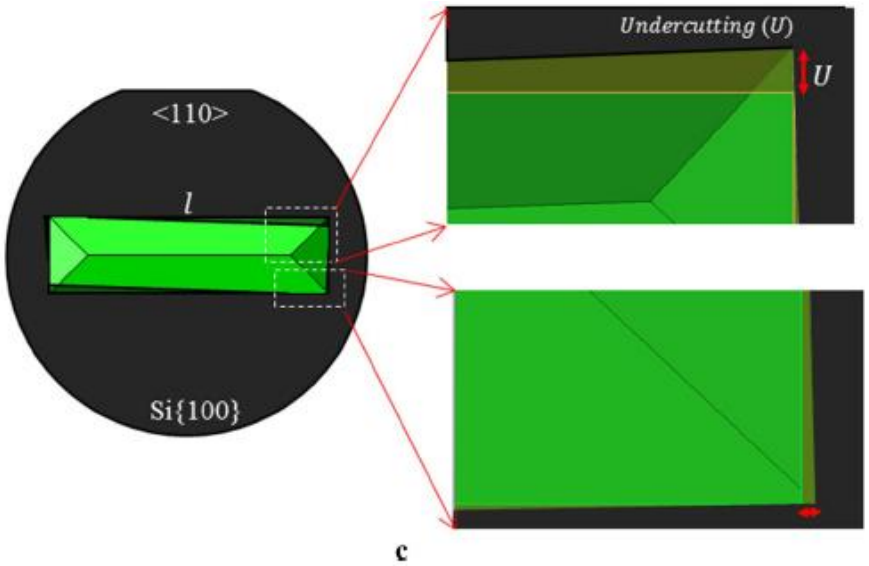


Figure 3.12. *Illustration of the mask undercutting that will occur following photomask misalignment. Image credit: ref* [38].

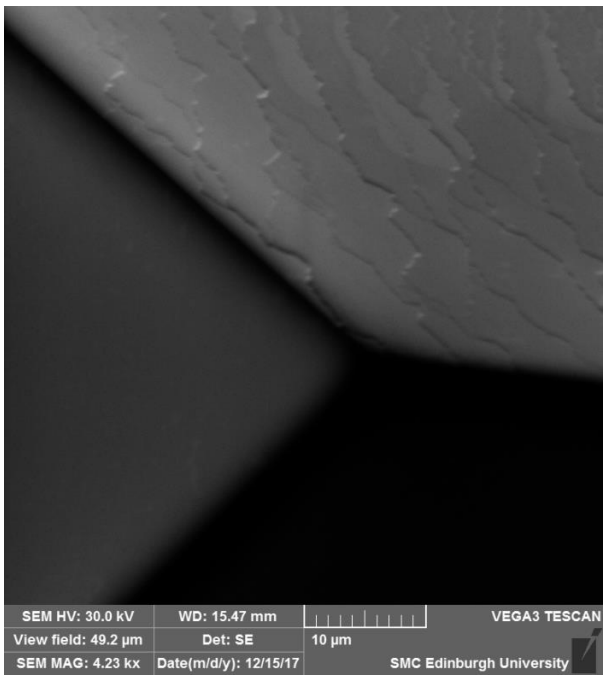


Figure 3.13. *SEM image of SIRE V-grooves. The steps on the silicon wall are crystal planes which have been mistakenly etched due to misalignment of the mask layer to the crystal planes, or by neglecting to stir the KOH solution enough during etching.*

3.3.2.2. Photoresist Defects and Scratch Defects

Any break in the photoresist layer, or hard mask layer, will result in massive areas etched out of the silicon located beneath the photoresist discontinuity (Figure 3.14.(a)). This is typically caused by particles that could not be removed through cleaning procedures before a photomask layer was applied. The manufacturing yield is about 85% for SIRE devices and the bulk of the 15% loses are attributed to photoresist defects. These imperfections are inevitable but can be mitigated through strict cleaning and handling protocols within the cleanroom. Scratch defects are quite self-explanatory arising from poor handling at any time during the product lifecycle (Figure 3.14.(b)). Silicon is very hard, but brittle so it is prone to severe fracture propagation and complete device failure is likely if the crystal is handled directly. For

this reason, the SIRE is nestled within a protective PLA receptacle as mentioned earlier. This has drastically improved device viability following the manufacturing process.

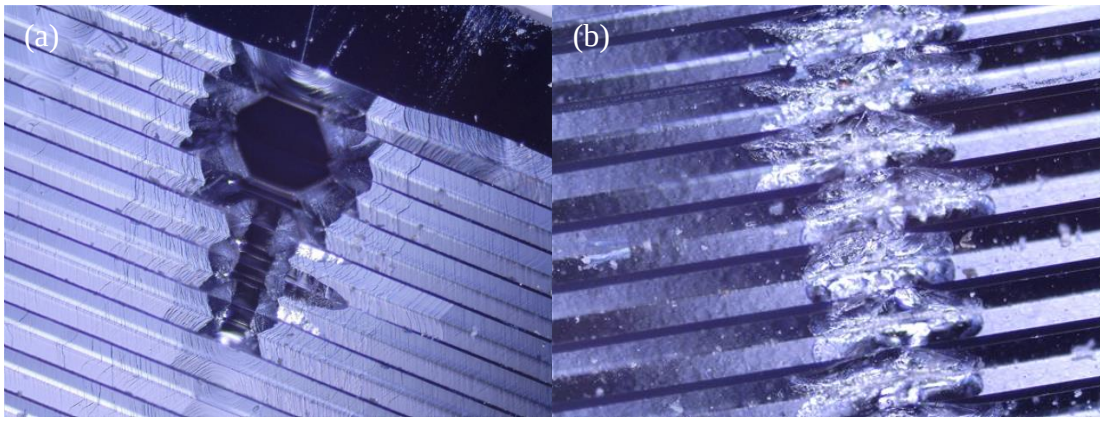


Figure 3.14.*(a) Microscope image of a defective SIRE. Large areas have been etched out due to a lack of protective hard mask. This is a consequence of debris that could not be cleaned off the wafer therefore preventing the photoresist layer from covering the underlying silicon. (b) Microscope image of a scratch spanning several grooves of a SIRE as a result of poor handling using tweezers.*

3.3.2. SIRE Sample Slide Assessment

Automated SIRE handling would be difficult to implement without the receptacle component due to the brittleness of silicon and sample contamination concerns. The PLA receptacle conforms to the same dimensions as standard microscope slides (75x25x1mm) (Figure 3.15.(a)) and has an off-centre rectangular cavity (33.4x6.2x0.4mm) with a ledge upon which the SIRE sits (Figure 3.15.(b)).

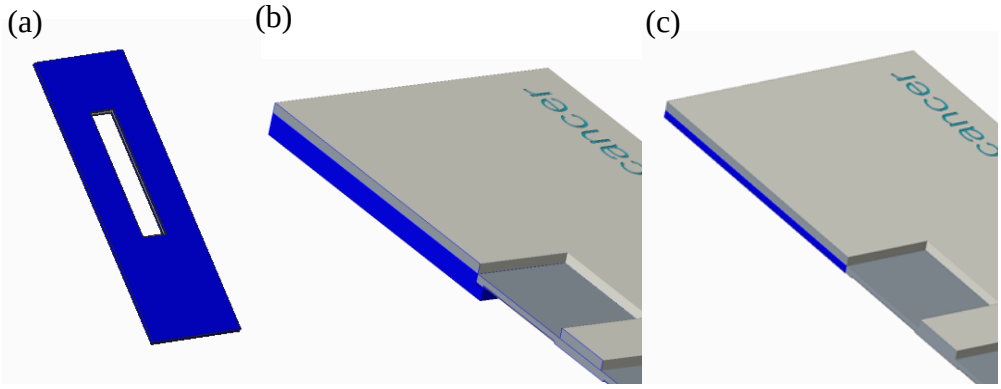


Figure 3.15. *PLA receptacle and cross-sectional views of fully assembled multi-compartmental SIRE. (a) PLA receptacle that supports SIRE. (b) SIRE set within a 1mm thick PLA receptacle. It is suspended off the ground by a ledge within the receptacle inner cavity. (c) SIRE set within a 0.4mm thick PLA receptacle. The SIRE can be made flush with beam apertures to maximise beam throughput while still benefitting from the protection of the PLA receptacle.*

Early receptacle designs suspended the SIRE above the specular reflectance aperture by about 0.6mm to prevent the SIRE contacting the puck which could scratch or contaminate the SIRE. However, this was found to affect the beam throughput. Table 3.6 shows the beam throughputs achieved from a SIRE with and without a PLA receptacle shown in Figure 3.15.(a), and from a diamond IRE. This prompted the move from thick PLA receptacles (Figure 3.15.(b)) to thin 0.4mm PLA receptacles (Figure 3.15.(c)) which removes the 'ledge' to keep the silicon flush with the beam aperture while still protecting the SIRE from damage. The beam throughput achieved with the thin PLA receptacles is identical to the throughput achieved by the SIRE on its own (i.e., 85.5).

Table 3.6. *Energy throughput measured using diamond IREs and silicon with/without a 1mm thick PLA receptacle.*

	Average throughput (absorbance)
SIRE	85.5
SIRE w/1mm thick receptacle	37
Diamond IRE	83

The dimensions of the inner cavity need to be precise otherwise SIREs either fit too loosely within the receptacle or protrude above which can lead to low quality spectra or seepage of liquid samples out of the sample compartment. The 3D printing FFF technique is adequate in the context of proof-of-concept and prototyping work, but variations in PLA receptacles can impact beam throughput and by extension sample slide reproducibility and the low fabrication rate is not suitable for scaled production. Table 3.7 shows the tolerances of SIREs and the inner cavity of 3D printed PLA receptacles demonstrating dimension mismatching is bound to occur, the effects of which would magnify upon scaled production.

Table 3.7. *SIRE dimensions and tolerances and PLA receptacle dimensions and tolerances using the Fused Filament Fabrication method.*

		Width (mm)	Length (mm)
SIRE	Average	6.07	33.10
	Tolerance (±)	0.02	0.09
PLA Receptacle Inner Cavity	Average	6.66	33.40
	Tolerance (±)	0.29	0.5

Injection moulding is an ideal candidate for scaled receptacle production, delivering precisely fabricated components quickly, although high tooling costs are prohibitive in early design and testing stages. Simulations performed to determine if a sample slide receptacle could be fabricated using standard injection moulding techniques have been presented in Figure 3.16. The simulations display the injection pin locations that produced the best outcomes in terms of confidence of fill and predicted part quality for each material. Both PP and ABS simulations returned a 100% confidence of fill across the entire component. The PP part took 0.68 seconds to fill while the ABS part took 0.33 seconds. PP showed a better chance of producing high

quality parts than ABS. However, the ABS part utilised two injection pins compared to a single injection pin needed for the PP part. It was determined that all features could be manufactured to a reasonable level of quality with about 68.5% of the receptacle produced with medium quality and 31.5% of high quality. MoldFlow indicated the reason the simulation predicted the ABS part may produce areas of medium quality is because the ABS may experience sheer stresses too high during injection. To reduce sheer stress the melt viscosity can be lowered by increasing the melt and mould temperature [39]. Increasing the melt temperature to 240°C and mould temperature to 80°C improved the predicted part quality to 55.7% of the part being high quality and 44.3% being of medium quality (Figure 3.16 (c)). For an ABS receptacle, the areas of concern are the right and left flanks which need to be the focus of future design improvements of this component. These findings indicate that either ABS or PP may be suitable for scaled production of the sample slide receptacle, albeit with different injection gate configurations required for each material.

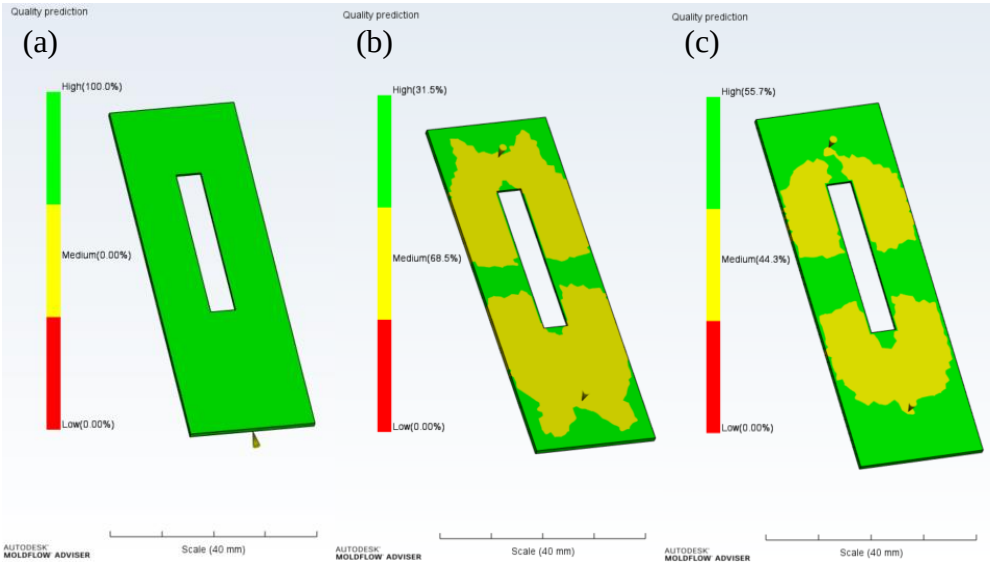


Figure 3.16. *Results from a simulation of the injection moulding the receptacle in PP (a) and ABS (b) using recommended MoldFlow parameters and improved ABS parameters (c). It is predicted that 100% of the component will be of high quality when made form PP and 31.5% will be of high quality when made from ABS with the remaining 68.5% of medium quality. Simulations indicate that by increasing the mould and melt temperature for the ABS part the probability of higher part quality improved to 55.7% high and 44.3% medium quality. The small cones on the parts indicate the injection pin location.*

3.4. Discussion

Lowering pitch and increasing groove width aimed to reduce the chance of multiple internal reflections occurring in order to limit the presence of silicon lattice vibration artefacts in the spectra whilst also removing the need for beam collimation by increasing the beam coupling surface area. Spectral artefacts may occur when spatial features match in size to the wavelength of light used to interrogate the sample [40]. The 25µm lower limit selected for pitch, P, matches the wavelength of mid-infrared light to see if this resulted in any interference in the form of scattering

artefacts. No evidence of this type of scattering was observed, likely because the 25µm features lie at the extreme upper end of the mid-infrared wavelength range (2.5-25µm). Smaller distances separating each consecutive groove allows for more grooves to be etched within a defined area providing more opportunities for IR light to properly couple into the SIRE and follow the intended beam path. It therefore benefits the SIRE design to reduce this value by as far as possible, especially since no scattering could be observed when P = 25µm. The depth of etched grooves is directly proportional to their width. Deeper grooves give more area for the IR beam to couple into the crystal, but also weaken the SIRE by intensifying stress concentrations located at the trough of the grooves. The upper limit (W = 250µm) was selected so that the groove etched was roughly halfway through the SIREs. At the lower limit (W = 100µm) there is a 40% reduction in area of beam coupling surface exposed to the IR beam in comparison to the 250µm wide grooves. This lower limit was selected arbitrarily to determine if a reduction in beam coupling area would have an impact on spectral intensity. It could not be conclusively determined if changing the size of pitch of the grooves had any impact on average SNR, but wider grooves with smaller pitch did improve SNR RSD%, and average signal intensity and signal intensity RSD% of amide I and amide II bands (Table 3.2 (bottom), Table 3.3 and Table 3.4). There is also a general trend that suggests thinner SIREs improve SNR of resultant spectra (Table 3.2). The SNR RSD% of thinner SIREs also indicates better reproducibility of thinner SIRE designs. 250µm groove width, 25µm pitch, 380µm crystal thickness has been selected as the optimal SIRE design parameters for future analysis because it produced the most reliable spectra (2.1 amide I RSD%, 1.5 amide II RSD%) with good signal quality (136.7 SNR).

It is essential that DSP wafers are used in this application because both sides of the wafer will be in active use. The upper surface (or the sample side) needs to be extremely flat to avoid light scattering at the interface between the sample and the beam. The lower surface (the optical side) also needs to be highly polished as this face is to be etched. When carrying out chemical wet etching processes it is easier to mark patterns to be etched upon polished surfaces thereby helping to ensure more reliable results [41]. All wafers used within this project have been made P-type by lightly doping the silicon with Boron. Un-doped silicon is a difficult specification to fulfil when sourcing wafers due to the added difficulty in producing higher purity silicon. Fortunately, the vibrational bands for boron dopants in silicon are located at wavenumbers 668.8cm^{-1} and 319cm^{-1} which is outside the biological fingerprint region of interest [6], [9]. Pure FZ SIREs are a sensible upgrade for future directions, but the Cz SIREs are of high enough quality that the higher costs of FZ wafers cannot be justified in prototyping and proof-of-concept stages.

It is feasible that all components of the SIRE sample slide are suitable for economical scaled production. Silicon microfabrication techniques are well established and widely used the world over in silicon microfabrication foundries for the production of microchips [42]. The receptacle can be fabricated using standard injection moulding techniques, but some design alterations may be required in the future to improve manufacturing yield of ABS receptacles as generic injection moulding techniques may incur shear stresses too high to reliably fabricate high quality receptacles [43]. However, simulations have shown that PP receptacles could be reliably fabricated at high quality with no design modifications. Labels are easily printed in bulk. Together the SIRE, label, and plastic receptacle could easily be

assembled using pick and place automated assembly line techniques giving a clear and achievable pathway to scaled production of sample slides. The custom accessory module is not a consumable item, so does not necessarily require high-throughput scaled production, but since no advanced fabrication techniques are required in its construction it could be manufactured and assembled quickly and at relatively low cost. However, consideration must be given to other costs associated with the transition from development to clinical market such as implementation of a quality management system, maintenance, and distribution. The implications of increased costs for the application of SIREs as a brain cancer test platform has already been investigated through a health economic analysis [23], but a separate economic analysis would need to be carried out to elucidate the consequence of increased costs of SIREs when used as a sepsis testing platform.

3.5. Conclusion

The design process of a disposable Silicon Internal Reflection Element sample slide and an automated slide indexing unit for a Quest ATR universal accessory has been described. Sample slide design rationale was informed by resultant spectral quality and reproducibility and on the basis of compliance to strict financial constraints. Good signal quality (SNR) was achieved, and device reproducibility was optimised through modulating both SIRE design parameters and receptacle design parameters. It has been demonstrated that the SIRE sample slide is capable of delivering a high-throughput ATR-FTIR workflow for serum diagnostics with the

construction of a custom ATR accessory unit. Consideration has been given to how sample slide design fits into economic models as a brain cancer diagnostics platform by considering production costings and feasibility of scaled manufacture. Some of the potential difficulties with regards to serum sample analysis and fabrication of the SIRE sample slides have also been discussed. SIRE performance characterisation in the context of serum analysis will be established in subsequent chapters.

3.6. References

[1] J. R. Hands *et al.*, "Attenuated Total Reflection Fourier Transform Infrared (ATR-FTIR) spectral discrimination of brain tumour severity from serum samples," *J. Biophotonics*, vol. 7, no. 3–4, pp. 189–199, 2014, doi: 10.1002/jbio.201300149.

[2] M. J. Baker, E. Gazi, M. D. Brown, J. H. Shanks, P. Gardner, and N. W. Clarke, "FTIR-based spectroscopic analysis in the identification of clinically aggressive prostate cancer," *Br. J. Cancer*, vol. 99, no. 11, pp. 1859–1866, 2008, doi: 10.1038/sj.bjc.6604753.

[3] B. R. Wood, M. A. Quinn, F. R. Burden, and D. Mcnaughton, "An Investigation into FTIR Spectroscopy as a Biodiagnostic Tool for Cervical Cancer," *Biospectroscopy*, vol. 2, pp. 143–153, 1996.

[4] H. J. Byrne *et al.*, "Spectropathology for the next generation: Quo vadis?," *Analyst*, vol. 140, no. 7, pp. 2066–2073, 2015, doi: 10.1039/C4AN02036G.

[5] C. Hughes *et al.*, "Assessing the challenges of Fourier transform infrared spectroscopic analysis of blood serum," *J. Biophotonics*, vol. 7, no. 3–4, pp. 180–188, 2014, doi: 10.1002/jbio.201300167.

[6] M. J. Baker *et al.*, "Using Fourier transform IR spectroscopy to analyze biological materials," *Nat. Protoc.*, vol. 9, no. 8, pp. 1771–1791, 2014, doi: 10.1038/nprot.2014.110.

[7] H. Schumacher, U. Kunzelmann, B. Vasilev, K. J. Eichhorn, and J. W. Bartha, "Applications of microstructured silicon wafers as internal reflection elements in attenuated total reflection Fourier transform infrared spectroscopy," *Appl. Spectrosc.*, vol. 64, no. 9, pp. 1022–1027, 2010, doi: 10.1366/000370210792434404.

[8] D. Widmann, H. Mader, and H. Friedrich, "Film Technology," in *Technology of Integrated Circuits*, 2000, pp. 14–95.

[9] J. W. Medernach, “Infrared Characterization of Device-Quality Silicon,” in *Handbook of Vibrational Spectroscopy*, 2006, pp. 2621–2636.

[10] W. Zulehner, “Historical overview of silicon crystal pulling development,” *Mater. Sci. Eng.*, vol. B, no. 73, pp. 7–15, 2000.

[11] P. Lasch, “Spectral pre-processing for biomedical vibrational spectroscopy and microspectroscopic imaging,” *Chemom. Intell. Lab. Syst.*, vol. 117, pp. 100–114, 2012, doi: 10.1016/j.chemolab.2012.03.011.

[12] L. Lovergne, G. Clemens, V. Untereiner, R. A. Lukaszweski, G. D. Sockalingum, and M. J. Baker, “Investigating Optimum Sample Preparation for Infrared Spectroscopic Serum Diagnostics,” *Anal. Methods*, vol. 7, no. 17, pp. 6921–7382, 2015, doi: 10.1039/C5AY00502G.

[13] Specac, “The Quest Single Reflection ATR Accessory User Manual,” 2018.

[14] H. J. Gulley-stahl, J. A. Haas, K. A. Schmidt, A. P. Evan, and A. J. Sommer, “Attenuated Total Internal Reflection Infrared Spectroscopy (ATR-FTIR): A Quantitative Approach for Kidney Stones Analysis,” *Appl. Spectrosc.*, vol. 63, no. 7, pp. 759–766, 2009, doi: 10.1366/000370209788701044.Attenuated.

[15] M. Pilling and P. Gardner, “Fundamental developments in infrared spectroscopic imaging for biomedical applications,” *Chem. Soc. Rev.*, vol. 45, no. 7, pp. 1935–1957, Mar. 2016, doi: 10.1039/C5CS00846H.

[16] T. G. Mayerhöfer and J. Popp, “Electric field standing wave effects in internal reflection and ATR spectroscopy,” *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.*, vol. 191, pp. 165–171, 2018, doi: 10.1016/j.saa.2017.10.007.

[17] D. F. Swinehart, “The Beer-Lambert Law,” *J. Chem. Educ.*, vol. 39, no. 7, pp. 333–335, 1962, doi: 10.1021/ed039p333.

[18] M. M. Pradhan, R. K. Garg, and M. Arora, “Multiphonon infrared absorption in silicon,” *Infrared Phys.*, vol. 27, no. 1, pp. 25–30, 1987, doi: 10.1016/0020-0891(87)90046-7.

[19] G. Azarfar *et al.*, “Estimating and correcting interference fringes in infrared spectra in infrared hyperspectral imaging,” *Analyst*, vol. 143, no. 19, pp. 4674–4683, 2018, doi: 10.1039/c8an00093j.

[20] J. M. Chalmers, “Mid-Infrared Spectroscopy: Anomalies, Artifacts and Common Errors,” in *Handbook of Vibrational Spectroscopy*, 2006, pp. 2327–2347.

[21] F. Bonnier, M. J. Baker, and H. J. Byrne, “Vibrational spectroscopic analysis of body fluids: Avoiding molecular contamination using centrifugal filtration,” *Anal. Methods*, vol. 6, no. 14, pp. 5155–5160, 2014, doi: 10.1039/c4ay00891j.

[22] C. Petibois, V. Rigalleau, A. Melin, A. Perromat, G. Cazorla, and H. Gin, “Determination of Glucose in Dried Serum Samples by Fourier-Transform Infrared Spectroscopy,” *Clin. Chem.*, vol. 45, no. 9, pp. 1530–1535, 1999.

[23] E. Gray *et al.*, “Health economic evaluation of a serum-based blood test for

brain tumour diagnosis: Exploration of two clinical scenarios," *BMJ Open*, vol. 8, no. 5, pp. 1–9, 2018, doi: 10.1136/bmjopen-2017-017593.

[24] H. Sohn, J. Minion, H. Albert, K. Dheda, and M. Pai, "TB diagnostic tests: How do we figure out their costs?," *Expert Rev. Anti. Infect. Ther.*, vol. 7, no. 6, pp. 723–733, 2009, doi: 10.1586/ERI.09.52.

[25] J. R. Genzen, "Regulation of Laboratory-Developed Tests," *Am. J. Clin. Pathol.*, vol. 152, no. 2, pp. 122–131, 2019, doi: 10.1093/ajcp/aqz096.

[26] P. Wang, "Developing In Vitro Diagnostics for Commercialization and Clinical Implementation," in *Intellectual Property-Regulator*, vol. 1, no. 3, 2019, pp. 1–6.

[27] N.-T. Nguyen, "Fabrication technologies," in *Micromixers (Second Edition)*, 2012, pp. 113–161.

[28] A. J. Nijdam, J. G. E. Gardeniers, J. W. Berenschot, E. van Veenendaal, J. van Suchtelen, and M. Elwenspoek, "Influence of the angle between etched (near) Si { 111 } surfaces and the substrate orientation on the underetch rate during anisotropic wet-chemical etching of silicon," *J. Micromechanics Microengineering*, vol. 11, pp. 499–503, 2001.

[29] H. B. H. Seidel, L. Csepregi, A. Heuberger, "Anisotropic Etching of Crystalline Silicon in Alkaline Solutions: Orientation Dependence and Behaviour of Passivation Layers," *J. Electrochem. Soc.*, vol. 137, no. 11, pp. 3612–3626, 1990.

[30] W. Dong, X. Zhang, C. Liu, M. Li, B. Xu, and W. Chen, "Mechanism for convex corner undercutting of (110) silicon in KOH," vol. 35, pp. 417–419, 2004, doi: 10.1016/j.mejo.2004.01.005.

[31] J. Henrie, S. Kellis, S. M. Schultz, and A. Hawkins, "Electronic color charts for dielectric films on silicon," *Opt. Soc. Am.*, vol. 12, no. 7, pp. 1464–1469, 2004.

[32] W. Kern, "The Evolution of Silicon Wafer Cleaning Technology," *J. Electrochem. Soc.*, vol. 137, no. 6, pp. 1887–1892, 1990.

[33] S. Tan, R. Boudreau, and M. L. Reed, "Effects of Mask Misalignment and Wafer Misorientation on Silicon V-Groove Etching," vol. 15, no. 2, pp. 101–112, 2003.

[34] R. E. Oosterbroek *et al.*, "Etching Methodologies in 111 -Oriented Silicon Wafers," *J. Microelectromechanical Syst.*, vol. 9, no. 3, pp. 390–398, 2000.

[35] M. A. Gosálvez, I. Zubel, and E. Viinikka, "Wet Etching of Silicon," in *Handbook of Silicon Based MEMS Materials and Technologies: Second Edition*, 2015, pp. 470–502.

[36] S. Processing, "Etching roughness for (100) silicon surfaces in aqueous KOH," *J. Appl. Phys.*, vol. 70, no. 6, pp. 3291–3300, 1991, doi: 10.1063/1.349263.

[37] I. Zubel, "Silicon anisotropic etching in alkaline solutions II on the influence of anisotropy on the smoothness of etched surfaces," *Sensors Actuators, A Phys.*,

vol. 70, no. 3, pp. 260–268, 1998, doi: 10.1016/S0924-4247(98)00142-3.

[38] S. S. Singh, P. Pal, A. K. Pandey, Y. Xing, and K. Sato, “Determination of precise crystallographic directions for mask alignment in wet bulk micromachining for MEMS,” *Micro Nano Syst. Lett.*, pp. 65–67, 2016, doi: 10.1186/s40486-016-0027-5.

[39] I. T. Barrie, “The Rheology of Injection Moulding,” in *Polymer Rheology*, no. 2, 1978, pp. 141–163.

[40] R. Lukacs, R. Blümel, B. Zimmerman, M. Baıcioılu, and A. Kohler, “Recovery of absorbance spectra of micrometer-sized biological and inanimate particles,” *Analyst*, vol. 140, no. 9, pp. 3273–3284, 2015, doi: 10.1039/c5an00401b.

[41] K. Pandey and P. M. Pandey, “Chemically Assisted Polishing of Monocrystalline Silicon Wafer Si (100) by DDMAF,” vol. 184, pp. 178–184, 2017, doi: 10.1016/j.proeng.2017.04.083.

[42] S. Franssila, “Economics of Microfabrication,” in *Introduction to Microfabrication*, 2010, pp. 457–468.

[43] O. Brandau, “Material Basics,” in *Stretch Blow Molding*, 2017, pp. 5–25.

Chapter 4: SIRE Sample Slide Performance Evaluation and Characterisation

4.1. Introduction

After producing a functioning and effective high-throughput ATR-FTIR system it was necessary to characterise the SIRE devices and formulate a standard operating procedure. Doing so will help to maximise the reproducibility of spectra acquired from SIREs. As noted in chapter 3 section 3.3.1, the spectra acquired using SIREs were highly sensitive to faults in serum pipetting technique, but reproducibility issues may arise within other aspects of the sample preparation protocol and also in the assembly of SIRE slides. Sources of error must be elucidated in order to properly manage them. These sources of error have been broadly categorized as either intrinsic (e.g., silicon lattice discrepancies), or extrinsic (e.g., pipetting method, sample drying method, SIRE cleaning protocols, sample collection protocols). The silicon itself may give rise to variations in the spectra from SIRE to SIRE as impurities, added either intentionally to bestow electrical or mechanical properties or unintentionally through imperfections during the silicon crystal growth process, are not evenly distributed throughout the lattice [1]. The impurities present in silicon wafers is a thoroughly researched area and is highly dependent on the method and conditions under which the silicon crystals are grown [2]. The two most common methods are the Czochraslki (Cz) and Float Zone (FZ) methods [3], [4]. The former generally produces wafers with a markedly higher impurity content but is the best and most widely used method for mass production of semi-conductor grade silicon wafers. Generally, impurities form swirl or concentric ring patterns across the diameter of silicon wafers which may give rise to heterogeneity of optical properties of SIREs fabricated from different parts on a wafer [5], [6].

A core argument for the high-throughput ATR approach is the ability to carry out batched sample processing. It then follows that variance introduced through batched sample preparation process could undermine the entire workflow. Sample preparation of serum involves drying the sample upon the IRE to reduce the water signal that would otherwise obscure pertinent spectral information [7]. Biomolecular constituents are more homogeneously distributed within liquid sera samples than in dried sera films due to the phenomena known as the 'coffee ring effect' which causes inhomogeneity in molecule distribution following droplet evaporation since molecules of certain molecular weights precipitate in concentric rings around the droplet deposition site [8]–[10]. This introduces the potential for biomolecular information to be misrepresented in the spectra [11]. Thus, automated pipetting and drying methods are desirable. Another extrinsic source of error may be from particles accumulating during storage or handling prior to sample deposition. Prior to dicing, the Si wafer is coated in a protective layer of photoresist and can remain coated in this protective layer before assembly into the sample slide as any particles captured in the photoresist can be easily removed [12]. However, conventional cleaning agents such as ethanol or acetone may not be sufficient since these solvents may alter surface properties or leave evaporation residues which are visible in the spectra [13]. A standardised cleaning regime that eliminates such effects would benefit the optical reproducibility of SIREs. Following assembly, hermetically sealed packaging will likely be a necessary first line defence to prevent contaminant ingress, although to reduce costs non-hermetic packaging needs to be explored [14]. One facet of the sample slide is the ability store samples post-analysis. In the case of biological samples, such as serum, this is only useful if the sample is not compromised during storage.

This chapter details the findings of multiple investigations aimed at characterising the various potential error sources of the high-throughput ATR system and remark on how these error sources may be mitigated. Sources of variance have been investigated from the construction of the SIREs and the preparation and treatment of serum samples. Ideally, sample collection method would also be optimised for SIRE analysis, but accessibility to participants would be challenging and has not been carried out within the scope of this project. Controlled drying techniques were investigated using fans, heating pads and an incubator to reduce drying times by as much as possible with the least impact on spectral quality. Cleaning procedures were also considered to harmonise the surface properties between SIRE chips. The severity of intrinsic silicon variance has been investigated by comparing spectra of the silicon lattice of SIREs either selected randomly or selected from known locations across a silicon wafer. Lastly, investigations were carried out to assess how the analytical capabilities of a SIRE weighed up against other IRE options. By characterising these factors, it is envisaged that the SIRE approach will be more reproducible for clinical applications where accuracy is paramount to successful regulatory and clinical approval.

4.2. Materials and Methods

4.2.1. Spectral Acquisition

All analysis was carried out on a Spectrum2 infrared spectrometer (PerkinElmer, USA) with a custom automated slide indexing unit (see chapter 3, section 3.2.3) and Quest ATR accessory unit (Specac, UK). Silicon Carbide (SiC) IRE

puck, Diamond (Di) IRE puck and specular reflectance pucks were used throughout various experiments. Spectra were acquired from the 4000-450cm^{-1} range using 16 co-added scans and a resolution set to 4cm^{-1} with 1cm^{-1} data spacing. Positions '1', '2', and '3' were sampling positions of the SIRE and background scans were acquired using either position '0' of the SIRE sample slide, or a gold mirror depending on the aspect under investigation.

4.2.2. Data Processing

Spectra have been presented both raw and processed throughout. When processed, spectra were rubber band baseline corrected then vector normalised. Remarks have been made at each point of the spectral processing for each study, as in many instances different information is elucidated throughout spectral processing. Spectra were not cut prior to processing unless otherwise stated.

To analyse the impact of extrinsic and intrinsic factors on variance, spectra standard deviations were plotted and PCA was conducted. A cut-off value of 90% explained variance was taken when selecting the number of components to investigate. Statistical analysis was conducted in Minitab (Pennsylvania, USA) with a value of $p < 0.05$ considered significant. Where necessary, absorbance and SNR values were calculated for comparison. SNR was calculated by dividing the absorbance intensity of the amide I band (1625-1675cm^{-1}) by the standard deviation of wavenumber range representative of baseline noise (1800-1850cm^{-1}). A Matlab (MathWorks, USA) based General User Interface (GUI) developed in-house was generally used for pre-processing and exploratory analysis.

4.2.3. Intrinsic Variance Investigations

The most prominent impurities in silicon wafers are oxygen, carbon, and boron [4]. Nitrogen may also be present in circumstances where the crystals are grown within a silicon nitride crucible [15], [16]. Impurities are usually a desired feature in the field of semiconductors adding structural integrity, bestowing electrical properties, permitting free-charge carrier movement, and altering optical properties [17], [18]. These impurities can form bonding environments that are clearly visible in the SIRE spectrum together with vibrations from the silicon lattice itself. Figure 4.1 shows a baseline corrected SIRE spectrum with annotations of notable peaks. These were the two different silicon phonon bands (610cm^{-1} and 819cm^{-1}), the Si-B band (668.8cm^{-1}), and the Si-O band (1107cm^{-1}) and can be used to estimate the variability of impurities across the SIREs. The most prominent of these is by far the silicon phonon at 610cm^{-1} which is overlapped by the Si-C band at 605cm^{-1} rendering the analysis of substitutional carbon content difficult. Variances in the silicon lattice itself are instead estimated using the 819cm^{-1} silicon phonon 2 band as this is the most faithful indicator of pure silicon interference [2]. The interstitial Si-O peak will distort from 1107cm^{-1} to a centre position at 1040cm^{-1} in the presence of various oxygen precipitate forms so is a good indicator of impurity homogeneity across the silicon wafer [19].

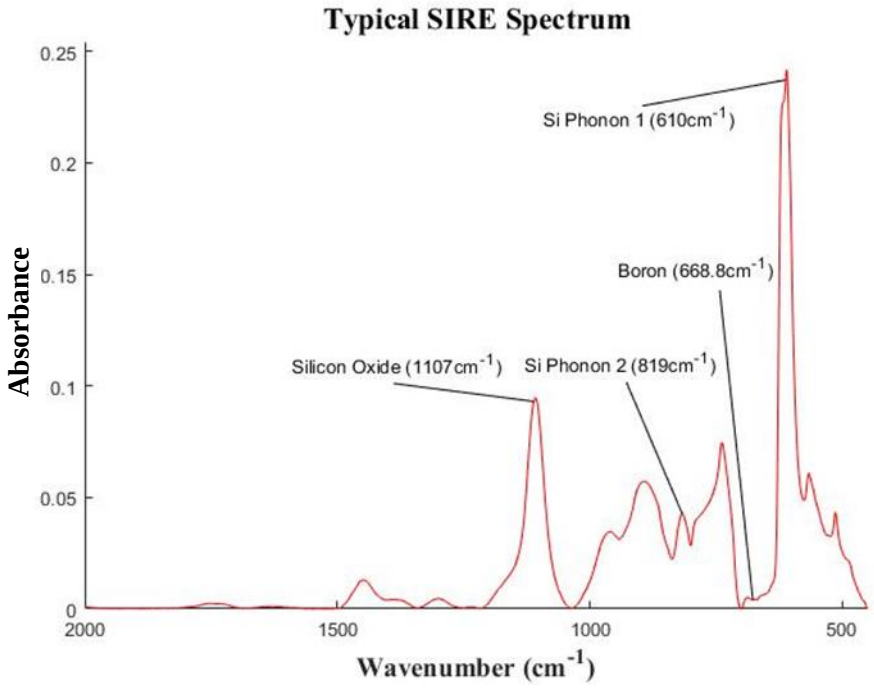


Figure 4.1. *Typical baseline corrected spectrum of silicon the SIREs are fabricated from. Significant peaks have been annotated.*

4.2.3.1. Silicon Variance

SIRE slides were selected at random (n=14). A gold mirror was placed on the specular reflectance puck and scanned for a background spectrum of the environment because gold is an excellent reflector of IR light and will not interfere with the reference signal. Positions ‘1’, ‘2’, and ‘3’ of each SIRE slide were then scanned on the specular reflectance puck (see chapter 3, section 3.2.2 for schematic of SIRE sample slide). A new background scan was acquired prior to analysing each slide. 126 spectra were collected. 3µl of HPS was then pipetted into positions ‘1’, ‘2’, and ‘3’, dried, then scanned again to see if general observations of variance in the silicon lattice spectra could be observed in the spectra of serum samples collected using the same

SIREs. Samples were scanned in the same way, but this time scanning position ‘0’ of the slide to acquire a background scan. A further 126 spectra were collected.

4.2.3.2. Silicon Intra-Wafer Variance

An experiment was set up to examine if the location on a silicon wafer from which a SIRE originated impacted the variance between SIREs given the non-homogeneous distribution of impurities. For the purposes of this study, a wafer was diced into individual SIRE chips as opposed to the finalised SIRE slide design described in the previous chapter. Individually cut SIREs were removed from 66 positions across the wafer (positions 1-66, Figure 4.2) and analysed using a gold mirror to acquire a background scan. Nine spectra were acquired from each of the locations for a total of 585 spectra. The spectra from each location were averaged then the entire dataset was baseline corrected and vector normalised. No meaningful spectra could be retrieved from position 66 as the wafer was heavily damaged in this area and so this data has been removed from the study. An Anderson-Darling normality test was performed on both the Si2 and the SiO peaks to determine if the dataset was normally distributed [20], [21]. This was then followed by a one-sample student t-test to determine if the peaks were significantly different from a hypothetical mean of 0.0276 for the Si2 peak and 0.062 for SiO.

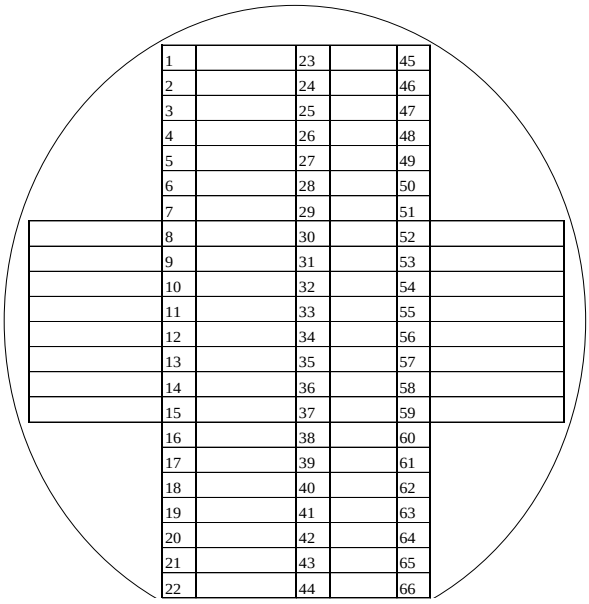


Figure 4.2. *Each location individual SIRE chips were taken from across the silicon wafer.*

4.2.4. Extrinsic Variance Investigations

4.2.4.1. Optimum Drying Conditions

A series of simple experiments were arranged to determine optimal drying conditions. In the first instance, a fan and heating pads, either alone or in combination, were used to force dry 3µl of serum samples pipetted onto the Di top-plate accessory module of the Spectrum2. A negative feedback heating system was set-up using an Arduino Mega 2560 (Arduino, Italy) microcontroller, two RS Pro 12V 2W circular adhesive silicone heating pads (RS components, UK), two RS Pro 10kΩ (at 25°C) 2.4mm diameter thermistors (RS components, UK), two TE Connectivity 10kΩ metal film fixed resistors (RS components, UK), two ON semi TIP120G NPN Darlington Pair 8A 60V transistors (RS components, UK), TMP102 digital temperature sensor breakout board (Cool Components, UK) and a small 5V DC Sunon® DC axial fan (RS components, UK). The heating pads were positioned on either side of the Di IRE and

the fan was mounted on 3D printed legs directly above the IRE. The set-up could accurately maintain temperatures of 25°C, 30°C, and 35°C and air flows of 5CFM and 9CFM. As an extra precaution, temperature was monitored visually with a mercury thermometer taped directly to a heating pad. Full details of the experimental set-up are presented in Appendix IV (chapter 8). To prevent any denaturation of proteins in the serum, the samples were kept strictly below 37°C. A total of six different drying regimes were tested: two with only fans (set to either 5CFM and 9CFM), two with only heating pads (set to either 30°C or 35°C), a combination of both fan and heating pads (set to 9CFM and 35°C respectively), and a control (ambient conditions, 25°C) (Table 4.1). In total, three sensing mechanisms were employed to monitor temperature: the thermistors (±0.9% tolerance = approx. ±0.315°C at 35°C), the temperature sensor breakout board (±0.5°C), and the mercury thermometer (±0.1°C). Thermistor and breakout board tolerances have been extracted from manufacturer datasheets and mercury thermometer tolerances have been approximated from a published clinical reliability study [22]. The 5V fan tolerances have been estimated based on the manufacturer's specification of ±15% fan speed tolerance at 7200RPM. The most conservative tolerance estimates have been presented in each case. Serum was allowed to dry for 0, 2, 4, 6, 8, 16, and 32 minute intervals taking three spectra after each interval.

Table 4.1. *Experimental airspeed and temperature settings with estimated tolerances.*

Drying Regime	Air flow (CFM)	Temperature (°C)
Fan 1	5 ±0.75	25 ±0.5°C
Fan 2	9 ±1.35	25 ±0.5°C
Heater 1	0 ±0.00	30 ±0.5°C
Heater 2	0 ±0.00	35 ±0.5°C
Combined	9 ±1.35	35 ±0.5°C
Control	0 ±0.00	25 ±0.5°C

Following this, a similar experiment was set up using SIREs. This particular experiment was conducted on early prototypes where only two identical SIREs were available. 3µl of serum samples were pipetted onto the SIREs which were cleaned after each spectral acquisition using a Virkon soaked cotton bud before preparing the SIRE with more serum. The exact same heating control system was used with the exception of the 5V fan being replaced with a 12V DC ARX CeraDyna series axial fan (RS components, UK). Each SIRE was also placed on top of a single 12V heating pad and under the fan. Temperatures could be kept stable at 25, 30, and 35°C (all ±0.5°C as previously) and fan air flows could be set to 9 ±0.9CFM, 60 ±6.0CFM and 70 ±7.0CFM. Again, temperature was monitored visually with a mercury thermometer as an extra precaution. 12V fan tolerances have been estimated from the manufacturer's specification of a 10% fan speed tolerance at 3600RPM. Serum samples were allowed to dry for 0, 2, 4, and 6 minute intervals with spectra acquired at each interval. This was repeated at least five times per fan and heater setting combinations and the spectra averaged and baseline corrected.

Later, a variable temperature incubator was acquired that could simultaneously dry 36 SIRE slides at once. To verify the incubator maintained similar, or better, performance benefits of the previous drying methods, 3µl serum droplets were pipetted onto individual SIREs and placed under a fan set to 60CFM within the incubator set at 30°C and allowed to dry for 0, 2, 4, 6, 8, 16, and 32 minute intervals taking a spectrum after each interval. Three Spectra were collected per interval which were averaged and baseline collected for analysis. A total of 456 spectra were collected for this study.

4.2.4.2. Serum Variance Investigations

A series of small experiments were set up to identify if and how experimental error can be introduced when analysing serum samples. 3µl samples of dried HPS was analysed on a group of 14 randomly selected SIREs (group 1). Three biological and three instrument repeats were performed per SIRE (n = 126 spectra). PCA was performed to determine if spectra could be separated by the SIREs they originated from. A one-sample student t-test was performed on the scores of the PC1 direction to determine if the mean PC1 score differed significantly from a hypothetical mean of zero. Since PC1 is the direction of greatest variance, reproducible spectra should not differ significantly from the centre point of the PC.

Group 1 SIREs were stored in a desiccator overnight with the dried HPS still intact. The group was then reanalysed the next day in the same manner. PCA was performed again to determine if separation could be observed between PC scores of spectra obtained on 'day 1' and 'day 2'. A paired student t-test was performed on the PC1 scores of the spectra acquired on each day to determine if dried sera films were stable overnight.

A second group of 14 randomly selected SIREs (group 2) was used to analyse 3µl dried sera samples again with three biological repeats and three instrument repeats per SIRE (n = 126 spectra). PCA was performed to determine if separation could be observed between PC scores of spectra obtained from two different groups of SIREs. An unpaired student t-test was performed on the PC1 scores of the spectra acquired from each group to determine if serum analysed on different days on different SIREs were reproducible.

Lastly, two SIRE slides were used to analyse 3µl of HPS. On one slide the serum was spread to fit the entire confinements of the sample area, while on the other slide the serum was spot in the centre of the sample area and left alone. This was to demonstrate how a difference in pipetting method can impact spectral quality.

4.2.4.3. SIRE Surface Treatments

Cleaning protocols using three different cleaning agents were investigated. Acetone (≥99.9%, Sigma-Aldrich, UK), Piranha and a Virkon (Fisher Scientific, UK) control group were selected for investigation. Acetone is a ubiquitous laboratory solvent, Piranha is a colloquial term for a powerful corrosive widely used to eliminate organic matter from laboratory glassware, and Virkon is a trademark name for a multi-purpose alkaline disinfectant. Formulations for Piranha vary, but here 50ml batches have been made in house from three parts (30ml) sulfuric acid (96% solution in water, Thermo Fisher Scientific, UK) to one part hydrogen peroxide (30% in water, Thermo Fisher Scientific, UK). For the Virkon solution, 0.5g of Virkon powder was measured out using an analytical balance then added to 50ml warm water to give a 1% Virkon solution. Virkon was selected for the control group because decontamination of biological samples using Virkon is a procedural requirement enforced by Strathclyde University safety standards meaning all SIRE groups needed to be cleaned with Virkon anyway. All formulations were made fresh each day of experimentation.

Five SIREs were used per protocol over a course of five days. On each day 3µl of serum was pipetted into wells ‘1’, ‘2’, and ‘3’ of each SIRE and dried before analysis. All bulk sample was then removed with a cotton swab soaked in Virkon as

per procedural requirements enforced by Strathclyde University safety standards. SIREs were then immersed in either acetone, Virkon, or Piranha for 30 minutes. The acetone group was then immediately transferred to 2-propanol (≥99.5%, Sigma-Aldrich, UK) to stop the formation of evaporation stains and salt residues. All SIREs were then rinsed in deionised (DI) water and dried using a nitrogen gun. SIREs were then left in a desiccator overnight and the process was repeated for each group the next day. Three spectra were collected from each SIRE per day for a total of 135 spectra.

4.2.5. SIRE Analytical Characteristics

4.2.5.1. SIRE comparison to other IRE options

SIREs were compared to four commercial and non-commercial IRE options. These were: A Diamond IRE (Di IRE), a Silicon Carbide (SiC) IRE, a blank piece of Double Side Polished (DSP) silicon, a blank piece of Single Side Polished (SSP) silicon. The intention is to firstly compare SIREs to commercially available systems (Di IRE and SiC IRE) and to highlight the importance of the optically engineered aspects of the SIREs by comparing them to blank, un-etched SSP and DSP silicon pieces. Ethanol (Absolute ≥ 99.5%, Sigma-Aldrich, UK) was selected for this purpose as it is a readily available substance and interacts strongly with IR light. A stock solution of 50% v/v ethanol in water was made for this study by diluting 1ml ethanol with 1ml water using a pipette. Each IRE type was used to analyse 3µl of the stock ethanol solution. A background scan to water was taken and subtracted from the spectra. The data was first inspected raw, then it was cut between 900-1350cm^{-1} (silicon phonon region) then baseline corrected.

4.2.5.2. Ethanol Concentration Study

In this part of the investigations only the Di IRE was compared to the SIRE. The ethanol stock solution was diluted into 2ml Eppendorf tubes to concentrations of 50, 25, 12.5, 6.25, 3.125, 2, 1, 0.5, 0.25, and 0% v/v ethanol in DI water. Dilutions were made by a serial dilutions method whereby concentrations down to 3.125% v/v were diluted by a factor of two and concentrations from 2 to 0.25% v/v were also diluted by a factor of two. The 2% v/v dilution was made by pipetting 20µl 50% v/v into 480µl DI water. The reason for staggering the serial dilutions in this way was to capture more datapoints at the low concentration range. All dilutions were thoroughly mixed through multiple inversions. An approximate error band of ±0.5% applies to each factor level based on a ≥99.5% (i.e., 0.5% uncertainty) purity of original solution. 3µl of each concentration was analysed three times with instrument triplicates on both the Di IRE and SIRE. A background scan of water was taken and subtracted from the spectra. Analysis was carried out immediately after solutions were made to prevent ethanol concentration loss by evaporation. A total of 180 spectra were collected. The data was first visually compared raw, then cut between 1020-1060cm^{-1} to only encapsulate the C-O peak of ethanol and then baseline corrected. The area under the C-O peak was then used plot calibration curves and to ascertain R^2 values, RMSE, LOD, and LOQ. Proper characterisation of LOD and LOQ along with the reproducibility of SIREs and high-throughput ATR will highlight any weaknesses and allows for effective quality control procedures to be formulated and may allude to specific applications SIREs are better suited to.

4.3. Results and Discussion

4.3.1. Intrinsic Variance Characterisation and Mitigation

4.3.1.1. Silicon Variance

Spectra of 14 SIRE slides are presented alongside spectra of serum acquired on the same slides in Figure 4.3. In Figures 4.3.(c) and (d) the average spectra have been indicated by a red line along with the standard deviation as represented by the shaded area. By visual inspection of Figures 4.3.(a) and (b) the baseline offsets do not seem to differ significantly from raw spectra of SIREs with or without serum on it indicating that this effect predominantly arises from differences in the crystal lattice itself. The spectra of SIREs curve up at the low wavenumber region starting from about 1700cm^{-1}. This is an artefact caused by free charge carrier absorption and can be eliminated by selecting silicon with a free carrier concentration of less than $15x10^{15}$atoms cm^{-3} [2], [23]. Pre-processing the spectra using a rubber band baseline correction effectively rectifies both baseline offset and free-carrier absorption artefacts, significantly reducing the standard deviation in both SIRE and serum spectra (Figures 4.3.(c) and (d)). To identify the specific spectral regions where variance is the highest, the standard deviation across the entire spectra has also been plotted. Standard deviation across the entire spectrum is in both cases very low roughly reaching a maximum of only $5x10^{-3}$, but slightly higher in the serum analysis. Figure 4.3.(f) shows many expected prominent peaks that are attributed to spectral variations of the biomolecular constituents of the serum samples with the exception of 1107cm^{-1} which is most likely due to variations in the Si-O band of the SIREs with which the serum was interrogated. Most surprising is the curve between 2000-2500cm^{-1} of the standard

deviation of serum (Figure 4.3.(f)), which is not visible in any other interpretation of the serum spectra. This curved spectral feature can be indicative of dispersion artefacts, and somewhat resembles resonant Mie scattering [11], [24]–[26]. It is possible that the close match between the size of the smallest features on the SIRE (25µm) and the mid-IR wavelength range (2.5-25µm) could give rise dispersion effects, but if this were the case then similar artefacts should be present in the corresponding SIRE spectra (Figure 4.3.(e)). Instead, It is more likely that micro-fissures in the dried serum films are of the same magnitude as the mid-IR wavelength range and therefore result in dispersion artefacts caused by scattering at the air to sample interface [11], [25]. In either case, the spectral contribution appears minimal and can be accounted for computationally [24]. Another notable feature present in the SIRE spectra (Figures 4.3.(a), (c), and (e)) is the prominent baseline noise located between 1500-1800cm^{-1} and 3500-4000cm^{-1}. This is strikingly similar to water vapour interference, a well-documented FTIR spectral artefact [27], and can be removed computationally and doing so would have little impact on the outcome of this particular analysis.

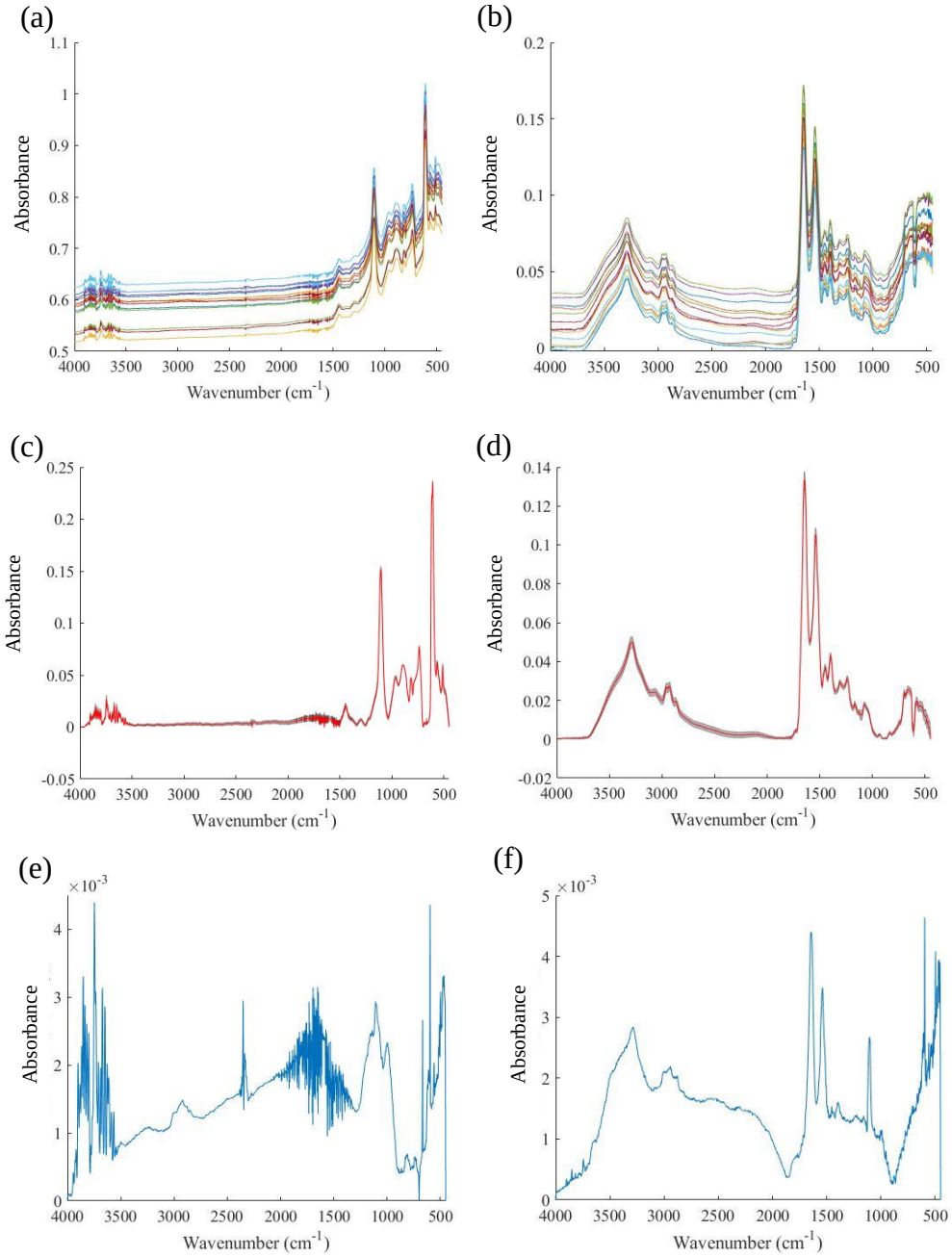


Figure 4.3. *Spectra of the silicon lattice of a group of 14 SIREs adjacent to spectra of serum taken using the same SIRE spectra highlights how problems with inconsistencies of the silicon lattice spectra translates into issues in the analysis of serum samples. (a) and (b) raw spectra demonstrates large baseline offsets between SIREs. (c) and (d) average spectra (red) with standard deviation of the spectra shown (shaded) following baseline correction. Standard deviation does not seem to increase a large amount when analysing serum. Complete attenuation of signal can be observed in the serum spectra at 610cm^{-1} which coincides with the Si1 phonon on the silicon lattice. (e) and (f) standard deviation of each case has been plotted which highlights low standard deviation across the entire silicon lattice spectrum.*

4.3.1.2. Silicon Intra-Wafer Variability

Impurities are distributed non-homogenously across silicon wafers in helical swirl, or ring-like patterns varying in character depending on the crystal growth conditions such as crystal rotation speed, pull rate, annealing process, dopants, and crucible material [1], [6], [28]–[30]. Ring distributions have been observed to range from microns to millimetres in width [5], [31]. However, impurity densities vary continuously across the wafer radius and generalizing the shape and size of these distributions is therefore difficult. SIREs cannot be fabricated such that they avoid such impurity distributions since the SIRE pattern must be aligned to the 110 crystal plane to produce v-shaped grooves as described in chapter 2 section 2.5.3. It is likely that SIREs may cross from regions of high impurity concentrations to low which will result in inhomogeneity across the SIRE crystal and thus produce variance across the three sample areas. If this variance can be quantified and understood, it may be possible to arrange spectral pre-processing protocols that negate it or apply boundaries that can identify high variance SIREs for quality control purposes. Spectra were acquired from 65 individual SIRE chips taken across the entire diameter of a wafer and scanned with reference to a Gold mirror. Technically this means the wafer was sampled at 6mm intervals and the exact position of each chip was recorded. Figure 4.4 shows the location on the silicon wafer each SIRE was sampled from along with a heat map displaying the average absorbance of the Si phonon 2 (Si2) at each corresponding point. No swirl pattern or concentric rings could be discerned from any spectral interpretation of the data gathered. While it is clear that there is some inhomogeneity of impurities across the wafer it does not appear to be in the form of the concentric ring pattern of oxygen precipitates described in the literature [19]. However, it may be

that a greater spatial resolution would be required to delineate such structures, or that the spectral contributions of the peaks of modulating impurity concentration (Si-O, Si-C, or Si-B) are not significant enough to be adequately spatially characterized using this mode of IR spectroscopy [32]. If this is the case it would suggest that impurity homogeneity has little impact on device reproducibility.

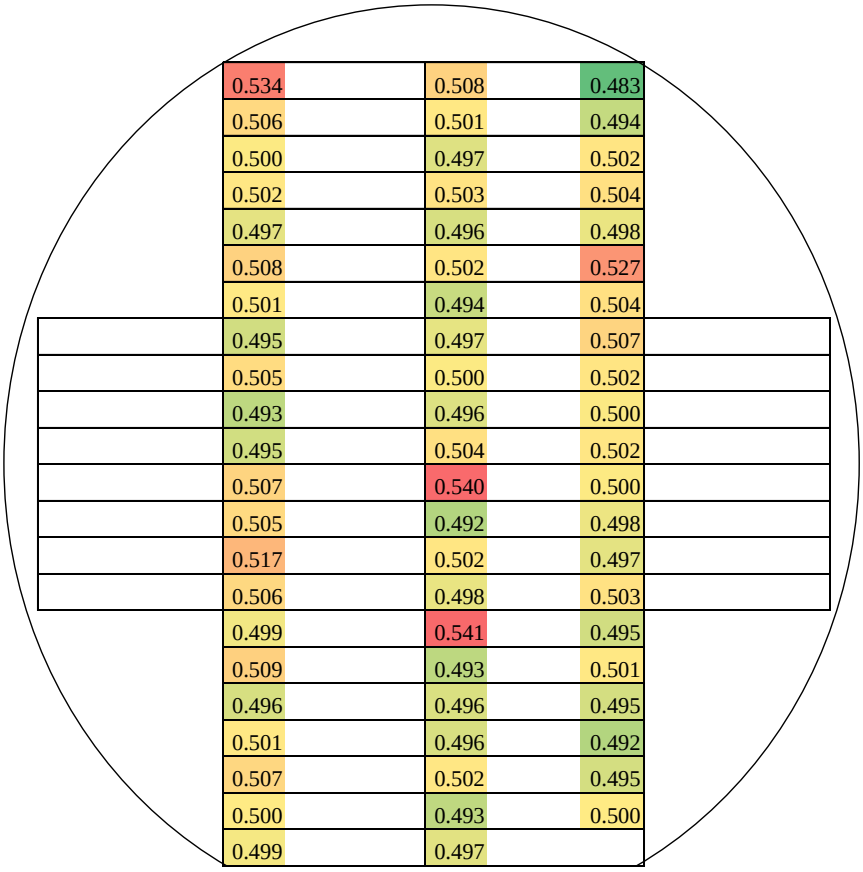


Figure 4.4. *Illustration of a silicon wafer with the points at which the silicon was sampled. Each box indicates a SIRE slide and boxes with numbers indicates a SIRE chip that was selected for spectroscopic analysis. Colours represent a heat map of the absorbance intensity of the silicon phonon 2 with green being lowest intensity and red the highest. No pattern could be distinguished that would indicate any kind of predictable spectral profile based on SIRE location from the source silicon wafer.*

The average spectra from all 65 SIREs of this study have been presented in Figure 4.5. In contrast to spectra previously displayed in Figure 4.3, spectra are highly

reproducible with a smaller standard deviation across the spectra as a whole and lack the water vapour artefacts located at 1500-1800cm^{-1} and 3500-4000cm^{-1}. The improvement in spectral standard deviation and lack of water vapour signal compared to the previous experiment is likely primarily due to factors relating to the experimental set-up, not the SIREs themselves. For instance, SIRE chips were placed directly on the spectrometer aperture for analysis whereas SIRE slides were placed within an old variant of the PLA receptacle that raised the SIRE ~0.6mm from the beam aperture, reducing beam throughput and lowering spectral quality (see chapter 1, section 3.3.2, for more details). This may also have increased the likelihood of water vapour interference by exposing the beam to the environment. The improvement of standard deviation will also be in part due to the increased number of repeat scans per optical area (i.e., 3x spectra per SIRE slide position in Si variance study, 9x spectra per SIRE chip in Si intra-wafer variance study). Small peaks visible just under 3000cm^{-1} is an abnormality attributed to CO_2 interference arising from an increase of atmospheric CO_2 possibly due to exhaled air from the operator obstructing the beam path between background and sample scans [33]. Standard deviations across the entire unprocessed spectra is quite low, but, of course, higher than that of the processed spectra and at 610cm^{-1} the most variance is observed whether or not the spectra have been processed.

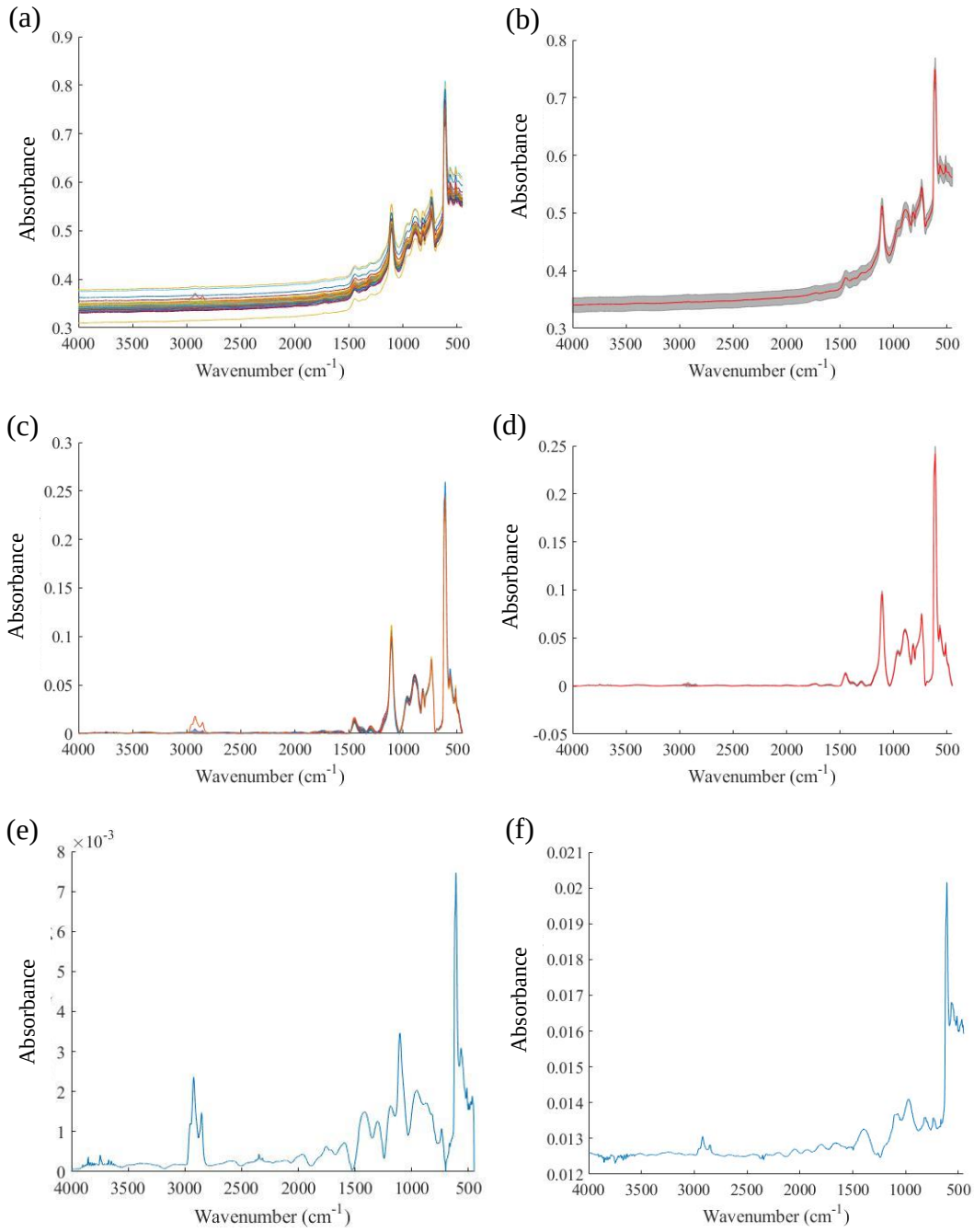


Figure 4.5. *(a) Raw spectra of all SIREs collected from known locations across a single silicon wafer. (b) Average raw spectra with standard deviation across the spectra indicated. (c) all spectra baseline corrected. The interference of atmospheric CO_2 at around 3000cm^{-1} is more noticeable in this interpretation of the spectra. (d) average baseline corrected spectra with standard deviation indicated. The shaded region indicating standard deviation is barely discernible following this simple spectral pre-processing step. (e) and (f) the standard deviation across the entire raw and processed spectra. Standard deviation is vastly reduced after baseline correction, but in each case there are significant variance contributions at wavenumbers 610cm^{-1} and 1107cm^{-1}.*

The baseline offset of silicon spectra presented here in Figure 4.5 and previously in Figure 4.3 may well only be a result of unavoidable atmospheric dynamics between measurements, but to offer an alternative explanation it is worth noting that the attenuation of light across the entire mid-IR range through the silicon lattice is in some part governed by the concentration of impurities within the lattice as described by the equation below [2]:

$$\alpha = \alpha_{im} + \alpha_{ph} + \alpha_{fe} + \alpha_{sc}$$

Equation 4.1. *Constituent parts of the absorbance coefficient.*

Where α = the total absorption coefficient, α_{im} = absorption from impurities, α_{ph} = phonon absorption, α_{fe} = free carrier absorption, α_{sc} = light scattering due to IRE surface roughness.

In the present, study the equation can be reduced to omit the α_{fe} and α_{sc} components because the dopant concentrations used here will lead to negligible values of free carrier absorption coefficient (α_{fe}), and the use of DSP wafers will provide such low levels of surface roughness that the internal IRE scattering term will also be considered negligible (α_{sc}) although scattering from the sample may still occur [2].

$$\alpha = \alpha_{im} + \alpha_{ph}$$

Equation 4.2. *Reduced absorbance coefficient equation.*

The significance of this is that non-homogeneity of impurities across a silicon wafer will not only result in variability at the impurity peaks, Si-O, Si-C and Si-B, but also in the degree to which each individual SIRE transmits light across the mid-IR spectrum as a whole resulting in the disparity in baseline offsets observed in Figures

4.3.(a) and 4.5.(a). Future alterations to the specifications of the silicon material used for the fabrication of SIREs should be mindful of this since changes to the wafers surface roughness, dopant concentrations, or dopant type, for example, could exacerbate SIRE inconsistencies. In the present findings, however, it appears that the impact of this is marginal and can be accounted for easily through simple pre-processing and quality control mechanisms. Figure 4.6 shows histograms of the Si2 and SiO absorbances from all 65 SIRE chips. Anderson-Darling normality tests found that the Si2 absorbances were normally distributed across the wafer ($p = 0.8137$). In contrast, SiO absorbances were not normally distributed due to four outliers with a peak absorbance of 0.065 or 0.068 ($p < 0.0001$). The 0.065 outliers correspond to wafer positions 22 and 65 and the 0.068 outliers correspond to wafer positions 1 and 65. A one-sample student t-test determined that no significant difference existed between the mean Si2 absorbance and a hypothetical mean of 0.0276 ($t(64) = 1.044$, $p = 0.3004$). A separate one-sample student t-test also found that no significant difference existed between the mean SiO absorbance and a hypothetical mean of 0.062 ($t(64) = 0.6365$, $p = 0.5267$). This indicates that both the impurity component and Si lattice component of IR absorbance is reproducible across the Si wafer. However, the locations of the four outliers detected indicates that high oxygen impurity exists at the periphery of the Si wafers. Si interstitials and other defects (e.g., photoresist adhesion failure) are common at the periphery of the wafer and device yield can be low at the wafer edge [34]–[38]. Hence, SIREs fabricated on the periphery of the wafer may be unreliable. The true number of SIRE slides that may be reliably fabricated from a single wafer is therefore estimated to be 52 of a total of 60 giving an estimated manufacturing yield of 86.67%. In practice, a quality control procedure whereby SIREs that possess an

average SiO peak absorbance of <0.065 should be applied to ensure reproducible SIRE slides are manufactured.

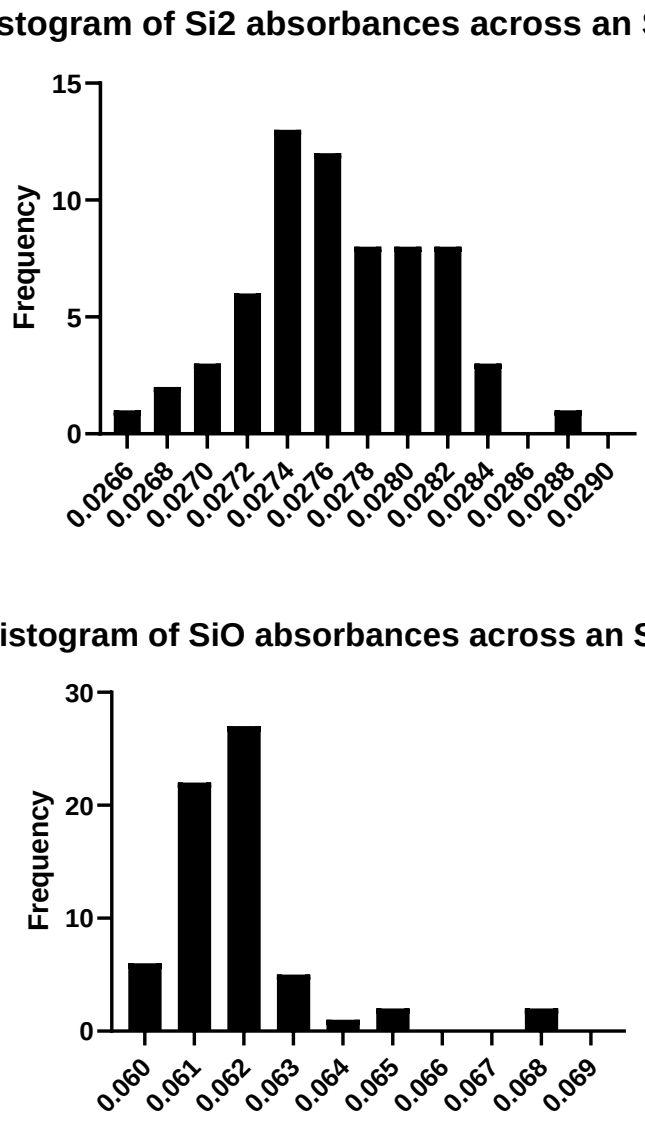


Figure 4.6. *Frequency distribution histograms of Si2 peak absorbances (top) and SiO peak absorbances (bottom).*

4.3.2. Extrinsic Variance Sources and Mitigation

4.3.2.1. Optimised Serum Drying

When interrogating serum using ATR-FTIR samples must be dried onto the IRE before any analysis can be carried out. With respect to sample preparation, this is an extremely simple practice relative to many other bioanalytical techniques that require many careful steps and expensive reagents. The reason serum should be dry is because water is a strong absorber of IR light and to optimise the sample to crystal contact profile (for more information see chapter 2, section 2.2.2) [7]. Water has three vibrational modes, all of which tend to obfuscate concurrent vibrational modes of analyte molecules. Furthermore, in ATR mode water gives rise to anomalous dispersion effects which distort the O-H band and skew quantitative analysis [39]. Serum spectra is no exception to this and there is obviously a huge water content in serum. Figure 4.7 shows the drastic spectral transformation between wet and dry serum after drying in ambient conditions for 6 minutes. Blue spectra indicate wet samples while red spectra indicate dry samples. The information rich 'fingerprint region' is clearly observed in dry serum spectra. In aqueous spectra the O-H stretch vibration will be most prominent, but once dried the amide I and II peaks within the fingerprint region will have the greatest intensity on the spectrum unless special treatments have been applied [40]. These peaks are attributed to molecular vibrations arising from proteins, carbohydrates, nucleic acids, and lipids that together display a unique biochemical pattern that can be used for diagnostic applications [41]. Specific important spectral regions that are obscured by water are between 3000-3500cm^{-1} where the NH and CH bonds of proteins and lipids respectively may have significant spectral contributions, and the absorption bands of the fingerprint region at 800-

1800cm^{-1} which are indicative of a wide variety of molecular bonding environments representative of proteins, nucleic acids, lipids and carbohydrates [41], [42]. With the water removed, it is imperative that a good contact profile of the sample can be made with the IRE crystal as to permit as much interaction with the evanescent wave and the sample contents as possible [43]. An effective drying protocol should be capable of reproducibly producing both physically and chemically homogenous sample films.

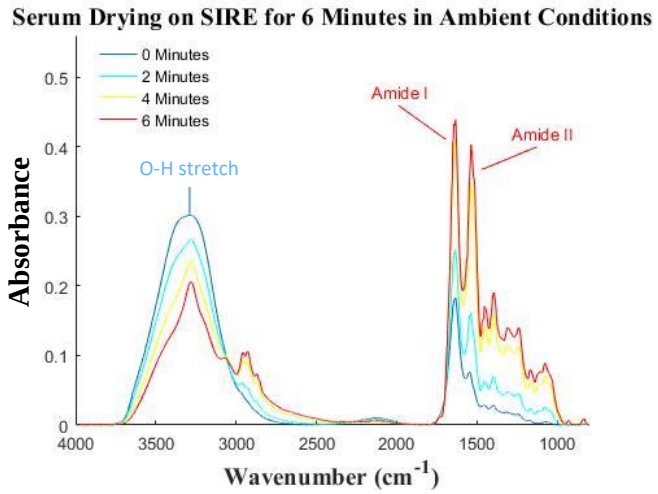


Figure 4.7. *The spectral transformation of HPS as it dries at two-minute intervals over a six-minute period. Blue indicates wet samples, red indicates dry. The O-H stretch curve, Amide I peak, Amide II peak and the fingerprint region have been denoted. Spectra have been acquired on a Di IRE.*

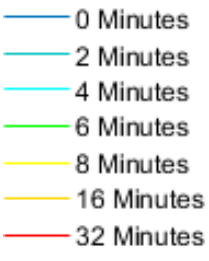
It has previously been shown that serum samples will be completely dry under ambient conditions after 8 minutes [44]. In order to meet clinical standards for sample throughput this drying speed needs to be reduced by as much as possible. To optimise a sample preparation regime, forced drying was executed in two ways: through gentle heating and through blow drying both separately and together. It was believed that moderate application of both strategies would provide the optimal drying conditions.

Figures 4.8.(a) – (f) display the spectra acquired at discrete time intervals for each of the six drying methods. Table 4.2 outlines all the different drying conditions tested on the Di IRE and specifies the time intervals each colour corresponds to. A

legend has been provided to specify at what time interval a spectra was acquired. Blue spectra indicate wet samples, and red spectra indicate dry samples. Spectra shown in Figure 4.8 have been cut and baseline corrected. The difference between wet and dry serum is immediately apparent through visual inspection of the spectra alone (Figure 4.7). The most prominent transformation occurs at 3000-3500cm^{-1} wavenumbers which indicates the O-H stretch vibration diminishing to reveal the NH and CH_3 vibrations of protein and lipid molecules respectively [42]. Overall, it is abundantly clear that any type of forced drying protocol tested here dries specimens quicker than if left to dry in ambient conditions as evidenced by the more rapid reduction in O-H stretch region of all drying methods compared to the control in Figure 4.8. There does not seem to be any substantial difference in drying rate from 5CFM blow dried serum to 9CFM blow dried serum. It is apparent, though, that fan drying alone leads to poor spectral quality evidenced by the large amount of baseline noise visible between approximately 1800cm^{-1}-2500 cm^{-1}. Heating alone does not appear to incur spectral noise and exhibits the largest increase in drying speed particularly at 35°C. There is not enough evidence to suggest that this drying speed is further improved upon by using the heating mats in combination with blow drying at this point.

Table 4.2. *Parameters of all drying methods tested on a Di IRE. The colours which have been used to specify which time interval the spectra in Figure 4.8 represent have also been presented here.*

Drying Regime	Air flow (CFM)	Temperature (°C)
Fan 1	5 ±0.75	25 ±0.5°C
Fan 2	9 ±1.35	25 ±0.5°C
Heater 1	0 ±0.00	30 ±0.5°C
Heater 2	0 ±0.00	35 ±0.5°C
Combined	9 ±1.35	35 ±0.5°C
Control	0 ±0.00	25 ±0.5°C



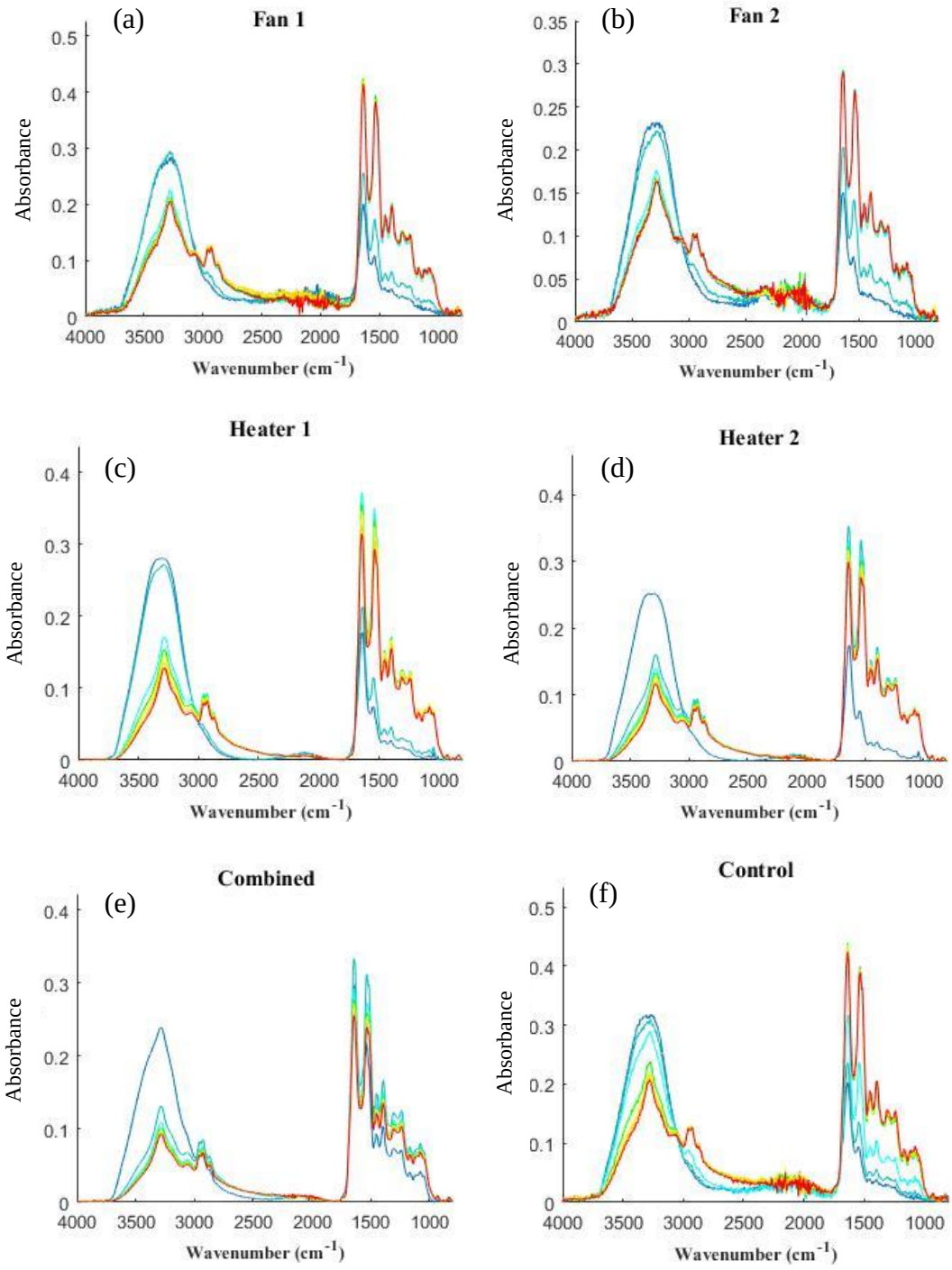


Figure 4.8. *(a)-(f) Transformation of HPS spectra as it dries on a Di IRE over a 32 minute period under different drying conditions. The parameters of each drying condition and the time intervals spectra were collected at have been presented in table 4.2.*

Individual sample preparation with conventional ATR-FTIR upon a fixed IRE crystal is a severely time limiting factor. Batched forced drying is therefore a highly promising solution. It needed to be determined if the drying methods described here could be applied to SIREs and if the drying speed may even be further improved upon. A more powerful fan was utilised for subsequent trials since the previous fan speeds yielded disappointing results. The average area under the O-H curve was calculated at four time intervals of each of the 12 drying regimes tested. These results have been presented in Table 4.3 as the reduction in O-H is a direct indicator of how dry a sample is and is a better visualisation of the differences in drying speeds between drying regimes than copious spectra in this case. Larger O-H curve areas indicate a higher presence of water in the sample. Again, blue values indicate wet samples, while red indicates dry samples. Temperature alone appears to have a much larger impact on drying rate than fan usage, although used in combination the effectiveness of each technique is greatly amplified. Samples can be dried in as little as 2 minutes with the more powerful airflow rate. While there appears to be significant improvements in drying time when increasing fan rates from 9CFM to 60CFM, there does not seem to be substantial benefit in increasing the fan speed to 70CFM. The greatest improvement in drying speed is still due to gentle heating, but contrary to earlier findings the only way to ensure complete drying in two minutes is to use a combination of fans and heating as observed by the miniscule changes in area under the O-H curve following the two-minute mark for droplets subjected to both heating above 30°C and forced air drying above 60CFM. Table 4.4 shows the RSD% of the area under the O-H stretch for each drying regime tested to help establish which method is most reproducible, or to reveal any methods that might inadvertently be introducing variance into the sample.

Here, red has been used to indicate a high RSD% while green has been used to indicate a low, and hence more reproducible, RSD%. A less discernible pattern has emerged making the distinction of the most repeatable method less clear cut. In general, wet samples have poorer reproducibility perhaps due to higher scattering, less control over sample biomolecular distribution and sample thickness, and dispersion effects [45]. RSD% values for tests carried out at 25°C and 9CFM appear unreliable as they are both very low at time zero (0.42) and very high at four minutes (17.66). Other values on Table 4.4 indicate that the use of a fan generally improves sample reproducibility with some of the lowest RSD% values obtained at 70CFM across all heater settings. It is not clear whether heating alone improves reproducibility of the dried films, but heating in combination with fans does seem to improve RSD% values. The droplet displacement caused by airflow from the fan may be responsible for improvements in reproducibility as this would result in greater as well as more consistent sample coverage within the well, forcing samples into uniform thicknesses and disrupting the normal mechanisms by which the ‘coffee-ring’ phenomena takes place [9].

Table 4.3. *Area under the O-H stretch curve at every time interval for every combination of drying conditions tested. Blue indicates wet samples, red indicates dry.*

Area under O-H stretch at different time intervals

		Air Flow (Cubic Feet per Minute)			
Temp (°C)	Time (Min)	0	9	60	70
25	0	28.73	30.65	27.91	29.27
	2	28.29	29.22	16.36	18.51
	4	22.79	19.71	15.98	15.81
	6	19.84	15.16	15.93	15.68
30	0	28.71	26.79	21.88	22.55
	2	16.94	13.95	13.00	12.86
	4	12.88	12.50	12.61	12.44
	6	12.43	12.28	12.65	12.37
35	0	27.86	21.70	19.71	18.11
	2	14.11	11.86	11.87	11.88
	4	12.33	11.16	11.81	11.81
	6	11.14	10.87	11.99	11.89

Table 4.4. *RSD% of area under the O-H stretch curve at every time interval for every combination of drying conditions tested. Red indicates high RSD%, green indicates low RSD%.*

RSD% of O-H stretch at different time intervals

		Air Flow (Cubic Feet per Minute)			
Temp (°C)	Time (Min)	0	9	60	70
25	0	17.93	0.42	8.49	3.76
	2	12.34	3.18	8.80	23.61
	4	32.16	17.66	5.94	4.17
	6	32.56	3.30	5.40	4.40
30	0	16.51	10.71	20.80	5.54
	2	32.23	21.79	5.62	7.31
	4	10.25	6.88	5.47	8.36
	6	7.32	6.60	6.01	8.97
35	0	6.93	21.01	10.55	11.93
	2	26.01	4.47	9.69	5.89
	4	11.19	5.38	9.65	5.76
	6	4.76	6.62	9.26	5.97

Figure 4.9 shows the drying rate of spectra under a fan outputting 60CFM. This is to demonstrate the exceptional increase in drying speed this airflow has produced in comparison to that previously demonstrated in Figure 4.8 at 9CFM. More importantly,

it shows that spectral quality was not visibly compromised through this approach. This is most likely due to greater stability of the experimental setup and more direct application of the airflow onto the specimens hence enhanced sample spreading profile. The spectra reach a steady state by the first two minutes as demonstrated by the overlapping spectra (2, 4, & 6 minute interval spectra are overlapping) in Figure 4.9 suggesting there is no more water to remove from the sample after this time. This is substantially faster than any of the drying methods previously tested. The spectra shown represent typical spectral transformation following serum dehydration [11], [40], [44].

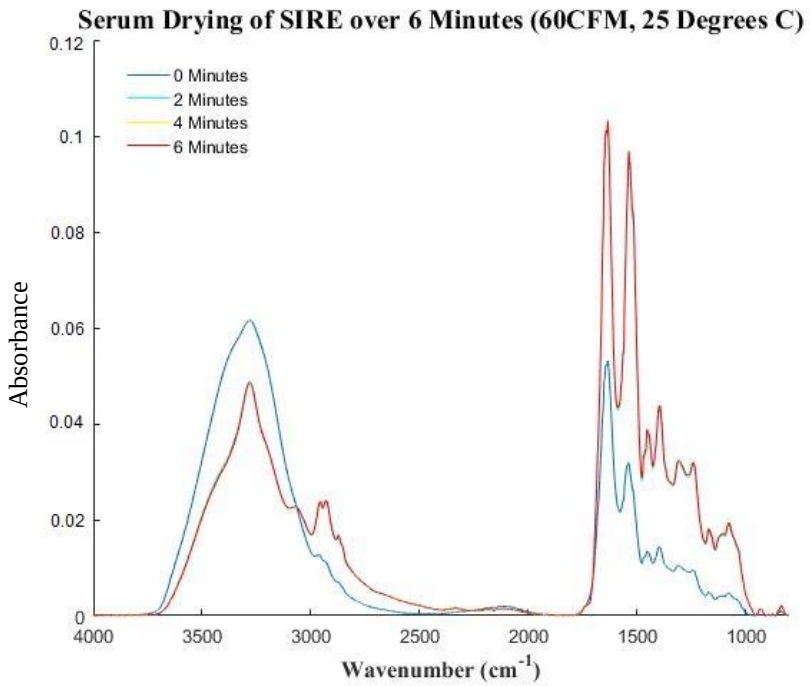


Figure 4.9. *HPS dried on a SIRE using a powerful 60CFM fan output. Spectra reaches a steady state after two minutes indicating the sample has been completely dried within this time. Note: spectra at 2, 4, and 6 minute time intervals are overlapping and are therefore not clearly visible.*

An incubator was later used as an environment that could uniformly heat numerous SIREs simultaneously. The effectiveness of this new approach was tested

by placing SIREs into the incubator set to 30°C and under a fan outputting 60CFM airflow for 0, 2, 4, 6, 8, 16, and 32 minutes. Figure 4.10 shows the spectral drying profile of 3µl of serum upon a SIRE under these conditions. As before, samples can be dried rapidly in at least two minutes. However, it will be challenging to ensure air flow is uniform across several SIREs drying simultaneously and so this approach may not be appropriate for batched drying. Table 4.5 shows the area under the O-H curve and the RSD% of the O-H curve at each time interval. Area under the O-H curve decreases rapidly after two minutes and remains relatively stable after this period. RSD% also improves dramatically after 2 minutes and remains low up until 32 minutes have elapsed.

Table 4.5. *Area under O-H stretch curve and RSD of O-H stretch curve of HPS spectra drying on a SIRE at various time intervals over a 32 minute period. Serum was dried within an incubator and under a fan outputting 60CFM airflow and at 30°.*

Time (Min)	0	2	4	6	7	16	32
Area under O-H stretch	12.93	7.39	7.22	7.06	7.10	6.90	6.96
RSD%	39.33	2.65	1.80	3.68	1.81	3.58	1.32

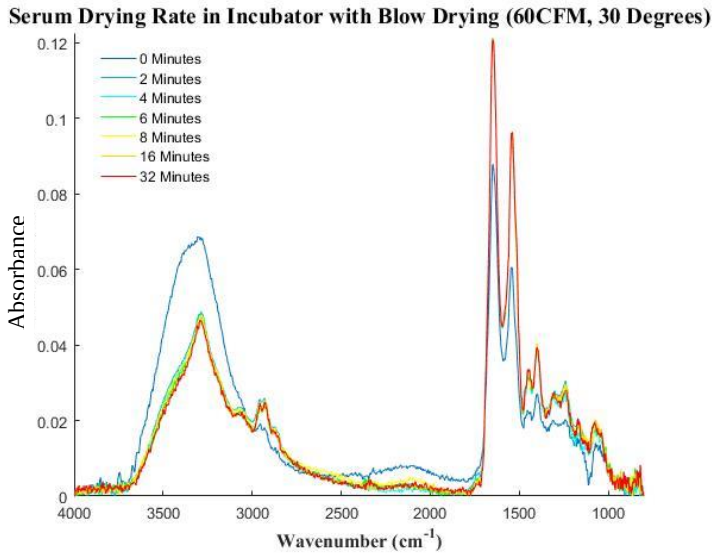


Figure 4.10. *Spectral drying profile of HPS in an incubator set to 30°C and under a fan outputting 60CFM airflow at various time intervals over a 32 minute period.*

Most drying strategies described here can rapidly remove water content from serum samples deposited upon SIRE sample sites with good reproducibility. In particular, moderate heating (30°C or 35°C) is capable of reducing drying wait times to at least 2 minutes, and fan assisted drying (60CFM or 70CFM airflow) is effective in maintaining and perhaps improving spectral quality of dried serum films. It is likely that these drying strategies can adequately prepare serum samples upon SIREs in less than two minutes, but to determine if greater drying speeds can be achieved tests must be performed at shorter time intervals. While still possessing good RSD% values, incubator dried serum samples exhibit a marked increase in baseline noise in comparison to previously trialled drying methods. The difference in relative humidity between the inside of the incubator (*ca.* 10%) and the ambient environment (*ca.* 30%) is likely the cause of this. Bou Zeid describes the differences in morphology of blood films which have dried under various relative humidity conditions [46]. Fracture patterns more prevalent in films dried in lower humidity may increase scattering prevalence in FTIR analysis thus increasing spectral noise. Thus, humidity is an important parameter that must be kept stable across batches of drying serum samples. A study has shown that Constant Temperature Humidity Chambers (CTHC) set to 50% humidity and 75°C can be used to homogeneously distribute droplet precipitates with rapid evaporation times, although optimal humidity conditions for serum would have to be independently investigated [47]. Numerous studies describe the impact drying patterns can have on the resultant spectra of samples studied using FTIR [8], [9], [48]. Non-uniformity in either sample thickness or biomolecular distribution are primary causes of poor spectral quality. Sustained shear forces from air flow generated by the fans will help smooth surface irregularities and maintain a uniform sample thickness.

Heating of droplets is also known to promote uniform precipitation of molecular constituents following evaporation [47]. The benefits both heat and air flow drying mechanisms have on spectral quality is evident in the findings presented here signifying that a combination of both fans and heaters can vastly reduce sample preparation time of serum for spectral analysis, but also reduce extraneous sample variance. Ultimately, this demonstrates the viability of batched sample preparation using SIREs for high-throughput ATR FTIR.

There are a multitude of alternative methods and techniques that may further improve drying rate and reduce sample variance. Serum dilutions with physiological water prior to sample drying may help reduce crack formation, and consequently light scattering, as shown by Lovergne *et al*. although an unfortunate repercussion of this is potentially exacerbating the coffee-ring effect [8]. Automated piezo-jetting is able to circumvent the coffee-ring effect by depositing several hundred sub-nanolitre droplets immediately adjacent to each other allowing droplet peripheries to merge and combine precipitate rings [45], [49]. A combination of serum dilution and piezo-jetting may present an attractive option for further improving spectral reproducibility of serum films for vibrational spectroscopy. Freeze-drying (Lyophilisation) removes water via sublimation in a reduced pressure environment to produce glassy films of serum [50]. The primary advantage of this is that the evaporated product can be stored in ambient conditions over much longer periods of time in comparison to conventionally dried serum [51]. This would greatly complement the SIRE slide as specific patient samples could be stored for reanalysis at a later date.

4.3.2.2. Serum Variance

To further explore how capable SIREs are in the analysis of complex media, a series of analysis were performed to examine how reliably HPS could be interrogated. 3µl of HPS taken from the same source was pipetted into 14 randomly selected SIREs (Group 1), dried, then scanned. Figure 4.11 shows a PCA plot of first two PCs of the data. No clear separation can be seen in these PCs, but some clustering of SIREs is visible. The first 13 PCs explain over 90% of the total variance in the dataset. Each PC displays scores distributed in a similar way to PC1 and PC2 showing no clear separation. There is an exception in PC 3 which indicates HPS spectra acquired from SIRE number 13 is distinct from the rest of the dataset (red circle Figure 4.11.(b)). PC3 loadings indicates variance in this PC is dominated by the 1900-2300cm^{-1} wavenumber region (Figure 4.11.(c)). This is confirmed through inspection of the HPS spectra acquired from each sampling position of SIRE number 13 which shows discrepancies in this wavenumber region (red circle, Figure 4.11.(d)). This is unusual because the 1900-2400cm^{-1} wavenumber region is normally a low signal baseline region of the HPS spectrum and is not useful for serum analysis. This signal is indicative of light scattering on sampling position 1 of SIRE number 13 likely due to sample preparation error rather than originating from the SIRE itself. A non-homogenous or broken sample film could lead to scattering artefacts resulting from a difference in optical density across the IRE at the sample to air interface like the artefact observed earlier in section 4.3.1.1 [11], [25]. This highlights the need for controlled sample preparation protocols. In future work, it may be possible to use this spectral region to construct a quality control algorithm and automate the removal of SIRE wells exhibiting scattering at the sample to air interface. A one-sample t-test determined that there was no

significant difference between the mean PC1 scores and a hypothetical mean of zero ($t(13) = 0.9962$, $p = 0.3373$). This together with the clustering of samples in the PCA scatterplot indicates that identical HPS samples can be analysed reliably using SIREs as the variation between slides is low provided sample preparation is performed correctly. Moreover, the low percentages of the loading directions suggest there is a low level of variability in the dataset.

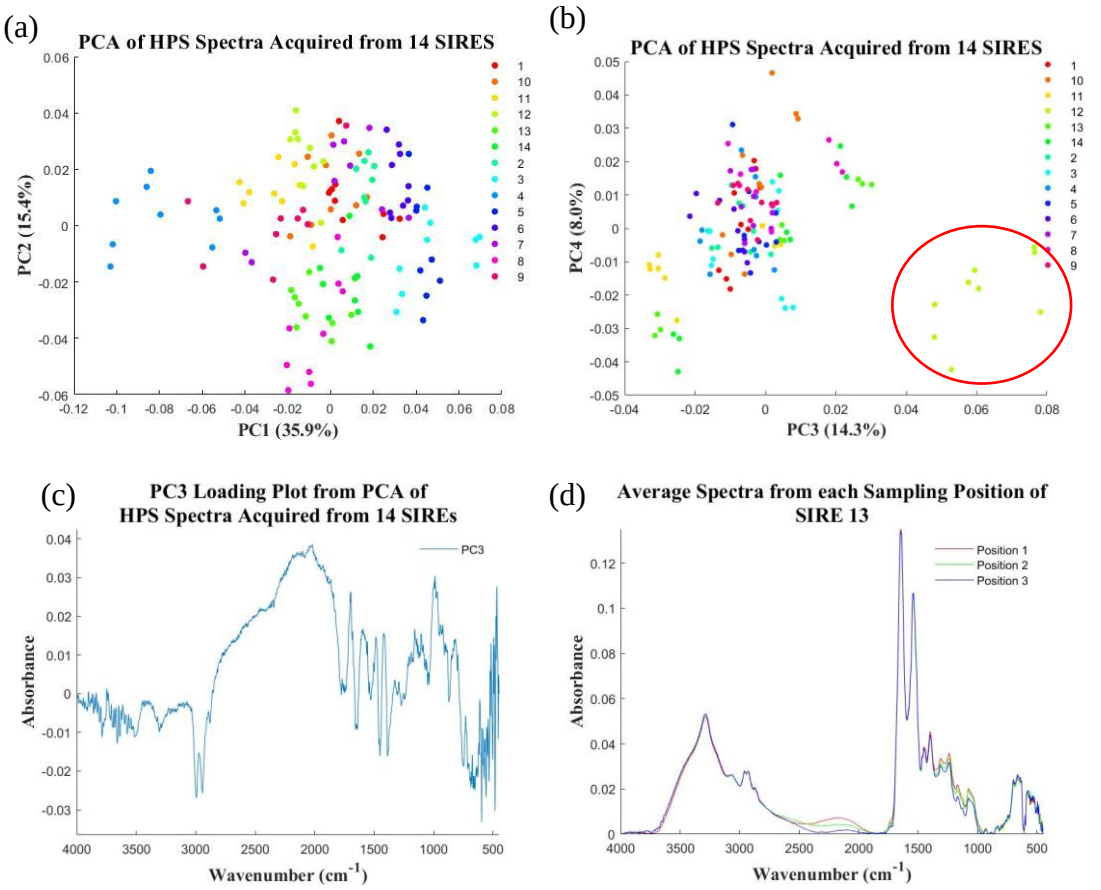


Figure 4.11. *(a) PC1 and PC2 scores plot of HPS spectra acquired from 14 randomly selected SIREs. Clustering of datapoints and low loading percentages indicates a high level of agreement between the samples. (b) PC3 and PC4 scores plot from same PCA revealing one outlier SIRE (number 13) as indicated by the red circle. (c) PC3 loading plot showing baseline wavenumbers (1900-2300cm^{-1}) to be the biggest contributors to separation in PC3. (d) Evidence of scattering in the baseline of SIRE 13 HPS spectra (1900-2300cm^{-1}) likely due to sample preparation error.*

Figure 4.12 shows the average spectrum with standard deviation of all serum spectra taken using SIREs on two consecutive days as well as the corresponding PCA plot of the first two PCs. There is no obvious difference in groups and percent explained variance is low. Spectra have been offset for comparison. Mean standard deviation is slightly higher in the day two group than the day 1 group (0.0014 and 0.0015 respectively). The first 14 PCs explained over 90% of the variance in the dataset with each PC displaying PC scores distributed in a similar way to PC1 and PC2 showing no clear separation or clustering of groups. PCA plots show a great deal of agreement between spectra acquired on each day indicating the serum has remained stable overnight and spectra are not significantly altered by differences in ambient conditions from one day to the next. However, a paired samples t-test confirmed a significant difference ($t(13) = 4.831$, $p < 0.05$) existed between the mean PC1 direction scores of day 1 ($M = -0.002028$, $SD = 0.005084$), and day 2 ($M = 0.00208$, $SD = 0.07396$). Standard deviation of group scores in PC1 are shown to be overlapping but means are not equivalent. These findings indicate that SIREs may not be suitable as a means to preserve dried serum samples for reanalysis at a later date. However, longer study periods would be needed to confirm this.

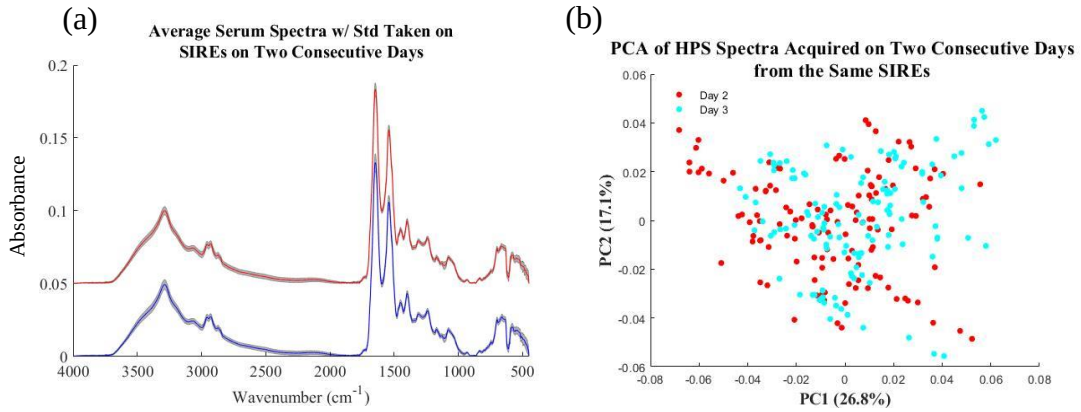


Figure 4.12. *Analysis of dry HPS using SIREs on two consecutive days. (a) dried HPS spectra acquired using the same samples on the same SIREs on Day 1 (red) and Day 2 (blue). The standard deviations have been indicated by the shaded region. (b) PCA plot of the spectra acquired on each day shows very little separation.*

The next day a second group of randomly selected SIREs (Group 2) were then used to analyse serum in exactly the same way and compared against Group 1. Again, spectra have been offset for comparison (Figure 4.13.(a)). A PCA scores plot of the first two PCs show differences, but no clear separation of the groups. The first 16 PCs explain over 90% of the variance in the data, each PC showing no clear separation of PC scores or clustering of groups. Again, the low PC loading percentages indicates a low amount of variance present in the dataset. Both groups are reproducible with mean standard deviation of Group 1 and Group 2 at 0.0014 and 0.0018 respectively. An unpaired two-sample t-test shows a significant difference ($t(26) = -3.016$, $p = 0.0057$) existed between the mean PC scores of Group 1 ($M = -0.00387$, $SD = 0.007722$), and Group 2 ($M = 0.00387$, $SD = 0.005707$). The standard deviation of the group scores in PC1 indicate groups are overlapping but means are not equivalent. This result suggests measurements taken on separate days may impact reproducibility although again longer study periods need to be undertaken to clarify this.

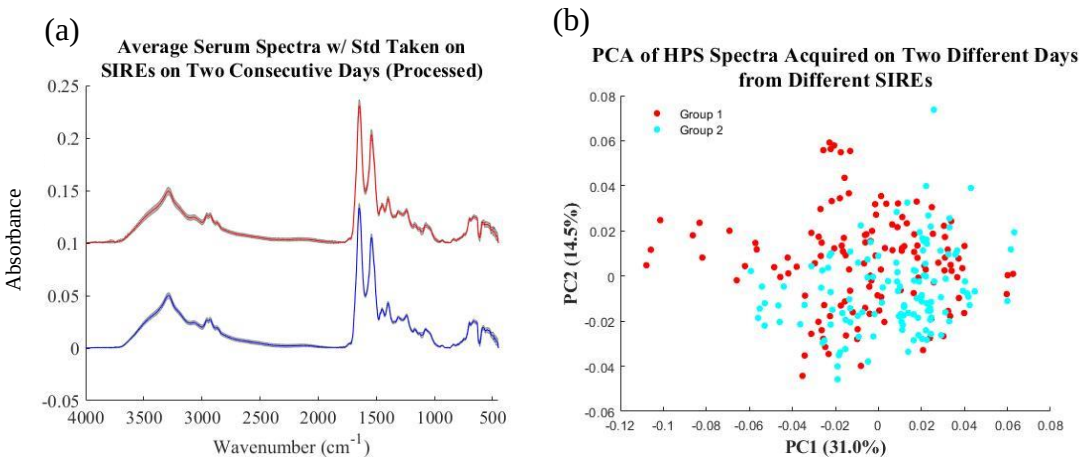


Figure 4.13. *Analysis of dried HPS on two different groups of 14 SIREs. (a) dried HPS spectra acquired using SIREs from Group 1 (red) and Group 2 (blue). Spectra have been baseline corrected and averaged. And the standard deviations have been indicated by the shaded region. (b) PCA plot showing no clear separation in data between Groups 1 and 2.*

To demonstrate how important pipetting consistency is in introducing variance, spectra were acquired of serum deposited into SIRE sample compartments in two different ways. 3µl Serum was either pipetted into the centre of the sample compartment and spread to cover the entire compartment area or pipetted into the centre of the sample compartments and left as a spot were the droplet was deposited (Figure 4.14). Only marginal differences in variance were observed with mean standard deviations of 0.0014 for spot serum and 0.0016 for spread serum. Figure 4.14 shows there is a significant increase in absorbance intensity across the entire spectrum when samples have been spread over the entire compartment. Spreading ensures the entire beam spot above the specular reflectance puck aperture is interacting with sample instead of open air and thus maximising the IR beam throughput that is being utilised to interrogate the sample. Spreading is also a far more reliable method to ensure consistent drying profiles across SIREs as it reduces the chance of random droplet shape and size due to subtle differences in surface chemistries between SIREs which in turn affects droplet formation and contact angles [52]. To further improve

this aspect of using SIREs automated pipetting systems would need to be used as this variance is largely owed to operator error.

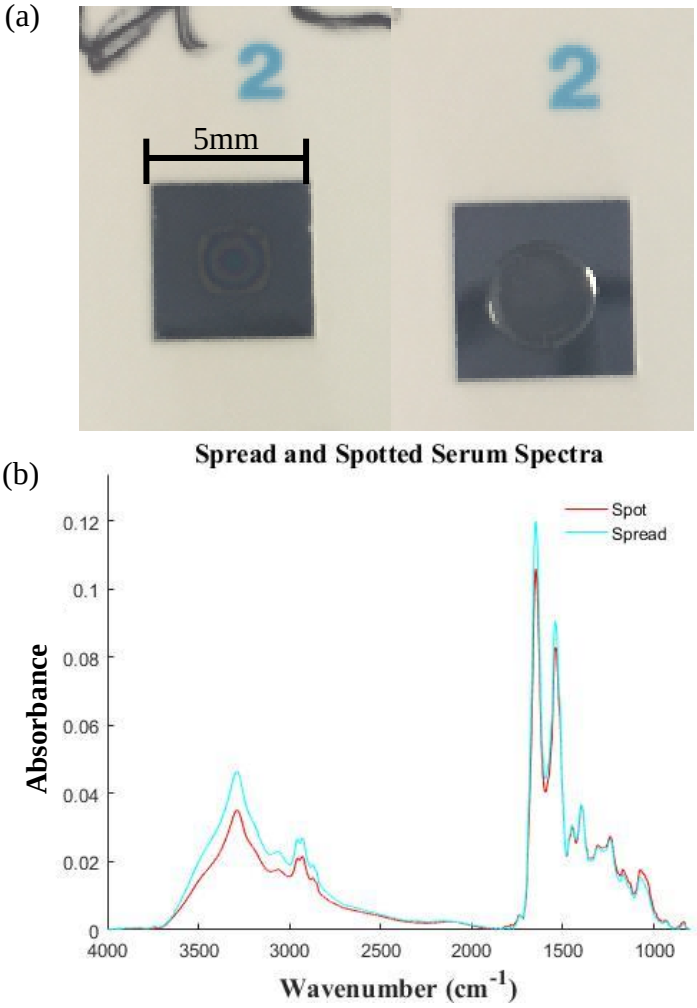


Figure 4.14. *Differences in the analysis of spot and spread HPS spectra. (a) 3µl of serum have been pipetted onto a SIRE and either spread around to fit the confinements of the sample area (left) or spotted into the centre of the sample area (right). (b) the resultant spectra of the two different pipetting methods show a substantial decrease in spectral intensity when serum has not been spread around the sample compartment.*

4.3.2.3. SIRE Surface Treatments

Methods to clean SIREs that had been used for serum analysis were investigated. SIREs were treated with either Piranha, acetone, or Virkon (control) solution over a course of five days and the signal quality of the devices monitored. The

SNR of all SIREs on each day was calculated to see if changes in spectral quality could be observed. All results for each treatment have been presented as a bar chart in Figure 4.15.

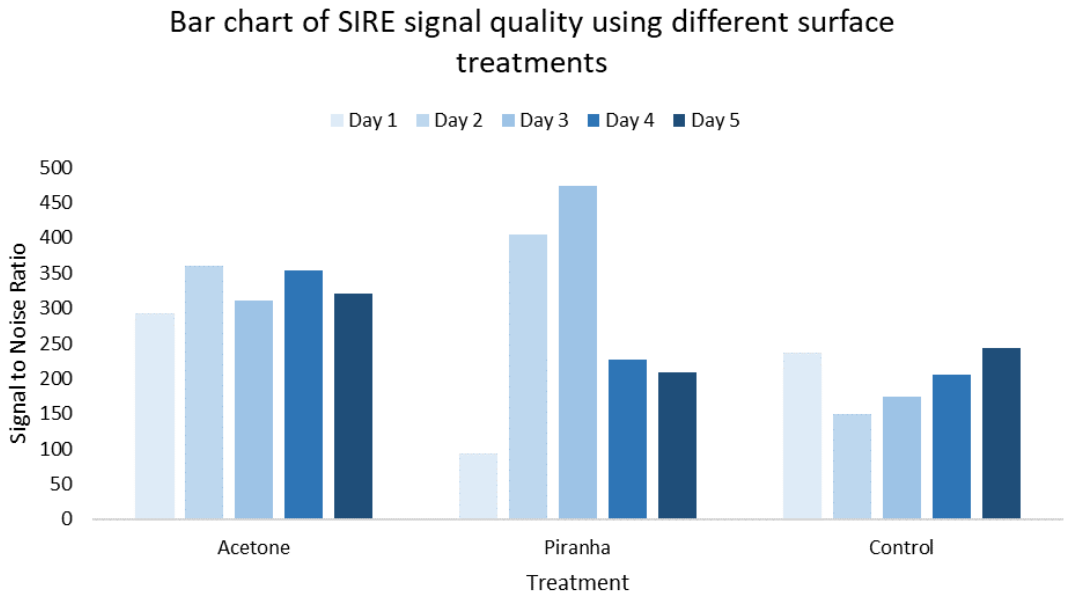


Figure 4.15. *Bar chart of spectral quality of SIREs over five days following different cleaning regimes.*

SIREs treated with acetone were observed to have the most reproducible SNR over the five days. The major difficulty in using acetone for the cleaning of SIREs is the accidental deposition of residue upon the silicon which follows the rapid evaporation of acetone [13]. Acetone coated SIREs must be rapidly transferred to 2-propanol then DI water to prevent contaminant residues which would otherwise be visible during reanalysis. Closer inspection of the spectra revealed that the control group suffered from artefacts as a result of residual sample present on the SIREs between cleaning. Virkon tends to create clumps of serum debris which could easily lodge into the v-grooves of SIREs undetected creating spectral artefacts such as the baseline deviation seen in Figure 4.16.

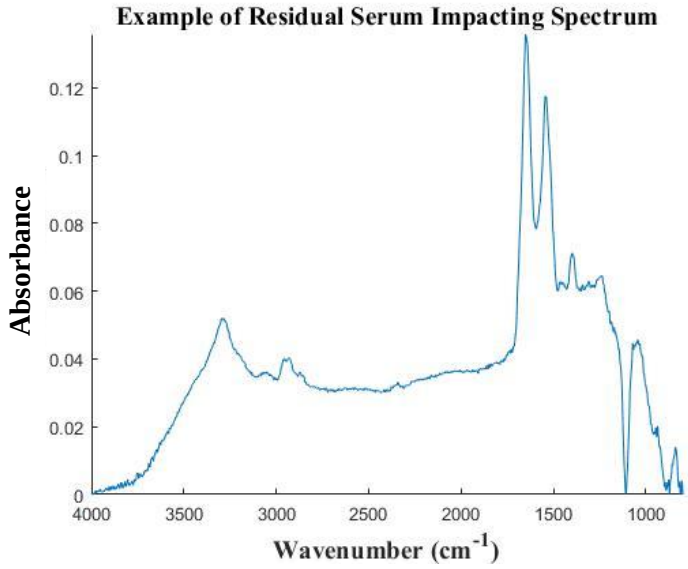


Figure 4.16. *HPS spectra acquired from a SIRE that has been improperly cleaned using Virkon. HPS from a previous analysis has not been successfully removed from this SIRE and has skewed the baseline of the spectra as a result.*

Days 2 and 3 following piranha treatment saw a marked improvement in SNR after which time SIREs began to break down due to the aggressiveness of the chemical and SNRs were difficult to maintain. Piranha is effective in removing organic contaminants which would otherwise occupy the silanol (Si-OH) groups on inorganic silicon surfaces. The amount of silanol groups and extent of water adsorption on the silicon surface is instrumental in determining the wettability characteristics of silicon. More silanol group availability and surface water adsorption leads to an increase in hydrophilicity of the surface [53]–[55]. Silicon oxide surfaces tend to have numerous hydroxyl terminated groups and therefore are more hydrophilic than pure silicon surfaces which tend to have hydrogen terminated groups. Over time silicon will grow native oxide layers in ambient conditions, but, as silanol groups are quite reactive, will also adsorb contaminants which may reduce hydrophilicity [53]. Treatments that either contaminate the native silicon oxide layer, such as acetone [56], or remove it entirely, such as hydrofluoric acid [55], should be avoided. Both acetone and hydrofluoric acid

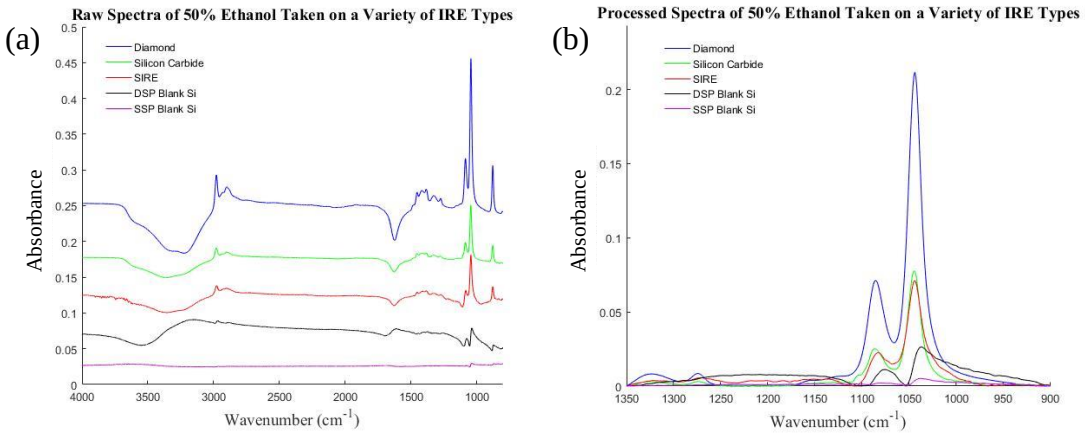
are used in the later stages of the SIRE fabrication process which implies SIREs do not possess optimised surface chemistries immediately following fabrication. Adding a final stage in the fabrication process that mimics well established treatments that are used to prepare soda lime glass for silanization (e.g., Piranha etching) [57] is a good option as this will minimise organic contaminants and maximise surface silanol group homogeneity hence improving sample spreading uniformity and SIRE quality. However, the surface chemistry may not be stable which may present issues after long storage periods. Plasma activation should therefore be explored as a more reliable alternative to bestow a long-term hydrophilic surface chemistry [58], [59].

4.3.3. SIRE Analytical Characteristics

4.3.3.1. SIRE Versus IRE Alternatives

To help gauge the efficacy of SIREs as FTIR optical components, SIREs have been compared to a Di IRE as well as other commercially available silicon FTIR utensils such as a fixed SiC IRE, a blank Single-Side Polished (SSP) un-etched piece of silicon, and blank Double-Side Polished (DSP) un-etched piece of silicon. For the sake of brevity all of these will collectively be referred to as IREs although strictly speaking the last two listed here are not IREs. Ethanol/water mixtures were used in each instance as this is a simple sample type and presents the best option for a fair comparison. A stock solution of 50% v/v ethanol in water was prepared and analysed on all IRE types. In each case, a background scan of water was subtracted from the acquired spectra. Figure 4.17 shows spectra taken using each of the different IRE types. Baselines have been offset for visibility in Figure 4.17 (a). In this case, subtracting a water background results in negative O-H peaks so to avoid errors the

data has been cut between 900-1350cm^{-1} before baseline correction (Figure 4.17 (b)). SNRs were calculated using the maximum absorbance at the C-O peak (1045cm^{-1}) to represent the signal portion of the SNR equation and the standard deviation between 1150-1200cm^{-1} was selected to represent noise. SNR and C-O peak absorbance have been reported in Table 4.6.



IRE Type	Diamond	Si C	SIRE	DSP	SSP
C-O Peak Absorbance	0.212	0.078	0.071	0.027	0.005
SNR	141.018	129.860	163.318	110.391	36.359

Figure 4.17. *Spectra of 50% v/v ethanol taken on a variety of IREs. (a) raw spectra with offset baselines for visibility. (b) processed spectra that has been cut then baseline corrected.*

Table 4.6. *C-O peak intensity and SNR values achieved using each IRE.*

As expected, the DSP silicon and the SSP silicon have very poor signal intensities and SNR, but never-the-less can communicate signal. This demonstrates two things: the importance of optical coupling grooves for SIRE functionality, and the impact of surface roughness on the attenuation of the IR beam. The near complete extinction of signal from SSP blank silicon chips is exemplary of the effects of the scattering component in Equation 4.1 in section 4.3.1.2 mentioned previously in this chapter. Clearly, the greatest signal intensity is achieved when using a Di IRE with almost three-fold the absorbance value of C-O peak compared to the SiC IRE, the next

highest absorbance from the IREs tested. It was shown previously (chapter 3, section 3.3.2) that beam throughput of SIREs and the Di IRE is approximately equivalent so it is not expected that differences in critical angle will contribute to lower C-O absorbance peaks in Si IREs. It is also not likely that differences in material transmittance of IR light is responsible as absorbance peaks appear to be greater across the Di spectrum and not just localised to 1000-1100cm^{-1}. Thus, the larger penetration depth of Di compared to Si is likely responsible for greater absorbance signal owed to Di's lower refractive index [60]. SIREs and Di both show good signal quality with high SNR values, while the SiC IRE has a notably lower SNR. Figure 4.18 shows the lattice spectra of Di, SiC, and SIRE. The Si1/Si2 ratio at 610cm^{-1} and 1108cm^{-1} for SiC and SIRE spectra are 2.14 and 1.6 respectively. This observation indicates the higher carbon impurity concentration in SiC compared to silicon since Si-C bond vibration (605cm^{-1}) overlaps the Si1 phonon band (610cm^{-1}) [2]. Stronger vibrations at the Si2 phonon band in SiC compared to SIRE is due to a longer beam pathlength. In general, analysis of vibrational modes below 1500cm^{-1} will be more difficult on SiC IREs compared to SIREs as evidenced by the lower SNR of SiC in Table 4.6. The spectrum of Di lattice is entirely different to silicon based IREs (Figure 4.18) indicating SIREs may hold some analytical advantage in high wavenumber regions, while being at a slight disadvantage in some low wavenumber (fingerprint) regions for the analysis of biological materials. In particular, SIREs may be more sensitive to lipid vibrational bands present between 3000-2700cm^{-1} while being less sensitive to carbohydrate vibrations at about 1100cm^{-1} [41].

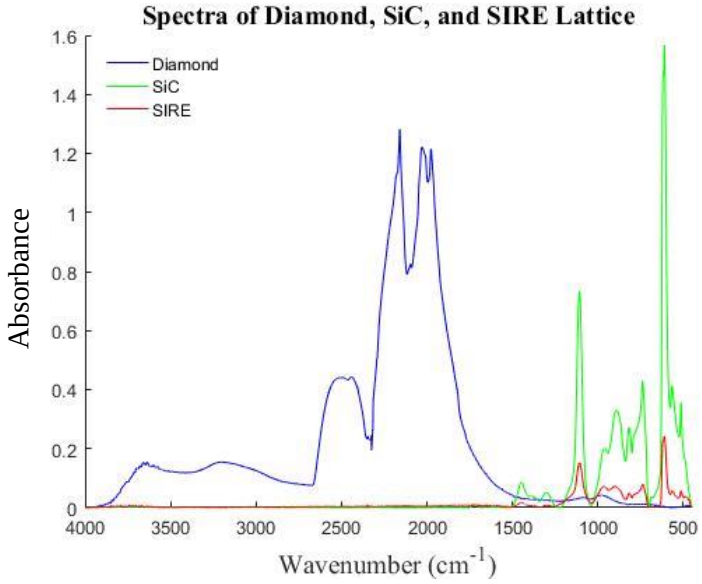


Figure 4.18. *Spectra of the crystal lattice of Di IRE, SiC IRE and SIRE. These spectra in addition to environmental signal from the air is the waveform that is subtracted from the sample scan by the background scan under normal analysis using these IREs. All spectra have been baseline corrected.*

4.3.3.2. Ethanol Concentration Study

50% v/v ethanol/water stock solution was then diluted to concentrations of 25, 12.5, 6.25, 3.125, 2, 1, 0.5, 0.25, and 0% v/v ethanol in water. 3µl of each concentration was analysed on SIREs and on a Di IRE. Nine spectra were acquired of each instance (three sample triplicates and three instrument triplicates) for a total of 180 spectra. The most notable difference in the raw data is the extent of baseline offset of samples analysed on SIREs compared to samples analysed on a Di IRE (Figure 4.19.(a) and (b)). The reason for this, as noted previously, is due to the different silicon used for background scans than for sample scans. Also notable from visual inspection of the raw data is the larger signal intensity and higher quality baseline of spectra from the Di IRE. For quantitative analysis, the C-O peak at 1045cm^{-1} was isolated and the spectra baseline corrected (Figure 4.19.(c) and (d)). The area under the curve at each

concentration was then calculated and plotted against the known concentration (Figure 4.19.(e) and (f)). A linear regression model was fitted to the data and statistical outputs have been reported in Table 4.7. R^2 values are comparable, but high (0.9981 and 0.9970 for SIRE and Di respectively) in both cases while the RMSE value is much more favourable for the linear regression model fitted to the SIRE data (0.0115 and 0.0542 %v/v ethanol for SIRE and Di respectively). This suggests both have a strong correlation of C-O peak area to ethanol concentration, but that a more accurate prediction of ethanol concentration may be made from C-O peaks calculated from SIREs. LOD and LOQ were estimated from the linear regression model and have also been reported in Table 4.7:

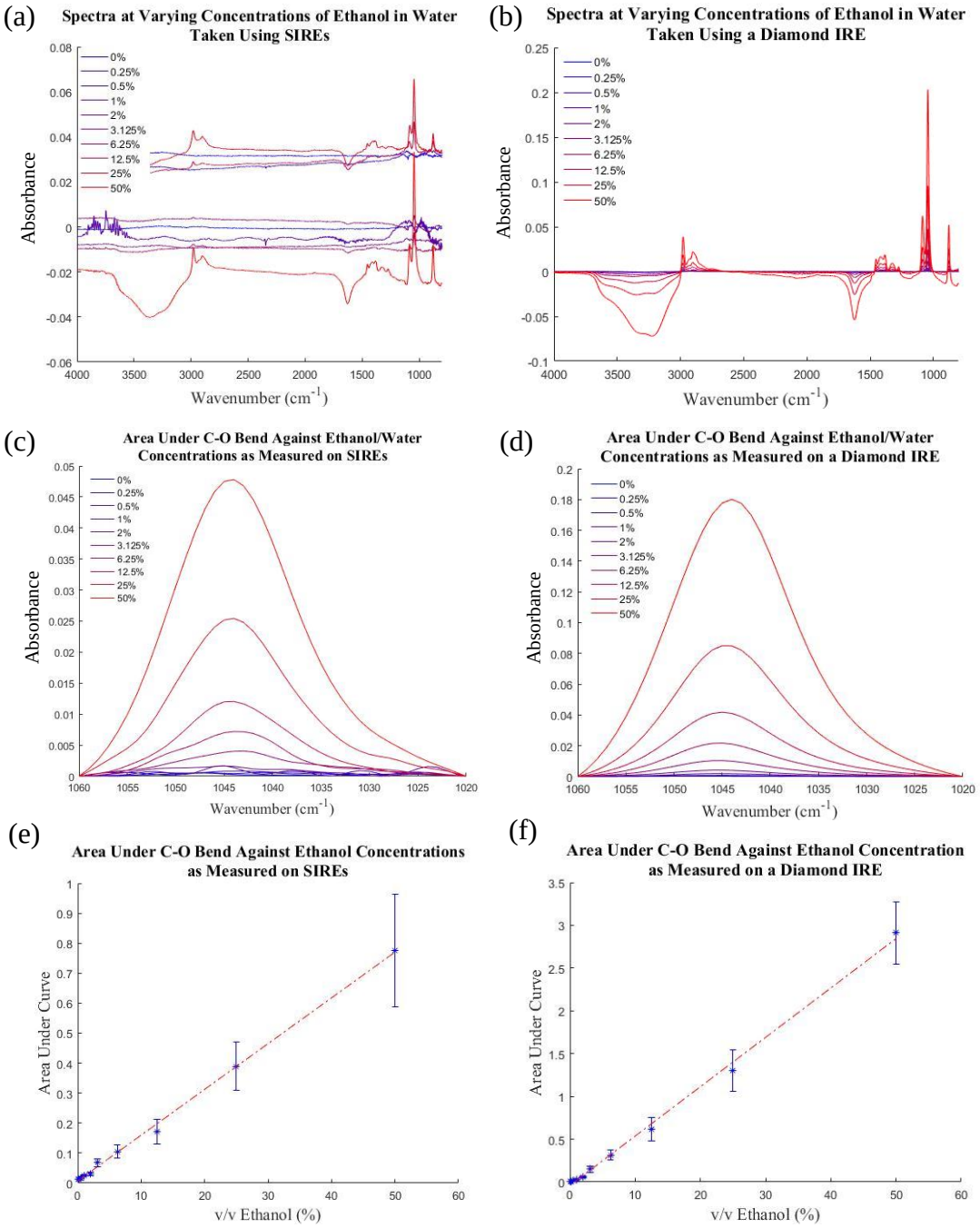


Figure 4.19. *Analysis of a range of ethanol concentrations on SIREs and on a Di IRE. (a) and (b) Raw spectra of ethanol concentrations. Baselines are very consistent on the Di IRE, but quite erratic on SIREs. (c) and (d) Processed spectra of ethanol concentrations on SIREs and on a Di IRE. Spectra have been cut around the C-O peak and baseline corrected. Maximum absorbance is almost 4-fold higher on the Di IRE than on SIREs, but there is no difference in the concentration at which peak the C-O peak is no longer discernible from baseline which occurs at about 3.125% ethanol. (e) and (f) Linear calibration curve where area under the C-O peak has been plotted against ethanol concentrations. Error bars indicate one standard deviation of results. Statistical outputs have been reported in Table 4.7.*

Table 4.7. *Statistical outputs calculated from the linear regression model fitted to the calibration curves presented in Figure 4.19.*

	R^2	RMSE (%Ethanol)	LOD (% Ethanol)	LOQ (% Ethanol)
Diamond	0.9970	0.0542	3.3503	11.1677
SIRE	0.9981	0.0115	2.6791	8.9305

The LOD is defined as the lowest analyte concentration that can be reliably distinguished from background noise whereas LOQ is defined as the lowest analyte concentration that can be both reliably detected and quantified [61]. Both the LOD and LOQ were lower in ethanol concentration studies carried out on SIREs than on Di IREs. This finding is consistent with the higher SNR value achieved by SIREs presented earlier in Table 4.6. These findings support the claim that SIREs are operating at an analytically competitive level to industry gold standard Di IREs for the analysis of simple mixtures. As for the analysis of more complex media (i.e., blood serum) there are some interesting differences that emerge and may illuminate key analytical advantages and disadvantages of SIREs over Di IREs.

Figure 4.20.(a) shows typical HPS spectra acquired from both a Di IRE and SIRE. Clearly, Di has a superior spectral intensity compared to SIREs due to its lower refractive index which allows a larger penetration depth of the evanescent wave into the sample [62]. However, the refractive index of the sample may be dispersed or vary slightly across a sample set and wavelength obviously varies across the IR spectrum. This may give rise to band distortions due to anomalous dispersion where spectral bands measured in transmission mode will deviate from spectral bands measured in ATR to varying degrees depending on both the sample and IRE refractive indices and angle of incidence used [63]. The closer the refractive index of IRE is to the refractive index of the sample, the more the spectral bands will be distorted [60]. Since Di has a

refractive index (n_{Di} = ~2.4) closer to serum (n_{serum} = ~1.34) than Si (n_{Si} = ~3.4), it can be expected that Di serum ATR-FTIR spectra will be more distorted than corresponding Si spectra [60], [64]. Vector normalising the data allows for closer inspection of the subtler differences the different reflection element types have on HPS spectra (Figures 4.20.(b), (c) and (d)). There is a great disparity in the ratio between the Amide I and Amide II peaks (indicated by AI:AII in Figure 4.20.(b)) of the two reflection elements. AI:AII ratios have been calculated as 0.7676 and 0.9362 for HPS analysed on SIRE and Di IREs, respectively. Figure 4.20.(c) shows an enlarged view of the Amide II peak to illustrate a band shift, Δw_n, (~8cm^{-1} difference) observed between HPS spectra acquired on a Di IRE and on a SIRE. Both discrepancies are a result of anomalous dispersion of the refractive index and may lead to errors during analysis. For example, if attempting to analyse the protein orientation or secondary structures peak position and AI:AII ratio is important [65]–[67]. However, in transmission mode the amide I peak occurs at 1656cm^{-1} [66], but it is shifted to 1648cm^{-1} when interrogated in ATR mode using a SIRE and again to 1640cm^{-1} when using a Di IRE and AI:AII ratio is also altered as noted earlier (Figure 4.20.(b)). It is easy to see, then, how protein analysis may become complicated as bands are further distorted from their true position and morphology. As shown from the crystal lattice spectra earlier (Figure 4.18), Di absorbs light at wavenumbers above 1600cm^{-1} where the Si lattice does not absorb which will in part contribute to the difference in AI:AII or the IREs. Furthermore, the lower refractive index of Di will cause greater frequency shifts and band distortions due to anomalous dispersion of the refractive index near absorption maxima in comparison to SIRE spectra [65], [66], [68], [69]. Penetration depth, and by extension absorption intensity, is dependent on the ratio of sample and

IRE refractive indices (Equation 2.13, chapter 2). Thus, there is greater distortion of the amide I and II peak using a Di IRE than when using a SIRE as evidenced by the difference in AI:AII ratio [66].

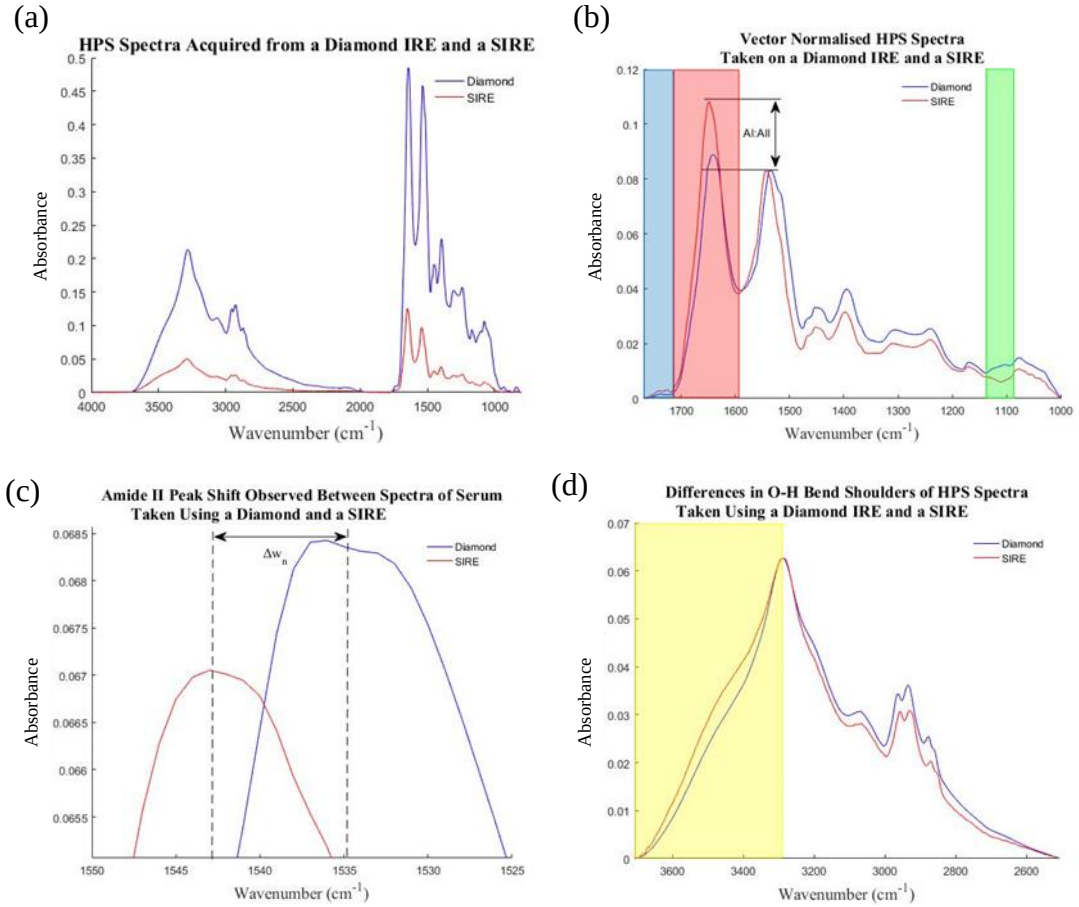


Figure 4.20. *Comparisons of HPS spectra acquired from a Di IRE (blue) and a SIRE (red). (a) baseline corrected spectra shows a significantly higher spectral intensity is achieved using a Di IRE. (b) Spectra has been cut around the fingerprint region and vector normalised to illuminate subtle differences in how HPS spectra presents using the two IREs. Blue and red highlights the higher sensitivity of SIREs to the lipid and amide 1 region of HPS, green indicates the sharp dip in the spectra taken on SIREs around 1107cm^{-1}. AI:AII highlights the disparity between the Amide I to Amide II ratio of the two spectra. (c) an enlarged view of the Amide II peak reveals the wavenumber shift, Δw_n, between spectra. (d) A vector normalisation around the O-H/N-H peak reveals a slightly raised shoulder of spectra acquired using a SIRE as highlighted by the yellow region.*

Figures 4.20.(b) highlights spectral ranges of the fingerprint region associated with lipids (blue), proteins (red), and carbohydrates (green) [70]. The blue and red

regions are some of the few areas that SIREs display a greater signal intensity of HPS over Di indicating greater sensitivity to the Amide I vibrations of proteins and C=O bonds of lipids [50], [66], [71], [72]. In the green region HPS spectra acquired from a SIRE dips while spectra from the Di IRE rises. This is a direct result from the silicon 2 phonon of intrinsic silicon vibrations that attenuate light in this specific region. This coincides with wavenumbers that have been attributed to the CO-O-C bonding environment [41]. This is spectroscopically significant in the detection and quantification of carbohydrates which may put SIREs at a disadvantage compared to Di IREs for carbohydrate centric applications [48], [73]. The yellow region of Figure 4.20.(d) denotes a part of the peak associated with the N-H bond of protein or O-H stretch of water which also seems to be slightly more sensitive on SIREs than on Di IREs [74].

The diagnostic implications of the spectral differences between SIREs and Di IREs for the analysis of serum are that Di IREs may be better suited to diagnostic applications where the detection or quantification of carbohydrate species (e.g., glucose levels in diabetics, lactate levels in multiple sclerosis) is needed, but SIREs may be better suited to clinical applications where lipid or protein determination is pertinent (e.g., triglyceride levels in cardiovascular disease, albumin levels in renal disease) [75]–[78]. SIREs may therefore be better equipped to detect sepsis in patients as sepsis is a condition characterised by release of a wide array of cytokine/chemokine proteins [79]. However, sepsis is also a highly metabolic condition where glucose and lactate levels play a key role in diagnosis and/or prognosis which may be more reliably detected using a Di IRE [80]. The theory that SIREs and Di IREs have advantages or

disadvantages for the analysis of carbohydrate, lipid, or protein macromolecule classes is explored further in chapter 5.

4.4. Conclusion

Characterizing the SIRE devices with respect to their spectroscopic response is important to determine specific biospectroscopic applications they may be suited toward. This has been done by first examining the spectral profile of the SIREs themselves and observing how the spectrum might change from SIRE to SIRE. It was found that SIREs will offset sample spectra baselines to unpredictable extents due to variations in impurity and free carrier concentrations in the silicon. Baseline correction largely corrects for this, but particularly large variability is still observed at 610cm^{-1} and 1107cm^{-1} which coincides with the most prominent silicon phonon band and with the Si-O band vibration, respectively. It may be more effective to remove any Si interference in serum spectra computationally by using EMSC with an Si reference spectrum [81]. The SIRE manufacturing yield has been estimated at 86.67% based on the reduced device quality seen in devices fabricated near the circumference of the silicon wafer. It is suggested that an acceptance criterion of <0.065 SiO peak absorbance is implemented as a SIRE quality control mechanism. From the remaining SIREs an 86.7% yield of viable devices was estimated. This figure may be improved with the use of low dopant concentrations and high purity Cz or FZ wafers.

In several instances the subject sample has been serum because it is important to see how the intrinsic variance of the silicon manifests in the analysis of a clinically

relevant sample medium. It was also necessary to establish how different approaches to handling this sample type might reduce procedural variance. A standardised serum preparation protocol was designed that utilised moderate forced air drying (60CFM) and moderate heating (30°C) that could completely dry serum upon SIREs in two minutes or less. This protocol was successfully implemented in an incubator to achieve batched drying of serum upon numerous SIREs, although it is recommended that efforts are made to maintain a relative humidity of approximately 30% in order to maintain spectral quality. It is also recommended that liquid samples are pipetted in sufficient volume (~3µl) and spread to cover the confinements of the sample area. Unfortunately, there is evidence to suggest that storage of dry serum samples on SIREs for reanalysis beyond a single day may not be possible as statistically significant differences in spectra over a 24-hour period was observed. This may be due to degradation of the serum in ambient conditions which may hopefully be remedied by refrigeration or freezing. Perhaps it is the method of drying, such as freeze drying [82], [83], that is key to establishing sample longevity and permit a sample reanalysis option. HPS spectra acquired from numerous randomly selected SIRE slides (n=14) showed a high level of reproducibility. However, sample preparation errors were observed to negatively impact spectral signature of a single SIRE slide further highlighting the need for controlled sample preparation protocols. Longer study periods are required to elucidate how reproducible SIREs are when used to analyse samples on different days. Surface treatment with Piranha was found to greatly improve sample area wettability and hence establish uniform sample coverage. Going forward, it may be advantageous to bestow a hydrophilic surface chemistry to all

SIREs using Piranha solution. However, hydrophilization by plasma activation may present a more stable alternative.

SIRE analytical abilities were demonstrated in a series of studies comparing the performance of SIREs to other IRE devices in the analysis of ethanol. Ethanol was selected at this stage to demonstrate SIREs as a competitive platform for FTIR analysis on a simple molecule. Although peak intensity was mediocre in SIREs compared to other IRE options, SIREs maintained a very high SNR. Furthermore, an ethanol concentration study found that SIREs had comparable R^2, RMSE, LOD and LOQ values to a Di IRE. Attention was given to the prominent differences between spectra of serum analysed on a SIRE and on a 'gold standard' Di IRE. These differences are primarily owed to anomalous dispersion artefacts that distort the spectrum acquired using a Di IRE to a greater degree. For example, a wavenumber shift of $8cm^{-1}$ and a change in amide I to amide II ratio is observed. It is speculated that these differences will give SIREs an advantage for the analysis of certain biological material or target biomolecules based on the observation that spectral regions associated lipid vibrational modes (e.g., C=O symmetric stretching vibration of lipids). Conversely, analysis of carbohydrates will likely be more difficult due to the attenuation of signal by the Si-O band at $1107cm^{-1}$ which coincides with CO-O-C symmetric stretching vibration common to some carbohydrates [41]. The aim of this chapter was to help accredit the practical advantages SIREs hold over conventional ATR methods by demonstrating that spectral quality and analytical capabilities can be maintained all while reaping the benefits of the high sample throughput capabilities of SIREs. To further accredit SIREs as a serum diagnostics platform, serum will be investigated more thoroughly as a complex and clinically relevant media in the next chapter. In particular, the ability to

discern key biomolecules within serum that are used in routine clinical diagnostics and decision making and that also broadly represent molecular features typical to carbohydrates, lipids and proteins is investigated.

4.5. References

[1] W. Zulehner, “Czochralski growth of silicon,” *J. Cryst. Growth*, vol. 65, no. 1–3, pp. 189–213, 1983, doi: 10.1016/0022-0248(83)90051-9.

[2] J. W. Medernach, “Infrared Characterization of Device-Quality Silicon,” in *Handbook of Vibrational Spectroscopy*, 2006, pp. 2621–2636.

[3] M. S. Kulkarni, “A selective review of the quantification of defect dynamics in growing Czochralski silicon crystals,” *Ind. Eng. Chem. Res.*, vol. 44, no. 16, pp. 6246–6263, 2005, doi: 10.1021/ie0500422.

[4] F. Shimura, “Springer Handbook of Electronic and Photonic Materials,” in *Springer Handbook of Electronic and Photonic Materials*, 2017, pp. 293–307.

[5] H. Ono and T. Ikarashi, “Ring-distribution of oxygen precipitates in Czochralski silicon revealed by low-temperature infrared absorption spectroscopy,” *Appl. Phys. Lett.*, vol. 63, no. 24, pp. 3303–3305, 1993, doi: 10.1063/1.110181.

[6] R. C. Newman, “Defects in silicon,” *Reports Prog. Phys.*, vol. 45, no. 10, pp. 1163–1210, 1982, doi: 10.1088/0034-4885/45/10/003.

[7] H. Fabian, P. Lasch, and D. Naumann, “Analysis of biofluids in aqueous environment based on mid-infrared spectroscopy,” *J. Biomed. Opt.*, vol. 10, no. 3, pp. 1–10, 2005, doi: 10.1117/1.1917844.

[8] L. Lovergne, G. Clemens, V. Untereiner, R. A. Lukaszweski, G. D. Sockalingum, and M. J. Baker, “Investigating Optimum Sample Preparation for Infrared Spectroscopic Serum Diagnostics,” *Anal. Methods*, vol. 7, no. 17, pp. 6921–7382, 2015, doi: 10.1039/C5AY00502G.

[9] J. M. Cameron, H. J. Butler, D. S. Palmer, and M. J. Baker, “Biofluid spectroscopic disease diagnostics: A review on the processes and spectral impact of drying,” *J. Biophotonics*, vol. 11, pp. 1–12, 2018, doi: 10.1002/jbio.201700299.

[10] E. Diessel, P. Kamphaus, K. Grothe, R. Kurte, U. Damm, and H. M. Heise, “Nanoliter serum sample analysis by mid-infrared spectroscopy for minimally invasive blood-glucose monitoring,” *Appl. Spectrosc.*, vol. 59, no. 4, pp. 442–451, 2005, doi: 10.1366/0003702053641360.

[11] C. Hughes *et al.*, “Assessing the challenges of Fourier transform infrared spectroscopic analysis of blood serum,” *J. Biophotonics*, vol. 7, no. 3–4, pp. 180–188, 2014, doi: 10.1002/jbio.201300167.

[12] Y. Fan, A. Arevalo, H. Li, and I. G. Foulds, “Low-cost silicon wafer dicing using a craft cutter,” *Microsyst. Technol.*, vol. 21, no. 7, pp. 1411–1414, 2015, doi: 10.1007/s00542-014-2198-4.

[13] A. J. Souna, J. S. Bender, and J. T. Fourkas, “How clean is the solvent you use to clean your optics? A vibrational sum-frequency-generation study,” *Appl. Opt.*, vol. 56, no. 13, p. 3875, 2017, doi: 10.1364/ao.56.003875.

[14] S. Madduri, B. G. Sammakia, B. Infantolino, and S. Chaparala, “A review of non-hermetic optoelectronic packaging,” *2008 11th IEEE Intersoc. Conf. Therm. Thermomechanical Phenom. Electron. Syst.*, pp. 913–919, 2008, doi: 10.1109/ITHERM.2008.4544364.

[15] R. C. Newman, “Light impurities and their interactions in silicon,” *Mater. Sci. Eng.*, vol. 36, no. B, pp. 1–12, 1996, doi: 10.1016/0921-5107(95)01271-0.

[16] J. Li, M. Liu, Y. Tan, and L. Wen, “Effect of the silicon nitride coating of quartz crucible on impurity distribution in ingot-cast multicrystalline silicon,” *Int. J. Appl. Ceram. Technol.*, pp. 1–5, 2012, doi: 10.1111/j.1744-7402.2012.02832.x.

[17] K. Takata, J. Fujise, B. Ko, and T. Ono, “Influence of impurities on mechanical strength of silicon crystal,” *Jpn. J. Appl. Phys.*, vol. 58, no. 45507, pp. 1–6, 2019, doi: 10.7567/1347-4065/ab0ebc.

[18] M. Auslender and S. Hava, “21. Single-Crystal Silicon: Electrical and Optical Properties,” in *Materials Today*, vol. 10, no. C, 2006, pp. 485–522.

[19] P. Torchio and R. Occelli, “Detection and Mapping of Oxygen in Silicon Wafers by Scanning Infrared Absorption,” *Int. J. Infrared Millimeter Waves*, vol. 18, no. 2, pp. 491–499, 1997.

[20] B. Yazici and S. Yolacan, “A comparison of various tests of normality,” *J. Stat. Comput. Simul.*, vol. 77, no. 2, pp. 175–183, 2007, doi: 10.1080/10629360600678310.

[21] M. Paraskevaidi *et al.*, “Attenuated total reflection Fourier-transform infrared (ATR-FTIR) spectroscopy to diagnose osteoarthritis in equine serum,” *Equine Vet. J.*, vol. 52, no. 1, pp. 46–51, 2020, doi: 10.1111/evj.13115.

[22] S. P. Davies, J. Y. K. BSC, A. J. Thrush, and P. H. S. S. BA, “A comparison of mercury and digital clinical thermometers,” *J. Adv. Nurs.*, vol. 11, no. 5, pp. 535–543, 1986, doi: 10.1111/j.1365-2648.1986.tb01285.x.

[23] D. Schroder, R. Thomas, and J. Swartz, “Free Carrier Absorption in Silicon,” *IEEE J. Solid-State Circuits*, vol. SC-13, no. 1, pp. 180–187, 1978.

[24] P. Bassan *et al.*, “FTIR microscopy of biological cells and tissue: Data analysis using resonant Mie scattering (RMieS) EMSC algorithm,” *Analyst*, vol. 137, no. 6, pp. 1370–1377, 2012, doi: 10.1039/c2an16088a.

[25] P. Bassan *et al.*, “Reflection contributions to the dispersion artefact in FTIR spectra of single biological cells,” *Analyst*, vol. 134, no. 6, pp. 1171–1175, 2009, doi: 10.1039/b821349f.

[26] K. Wehbe, J. Filik, M. D. Frogley, and G. Cinque, “The effect of optical substrates on micro-FTIR analysis of single mammalian cells,” *Anal. Bioanal. Chem.*, vol. 405, no. 4, pp. 1311–1324, 2013, doi: 10.1007/s00216-012-6521-6.

[27] H. Yang, S. Yang, J. Kong, A. Dong, and S. Yu, “Obtaining information about protein secondary structures in aqueous solution using Fourier transform IR spectroscopy,” *Nat. Protoc.*, vol. 10, no. 3, pp. 382–396, 2015, doi: 10.1038/nprot.2015.024.

[28] H. Ono, T. Ikarashi, S. Kimura, A. Tanikawa, and K. Terashima, “Anomalous Distribution of Oxygen Precipitates in a Silicon Wafer After Annealing,” *Early Stages Oxyg. Precip. Silicon*, pp. 509–516, 2012, doi: 10.1007/978-94-009-0355-5_42.

[29] M. Hasebe, Y. Takeoka, S. Shinoyama, and S. Naito, “Formation process of stacking faults with ringlike distribution in CZ-Si wafers,” *Jpn. J. Appl. Phys.*, vol. 28, no. 11 A, pp. L1999–L2002, 1989, doi: 10.1143/JJAP.28.L1999.

[30] L. Valek and J. Sik, “Defect Engineering During Czochralski Crystal Growth and Silicon Wafer Manufacturing,” in *Modern Aspects of Bulk Crystal and Thin Film Preparation*, 2012, pp. 43–70.

[31] R. Basnet, C. Sun, H. Wu, H. T. Nguyen, F. E. Rougieux, and D. Macdonald, “Ring defects in n-type Czochralski-grown silicon: A high spatial resolution study using Fourier-transform infrared spectroscopy, micro-photoluminescence, and micro-Raman,” *J. Appl. Phys.*, vol. 124, no. 24, pp. 1–7, 2018, doi: 10.1063/1.5057724.

[32] V. K. Malyutenko, V. I. Chernyakovsky, and T. Piotrowski, “Characterization of oxygen impurity concentration in silicon based on thermal emission measurements,” *Infrared Phys. Technol.*, vol. 37, no. 4, pp. 499–504, 1996, doi: 10.1016/1350-4495(95)00077-1.

[33] J. M. Chalmers, “Mid-Infrared Spectroscopy: Anomalies, Artifacts and Common Errors,” in *Handbook of Vibrational Spectroscopy*, 2006, pp. 2327–2347.

[34] W. von Ammon, A. Sattler, and G. Kissinger, “Defects in Monocrystalline Silicon,” in *Springer Handbook of Electronic and Photonic Materials*, 2017.

[35] K. Jami, I. Pollentier, S. Vedula, and G. Blumenstock, “Influence of Immersion Lithography on Wafer Edge Defectivity,” in *Lithography*, no. 1, 2010, pp. 34–40.

[36] T. Kim, J. Oh, and H. Yoo, “Criticality of Wafer Edge Inspection and Metrology Data to All-Surface Defectivity Root Cause and Yield Analysis,” in *Proceedings of the SEMI Advanced Semiconductors Manufacturing Conference*, 2016, pp. 140–146.

[37] O. Yavas, E. Richter, C. Kluthe, and M. Sickmoeller, “Wafer-edge yield engineering in leading-edge DRAM manufacturing,” *Semicond. Fabtech*, vol. 39, pp. 1–5, 2005.

[38] F. Burkeen, S. Vedula, and S. Meeks, “Visualizing the Wafer’s Edge,” *Yield Manag. Solut.*, pp. 18–20, 2007.

[39] M. Hancer, R. P. Sperline, and J. D. Miller, “Anomalous dispersion effects in the IR-ATR spectroscopy of water,” *Appl. Spectrosc.*, vol. 54, no. 1, pp. 138–143, 2000, doi: 10.1366/0003702001948222.

[40] F. Bonnier *et al.*, “Screening the low molecular weight fraction of human serum using ATR-IR spectroscopy,” *J. Biophotonics*, vol. 9, no. 10, pp. 1085–1097, 2016, doi: 10.1002/jbio.201600015.

[41] M. J. Baker *et al.*, “Using Fourier transform IR spectroscopy to analyze biological materials,” *Nat. Protoc.*, vol. 9, no. 8, pp. 1771–1791, 2014, doi: 10.1038/nprot.2014.110.

[42] N. Fatih *et al.*, “Mid-infrared spectroscopy of serum, a promising non-invasive method to assess prognosis in patients with ascites and cirrhosis,” *PLoS One*, vol. 90, pp. 1–15, 2017, doi: 10.6084/m9.figshare.4753468.Funding.

[43] E. Goormaghtigh, V. Raussens, and J. M. Ruysschaert, “Attenuated total reflection infrared spectroscopy of proteins and lipids in biological membranes,” *Biochim. Biophys. Acta - Rev. Biomembr.*, vol. 1422, no. 2, pp. 105–185, 1999, doi: 10.1016/S0304-4157(99)00004-0.

[44] J. R. Hands *et al.*, “Attenuated Total Reflection Fourier Transform Infrared (ATR-FTIR) spectral discrimination of brain tumour severity from serum samples,” *J. Biophotonics*, vol. 7, no. 3–4, pp. 189–199, 2014, doi: 10.1002/jbio.201300149.

[45] J. Ollesch, H. M. Heise, and T. Behrens, “FTIR spectroscopy of biofluids revisited: an automated approach to spectral biomarker identification,” *Analyst*, vol. 138, pp. 4092–4102, 2013, doi: 10.1039/c3an00337j.

[46] W. Bou Zeid and D. Brutin, “Influence of relative humidity on spreading, pattern formation and adhesion of a drying drop of whole blood,” *Colloids Surfaces A Physicochem. Eng. Asp.*, vol. 430, pp. 1–7, 2013, doi: 10.1016/j.colsurfa.2013.03.019.

[47] Y. Li, Q. Yang, M. Li, and Y. Song, “Rate-dependent interface capture beyond the coffee-ring effect,” *Sci. Rep.*, vol. 6, no. April, pp. 1–8, 2016, doi: 10.1038/srep24628.

[48] F. Bonnier *et al.*, “Ultra-filtration of human serum for improved quantitative analysis of low molecular weight biomarkers using ATR-IR spectroscopy,” *Analyst*, vol. 142, pp. 1285–1298, 2017, doi: 10.1039/c6an01888b.

[49] C. Hughes *et al.*, “Introducing Discrete Frequency Infrared Technology for High- Throughput Biofluid Screening,” *Nat. Publ. Gr.*, vol. 6, no. 20173, pp. 1–8, 2016, doi: 10.1038/srep20173.

[50] K. Z. Liu, R. A. Shaw, A. Man, T. C. Dembinski, and H. H. Mantsh, “Reagent-free, simultaneous determination of serum cholesterol in HDL and LDL by infrared spectroscopy,” *Clin. Chem.*, vol. 48, no. 3, pp. 499–506, 2002.

[51] W. Ohishi *et al.*, “Feasibility of freeze-dried sera for serological and molecular biological detection of hepatitis B and C viruses,” *J. Clin. Microbiol.*, vol. 44, no. 12, pp. 4593–4595, 2006, doi: 10.1128/JCM.00621-06.

[52] G. McHale, N. J. Shirtcliffe, and M. I. Newton, “Contact-angle hysteresis on super-hydrophobic surfaces,” *Langmuir*, vol. 20, no. 23, pp. 10146–10149, 2004, doi: 10.1021/la0486584.

[53] V. Silverio, P. A. G. Canane, and S. Cardoso, “Surface wettability and stability of chemically modified silicon, glass and polymeric surfaces via room temperature chemical vapor deposition,” *Colloids Surfaces A Physicochem. Eng. Asp.*, vol. 570, no. January, pp. 210–217, 2019, doi: 10.1016/j.colsurfa.2019.03.032.

[54] X. M. Yang, Z. W. Zhong, E. M. Diallo, Z. H. Wang, and W. S. Yue, “Silicon wafer wettability and aging behaviors : Impact on gold thin-film morphology,” *Mater. Sci. Semicond. Process.*, vol. 26, pp. 25–32, 2014, doi: 10.1016/j.mssp.2014.03.044.

[55] L. Chen, X. He, H. Liu, L. Qian, and S. H. Kim, “Water Adsorption on Hydrophilic and Hydrophobic Surfaces of Silicon,” *J. Phys. Chem. C*, vol. 122, pp. 11385–11391, 2018, doi: 10.1021/acs.jpcc.8b01821.

[56] R. Souda, “Adsorption, diffusion, dewetting, and entrapment of acetone on Ni(111), surface-modified silicon, and amorphous solid water studied by time-of-flight secondary ion mass spectrometry and temperature programmed desorption,” *J. Chem. Phys.*, vol. 135, no. 16, 2011, doi: 10.1063/1.3656071.

[57] J. J. Cras, C. A. Rowe-taitt, D. A. Nivens, and F. S. Ligler, “Comparison of chemical cleaning methods of glass in preparation for silanization,” *Biosens. Bioelectron.*, vol. 14, pp. 683–688, 1999.

[58] G. Kissinger and W. Kissinger, “Hydrophilicity of Silicon Wafers for Direct Bonding,” *Phys. Status Solidi*, vol. 123, no. 1, pp. 185–192, Jan. 1991, doi: 10.1002/pssa.2211230117.

[59] S. Kaya, P. Rajan, H. Dasari, D. C. Ingram, W. Jadwisienczak, and F. Rahman, “A Systematic Study of Plasma Activation of Silicon Surfaces for Self Assembly,” *ACS Appl. Mater. Interfaces*, vol. 7, no. 45, pp. 25024–25031, Oct. 2015, doi: 10.1021/acsami.5b08358.

[60] P. J. Larkin, “Instrumentation and Sampling Methods,” in *Infrared and Raman Spectroscopy*, 2011, pp. 27–54.

[61] D. A. Armbruster and T. Pry, “Limit of Blank, Limit of Detection and Limit of Quantification,” *Clin Biochem Rev*, vol. 29, no. I, pp. 49–52, 2008.

[62] L. A. Averett, P. R. Griffiths, and K. Nishikida, “Effective path length in attenuated total reflection spectroscopy,” *Anal. Chem.*, vol. 80, no. 8, pp. 3045–

3049, 2008, doi: 10.1021/ac7025892.

[63] S. G. Kazarian and K. L. A. Chan, “ATR-FTIR spectroscopic imaging: recent advances and applications to biological systems,” *Analyst*, vol. 138, no. 7, pp. 1940–1951, 2013, doi: 10.1039/C3AN36865C.

[64] S. Liu *et al.*, “Measurement of the refractive index of whole blood and its components for a continuous spectral region,” *J. Biomed. Opt.*, vol. 24, no. 3, pp. 1–5, 2019, doi: 10.1117/1.jbo.24.3.035003.

[65] J. Grdadolnik, “ATR-FTIR spectroscopy: Its advantages and limitations,” *Acta Chim. Slov.*, vol. 49, no. 3, pp. 631–642, 2002.

[66] M. Boulet-Audet, T. Buffeteau, S. Boudreault, N. Daugey, and M. Pézolet, “Quantitative determination of band distortions in diamond attenuated total reflectance infrared spectra,” *J. Phys. Chem. B*, vol. 114, no. 24, pp. 8255–8261, 2010, doi: 10.1021/jp101763y.

[67] S. E. Glassford, B. Byrne, and S. G. Kazarian, “Recent applications of ATR FTIR spectroscopy and imaging to proteins,” *Biochim. Biophys. Acta - Proteins Proteomics*, vol. 1834, no. 12, pp. 2849–2858, 2013, doi: 10.1016/j.bbapap.2013.07.015.

[68] D. . Edwards and E. Ochoa, “Infrared refractive index of silicon,” *Appl. Opt.*, vol. 19, no. 24, pp. 4130–4131, 1980.

[69] M. Miljković, B. Bird, and M. Diem, “Line shape distortion effects in infrared spectroscopy,” *Analyst*, vol. 137, no. 17, pp. 3954–3964, 2012, doi: 10.1039/c2an35582e.

[70] E. Gray *et al.*, “Health economic evaluation of a serum-based blood test for brain tumour diagnosis: Exploration of two clinical scenarios,” *BMJ Open*, vol. 8, no. 5, pp. 1–9, 2018, doi: 10.1136/bmjopen-2017-017593.

[71] A. Oleszko, J. Hartwich, A. Wójtowicz, M. Gąsior-Głogowska, H. Huras, and M. Komorowska, “Comparison of FTIR-ATR and Raman spectroscopy in determination of VLDL triglycerides in blood serum with PLS regression,” *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.*, vol. 183, pp. 239–246, 2017, doi: 10.1016/j.saa.2017.04.020.

[72] K. Spalding *et al.*, “Enabling quantification of protein concentration in human serum biopsies using attenuated total reflectance – Fourier transform infrared (ATR-FTIR) spectroscopy,” *Vib. Spectrosc.*, vol. 99, pp. 50–58, 2018, doi: 10.1016/j.vibspec.2018.08.019.

[73] C. Petibois, A. M. Melin, A. Perromat, G. Cazorla, and G. Déléris, “Glucose and lactate concentration determination on single microsamples by Fourier-transform infrared spectroscopy,” *J. Lab. Clin. Med.*, vol. 135, no. 2, pp. 210–215, 2000, doi: 10.1067/mlc.2000.104460.

[74] G. I. Dovbeshko, N. Y. Gridina, E. B. Kruglova, and O. P. Pashchuk, “FTIR spectroscopy studies of nucleic acid damage,” *Talanta*, vol. 53, no. 1, pp. 233–246, 2000, doi: 10.1016/S0039-9140(00)00462-8.

[75] Y. C. Shen, A. G. Davies, E. H. Linfield, T. S. Elsey, P. F. Taday, and D. D. Arnone, “The use of fourier-transform infrared spectroscopy for the quantitative determination of glucose concentration in whole blood,” *Phys. Med. Biol.*, vol. 48, no. 13, pp. 2023–2032, 2003, doi: 10.1088/0031-9155/48/13/313.

[76] A. M. Amorini *et al.*, “Serum lactate as a novel potential biomarker in multiple sclerosis,” *Biochim. Biophys. Acta - Mol. Basis Dis.*, vol. 1842, no. 7, pp. 1137–1143, 2014, doi: 10.1016/j.bbadis.2014.04.005.

[77] G. Yuan, K. Z. Al-Shali, and R. A. Hegele, “Hypertriglyceridemia: Its etiology, effects and treatment,” *Can. Med. Assoc. J.*, vol. 176, no. 8, pp. 1113–1120, 2007, doi: 10.1503/cmaj.060963.

[78] G. A. Kaysen, “Albumin turnover in renal disease,” *Miner. Electrolyte Metab.*, vol. 24, no. 1, pp. 55–63, 1997, doi: 10.1159/000057351.

[79] C. Pierrakos and J. L. Vincent, “Sepsis biomarkers: A review,” *Crit. Care*, vol. 14, no. 1, pp. 1–18, 2010, doi: 10.1186/cc8872.

[80] S. Y. Cho and J. H. Choi, “Biomarkers of Sepsis,” *Crit. Rev. Clin. Lab. Sci.*, vol. 50, no. 1, pp. 23–36, 2013, doi: 10.3947/ic.2014.46.1.1.

[81] R. Gautam, S. Vanga, F. Ariese, and S. Umapathy, “Review of multidimensional data processing approaches for Raman and infrared spectroscopy,” *EPJ Tech. Instrum.*, vol. 2, no. 8, pp. 1–38, 2015, doi: 10.1140/epjti/s40485-015-0018-6.

[82] F. Emami, A. Vatanara, E. J. Park, and D. H. Na, “Drying technologies for the stability and bioavailability of biopharmaceuticals,” *Pharmaceutics*, vol. 10, no. 3, pp. 1–22, 2018, doi: 10.3390/pharmaceutics10030131.

[83] Z. Kakooei, M. Mohebali, B. Akhoundi, M. Mohebali, and A. R. Foroshani, “Stability of freeze-dried sera stored at different temperatures for the detection of anti-leishmania infantum antibodies using direct agglutination test,” *Iran. J. Public Health*, vol. 43, no. 11, pp. 1557–1562, 2014.

Chapter 5: SIRE Sample Slide Biomolecule Detection and Quantification

5.1. Introduction

All Internal Reflection Elements (IREs) possess strengths and weaknesses that dictate the real-world applications they are suitable for. It has previously been mentioned that implementation of ATR-FTIR within the clinic is limited (chapter 1), but clinical non-implementation doesn't mean there hasn't been huge amounts of research surrounding the spectroscopic analysis of clinically relevant biological samples [1]–[3]. This research has provided a wealth of information detailing the band assignments for bonding environments indicative of specific biomolecules, particularly of the four major biomolecular classes: carbohydrates, lipids, proteins and nucleic acids [4]. These four classes, by definition, have characteristic bonding environments which are useful for distinguishing them from each other using IR spectroscopy. Given previous findings that SIREs and Di IREs differ quite significantly in their spectroscopic response (chapter 4, section 4.3.3.2), it stands to reason certain biomolecular classes will be more easily interrogated using one IRE substrate than the other. This in turn has implications concerning which applications within the clinical laboratory each IRE substrate is suitable for, putting aside the practical advantages of high-throughput FTIR SIREs. Capturing fingerprint region information from 600-1500cm^{-1} using a Silicon substrate has been demonstrated, but with the condition that at isolated and narrow wavenumbers the IR beam was strongly attenuated. In particular, the spectrum at 1107cm^{-1} was found to be affected in such a way that a univariate analysis on an analyte at this wavenumber is likely not possible. Conversely, it was observed that serum spectra acquired on Di IREs are more distorted than equivalent spectra acquired using SIREs due to Di's lower refractive index. This,

together with the crystal lattice absorptions of Di above 1500cm^{-1}, may make SIREs better suited to the analysis of carbonyl peaks adjacent to the Amide I band. These differences need to be understood to fully appreciate and characterise analytical capabilities of the SIRE with reference to Diamond as a 'gold standard' technique. To do this, archetypal biomolecules of the major classes have been selected for comparative analysis on both IRE substrates. Each biomolecule has major clinical utility and have been subject to previous investigation by infrared spectroscopy so that analytical capabilities cited in the literature may also be used as a reference point. The aim of this chapter is to demonstrate that limits of detection and reliability comparable to conventional ATR-FTIR methods can be achieved using SIREs.

Four biomolecules were spiked into Human Pooled Serum (HPS) at concentrations that reflect physiological ranges in disease. The biomolecules selected were glucose, triglyceride, lactate, and albumin. Nucleic acids have been neglected from this study since they are only present in serum in the ng/ml range, which is well below detection limits for ordinary benchtop ATR-FTIR [5].

Glucose is ubiquitous and abundant in almost all physiological environments and is essential to metabolic processes as the primary energy currency in most organisms [6]. Within this respect it is a vital diagnostic and prognostic marker in a wide variety of pathological states. Glucose monitoring is crucial in control of diabetes, and monitoring of patients in critical care including premature babies and patients with severe sepsis and septic shock [7], [8]. The most prominent example is the measurement of blood glucose for monitoring treatment in diabetes mellitus, although tests for this application are well established and future developments seek to develop non-invasive alternatives [9]. It has also been suggested that frequent

monitoring of blood glucose will improve mortality rates in sepsis patients [10]. The standard mode of measurement is by enzymatic or electrochemical means using a glucose oxidase, hexokinase, or glucose dehydrogenase which necessitates many more preparation steps than FTIR approaches and are selective only to glucose [11]. Typically, the physiological range of glucose is between 0.8 and 1.2 mg/ml [12], but this can be expected to be anywhere around 0.3 to 3.0 mg/ml in pathological conditions [10]. In sepsis, for instance, hyperglycaemia can elevate blood glucose levels to 3.43 mg/ml [13].

Lactate is a prominent metabolite and is therefore another key marker in altered metabolic states such as extreme exercise, multiple sclerosis, and sepsis [14], [15]. In fact, lactate determination is the most requested test in the monitoring and diagnosis of sepsis [16] and is more generally associated with higher mortality at elevated levels [17]. It is a small, polar, organic molecule that is present in increased concentrations under anaerobic conditions and as such is used to indicate tissue hypoxia. Lactate is also measured enzymatically using lactate oxidase which has been translated into central laboratory, point-of-care, and handheld devices [18]. The metabolite exists in blood at a concentration less than 0.25 mg/ml in healthy individuals, and can be observed at over 1 mg/ml in abnormal conditions and in extreme conditions over 1.5 mg/ml, although this can be subject to sample collection site as tissue hypoxia may be localized [19]–[21].

Triglycerides are the basic unit of lipids and comprise three fatty acid chains of variable length adjoined via a glyceride backbone. They are useful markers in cardiovascular disease, sepsis, and pancreatitis [22], [23]. Ordinarily triglycerides are found in serum at concentrations lower than 1.5mg/ml, but in hypertriglyceridemia can

elevate in excess of 5mg/ml [24][25]. In clinical determinations triglyceride is hydrolysed to produce glycerol which is subsequently measured through enzymatic means [26].

Albumin is the most prevalent biomolecule in blood serum, accounting for approximately 50% of the total protein content and at a concentration anywhere between 35 to 50mg/ml in healthy individuals [27]. Hypoalbuminemia is typically defined as serum albumin concentrations below 35mg/ml which progresses to severe hypalbuminaemia below 25mg/ml [28]. Depreciating serum albumin levels is broadly associated with increased mortality rates having clinical indications for sepsis, malnutrition, and renal disease among numerous other pathologies [29]–[31]. There are a wide variety of methods available to determine albumin levels which each hold certain advantages and disadvantages, but in general determination is performed via reagent method using Bromocresol Purple [32].

The chemical structures of each of the biomolecules under investigation is discussed and related to the spectroscopic response of each substance in its pure form. This information was used to elucidate spectral features that are specific to each biomolecule and identify peaks that can be used to discern the biomolecules presence within complex biological media, in this case: serum. The Area Under the Curve (AUC) of these features was calculated and used to fit linear regression models to the spiked HPS concentration models. Many highly specific spectral peaks present within the fingerprint region can generally be used to pinpoint exact features typical to the analyte of interest so long as the user knows what they are looking for. In complex media such as serum, it can be expected that many fingerprint regions overlap therefore the datasets lend themselves to interpretation by way of multivariate analysis [33].

Partial Least Squares (PLS) models were thus constructed to see if differences could be reconciled by way of multivariate analysis. Comments are made where calibration curves may be improved by altering sample preparation protocols or instrumentation drawing from pertinent examples in the literature. Both the univariate (AUC) and multivariate approaches are used to compare and contrast Di IRE and SIRE performance both at localized wavenumbers and across large wavenumber regions.

5.2. Materials and Method

A full list of the biomolecules under investigation along with details of the physical form they were received in is presented in Table 5.1. All biomolecules were acquired from Sigma Aldrich.

Table 5.1. *List of biomolecules under investigation.*

Biomolecule	Description
Albumin	1mg, lyophilized powder, >96%
Glucose	100g, D-(+)-glucose (powder), >99.5%
Lactate	10g, L-(+)-lactic acid (powder), >98%
Triglycerides	100mg, triglyceride mix (liquid), C2-C10

Biomolecules were spiked into three different media: distilled water, phosphate buffered saline (PBS), and human pooled serum (HPS). In most cases acquiring spectra of spiked distilled water or PBS was either difficult or impossible because the media would simply evaporate when dried leaving insufficient amounts of residue on the IRE to be reliably scanned. Spiked HPS concentrations need to account for a concentration offset of native biomolecular species. A full list of the known biomolecule

concentrations for the HPS used in these experiments can be found in Table 5.2. Information regarding native lactate concentration was not available so has not been included. Albumin typically constitutes 50% of the total protein content and so baseline albumin content has been calculated at 25.5mg/ml [34], [35].

Table 5.2. *List of known native biomolecule concentrations.*

Human serum type AB (male) from male AB (All values in mg/ml unless otherwise stated)	
Source	Male plasma from donors within the USA
Iron	0.00099
Cholesterol	1.49
Triglycerides	1.05
Glucose	0.92
Sodium (Na)	140mmol/l
Total protein	51
pH	7.6
Endotoxin level	1EU/ml
Hemoglobin	0.7

All experiments have been carried out using both SIREs and a Di IRE. Each biomolecule has been investigated within a clinically relevant concentration range taking slightly above and below maximum and minimum values found in the literature and incrementing between these values. These ranges have been presented in Table 5.3. In some cases, the extremities of the concentration range represent extreme physiological circumstances (e.g., in severe sepsis, lactate levels can rise to excess of 1mg/ml [19]). All concentration ranges tested and the revised concentrations with offset values have been presented in Tables 5.3.(a) and 5.3.(b) respectively. All analysis was carried out on the concentration ranges presented in Table 5.3.(b).

Table 5.3.(a). Initial concentrations of each biomolecule in spiked serum models.

Biomolecule	Concentration (mg/ml)															
Albumin	0	0.5	1	2.5	5	10	20	30	40	50	60					
Glucose	0	0.3	0.6	0.9	1.2	1.5	1.8	2.1								
Lactate	0	0.025	0.05	0.1	0.15	0.2	0.3	0.4	0.5	0.75	1	2	3	4	5	10
Triglycerides	0	0.1	0.2	0.5	1	1.5	2	2.5	3	4	5	5.5	11			

Table 5.3.(b). Revised concentrations of each biomolecule in spiked serum models.

Biomolecule	Concentration (mg/ml)											
Albumin	25.5	26	26.5	28	30.5	35.5	45.5	55.5	65.5			
Glucose	0.92	1.22	1.52	1.82	2.12	2.42	2.72	3.02				
Lactate	0	0.025	0.05	0.1	0.15	0.2	0.3	0.4	0.5	0.75	1	2
Triglycerides	1.05	1.15	1.25	1.55	2.05	2.55	3.05	3.55	4.05	5.05	6.05	6.55

The same Di IRE was used across all experiments and was cleaned by transferring the Specac Quest ATR interchangeable puck from the accessory unit to a sink and using an ethanol-soaked cotton swab to ensure all sample was removed from the crystal. The sample side of the puck was then rinsed using a squeeze bottle of DI water, then an ethanol (Absolute ≥99.5%, Sigma-Aldrich, UK) squeeze bottle, then DI water again before drying thoroughly with a paper towel. In contrast, different SIREs were used for every concentration under investigation thereby removing the need to wash SIREs between analysis. SIREs were pre-treated with Piranha before assembly. Piranha was prepared fresh as described in chapter 4 section 4.2.4.3 for every use.

For SIREs, 3µl of HPS was pipetted into positions '1', '2', and '3' of the SIRE slide by first depositing the droplet onto the centre then, without removing the pipette tip, using the pipette tip to spread the droplet to the corners and edges until all empty areas of the sample well has been covered in a thin film. For the Di IRE, 1.5µl of sample was pipetted onto the IRE. The reason a smaller sample volume was used on the Di IRE was because the Di IRE has a much smaller sampling area than the SIRE. Samples would spill over the edge of the Di IRE crystal when using 3µl serum and form molecular weight-dependant drying patterns due to the coffee ring effect. Thus, sample constituents interrogated by the Di IRE were not representative of the whole sample. Therefore, sample volume was reduced to 1.5µl, but remained 3µl for SIREs so that enough sample could cover the SIREs sampling area. All samples were dried in an incubator at 35°C. The incubator was employed to take advantage of the batched processing method of SIREs described previously (chapter 4, section 4.3.2.1). Serum was dried on SIREs for one hour, but only dried on the Di IRE for ten minutes because one sample at a time on the Di IRE was too time consuming. It was previously

demonstrated (chapter 4, section 4.3.2.1) that samples would be completely dry well within these time frames as indicated by the O-H stretch band reaching steady-state intensity in agreement with other researchers' findings under ambient air drying conditions [1], [36].

All analysis was carried out on a Spectrum2 infrared spectrometer (PerkinElmer, USA) with a custom automated slide indexing unit (see chapter 3, section 3.2.3). Spectra were acquired from the 4000-450cm^{-1} range using 16 co-added scans and a resolution set to 4cm^{-1} with 1cm^{-1} data spacing. All sample droplets were scanned in triplicate (three instrument repeats) and all trials were carried out twice to provide at least two biological repeats. In its intended use, one sample is analysed per SIRE. This means that 18 spectra were collected using SIREs (6 spectra per well), and 6 were collected using the Di IRE for each concentration level. In all cases, a blank sample of HPS was used to serve as a control. A Matlab (MathWorks, USA) based General User Interface (GUI) developed in-house was used for pre-processing and exploratory analysis. PLSR models were constructed using Matlab code developed by the Bonnier group at Université de Tours [37].

5.2.1. Quantitative Analysis

Analysis for all four biomolecules have been carried out following a general convention. All spectra were pre-processed by rubber band baseline correction and then vector normalisation. Spectra from the extremes of the concentration ranges investigated were subtracted from one another (i.e., a blank spectrum was subtracted from the spectrum of the highest analyte concentration) to elucidate where the greatest

spectral change occurs. This region was isolated for univariate analysis to establish if a linear relationship could be made between area under the curve and analyte concentration. A linear regression model was fitted to the data and r^2, RMSE, Limit of Detection (LOD), and Limit of Quantification (LOQ) calculated. Multivariate analysis was then carried out on the same datasets using the PLS method. 50% of the data was randomly selected to comprise the calibration set and construct predictive models and the other 50% was used in the validation set to carry out quantitative predictions on. This process was repeated over 100 iterations of a cross validation loop to assess how robust the models were. The average RMSE of the Validation set (RMSEV) and r^2 values are calculated at each iteration to inform how precise the model is and how well the known concentrations fit to the predicted concentrations. The standard deviation of both RMSEV and r^2 at each iteration is also calculated to inform how repeatable the models are. The PLS loadings of the first two dimensions have been displayed to show were the most and second most variance in the spectra exists. In each case it is expected that this will correspond to the regions selected in the area under the curve analysis. How the RMSEV value evolves over 20 PLS dimensions has also been shown with error bars indicating the standard deviation of RMSEV to show how repeatable the models are at each dimension. This is to inform the user how many dimensions are required to best describe the data. Once the RMSEV has reached the lowest value the optimal number of dimensions has been reached. The RMSEV and r^2 values of PLS models using this number of dimensions was taken for subsequent discussion. Finally, the average predicted concentrations were plotted against the known concentrations and a regression line fit to the data. The error bars represent one standard deviation of

the predicted concentration values. From this, the LOD and LOQ were again calculated.

5.3. Results and Discussion

5.3.1. Pure Compound Spectra and Band Assignments

Each biomolecule in the raw form they were received in was analysed on both the SIREs and the Di IRE. Resultant spectra have been presented in Figures 5.1.(a) – (d). Spectra have been baseline corrected then offset for clarity.

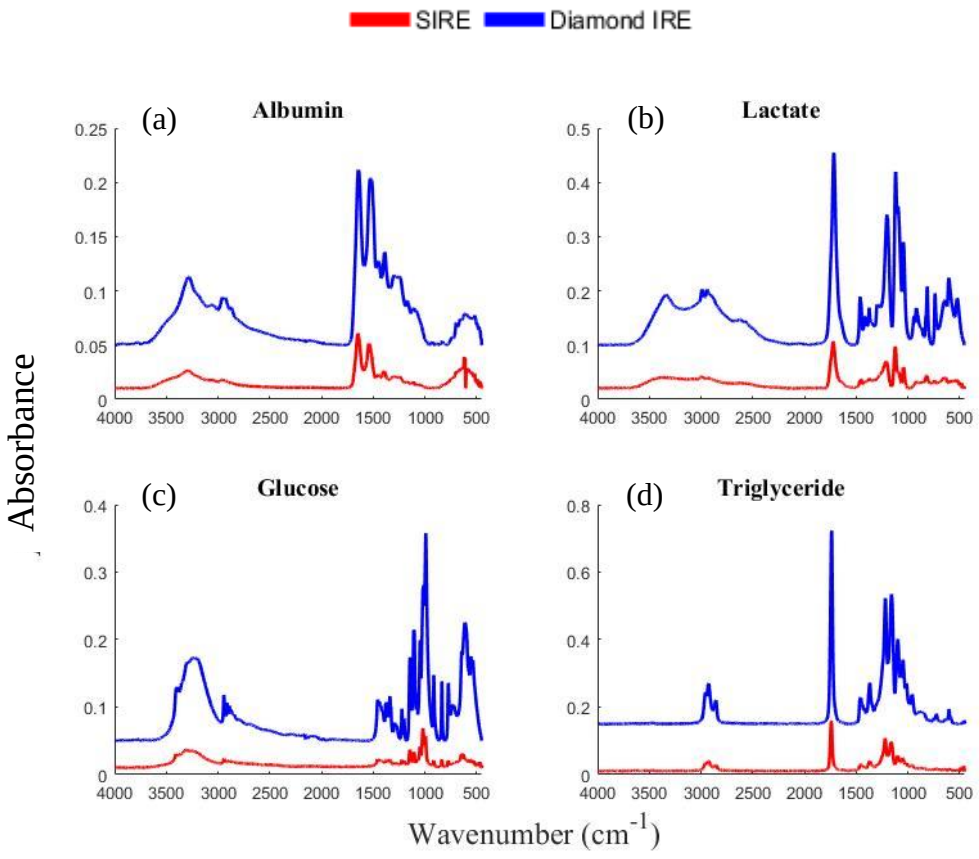


Figure 5.5. *Spectra of pure biomolecules as measured on a Di IRE (blue) and a SIRE (red). Spectra have been offset for clarity.*

In terms of overall form, spectra of samples scanned on a SIRE or a Di IRE do not differ considerably barring a fourfold increase in spectral intensity achieved on the Di IRE and a *ca*. ~8cm^{-1} wavenumber shift consistent with findings reported in chapter 4 section 4.3.3. Spectra of pure albumin bears a striking resemblance to spectra of dried HPS. Albumin dominates the HPS spectra as it is the most abundant protein found in serum accounting for roughly 50% of total protein content [38], [39]. Feature selection should therefore be straightforward as the entirety of the fingerprint region from roughly 1700-900cm^{-1} could be said to represent changes in albumin concentration. Indeed, fingerprint regions in all cases displayed in Figure 5.1 are clearly highly specific suggesting multivariate analysis techniques are likely an effective tool for quantitative models. Inspecting each biomolecule spectrum individually it is clear there are individual peaks useful for specific qualitative and quantitative identification of each. To highlight this, spectra have been overlaid in Figure 5.2.(a), then overlapping regions have been omitted in Figure 5.2.(b) leaving only biomolecule-specific peaks. Tentative peak assignments, arrived at through extensive assessment of the literature, have been tabulated in Table 5.4. Figure 5.3.(a) – (d) shows the chemical structures of glucose, lactate, triglyceride, and an amide bond for reference. The full structure of albumin is too large to present here, so this amide bond which is prominent in the IR spectrum of albumin has been presented in its place. Salient band assignments have been highlighted in colours corresponding to analyte spectra of Figure 5.2. The wavenumbers for each feature have been listed two times; once for Di IREs and once for SIREs to account for disparities in peak position arising from anomalous dispersion (chapter 4 section 4.3.3.2). This information will be used to pinpoint the most useful regions for quantitation in the analysis of spiked serum

models. Peaks in the 3500-3000cm^{-1} region do not appear to possess sufficiently specific information for detailed analysis due to confounding signal from O-H vibrations of water or carbohydrates, and N-H vibrations from proteins [40]. However, an adjacent triplet of C-H stretch related peaks (acyl chains) in the 3000-2800cm^{-1} present strongly in the spectrum of triglyceride, and have been found useful in quantitative analysis of lipids elsewhere [41]. Consistent with the visual observations made previously, the fingerprint region is revealed to contain several analyte specific peaks (Figure 5.2.(b)). Vibrations of the carbonyl (C=O) group are highly specific for the identification of albumin, lactate, and triglycerides. In the case of albumin, amide I is primarily influenced by the C=O stretch and only weakly influenced by C-N and N-H bending vibrations [4]. Carbonyl groups are strong IR absorbers owing to the highly polar (large dipole moment) bonding environment. The reason the same chemical bond presents in three distinct wavenumber regions is because the carbonyl group's polarity is influenced by its neighbouring bonding environment [42], which in turn alters the dipole moment. For instance, the carbonyl band of esters (triglycerides) will be located at a higher wavenumber than carbonyl groups of carboxylic acids (lactate), and again of amines (albumin) because in each case the conjugates have weakened the C=O bond thereby altering vibrational response [33]. The C-O bonds attaching numerous hydroxyl (O-H) groups to the lactate and glucose structure as well as the C-H bands of triglycerides and albumin have many overlapping bands in the fingerprint region, too numerous to list in Table 5.4. It is likely that these wavenumbers will not be useful in univariate analysis, except in circumstances where these shared wavenumbers are overwhelmingly indicative of analyte presence (e.g., 1110cm^{-1} of C-O bonds in glucose).

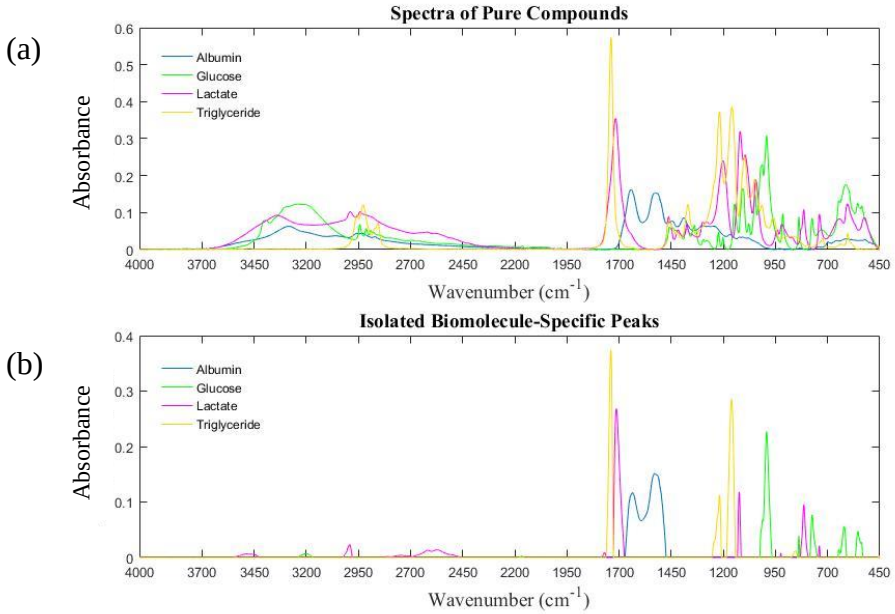


Figure 5.2. *Spectra of each biomolecule in their pure form (a) overlaid and (b) with overlapping bands removed to highlight unique peaks.*

Wavenumber (cm^{-1})		Assignments			
Di IRE	SIRE	Albumin	Glucose	Lactate	Triglycerides
3393	3393		ν OH	ν OH	
3285	3290	Amide A (ν NH)	ν OH	ν OH	
3250	3250		ν_s OH + ν_{as} OH		
3058	3065	Amide B			
2957	2956	ν_{as} CH_3	ν_s CH		ν_{as} CH_3
2990	2990			ν_s CH_3	
2943	2943		ν_s CH	ν_{as} CH_3	
2932	2935	ν_{as} CH_3			
2891	2890		ν_s CH	ν_s CH	
2872	2871	ν_s CH_3			ν_s CH_3
1738	1741				ν_s C=O
1717	1722			ν C=O	
1645	1650	Amide I (ν C=O + δ CN + δ NH)			
1527	1543	Amide II (ν CN + δ NH + δ CHN)			
1458	1458		δ CH_2 + δ OCH + δ CCH		δ_{as} CH_2
1455	1463	δ_s CH_2		δ_{as} CH_3	
1419	1420			δ CO + δ OH	δ_s CH_2
1380	1380		δ CH_2 + δ OCH + δ CCH	δ_s CH_3 + δ OH	
1270	1270	Amide III (ν CN + δ NH)			
1219	1220				ν_s CO
1159	1163				ν_s CO
1119	1123			ν CO + γCH_3	
1109	1110		ν CO		
1014	1020		ν CO + δ CO		
993	995		ν CO + ν CC		
775	775		ν CCO + ν CCH		

Table 5.4. *List of band assignments for each biomolecule. Band assignments have been highlighted if they are unique to glucose (green), lactate (pink), triglyceride (yellow), and albumin (blue). ν = stretching, δ = in-plane bending, γ = out-plane bending. References:* [4], [43]–[60].

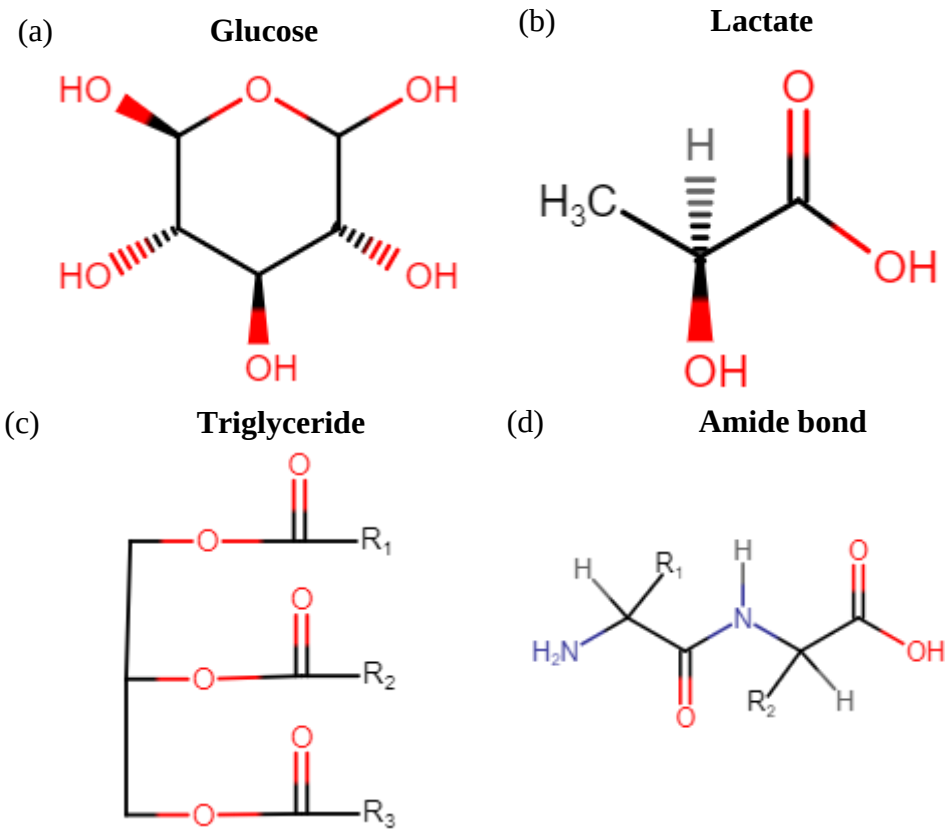


Figure 5.3. *Chemical structures of all four biomolecules analysed in the present study.*

5.3.2. PBS Concentration Studies

In all instances the concentration ranges were used to ascertain calibration curves and subsequently determine statistically relevant information (e.g., r^2, RMSE, Limit of Detection (LOD), and Limit of Quantification (LOQ)). Where possible the target biomolecule was analysed in its pure form, then spiked into either water or PBS at varying concentrations. This was to assess the ability to discern all levels of biomolecule concentration in a simple media before attempting analysis in more complex media (i.e., serum). It was hoped that information regarding prominent band assignments could be applied in aid of the analysis of spiked sera models. Unfortunately, several unforeseen complications were encountered, particularly for

the smaller molecular weight molecules: glucose and lactate. There is little or no detectable residue left on the IRE surface following evaporation of water or PBS solutions spiked with glucose or lactate even at high analyte concentrations. This is particularly troublesome in SIRE analysis due to the comparatively large sampling area which requires more analyte residue for discernible signal to be acquired. In the case of PBS, phosphate salts would crystallise on the IRE in seemingly random fashion which would interfere with spectra to an unpredictable degree due to the nonhomogeneous distribution of sample film (Figure 5.4.(a) and (b)). This made analysis difficult and as such has not been included in detailed analysis. In albumin spiked models these issues are not as prominent because the relatively higher concentration and large polar molecules decrease the contact angle and improve hydrophilicity hence promoting good sample coverage leading to better signal quality (Figure 5.5) [61].

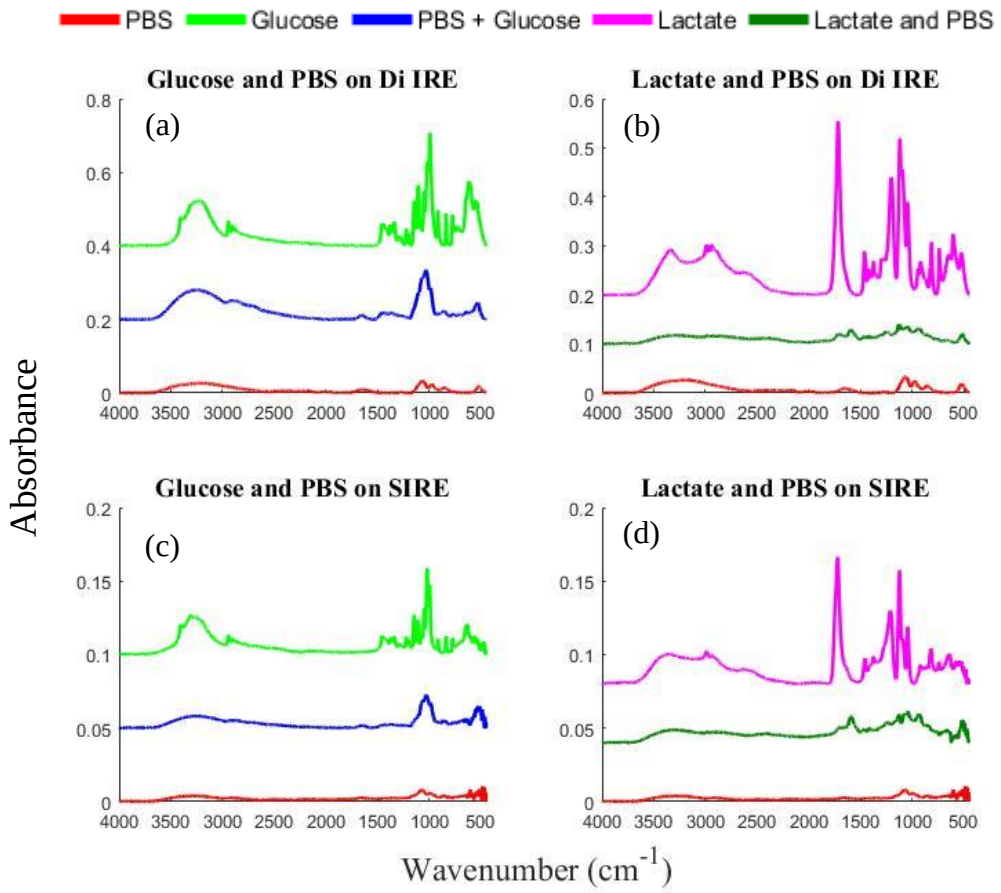


Figure 5.4. *Spectra of Glucose and lactate both in their pure form and in solution of PBS along with spectra of pure PBS. Spectra have been offset for clarity.*

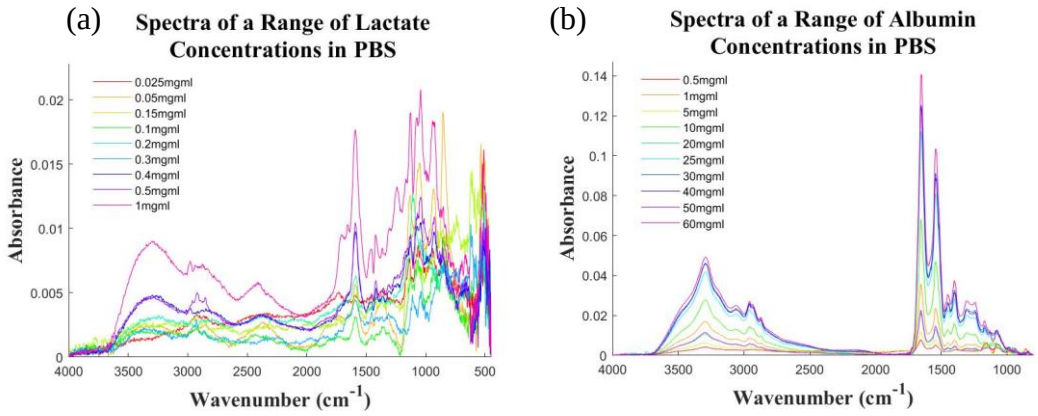


Figure 5.5. *Spectra of (a) lactate and (b) albumin spiked into PBS at a range of concentrations. It was difficult to get good quality spectra from the lactate samples because the low analyte concentrations and molecular weight caused samples to mostly evaporate upon drying. The concentration dependent spectral response is therefore clearer in the albumin group than the lactate group.*

5.3.3. Glucose Detection and Quantification

Figure 5.6 shows a step-by-step feature selection and linear regression model of glucose concentrations in spiked serum models both on SIREs and on a Di IRE. A 0.92mg/ml offset has been applied to account for the native glucose concentration present in the human pooled serum prior to spiking. There is agreement in SIRE and Di IRE analysed samples that a substantial spectral change is observed between the 2.1mg/ml (3.02mg/ml adjusted concentration) HPS spectrum and a blank (0.92 mg/ml adjusted concentration) HPS spectrum was in the 1000-1100cm^{-1} region. This arises from an increase in C-O present in the sample with a particularly pronounced signal at 1020cm^{-1} as noted previously in Table 5.4. This region was selected for area under the curve analysis and a linear regression model was constructed. Statistical findings have been reported in Table 5.5. Differences between the SIRE and Di IRE analysis are abundantly clear, with improved linearity and precision (higher r^2 and lower RMSE) observed in glucose concentration ranges analysed on the latter substrate. The regression profile shows that both IREs can reliably quantify high concentrations of glucose in HPS, but not mid-range or low concentrations as error increases toward lower range. This trend is exacerbated in the SIRE analysis potentially because of Si lattice absorption at 1107cm^{-1} leading to lower sample signal. The Di IRE is able to reliably detect, but not quantify clinically relevant ranges of glucose in serum. This is possibly due to the presence of large molecular weight proteins (e.g., albumin) that dominate the collected signature at low analyte concentrations [62]. Furthermore, this analysis shows that SIREs are not able to reliably detect nor quantify clinically relevant glucose concentrations using a simple univariate analysis technique. This is an entirely anticipated outcome, as it has already been made clear that SIREs have an analytical

weakness in this wavenumber range which explains the high level of variance across the SIRE concentration dataset as indicated by the wide error bars in Figure 5.6.(g). It is hoped that multivariate analysis may circumvent this shortcoming and can bolster the quantitative abilities of SIREs with regards to glucose analysis.

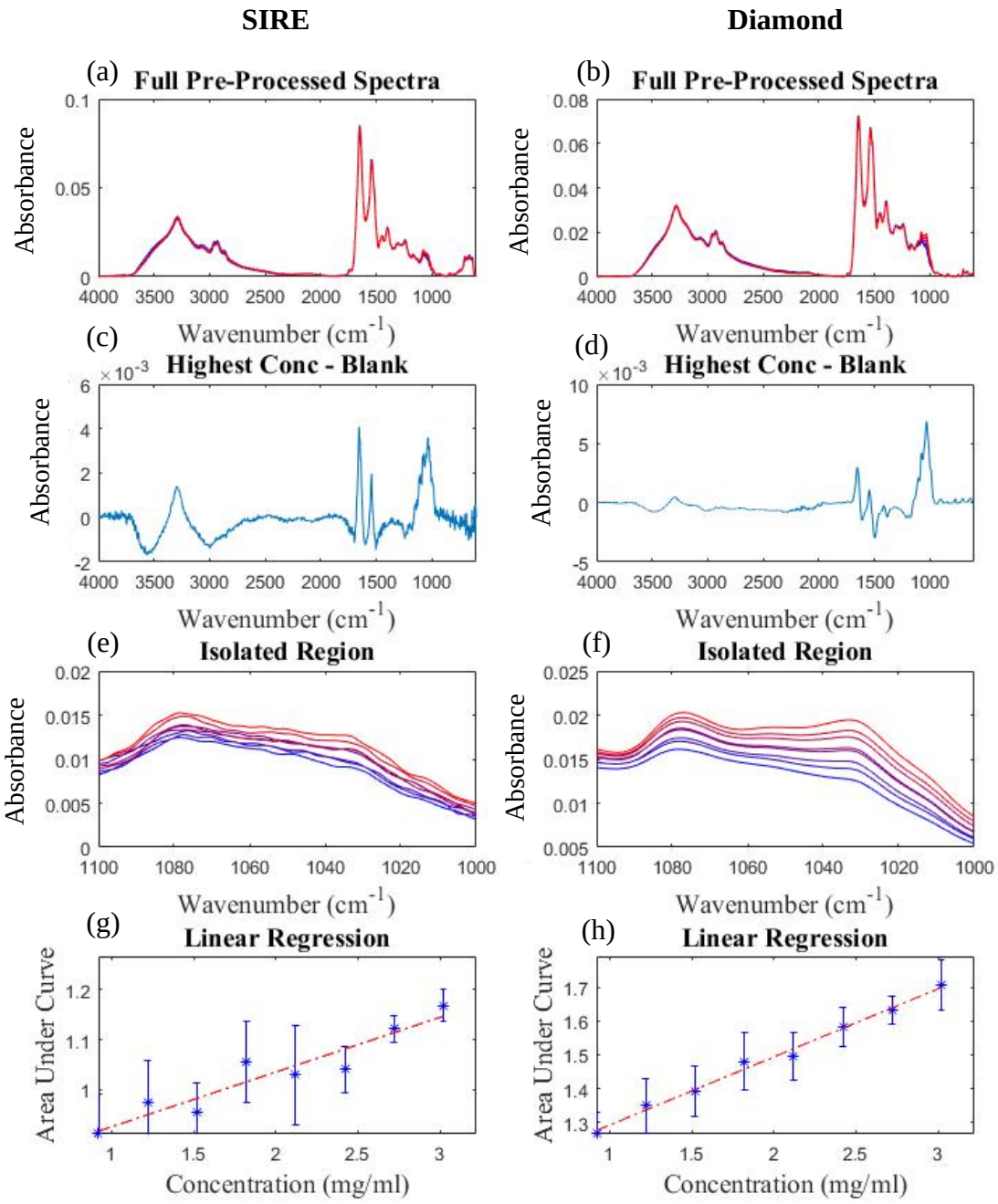


Figure 5.6. *Step-by-step quantification of glucose in spiked serum models measured on a SIRE (left) and on a Di IRE (right). (a) and (b) full pre-processed spectra averaged at each concentration level. (c) and (d) spectrum showing the where the greatest difference in absorption lies (1000-1100cm^{-1}) between the most heavily spiked samples and a control. (e) and (f) close up of the curve area where most spectral variance lies. (g) and (h) linear regression of the area under the curve against concentration of the spiked serum models. Error bars indicate a highly variable SIRE dataset and moderately variable Si IRE dataset.*

Table 5.5. *Statistical outputs of area under the curve versus glucose spiked HPS concentration linear regression model. Minimum and maximum clinical ranges of blood serum glucose concentrations have been added for reference.*

IRE Type	R^2	RMSE (mg/ml)	LOD (mg/ml)	LOQ (mg/ml)	Clinical Range (mg/ml)
Di	0.9904	0.0158	0.7116	2.3720	0.3-3.0
SIRE	0.8929	0.0300	2.5034	8.3447	

FTIR techniques can be quite sensitive to changes in glucose composition via multivariate approaches, even being able to determine glucose concentrations in whole blood [7], oral mucosa [63], and other interstitial fluids [64]. In each case, it is expressed that the use of single wavenumbers is not sufficient for appropriately sensitive infrared quantification of glucose in biological media.

A PLS model was constructed using the fingerprint region of the spectrum from 900-1500cm^{-1} so as to avoid the large signal of the Amide I and Amide II peaks interfering with the model by dominating the spectrum. Figure 5.7 shows outputs from the PLS model generation and Table 5.6 presents the corresponding statistical outputs. Plots on the left are from measurements taken using SIREs and plots on the right are from the Di IRE. In each case, the loadings show most variation to arise from C-O band vibrations in the 1000-1100cm^{-1} region consistent with earlier findings. Linearity of the regression model is marginally improved for Di IRE analysed samples, but greatly improved on the SIRE dataset as evidenced by an increase of r^2 value from 0.8929 to 0.9904. For both IRE types the RMSE value has increased indicating a decrease in model precision. Despite this, the RMSEV and standard deviation of the RMSEV is still lower than the smallest analyte concentration investigated meaning the model is suitably precise at the lower end of a clinically significant range. Both methods achieved a low LOD enabling reliable detection of glucose concentrations of

at least 0.3mg/ml. However, the LOQ achieved was 0.5342 and 0.8863mg/ml for the Di IRE and SIREs respectively, meaning concentrations can be reliably quantified within a clinically significant range, but not below these values. The datasets for SIRE and Di concentrations are less variable when using PLS in comparison to a univariate approach as indicated by the narrow error bars on the regression models (Figure 5.7.(g) and (h)). This error is consistent at all concentration levels for both IREs. Glucose has a low molecular weight and concentrations under investigation are also low so absorbance at discrete wavenumbers is likely low. Hence, by building models that are influenced by a broadband of wavenumbers the low signal can be detected even at very low concentrations.

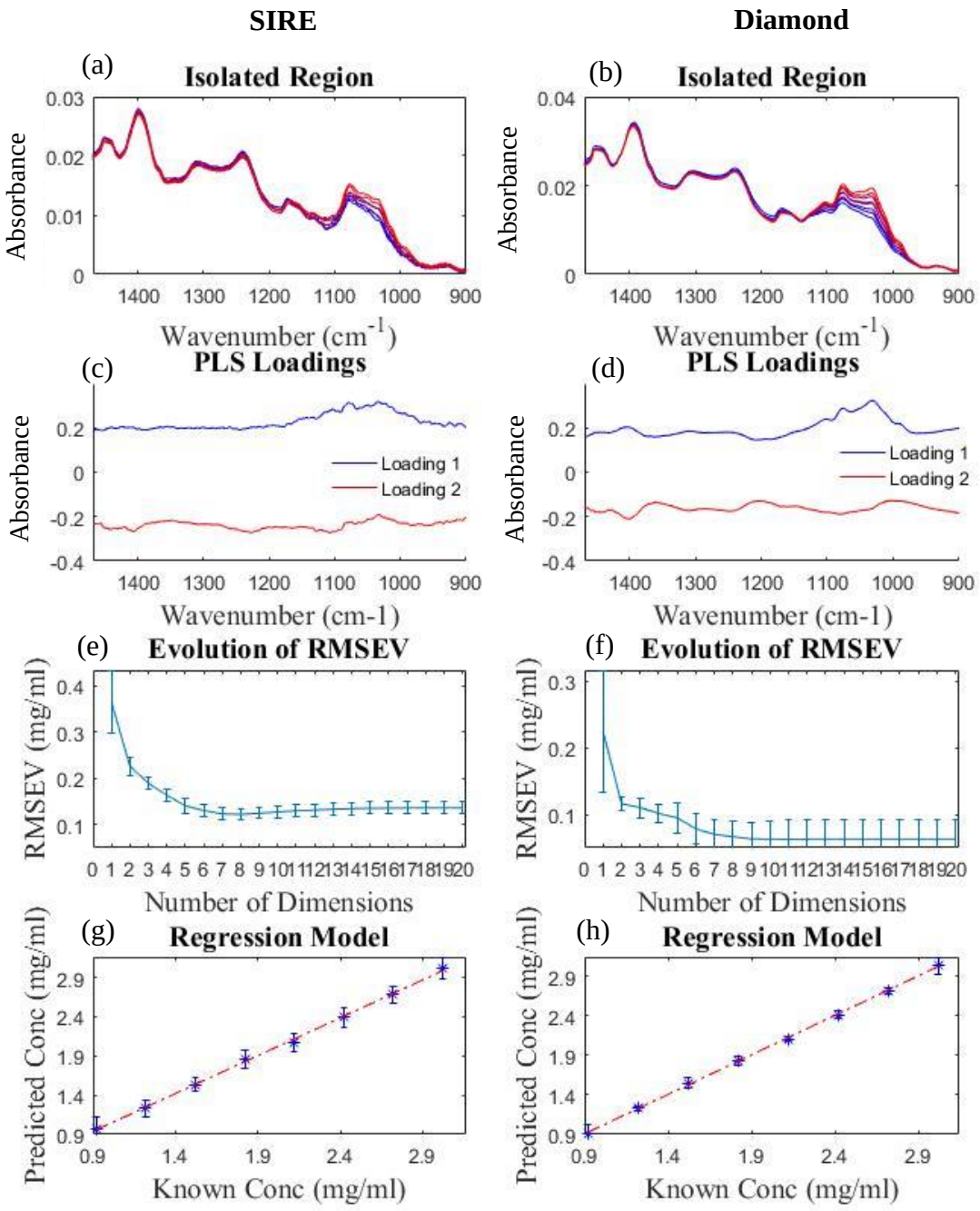


Figure 5.7. *Step-by-step quantification of glucose in spiked serum models measured on a SIRE (left) and on a Di IRE (right) by PLS method. (a) and (b) cut pre-processed spectra averaged at each concentration level. (c) and (d) vector of PLS loadings 1 and 2 indicating where the most variance in the spectra lies (1000-1100cm^{-1}). (e) and (f) evolution of the RMSEV with increasing number of dimensions in the PLS models. (g) and (h) PLS regression of the concentrations as predicted by the optimum PLS models and the known concentrations of the spiked serum models.*

Table 5.6. *Statistical outputs of PLS regression models for the prediction of glucose spiked serum models. Minimum and maximum clinical ranges of blood serum glucose concentrations have been added for reference.*

IRE Type	Dimensions No.	R^2	R^2 Std. Dev.	RMSEV (mg/ml)	RMSEV Std. Dev. (mg/ml)	LOD (mg/ml)	LOQ (mg/ml)	Clinical Range (mg/ml)
Di	11	0.9925	0.0070	0.0626	0.0290	0.1603	0.5342	0.3-3.0
SIRE	8	0.9702	0.0060	0.1219	0.0120	0.2659	0.8863	

Numerous studies have demonstrated the ability of FTIR spectroscopy to determine clinically accurate glucose concentrations [7], [9], [13], [39], [63]. It is difficult to directly compare the present findings to these studies as many predict concentrations of patient samples to model a closer likeness to a real clinical scenario. Furthermore, many of the studies employ additional techniques to enhance prediction capabilities such as improved sample preparation or instrumentation. It stands to reason that these techniques could be applied to SIREs to further improve prediction capabilities for glucose determination. The use of a Savitzky-Golay smoothing filter may be used to eliminate random fluctuations from baseline noise, enhancing analyte signal [7]. Bonnier *et al.* describe a method for eliminating the spectral signature of high-molecular weight molecules in serum by centrifuging samples in a 10kDa filter before analysis reporting a 5-fold improvement in RMSEV for glucose determination in patients [39]. Quantum Cascade Laser (QCL) mid-IR sources have also been used in the determination of glucose concentrations in patient samples with clinically accurate predictions down to 0.4mg/ml [9]. The added advantage of QCL is the ability to isolate specific wavenumber regions and the potential to adapt the technology for point-of-care diagnostics [64], [65]. Even without these advancements, it has been shown that SIREs, despite a hindered sensitivity to signal at 1107cm^{-1} where

carbohydrates are distinctly IR active, have comparable analytical capabilities to a Di IRE for the detection and quantification of glucose concentrations in serum.

5.3.4. Triglyceride Detection and Quantification

Triglycerides comprise three fatty acids adjoined via ester bonds to a glycerol backbone. Fatty acids themselves are hydrocarbon chains terminated by a carboxylic acid group and may vary in length from roughly 2-30 carbons. These features present three distinct bond interactions where analysis by vibrational spectroscopy is most likely. These are: C-O bonds (1150-1230cm^{-1}) binding fatty acids to glycerol within the O-C-O ester group formation, the C=O (1720-1770cm^{-1}) bond characteristic to ester groups, and the numerous C-H bonds (1420-1460cm^{-1} and 2700-3000cm^{-1}) lining the entirety of and terminating the fatty acid hydrocarbon chains. In addition to this, fatty acids might be present in saturated and unsaturated forms, a distinction which is detectable via IR spectroscopy through observation of =CH and C=C vibrations at 3008cm^{-1} and 1654cm^{-1} respectively [66][45]. In the present study, equal quantities of five short chain triglycerides were used with carbon chains measuring 2, 4, 6, 8, and 10 carbons (Triacetin, Tributyrin, Tricaproin, Tricaprylin, and Tricaprin). These are saturated as indicated by the lack signal contribution from the aforementioned double carbon bond peaks. Furthermore, due to the lower number of CH_2 bonds, the short chain hydrocarbons generate fairly weak signal within the 1420-1460cm^{-1} range [67]. This means that useful triglyceride predictors, while potent in the case of the carbonyl (C=O) stretch, are confined to limited spectral regions. Figures 5.8.(a) – (h) details feature selection and displays a linear regression model for an area under the curve analysis of triglyceride spiked serum models. A 1.05mg/ml offset has been applied to

account for native HPS triglyceride content. As before, the spectrum of the maximum additive concentration was subtracted from a blank to identify regions of greatest change (Figure 5.8.(c) and (d)). Unsurprisingly, the C=O band proved the most useful in ascertaining a linear relationship between concentration and spectral intensity. Reasonable r^2 values are obtained in both cases and RMSE values are low and within an acceptable range. LOD and LOQ, however, indicate neither method is suitable to detect or quantify clinical ranges of serum triglycerides. The regression profiles show the SIRE dataset to be more accurate, but less precise than the Di IRE dataset due to a high amount of error. Triglycerides are viscous and nonpolar and are therefore insoluble [68]. Consequently, consistent spreading of the spiked HPS samples across the SIRE sampling area is more difficult that on the smaller Di IRE crystal leading to greater variance. Nonpolar molecules are generally difficult to detect using IR spectroscopy. Nevertheless, a change in dipole moment should be possible and thus a signal can be obtained. Large overlapping error bars in both regression models suggest incorrect predictions could be made and unfortunately simple univariate feature analysis is not adequate in the analysis of triglycerides in this case.

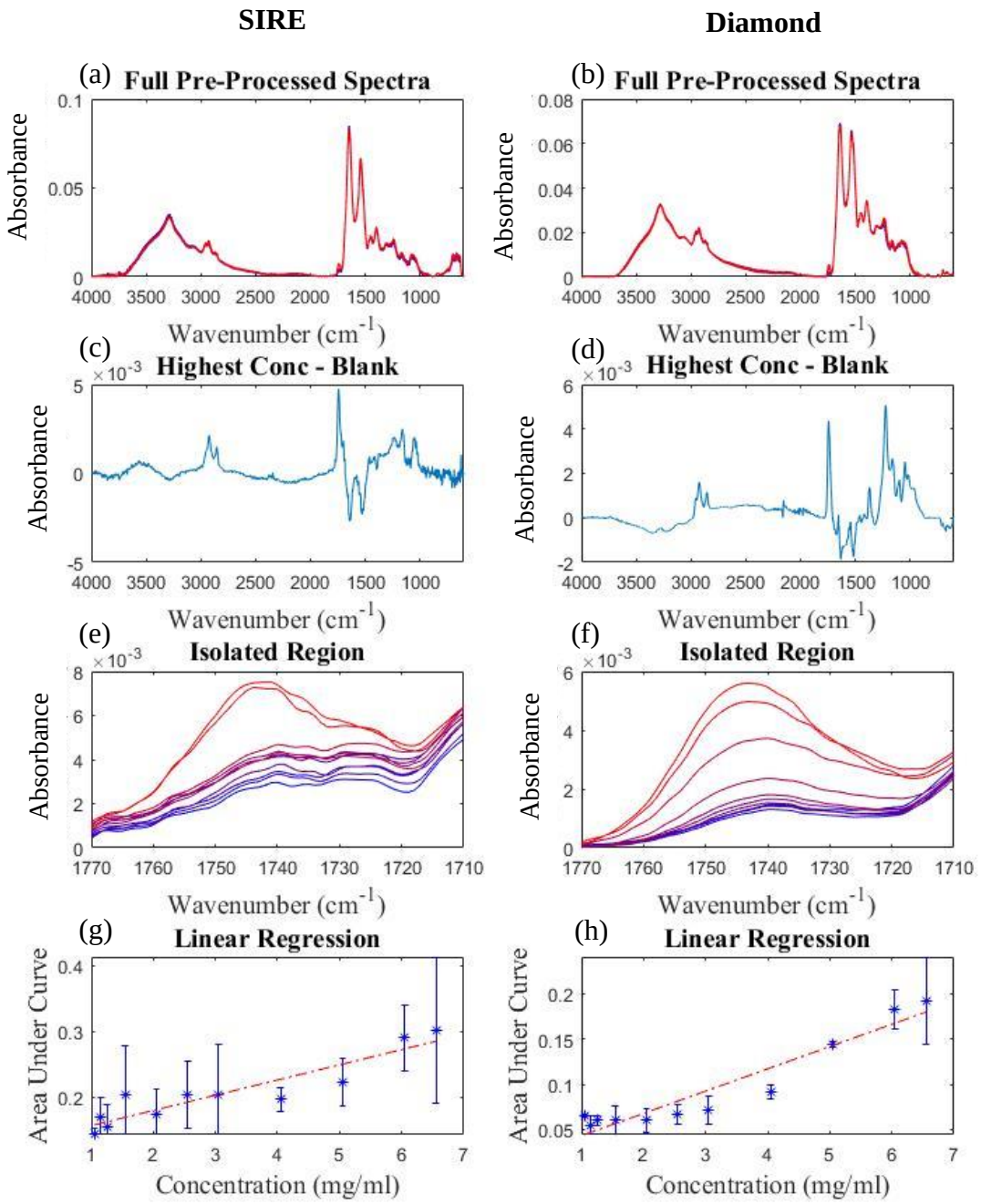


Figure 5.8. *Step-by-step quantification of triglyceride in spiked serum models measured on a SIRE (left) and on a Di IRE (right). (a) and (b) full pre-processed spectra averaged at each concentration level. (c) and (d) spectrum showing the where the greatest difference in absorption lies (1710-1770cm^{-1}) between the most heavily spiked samples and a control. (e) and (f) close up of the curve area where most spectral variance lies. (g) and (h) linear regression of the area under the curve against concentration of the spiked serum models. Error bars show the data to be highly variable across all concentrations using both SIREs and the Di IRE.*

Table 5.7. *Statistical outputs of area under the curve versus triglyceride spiked HPS concentration linear regression model. Minimum and maximum clinical ranges of blood serum triglyceride concentrations have been added for reference.*

IRE Type	R^2	RMSE (mg/ml)	LOD (mg/ml)	LOQ (mg/ml)	Clinical Range (mg/ml)
Di	0.9078	0.0167	3.8750	12.9167	1.5-5.0
SIRE	0.8455	0.0209	5.1968	17.3226	

Multivariate analysis was carried out following the described method (Figure 5.9. and Table 5.8). PLS models were constructed using the 900-1770cm^{-1} wavenumber region so as to include the C=O band as there is clearly valuable information to be gleaned despite poor outcomes in the previous univariate analysis. This is confirmed by the slight peaks observed in both loadings one and two of SIRE and Di IRE PLS models at the C=O band (1740cm^{-1}) (Figure 5.9.(c) and (d)). There also seems to be variance arising from the Amide I and II peaks as well as the region encapsulating the C-O bands at 1150-1230cm^{-1}. R^2 values show good linearity, but standard deviation of r^2 indicates the PLS simulation is more repeatable over the 100 iterations for samples analysed on Di. The model is fitted well to the data as indicated by low RMSEV values and RMSEV standard deviations comparable to those found in the literature [69]. Narrow error bars of Di IRE measurements are indicative of good model precision. Conversely, some overlapping error bars indicate lower precision in the SIRE PLS model. This difference in error between the two IREs can be attributed to the need to collect data across multiple SIREs instead of just one Di IRE. For both substrates LOD is within a suitable range to detect the lower end of clinically relevant quantities, however LOQ is within the intermediate clinical range so potentially still useful for the screening of at-risk patients where triglyceride content will preside in the higher end of the clinical range.

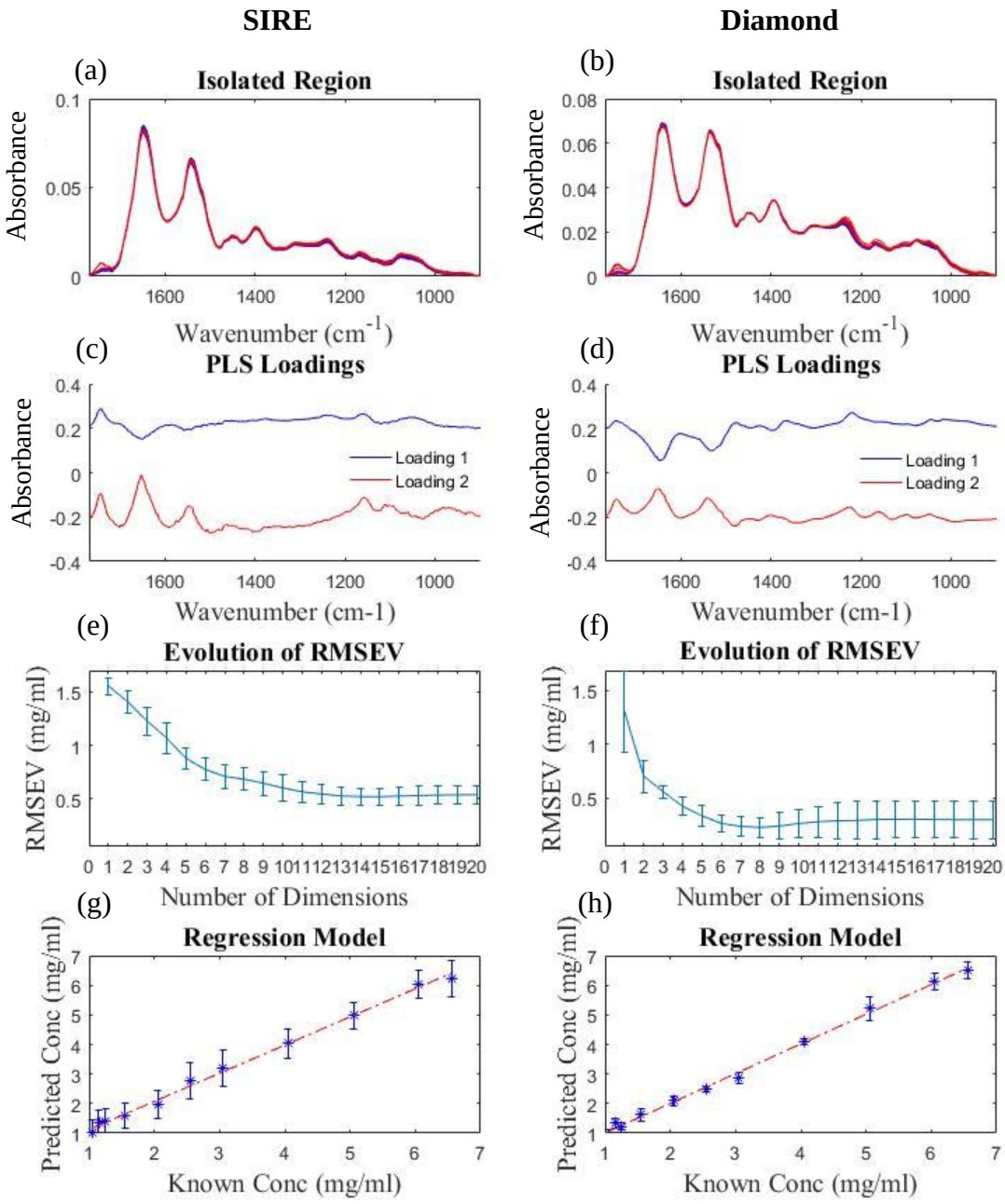


Figure 5.9. *Step-by-step quantification of triglyceride in spiked serum models measured on a SIRE (left) and on a Di IRE (right) by PLS method. (a) and (b) cut pre-processed spectra averaged at each concentration level. (c) and (d) vector of PLS loadings 1 and 2 indicating where the most variance in the spectra lies (1710-1770cm^{-1}). (e) and (f) evolution of the RMSEV with increasing number of dimensions in the PLS models. (g) and (h) PLS regression of the concentrations as predicted by the optimum PLS models and the known concentrations of the spiked serum models. Error bars indicated both datasets are reproducible, but more so in the case of Di IRE concentrations.*

Table 5.8. *Statistical outputs of PLS regression models for the prediction of triglyceride spiked serum models. Minimum and maximum clinical ranges of blood serum triglyceride concentrations have been added for reference.*

IRE Type	Dimensions No.	R^2	R^2 Std. Dev.	RMSEV (mg/ml)	RMSEV Std. Dev. (mg/ml)	LOD (mg/ml)	LOQ (mg/ml)	Clinical Range (mg/ml)
Di	8	0.9876	0.0083	0.2278	0.0928	0.7011	2.3371	1.5-5.0
SIRE	14	0.9251	0.0244	0.5308	0.0898	0.7611	2.5371	

Detection of clinically relevant concentrations of triglycerides in human pooled serum was achieved through multivariate analysis. Precision of PLS models need to be improved to realise clinically acceptable prediction capabilities. There is clearly room for this improvement given the patently observable spectral changes emanating from strong IR absorption of the ester carbonyl group. As with other low molecular weight fraction components of serum, triglycerides may also be more accurately determined through ultracentrifugation of the bulk serum sample [70]. The gold standard for lipid determination is a time-intensive multi-step ultracentrifugation method, presenting the opportunity for a simpler IR spectroscopy based method to exploit existing workflows while surpassing analysis times [71]. In the present study, the reason for poor predictive capabilities is probably due to mistakenly repeating freeze thaw cycles of the triglyceride fraction as this has been shown to severely impact lipid content [72], [73] as well as serum integrity as a whole for FTIR analysis [74]. This highlights another reason why sample reanalysis of SIREs would hold such a significant benefit over conventional IREs as samples can be revisited if a patient stock serum has been refrozen, although suitable storage methods still need to be considered to prevent degradation of the dry serum films.

5.3.5. Lactate Detection and Quantification

From a spectroscopic perspective, lactate exhibits both lipid and carbohydrate associated traits and is therefore of great interest to test the dissimilarities between SIREs and Di IREs. Having numerous hydroxyl groups bound to carbon (C-O) as well as various C-H bonds gives rise to fingerprint region activity with a likeness to that of glucose [75], while the carbonyl group produces a band characteristic of lipids. As mentioned previously (section 5.3.1), the peak shift of carbonyl groups due to the altering of bond energy of different conjugate forms is the hallmark of discrimination between carbonyl containing compounds [33]. Unfortunately, the carbonyl peak of lactate is closely aligned with that of the Amide I peak meaning it is difficult to resolve in serum samples. Observing a combination of peaks is obviously the wisest strategy, but for arguments sake both regions were analysed using the area under the curve. Of the two regions, 1110-1140cm^{-1} showed greatest differences between the highest HPS lactate concentration and a HPS blank (Figure 5.10.(b) and (d)) and provided a much better regression model than the alternative 1710-1770cm^{-1} region. Full step-by-step analysis process for the former is shown in Figure 5.10. No linear relationship between lactate concentration and area under the curve could be found using either SIREs or Di IREs from this wavenumber range. Table 5.9 displays the statistical findings for area under the curve regression analysis of the 1110-1140cm^{-1} region. In this case, the native lactate content of the serum was not known, so no offset has been applied. The spectra of SIRE samples can be seen to dip as silicon oxide attenuation begins to take over the absorbance profile at 1107cm^{-1} (Figures 5.10.(e) and (f)). Model fit and linearity is better on Di IREs compared to SIREs, and Di IREs are able to detect clinically significant concentrations of Lactate. Neither approach is suitable to quantify

clinically significant concentrations at this stage. However, it is elevated lactate levels that are of clinical interest, not low in which case only the Di IRE exhibits good predictive qualities. Both show a lot of error at low concentration ranges. Like glucose, lactate is low molecular weight and at low concentrations discrete absorbance bands will be weak. Multivariate analysis of a broadband IR spectral region will again be required.

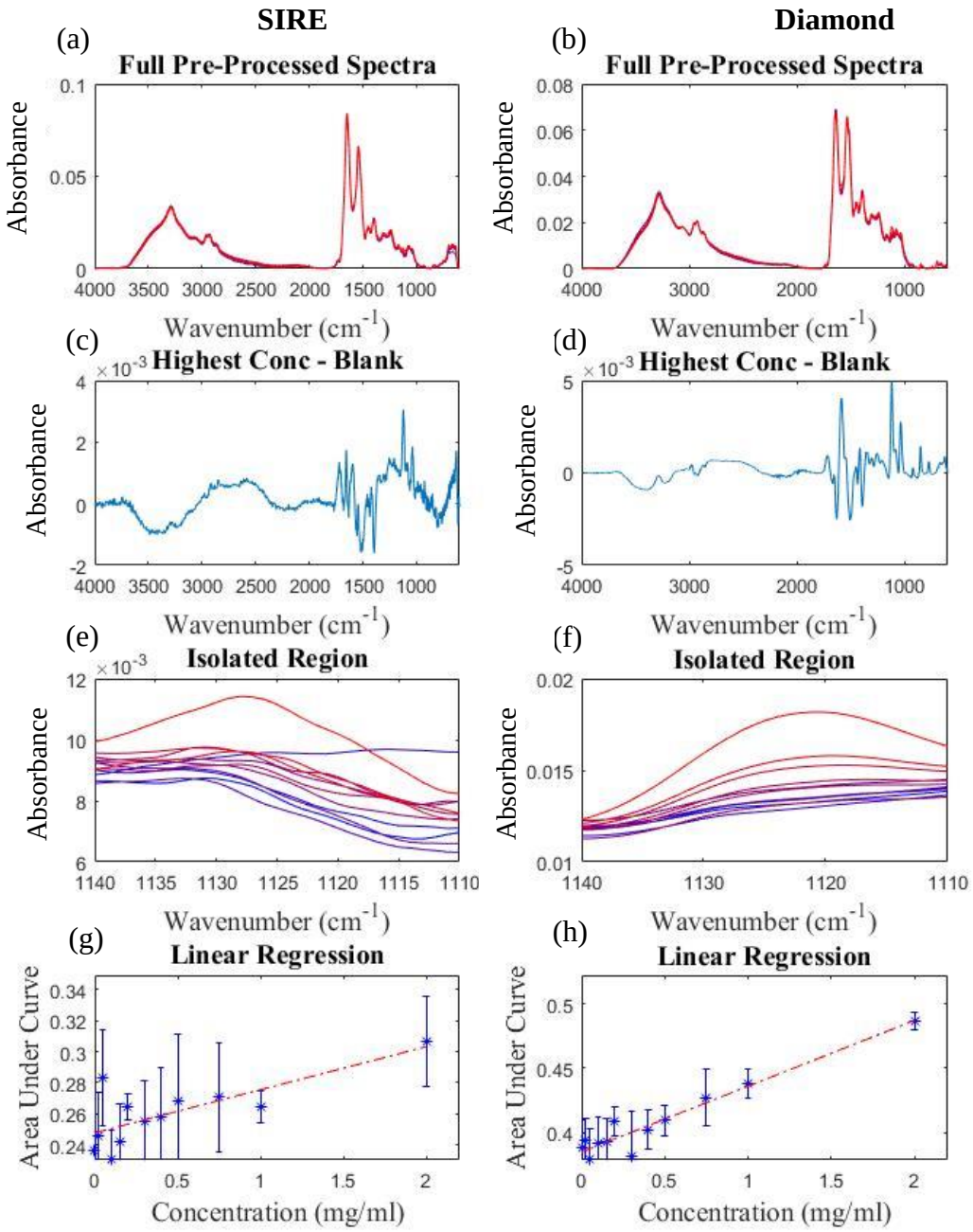


Figure 5.10. *Step-by-step quantification of lactate in spiked serum models measured on a SIRE (left) and on a Di IRE (right). (a) and (b) full pre-processed spectra averaged at each concentration level. (c) and (d) spectrum showing the where the greatest difference in absorption lies (1110-1140cm^{-1}) between the most heavily spiked samples and a control. (e) and (f) close up of the curve area where most spectral variance lies. (g) and (h) linear regression of the area under the curve against concentration of the spiked serum models. Error bars indicate highly variable SIRE data and moderately variable Di IRE data.*

Table 5.9. *Statistical outputs of area under the curve versus lactate spiked HPS concentration linear regression model. Minimum and maximum clinical ranges of blood serum lactate concentrations have been added for reference.*

IRE Type	R^2	RMSE (mg/ml)	LOD (mg/ml)	LOQ (mg/ml)	Clinical Range (mg/ml)
Di	0.9293	0.0085	0.6481	2.1602	0.25-1.5
SIRE	0.5681	0.0146	2.0492	6.8305	

PLS regression found a vast improvement in model linearity and fit to the data (Figure 5.11, Table 5.10). The 900-1500cm^{-1} region was selected to exclude influence from the Amide bands. The carbonyl band is too coincident with the Amide I band to be properly distinguished. Both SIRE and Di IRE PLS loadings reveal the greatest variation is located at 1040cm^{-1} and 1120cm^{-1} corresponding to the C-C and C-H + C-O bands respectively (Figure 5.11.(c) and (d)) [58]. In contrast to area under the curve regression, the linearity achieved from both IRE substrates is excellent with good r^2 values and standard deviation of the r^2 value. RMSEV is low and reproducible over the PLS iterations indicated by a low standard deviation of RMSEV. The regression profile is consistent at all concentration levels for both IREs with low error as indicated by narrow error bars. The LOD and LOQ is within clinically significant ranges which is further expressed by the narrow error bars of the regression model although the model becomes unreliable below roughly 0.2mg/ml. This would be acceptable as pathology is always indicated by elevated lactate levels. 0.2mg/ml (0.22mmol/L) would typically be a very low serum lactate concentration [17]. This is a very optimistic finding as this low detection and quantification capability has been demonstrated on a low molecular weight target in whole serum. Furthermore, it alludes to a sensitivity of ATR-FTIR spectroscopy as a powerful tool in the detection of pathologies where fluctuating lactate levels are a key predictor such as in sepsis diagnosis [76].

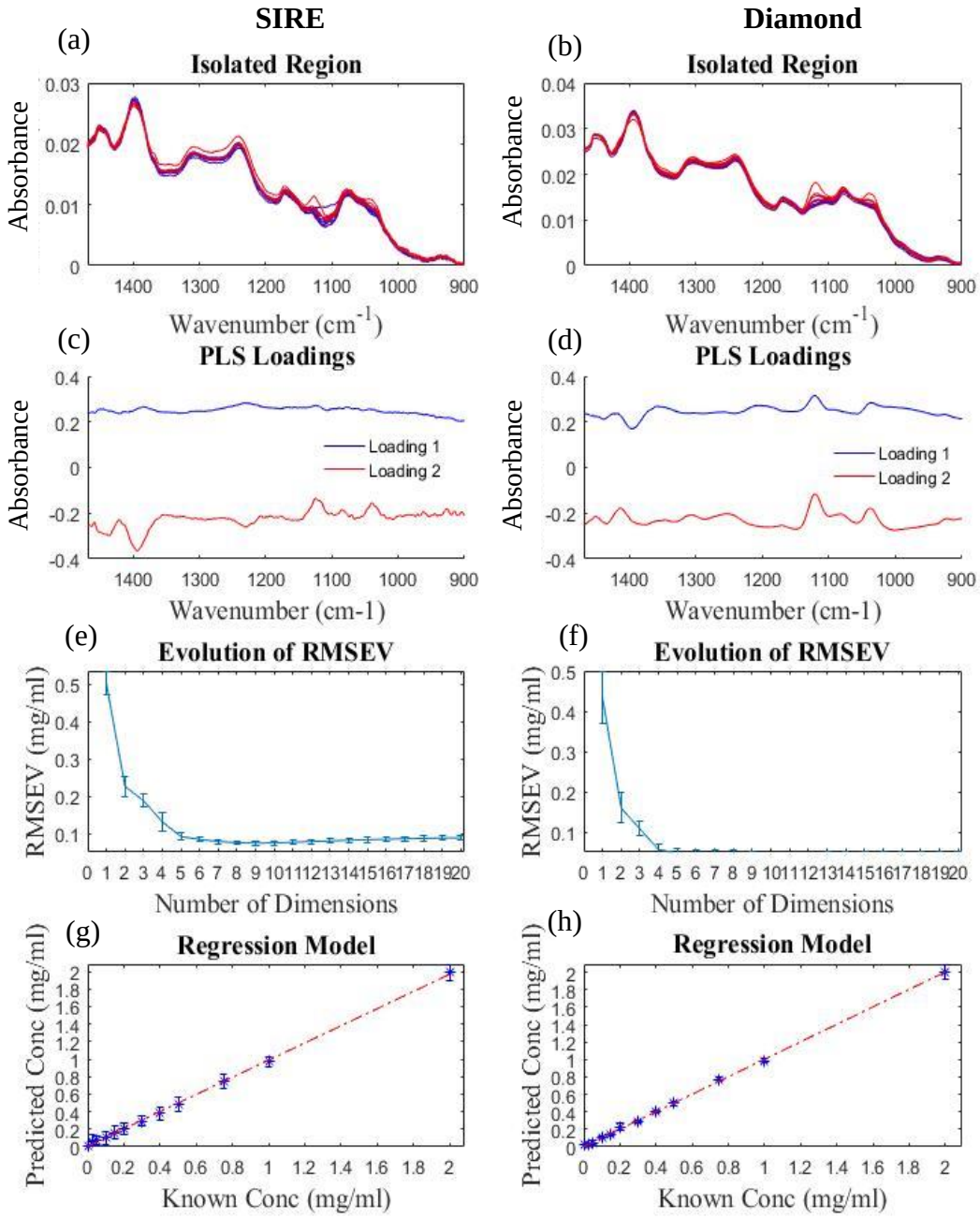


Figure 5.11. *Step-by-step quantification of lactate in spiked serum models measured on a SIRE (left) and on a Di IRE (right) by PLS method. (a) and (b) cut pre-processed spectra averaged at each concentration level. (c) and (d) vector of PLS loadings 1 and 2 indicating where the most variance in the spectra lies (1110-1140cm^{-1}). (e) and (f) evolution of the RMSEV with increasing number of dimensions in the PLS models. (g) and (h) PLS regression of the concentrations as predicted by the optimum PLS models and the known concentrations of the spiked serum models. Error bars indicated low error for both SIREs and Di IREs.*

Table 5.10. *Statistical outputs of PLS regression models for the prediction of lactate spiked serum models. Minimum and maximum clinical ranges of blood serum lactate concentrations have been added for reference.*

IRE Type	Dimensions No.	R^2	R^2 Std. Dev.	RMSEV (mg/ml)	RMSEV Std. Dev. (mg/ml)	LOD (mg/ml)	LOQ (mg/ml)	Clinical Range (mg/ml)
Di	11	0.9959	0.0040	0.0395	0.0198	0.0527	0.1757	0.25-1.5
SIRE	9	0.9817	0.0037	0.0742	0.0050	0.0637	0.2122	

Lactate is the smallest analyte under investigation in the present study (89.07Da) [77]. Despite this, determination of lactate in serum at clinically relevant concentration levels is achievable through a rapid, simple, and easy to operate high-throughput FTIR method. Fractionating the serum specimens would likely improve regression model characteristics, but this may not be a desirable method in practice as classification of disease states may benefit from the presence of high molecular weight serum components. For example, this would eliminate C-reactive protein (120kDa) from the sample which is an important indicator of sepsis and would likely play a key role in influencing machine learning algorithms for the development of classification regimes [78]–[80]. Simultaneous determination of relative concentrations of lactate and other chemical species in serum and also in histological studies using second derivative spectra has been demonstrated elsewhere [81]. This method could improve diagnosis and monitoring of highly metabolic physiological states through concurrent measurements of lactate and other serum components, such as CRP or glucose in sepsis monitoring which relate to the different phases of sepsis progression [16]. The r^2 values and RMSEV values achieved here are comparable to those found in the literature for commercially available hospital lactate sensors and alternative FTIR approaches, suggesting SIREs are capable of delivering clinically acceptable accuracies for the determination of lactate [82]–[84].

5.3.6. Albumin Detection and Quantification

Univariate analysis of albumin proved difficult (Figure 5.12.(a) – (h)). Spectral profile of pure albumin is complex, visually appearing identical to a pure HPS spectrum. The largest difference between spiked and blank HPS samples seems to occur at the Amide I peak (Figure 5.12.(c) and (d)), which is expected as it is a peak frequently selected for detailed protein analysis [85]–[88]. The Amide I peak was selected at different wavenumbers to account for differences in peak position due to anomalous dispersion effects and the area under the curve calculated (Figure 5.12(e) and (f)). A linear regression model was fitted to the data (Figure 5.12.(g) and (h)). The r^2 value for SIREs is low but reveal some kind of relationship between the Amide I peak and Albumin concentration. A linear relationship could not be found using the Di IRE, however. In both cases the LOD and LOQ are negligible as the data is too uncorrelated to make these estimations as indicated by the large spread of error bars at each data point. The solubility limit of Albumin in water is 60mg/ml. The maximum level included in the analysis slightly exceeded this value at 65.5mg/ml, although it is not known what the exact albumin offset is for these experiments. If only a little albumin was properly dissolved into the HPS stock solution, then this could be the reason for the poor linearity observed in this experiment. Furthermore, the hydrophilicity of SIREs due to Piranha pre-treatment may vastly improve albumin surface adsorption in comparison to the Di IRE hence enhancing IR absorption characteristics [89].

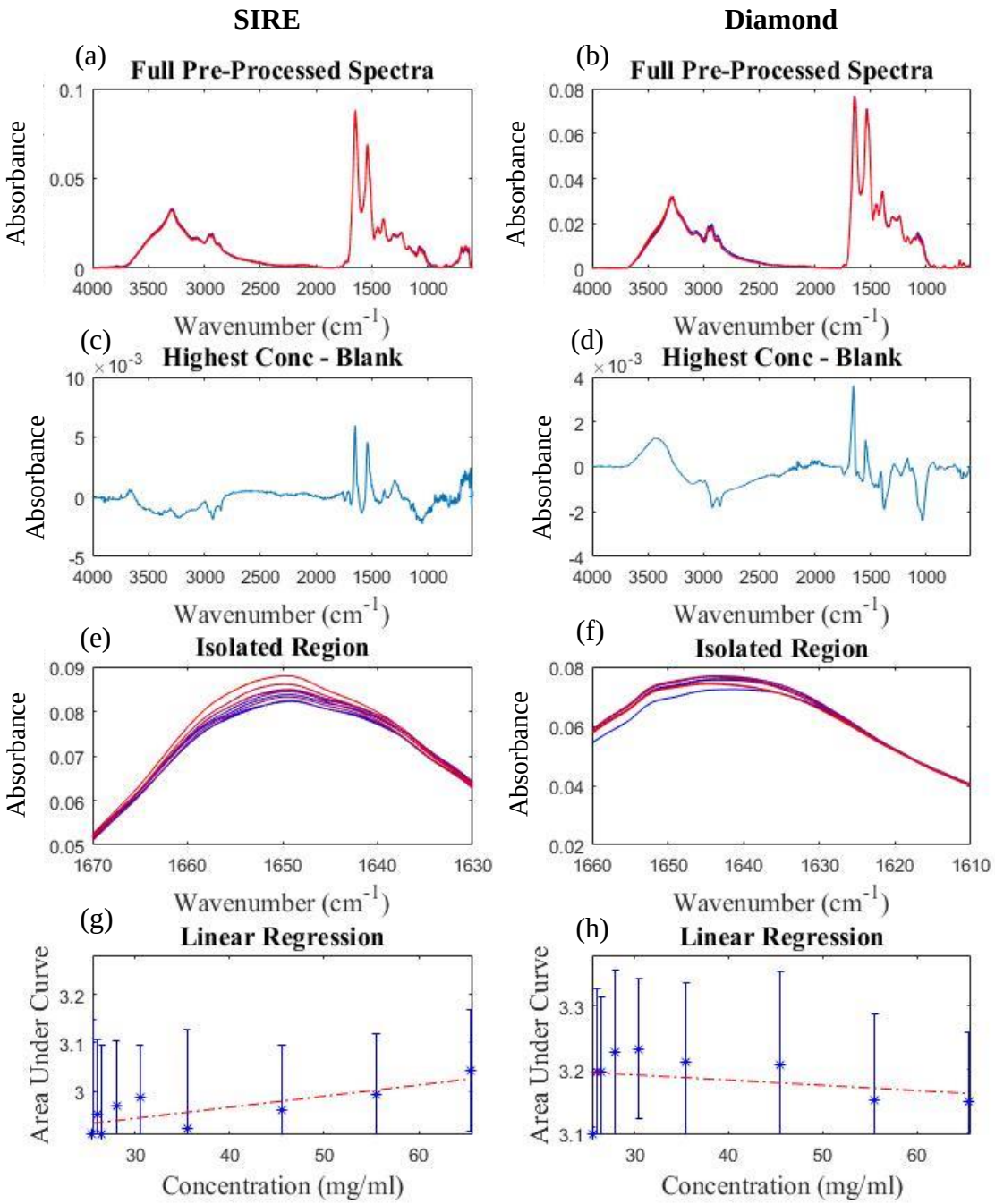


Figure 5.12. *Step-by-step quantification of albumin in spiked serum models measured on a SIRE (left) and on a Di IRE (right). (a) and (b) full pre-processed spectra averaged at each concentration level. (c) and (d) spectrum showing where the greatest difference in absorption lies (1630-1670cm^{-1}) between the most heavily spiked samples and a control. (e) and (f) close up of the curve area where most spectral variance lies. (g) and (h) linear regression of the area under the curve against concentration of the spiked serum models. Error bars indicate a high error across both datasets.*

Table 5.11. *Statistical outputs of area under the curve versus albumin spiked HPS concentration linear regression model. Minimum and maximum clinical ranges of blood serum albumin concentrations have been added for reference.*

IRE Type	R^2	RMSE (mg/ml)	LOD (mg/ml)	LOQ (mg/ml)	Clinical Range (mg/ml)
Di	0.0747	0.0452	-479.7348	-1599.1158	25.0-50.0
SIRE	0.6021	0.0292	110.7798	369.2660	

PLS regression was more successful, achieving r^2 values of 0.9394 and 0.9823 for Di and SIRE analysed samples respectively. Table 5.12 and Figure 5.13 show regression models and accompanying statistics for albumin spiked serum models. As expected, the loadings for each case reveal the Amide I and II peaks to contribute the most variance across the datasets (Figure 5.13.(c) and (d)). Since albumin concentration is the only physiological parameter changed between samples, it is assumed that variations in amide I and II bands is due to changes in albumin content, not other proteins. Model linearity and RMSEV of spiked sample sets measured on SIREs is comparable to values found in the literature [90], [91]. R^2 standard deviation is lower in the SIRE group than in the Di IRE group by just under a factor of 10. RMSEV and RMSEV standard deviation are also considerably lower indicating it is more likely for concentrations to be determined unambiguously using SIREs. This is reflected by the narrow error bars observed in the SIRE dataset (Figure 5.13.(g) and (h)). LOD and LOQ are within clinically acceptable ranges for both IRE substrates, but significantly better for analysis performed on SIREs. Both elevated and lowered albumin levels indicate pathology, so precision across the entire range is desirable. Albumin concentrations lower than 10mg/ml is rare, with one article claiming only 0.1% of their patient cohort fell below this value, and values below 25mg/ml would also be considered very low [92]. It is hopeful then that high-throughput SIRE ATR-

FTIR could predict serum albumin concentrations at both extremes of the clinical range given an LOQ of ~22mg/ml.

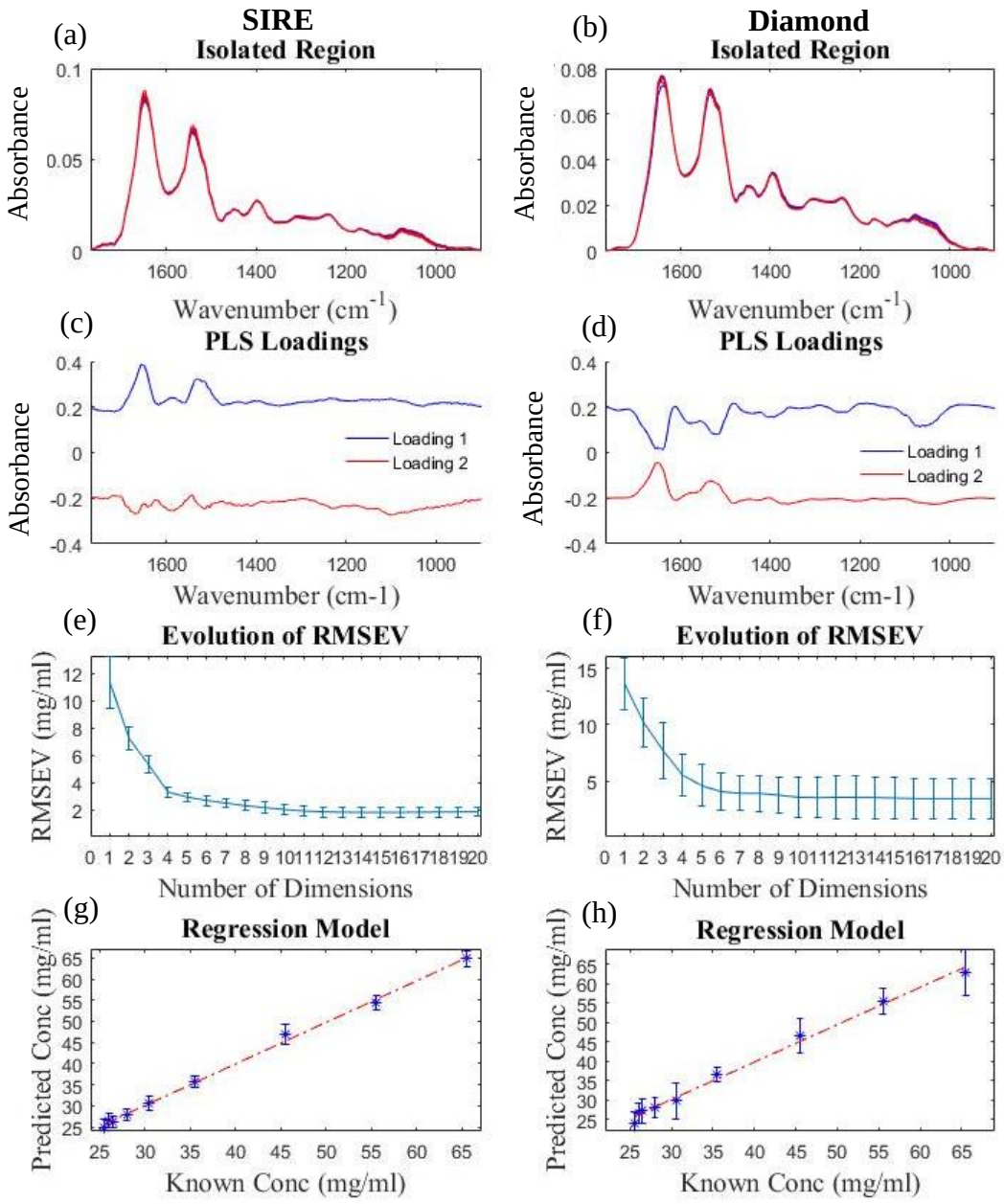


Figure 5.13. *Step-by-step quantification of albumin in spiked serum models measured on a SIRE (left) and on a Di IRE (right) by PLS method. (a) and (b) cut pre-processed spectra averaged at each concentration level. (c) and (d) vector of PLS loadings 1 and 2 indicating where the most variance in the spectra lies (1630-1670cm^{-1}). (e) and (f) evolution of the RMSEV with increasing number of dimensions in the PLS models. (g) and (h) PLS regression of the concentrations as predicted by the optimum PLS models and the known concentrations of the spiked serum models. Error bars indicated a low level of error in the SIRE dataset and a moderate level of error in the Di IRE dataset.*

Table 5.12. *Statistical outputs of PLS regression models for the prediction of albumin spiked serum models. Minimum and maximum clinical ranges of blood serum albumin concentrations have been added for reference.*

IRE Type	Dimensions No.	R^2	R^2 Std. Dev.	RMSEV (mg/ml)	RMSEV Std. Dev. (mg/ml)	LOD (mg/ml)	LOQ (mg/ml)	Clinical Range (mg/ml)
Di	19	0.9394	0.0713	3.4377	1.8014	11.0659	36.8862	25.0-50.0
SIRE	15	0.9823	0.0080	1.8256	0.3503	6.6147	22.0491	

Albumin determination is routinely carried out and there are several methods utilised in clinical laboratories to quantify albumin such as Bromocresol Green (BCG), Bromocresol Purple (BCP), and immunological assays [32]. All of these use reagents and are comparatively slow and expensive with respect to ATR-FTIR which requires no reagents and minimal sample preparation. BCG in particular is a popular method that suffers from decreased specificity if other serum proteins are present in large quantities [93]. FTIR determination of serum proteins has a similar limitation in that the overwhelming concentration of high molecular weight proteins can saturate the IR absorbance. There are thousands of protein species contained within a serum sample, yet albumin and globulin account for 57-71% and 8-26% of the total mass respectively [90]. Measurement linearity has been shown to be compromised at very high analyte concentrations (>25mg/ml) due to saturation of the ATR-FTIR spectra, although this effect is compensated by vector normalisation [62], [94]. Vector normalisation equalizes the spectral contributions of serum constituents and is suitable for enhancing systematic spectral changes incurred by varying analyte concentration [94]. Thus, peak normalisation approaches (e.g., min-max normalisation) may be better suited to this type of study where a single physiological parameter is being modulated since spectral peaks not associated with the analyte can effectively serve as an internal standard [39], [95]. However, peak normalisation may not translate from spike studies to patient

studies well where the general form of the spectra cannot be assumed to be consistent across patients [39].

For serum investigations of protein species it would undoubtedly be beneficial to dilute samples before analysis to avoid saturation of IR absorbance, particularly to investigate spectral features coincident with the Amide bands where saturation due to high albumin content is likely [90]. In regards to protein quantification specifically, it has been found that serum samples may be diluted down to 10% serum in deionised water to improve predictive capabilities of different protein species [90]. To further improve specificity and sensitivity, SIRE surfaces could be functionalised with target specific receptors similar to antibody immobilisation widely used in immunoassays [96]. This has already been demonstrated with promising detection limits (10^{-12} M of biotin) on hydrophobic silicon and germanium ATR crystals by grafting a thin protein-repelling organic layer (e.g., covalently immobilized DNA probes) that can selectively bind target molecules and repel unwanted proteins [97]–[99]. Excess sample is then washed away before analysis. This comes with the caveat that the surface modifications are only stable for a matter of hours and the evanescent wave must penetrate the organic layer before interrogating the bound molecules. Background scans would account for any interference, but spectral intensity could be marginally impacted. Translating this development to SIREs presents the opportunity for low-cost and highly selective ATR slides.

5.4. Conclusion

It was demonstrated that all analytes under consideration could be quantified at clinically relevant concentrations in complex media (serum) using SIREs except glucose which could not be quantified at the lower end of the clinical range using either SIREs or a Di IRE. Triglycerides can be quantified in clinically relevant concentrations as disease is indicated by high triglyceride concentrations. However, physiologically normal concentrations of triglycerides could not be reliably quantified using either approach. For each analyte, SIREs performed with comparable or better precision and accuracy as a gold standard Di IRE provided the right analytical approaches are used. There is the added caveat that a different SIRE was used for every concentration measured in contrast to the same Di IRE used for all experiments. This should be considered for a balanced view of how these IRE substrates perform relative to each other. Regardless, SIREs have proved to have predictive capabilities competitive with a gold standard Di IRE provided the right analytical approaches are used whilst also enabling rapid sample analysis. SIREs fill the unresolved limitations of conventional ATR-FTIR spectroscopy that inhibit translation into the clinical laboratory by permitting implementation of low-cost, automated, and high-throughput ATR clinical workflows. Trials carried out on SIREs consumed a third of the time and allowed all samples to be prepared and scanned under the same conditions since samples are dried simultaneously and relative humidity fluctuates from day to day. Compared to the Di IRE, SIREs exhibited an appreciable improvement in performance using both univariate and multivariate approaches to quantify albumin in HPS possibly due to enhanced surface properties of SIREs that enable surface adsorption. Univariate

analysis of glucose and lactate concentrations using spectral features coincident with the intrinsic absorption in SIREs by SiO at 1107cm^{-1} revealed SIREs to have impaired performance. In both cases, this shortcoming was overcome by using multivariate analysis using subtle spectral perturbations across wide wavenumber ranges to quantify the biomolecules. Multivariate methods are more likely to be implemented in real time clinical spectroscopy approaches as pathologies alter serum biochemistry with greater complexity than can be fully described by a stand-alone biomarker.

All analytes are associated to some degree with the diagnosis and prognosis of sepsis, severe sepsis, and septic shock. Since all these biomolecules are typically determined through use of target specific reagents or enzymes, FTIR holds the benefit that it is possible to detect and quantify biomolecules in serum simultaneously, but further investigations are required to determine if this is possible using SIREs [100]. Having shown each can be well resolved at clinically relevant concentrations within spiked serum models it is evident that sepsis could be diagnosed and classified using SIREs. However, spiked serum models are not adequate to represent the huge variation in physiological conditions found across real patient samples. This has motivated a 148 patient study to investigate if SIREs can discriminate between healthy patients, patients with sepsis, and patients with Systematic Inflammatory Response Syndrome (SIRS). The findings of this study shall be discussed in the next chapter.

5.5. References

[1] J. R. Hands *et al.*, "Attenuated Total Reflection Fourier Transform Infrared (ATR-FTIR) spectral discrimination of brain tumour severity from serum samples," *J. Biophotonics*, vol. 7, no. 3–4, pp. 189–199, 2014, doi:

10.1002/jbio.201300149.

[2] H. Fabian, P. Lasch, M. Boese, and W. Haensch, “Mid-IR microspectroscopic imaging of breast tumor tissue sections,” *Biopolymers*, vol. 67, no. 4–5, pp. 354–357, Jan. 2002, doi: 10.1002/bip.10088.

[3] M. A. Cohenford and B. Rigas, “Cytologically normal cells from neoplastic cervical samples display extensive structural abnormalities on IR spectroscopy: Implications for tumor biology,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 95, no. 26, pp. 15327–15332, 1998, doi: https://doi.org/10.1073/pnas.95.26.15327.

[4] Z. Movasaghi, S. Rehman, and I. U. Rehman, “Fourier transform infrared (FTIR) spectroscopy of biological tissues,” *Appl. Spectrosc. Rev.*, vol. 43, no. 5, pp. 134–179, 2008, doi: 10.1080/05704928.2016.1230863.

[5] P. B. Gahan, “Circulating nucleic acids in plasma and serum: Diagnosis and prognosis in cancer,” *EPMA J.*, vol. 1, no. 3, pp. 503–512, 2010, doi: 10.1007/s13167-010-0021-6.

[6] H. M. Heise, “Glucose Measurements by Vibrational Spectroscopy,” in *Handbook of Vibrational Spectroscopy*, 2006, pp. 3280–3294.

[7] Y. C. Shen, A. G. Davies, E. H. Linfield, T. S. Elsey, P. F. Taday, and D. D. Arnone, “The use of fourier-transform infrared spectroscopy for the quantitative determination of glucose concentration in whole blood,” *Phys. Med. Biol.*, vol. 48, no. 13, pp. 2023–2032, 2003, doi: 10.1088/0031-9155/48/13/313.

[8] H. Hirasawa, S. Oda, and M. Nakamura, “Blood glucose control in patients with severe sepsis and septic shock,” *World J. Gastroenterol.*, vol. 15, no. 33, pp. 4132–4136, 2009, doi: 10.3748/wjg.15.4132.

[9] S. Liakat, K. A. Bors, T.-Y. Huang, A. P. M. Michel, E. Zanghi, and C. F. Gmachl, “In vitro measurements of physiological glucose concentrations in biological fluids using mid-infrared light,” *Biomed. Opt. Express*, vol. 4, no. 7, pp. 1083–1090, 2013, doi: 10.1364/boe.4.001083.

[10] B. Wernly, M. Lichtenauer, U. C. Hoppe, and C. Jung, “Hyperglycemia in septic patients: an essential stress survival response in all, a robust marker for risk stratification in some, to be messed with in none,” *J. Thorac. Dis.*, vol. 8, no. 7, pp. E621–E624, 2016, doi: 10.21037/jtd.2016.05.24.

[11] Z. Dai, A. Yang, X. Bao, and R. Yang, “Facile Non-Enzymatic Electrochemical Sensing,” *Sensors*, vol. 19, no. 2824, pp. 1–13, 2019.

[12] J. McMillin, “Chapter 141: Blood Glucose,” in *Clinical Methods: The History, Physical, and Laboratory Examinations. 3rd edition*, 1990, pp. 662–665.

[13] J. Krinsley, K. Bochicchio, C. Calentine, and G. Bochicchio, “Glucose measurement of intensive care unit patient plasma samples using a fixed-wavelength mid-infrared spectroscopy system,” *J. Diabetes Sci. Technol.*, vol. 6, no. 2, pp. 294–301, 2012, doi: 10.1177/193229681200600212.

[14] A. M. Amorini *et al.*, “Serum lactate as a novel potential biomarker in multiple sclerosis,” *Biochim. Biophys. Acta - Mol. Basis Dis.*, vol. 1842, no. 7, pp. 1137–

1143, 2014, doi: 10.1016/j.bbadis.2014.04.005.

[15] M. L. Goodwin, J. E. Harris, M. Ed, A. Hernández, L. B. Gladden, and D. Ph, "Blood Lactate Measurements and Analysis during Exercise: A Guide for Clinicians," *J. Diabetes Sci. Technol.*, vol. 1, no. 4, pp. 558–569, 2007.

[16] S. Y. Cho and J. H. Choi, "Biomarkers of Sepsis," *Crit. Rev. Clin. Lab. Sci.*, vol. 50, no. 1, pp. 23–36, 2013, doi: 10.3947/ic.2014.46.1.1.

[17] R. Bou Chebl *et al.*, "Serum lactate is an independent predictor of hospital mortality in critically ill patients in the emergency department: A retrospective study," *Scand. J. Trauma. Resusc. Emerg. Med.*, vol. 25, no. 69, pp. 1–7, 2017, doi: 10.1186/s13049-017-0415-8.

[18] J. Bakker, M. W. N. Nijsten, and T. C. Jansen, "Clinical use of lactate monitoring in critically ill patients," *Ann. Intensive Care*, vol. 3, no. 1, pp. 1–8, 2013, doi: 10.1186/2110-5820-3-12.

[19] S. M. Lee and W. S. An, "New clinical criteria for septic shock: Serum lactate level as new emerging vital sign," *J. Thorac. Dis.*, vol. 8, no. 7, pp. 1388–1390, 2016, doi: 10.21037/JTD.2016.05.55.

[20] P. Wacharasint, T. A. Nakada, J. H. Boyd, J. A. Russell, and K. R. Walley, "Normal-range blood lactate concentration in septic shock is prognostic and predictive," *Shock*, vol. 38, no. 1, pp. 4–10, 2012, doi: 10.1097/SHK.0b013e318254d41a.

[21] M. S. El-Sayed, K. P. George, and K. Dyson, "The influence of blood sampling site on lactate concentration during submaximal exercise at 4 mmol/l lactate level," *Eur. J. Appl. Physiol. Occup. Physiol.*, vol. 67, no. 6, pp. 518–522, 1993, doi: 10.1007/BF00241648.

[22] G. Yuan, K. Z. Al-Shali, and R. A. Hegele, "Hypertriglyceridemia: Its etiology, effects and treatment," *Can. Med. Assoc. J.*, vol. 176, no. 8, pp. 1113–1120, 2007, doi: 10.1503/cmaj.060963.

[23] A. Cetinkaya *et al.*, "Is hypertriglyceridemia a prognostic factor in sepsis?," *Ther. Clin. Risk Manag.*, vol. 10, no. 1, pp. 147–150, 2014, doi: 10.2147/TCRM.S57791.

[24] P. P. Toth, "Triglyceride-rich lipoproteins as a causal factor for cardiovascular disease," *Vasc. Health Risk Manag.*, vol. 12, pp. 171–183, 2016, doi: 10.2147/VHRM.S104369.

[25] S. G. Klotzsch and J. R. Mcnamara, "Triglyceride Measurement: a Review of Methods and Interferences," *Clin. Chem.*, vol. 36, no. 9, pp. 1605–1613, 1990.

[26] T. P. Carr, C. J. Andresen, and L. L. Rudel, "Enzymatic Determination of Triglyceride, Free Cholesterol and Total Cholesterol in Tissue Lipid Extracts," *Clin. Biochem.*, vol. 26, pp. 39–42, 1993.

[27] G. Weaving, G. F. Batstone, and R. G. Jones, "Age and sex variation in serum albumin concentration: an observational study," *Ann. Clin. Biochem.*, vol. 53, no. 1, pp. 106–111, 2016, doi: 10.1177/0004563215593561.

[28] M. Padkins, T. Breen, N. Anavekar, G. Barsness, K. Kashani, and J. C. Jentzer, "Association Between Albumin Level and Mortality Among Cardiac Intensive Care Unit Patients," *J. Intensive Care Med.*, pp. 1–8, 2020, doi: 10.1177/0885066620963875.

[29] M. Yin *et al.*, "Predictive Value of Serum Albumin Level for the Prognosis of Severe Sepsis Without Exogenous Human Albumin Administration: A Prospective Cohort Study," *J. Intensive Care Med.*, vol. 33, no. 12, pp. 687–694, 2018, doi: 10.1177/0885066616685300.

[30] G. A. Kaysen, "Albumin turnover in renal disease," *Miner. Electrolyte Metab.*, vol. 24, no. 1, pp. 55–63, 1997, doi: 10.1159/000057351.

[31] V. Gounden and I. Jialal, *Hypoalbuminemia*. 2018.

[32] D. Kumar and D. Banerjee, "Methods of albumin estimation in clinical biochemistry: Past, present, and future," *Clin. Chim. Acta*, vol. 469, pp. 150–160, 2017, doi: 10.1016/j.cca.2017.04.007.

[33] N. Mobaraki and B. Hemmateenejad, "Structural characterization of carbonyl compounds by IR spectroscopy and chemometrics data analysis," *Chemom. Intell. Lab. Syst.*, vol. 109, no. 2, pp. 171–177, 2011, doi: 10.1016/j.chemolab.2011.08.011.

[34] W. M. Haschek, C. G. Rousseaux, and M. A. Wallig, "Clinical Pathology," in *Fundamentals of Toxicologic Pathology*, Elsevier, 2010, pp. 43–65.

[35] J. T. Busher, "Serum Albumin and Globulin," in *Clinical Methods: The History, Physical, and Laboratory Examinations*, 1990, pp. 497–499.

[36] V. Szentirmai *et al.*, "Reagent-free total protein quantification of intact extracellular vesicles by attenuated total reflection Fourier transform infrared (ATR-FTIR) spectroscopy," *Anal. Bioanal. Chem.*, vol. 412, pp. 4619–4628, 2020, doi: 10.1007/s00216-020-02711-8.

[37] L. Miloudi *et al.*, "ATR-IR coupled to partial least squares regression (PLSR) for monitoring an encapsulated active molecule in complex semi-solid formulations," *Analyst*, vol. 143, no. 10, pp. 2377–2389, 2018, doi: 10.1039/c8an00547h.

[38] G. S. Smith, G. L. Walter, and R. M. Walker, "Clinical Pathology in Non-Clinical Toxicology Testing," in *Haschek and Rousseaux's Handbook of Toxicologic Pathology, Third Edition*, Third Edit., Elsevier, 2013, pp. 565–594.

[39] F. Bonnier *et al.*, "Ultra-filtration of human serum for improved quantitative analysis of low molecular weight biomarkers using ATR-IR spectroscopy," *Analyst*, vol. 142, pp. 1285–1298, 2017, doi: 10.1039/c6an01888b.

[40] A. Galus *et al.*, "Hexagonal-shaped chondroitin sulfate self-assemblies have exalted anti-HSV-2 activity," *Carbohydr. Polym.*, vol. 136, pp. 113–120, 2016, doi: 10.1016/j.carbpol.2015.08.054.

[41] V. Shapaval *et al.*, "Biochemical profiling, prediction of total lipid content and fatty acid profile in oleaginous yeasts by FTIR spectroscopy," *Biotechnol.*

Biofuels, vol. 12, no. 140, pp. 1–12, 2019, doi: 10.1186/s13068-019-1481-0.

[42] A. Barth, "Infrared spectroscopy of proteins," *Biochim. Biophys. Acta*, vol. 1767, no. 9, pp. 1073–1101, 2007, doi: 10.1016/j.bbabio.2007.06.004.

[43] M. Ibrahim, M. Alaam, H. El-Haes, A. F. Jalbout, and A. De Leon, "Analysis of the structure and vibrational spectra of glucose and fructose," *Eclet. Quim.*, vol. 31, no. 3, pp. 15–21, 2006, doi: 10.1590/S0100-46702006000300002.

[44] K. Brandenburg and U. Seydel, "Vibrational Spectroscopy of Carbohydrates and Glycoconjugates," in *Handbook of Vibrational Spectroscopy*, 2006, pp. 3481–3507.

[45] A. Rohman, "Infrared spectroscopy for quantitative analysis and oil parameters of olive oil and virgin coconut oil: A review," *Int. J. Food Prop.*, vol. 20, no. 7, pp. 1447–1456, 2017, doi: 10.1080/10942912.2016.1213742.

[46] F. Bonnier, M. J. Baker, and H. J. Byrne, "Vibrational spectroscopic analysis of body fluids: Avoiding molecular contamination using centrifugal filtration," *Anal. Methods*, vol. 6, no. 14, pp. 5155–5160, 2014, doi: 10.1039/c4ay00891j.

[47] D. R. Parachalil *et al.*, "Analysis of bodily fluids using vibrational spectroscopy: A direct comparison of Raman scattering and infrared absorption techniques for the case of glucose in blood serum," *Analyst*, vol. 144, no. 10, pp. 3334–3346, 2019, doi: 10.1039/c9an00125e.

[48] M. Huleihel, A. Salman, V. Erukhimovich, J. Ramesh, Z. Hammody, and S. Mordechai, "Novel optical method for study of viral carcinogenesis in vitro.," *J. Biochem. Biophys. Methods*, vol. 50, pp. 111–121, 2002.

[49] K. Forfang, B. Zimmermann, G. Kosa, A. Kohler, and V. Shapaval, "FTIR spectroscopy for evaluation and monitoring of lipid extraction efficiency for oleaginous fungi," *PLoS One*, vol. 12, no. 1, pp. 1–17, 2017, doi: 10.1371/journal.pone.0170611.

[50] W. G. De Ruig, "Infrared spectra of monoacid triglycerides; with some applications to fat analysis," *Agric. Res. Reports*, vol. 759, no. 7, p. 117, 1971.

[51] J. Vongsvivut, M. R. Miller, D. McNaughton, P. Heraud, and C. J. Barrow, "Rapid Discrimination and Determination of Polyunsaturated Fatty Acid Composition in Marine Oils by FTIR Spectroscopy and Multivariate Data Analysis," *Food Bioprocess Technol.*, vol. 7, no. 8, pp. 2410–2422, 2014, doi: 10.1007/s11947-013-1251-0.

[52] H. Fabian, M. Jackson, L. Murphy, P. H. Watson, I. Fichter, and H. H. Mantsch, "A Comparative Infrared Spectroscopic Study of Human Breast Tumors and Breast Tumor Cell Xenografts," *Biospectroscopy*, vol. 1, pp. 37–45, 1995.

[53] F. Kucuk Baloglu, S. Garip, S. Heise, G. Brockmann, and F. Severcan, "FTIR imaging of structural changes in visceral and subcutaneous adiposity and brown to white adipocyte transdifferentiation," *Analyst*, vol. 140, no. 7, pp. 2205–2214, 2015, doi: 10.1039/c4an02008a.

[54] J. N. Murdock and D. L. Wetzel, "FT-IR microspectroscopy enhances

biological and ecological analysis of algae," *Appl. Spectrosc. Rev.*, vol. 44, no. 4, pp. 335–361, 2009, doi: 10.1080/05704920902907440.

[55] C. Paluszkiewicz and W. M. Kwiatek, "Analysis of human cancer prostate tissues using FTIR microspectroscopy and SRIXE techniques," *J. Mol. Struct.*, vol. 565–566, pp. 329–334, 2001, doi: 10.1016/S0022-2860(01)00527-0.

[56] R. Bandikari, V. Poondla, and V. S. R. Obulam, "Enhanced production of xylanase by solid state fermentation using Trichoderma koeningi isolate: effect of pretreated agro-residues," *3 Biotech*, vol. 4, no. 6, pp. 655–664, 2014, doi: 10.1007/s13205-014-0239-4.

[57] M. F. K. Fung, M. K. Senterman, N. Z. Mikhael, S. Lacelle, and P. T. T. Wong, "Pressure-tuning fourier transform infrared spectroscopic study of carcinogenesis in human endometrium," *Biospectroscopy*, vol. 2, no. 3, pp. 155–165, 1996, doi: 10.1002/(sici)1520-6343(1996)2:3<155::aid-bspy2>3.3.co;2-f.

[58] G. Cassanas, M. Morssli, E. Fabregue, and L. Bardet, "Vibrational spectra of lactic acid and lactates," *J. Raman Spectrosc.*, vol. 22, no. 7, pp. 409–413, 1991, doi: 10.1002/jrs.1250220709.

[59] D. Borchman, R. Stimmelmayr, and J. C. George, "Whales, lifespan, phospholipids, and cataracts," *J. Lipid Res.*, vol. 58, no. 12, pp. 2289–2298, 2017, doi: 10.1194/jlr.m079368.

[60] T. Ube, Y. Yoneyama, and T. Ishiguro, "In situ Measurement of the pH-dependent Transmission Infrared Spectra of Aqueous Lactic Acid Solutions," *Anal. Sci.*, vol. 33, no. 12, pp. 1395–1400, 2017, doi: 10.2116/analsci.33.1395.

[61] F. G. Tseng, C. Y. Huang, C. C. Chieng, H. Huang, and C. S. Liu, "Size effect on surface tension and contact angle between protein solution and silicon compound, PC, and PMMA substrates," *Microscale Thermophys. Eng.*, vol. 6, no. 1, pp. 31–53, 2002, doi: 10.1080/108939502753428220.

[62] H. J. Byrne, F. Bonnier, J. McIntyre, and D. R. Parachalil, "Quantitative analysis of human blood serum using vibrational spectroscopy," *Clin. Spectrosc.*, vol. 2, no. March, p. 100004, 2020, doi: 10.1016/j.clispe.2020.100004.

[63] R. Kasahara, S. Kino, S. Soyama, and Y. Matsuura, "Noninvasive glucose monitoring using mid-infrared absorption spectroscopy based on a few wavenumbers," *Biomed. Opt. Express*, vol. 9, no. 1, pp. 289–302, 2018, doi: 10.1364/boe.9.000289.

[64] M. Pleitez, H. Von Lilienfeld-Toal, and W. Mäntele, "Infrared spectroscopic analysis of human interstitial fluid in vitro and in vivo using FT-IR spectroscopy and pulsed quantum cascade lasers (QCL): Establishing a new approach to non invasive glucose measurement," *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.*, vol. A, no. 85, pp. 61–65, 2012, doi: 10.1016/j.saa.2011.09.007.

[65] A. Schwaighofer, M. Brandstetter, and B. Lendl, "Quantum cascade lasers (QCLs) in biomedical spectroscopy," *Chem. Soc. Rev.*, vol. 46, pp. 5903–5924,

2017, doi: 10.1039/C7CS00403F.

[66] V. Shapaval, N. K. Afseth, G. Vogt, and A. Kohler, “Fourier transform infrared spectroscopy for the prediction of fatty acid profiles in Mucor fungi grown in media with different carbon sources,” *Microb. Cell Fact.*, vol. 13, no. 86, pp. 1–11, 2014, doi: 10.1186/1475-2859-13-86.

[67] R. N. Jones, “the Effects of Chain Length on the Infrared Spectra of Fatty Acids and Methyl Esters,” *Can. J. Chem.*, vol. 40, no. 2, pp. 321–333, 1961, doi: 10.1139/v62-050.

[68] R. A. Cox and M. R. García-Palmieri, “Cholesterol, Triglycerides, and Associated Lipoproteins,” in *Clinical Methods: The History, Physical, and Laboratory Examinations*, 1990, pp. 153–160.

[69] D. Perez-Guaita, J. Kuligowski, G. Quintás, S. Garrigues, and M. D. La Guardia, “Modified locally weighted - Partial least squares regression improving clinical predictions from infrared spectra of human serum samples,” *Talanta*, vol. 107, pp. 368–375, 2013, doi: 10.1016/j.talanta.2013.01.035.

[70] K. Sato, M. Seimiya, Y. Kodera, A. Kitamura, and F. Nomura, “Application of Fourier-transform infrared (FT-IR) spectroscopy for simple and easy determination of chylomicron-triglyceride and very low density lipoprotein-triglyceride,” *Clin. Chim. Acta*, vol. 411, no. 3–4, pp. 285–290, 2010, doi: 10.1016/j.cca.2009.11.026.

[71] E. Schaefer, F. Tsunoda, M. Diffenderfer, E. Polisecki, N. Thai, and B. Asztalos, “Lipoprotein Assessment by Ultracentrifugation,” in *The Measurement of Lipids, Lipoproteins, Apolipoproteins, Fatty Acids, and Sterols, and Next Generation Sequencing for the Diagnosis and Treatment of Lipid Disorders*, 2016, pp. 6–8.

[72] S. L. Rebholz, J. T. Melchoir, J. A. Welge, A. T. Remaley, W. S. Davidson, and L. A. Woollett, “Effects of Multiple Freeze/Thaw Cycles on Measurements of Potential Novel Biomarkers Associated With Adverse Pregnancy Outcomes,” *J. Clin. Lab. Med.*, vol. 2, no. 1, pp. 1–22, 2017, doi: 10.16966/2572-9578.107.

[73] A. M. Zivkovic, M. M. Wiest, U. T. Nguyen, R. Davis, S. M. Watkins, and J. B. German, “Effects of sample handling and storage on quantitative lipid analysis in human serum,” *Metabolomics*, vol. 5, no. 4, pp. 507–516, 2009, doi: 10.1007/s11306-009-0174-2.

[74] L. Lovergne, P. Bouzy, V. Untereiner, and R. Garnotel, “Biofluid infrared spectro-diagnostics: pre-analytical considerations for clinical applications,” *Faraday Discuss.*, vol. 187, pp. 521–537, 2016, doi: 10.1039/c5fd00184f.

[75] C. Petibois, A. M. Melin, A. Perromat, G. Cazorla, and G. Déléris, “Glucose and lactate concentration determination on single microsamples by Fourier-transform infrared spectroscopy,” *J. Lab. Clin. Med.*, vol. 135, no. 2, pp. 210–215, 2000, doi: 10.1067/mlc.2000.104460.

[76] Y. J. Park *et al.*, “Serum lactate upon emergency department arrival as a predictor of 30-day in-hospital mortality in an unselected population,” *PLoS*

One, vol. 13, no. 1, pp. 1–14, 2018, doi: 10.1371/journal.pone.0190519.

[77] P. Database, "L-Lactate, CID=5460161," *National Center for Biotechnology Information*. https://pubchem.ncbi.nlm.nih.gov/compound/L-Lactate (accessed Aug. 20, 2019).

[78] R. Gautam *et al.*, "Molecular profiling of sepsis in mice using Fourier Transform Infrared Microspectroscopy," *J. Biophotonics*, vol. 9, no. 1–2, pp. 67–82, 2016, doi: 10.1002/jbio.201400089.

[79] M. Moutachakkir, A. Lamrani Hanchi, A. Baraou, A. Boukhira, and S. Chellak, "Immunoanalytical characteristics of C-reactive protein and high sensitivity C-reactive protein," *Ann. Biol. Clin. (Paris).*, vol. 75, no. 2, pp. 225–229, Mar. 2017, doi: 10.1684/abc.2017.1232.

[80] C. Palmiere and M. Augsburger, "Markers for sepsis diagnosis in the forensic setting: State of the art," *Croat. Med. J.*, vol. 55, no. 2, pp. 103–114, 2014, doi: 10.3325/cmj.2014.55.103.

[81] M. J. Hackett *et al.*, "Concurrent glycogen and lactate imaging with FTIR spectroscopy to spatially localize metabolic parameters of the glial response following brain ischemia," *Anal. Chem.*, vol. 88, no. 22, pp. 10949–10956, 2016, doi: 10.1021/acs.analchem.6b02588.

[82] M. Brandstetter, L. Volgger, A. Genner, C. Jungbauer, and B. Lendl, "Direct determination of glucose, lactate and triglycerides in blood serum by a tunable quantum cascade laser-based mid-IR sensor," *Appl. Phys. B Lasers Opt.*, vol. 110, no. 2, pp. 233–239, 2013, doi: 10.1007/s00340-012-5080-z.

[83] R. K. Tanner, K. L. Fuller, and M. L. R. Ross, "Evaluation of three portable blood lactate analysers: Lactate Pro, Lactate Scout and Lactate Plus," *Eur. J. Appl. Physiol.*, vol. 109, no. 3, pp. 551–559, 2010, doi: 10.1007/s00421-010-1379-9.

[84] A. Păucean *et al.*, "Monitoring lactic acid concentrations by infrared spectroscopy: A new developed method for lactobacillus fermenting media with potential food applications," *Acta Aliment.*, vol. 46, no. 4, pp. 420–427, 2017, doi: 10.1556/066.2017.0003.

[85] S. A. Tatulian, "Structural characterization of membrane proteins and peptides by FTIR and ATR-FTIR spectroscopy," in *Lipid-Protein Interactions: Methods and Protocols*, vol. 974, 2012, pp. 177–218.

[86] D. M. Byler and H. Susi, "Examination of the secondary structure of proteins by deconvolved FTIR spectra," *Biopolymers*, vol. 25, no. 3, pp. 469–487, 1986, doi: 10.1002/bip.360250307.

[87] P. I. Haris and F. Severcan, "FTIR spectroscopic characterization of protein structure in aqueous and non-aqueous media," *J. Mol. Catal. - B Enzym.*, vol. 7, no. 1–4, pp. 207–221, 1999, doi: 10.1016/S1381-1177(99)00030-2.

[88] K. A. Kristoffersen *et al.*, "Fourier-transform infrared spectroscopy for monitoring proteolytic reactions using dry-films treated with trifluoroacetic

acid," *Sci. Rep.*, vol. 10, no. 1, pp. 1–10, 2020, doi: 10.1038/s41598-020-64583-3.

[89] Y. L. Jeyachandran, E. Mielczarski, B. Rai, and J. A. Mielczarski, "Quantitative and qualitative evaluation of adsorption/desorption of bovine serum albumin on hydrophilic and hydrophobic surfaces," *Langmuir*, vol. 25, no. 19, pp. 11614–11620, 2009, doi: 10.1021/la901453a.

[90] K. Spalding *et al.*, "Enabling quantification of protein concentration in human serum biopsies using attenuated total reflectance – Fourier transform infrared (ATR-FTIR) spectroscopy," *Vib. Spectrosc.*, vol. 99, pp. 50–58, 2018, doi: 10.1016/j.vibspec.2018.08.019.

[91] C. Petibois, G. Cazorla, A. Cassaigne, and G. Deleris, "Plasma Protein Contents Determined by Fourier-Transform Infrared Spectrometry Cyril," *Clin. Chem.*, vol. 47, no. 4, pp. 730–738, 2001.

[92] J. Helmersson-Karlqvist, M. Flodin, A. M. Havelka, X. Y. Xu, and A. Larsson, "The Roche Immunoturbidimetric Albumin Method on Cobas c 501 Gives Higher Values Than the Abbott and Roche BCP Methods When Analyzing Patient Plasma Samples," *J. Clin. Lab. Anal.*, vol. 30, no. 5, pp. 677–681, 2015, doi: 10.1002/jcla.21921.

[93] B. T. Doumas and T. Peters, "Serum and urine albumin: A progress report on their measurement and clinical significance," *Clin. Chim. Acta*, vol. 258, no. 1, pp. 3–20, 1997, doi: 10.1016/S0009-8981(96)06446-7.

[94] F. Bonnier *et al.*, "Screening the low molecular weight fraction of human serum using ATR-IR spectroscopy," *J. Biophotonics*, vol. 9, no. 10, pp. 1085–1097, 2016, doi: 10.1002/jbio.201600015.

[95] A. Rohman, A. Windarsih, E. Lukitaningsih, M. Rafi, K. Betania, and N. A. Fadzillah, "The use of FTIR and Raman spectroscopy in combination with chemometrics for analysis of biomolecules in biomedical fluids: A review," *Biomed. Spectrosc. Imaging*, vol. 8, no. 3–4, pp. 55–71, 2020, doi: 10.3233/bsi-200189.

[96] S. K. Vashist and J. H. T. Luong, *Antibody Immobilization and Surface Functionalization Chemistries for Immunodiagnostics*. Elsevier Inc., 2018.

[97] M. Voue *et al.*, "Biochemical Interaction Analysis on ATR Devices : A Wet Chemistry Approach for Surface Functionalization," *Langmuir*, vol. 23, no. 2, pp. 949–955, 2007, doi: 10.1021/la061627j.

[98] S. Devouge *et al.*, "Surface functionalization of germanium ATR devices for use in FTIR-biosensors," *J. Colloid Interface Sci.*, vol. 332, no. 2, pp. 408–415, 2009, doi: 10.1016/j.jcis.2008.12.045.

[99] W. Liao, F. Wei, D. Liu, M. X. Qian, G. Yuan, and X. S. Zhao, "FTIR-ATR detection of proteins and small molecules through DNA conjugation," *Sensors Actuators, B Chem.*, vol. 114, no. 1, pp. 445–450, 2006, doi: 10.1016/j.snb.2005.06.021.

[100] R. A. Shaw, S. Kotowich, M. Leroux, and H. H. Mantsch, “Multianalyte serum analysis using mid-infrared spectroscopy,” *Ann. Clin. Biochem.*, vol. 35, no. 5, pp. 624–632, 1998, doi: 10.1177/000456329803500505.

Chapter 6: Sepsis and SIRS Discrimination by High-Throughput ATR-FTIR

6.1. Introduction

It is a common view that there is a dire unmet clinical need for a rapid and reliable clinical test for sepsis diagnosis [1]–[3]. Its prevalence paired with quick onset and low detectability render it undeniably challenging and lethal, not to mention the staggering economic impact sepsis has on healthcare institutions [4], [5]. Sepsis is almost identical in presentation to Systemic Inflammatory Response Syndrome (SIRS). The singular, and major, defining factor is that sepsis is a syndromic response in the presence of infection, while SIRS may manifest as a consequence of a wide variety of non-infectious insults (e.g., trauma, pancreatitis, or high risk surgery) [6], [7]. This crucial detail in turn dictates the ensuing healthcare pathway which may include administration of vasopressors, antibiotics, resuscitation, admission to ICU and other clinical actions as appropriate [8].

Currently, both sepsis and SIRS diagnoses are normally made when two or more of the four SIRS criteria are met and a variety of clinical observations and tests, such as heart rate, blood pressure, C-reactive Protein (CRP), PCT, glucose, or lactate levels, are abnormal as well as the identification of an infection in the case of sepsis. The four SIRS criteria are [9]:

1. Temperature >38°C or <36°C;
2. Heart rate >90 beats per minute;
3. Respiratory rate >20 breaths per minute or $PaCO_2$ <32mmHg;
4. White blood cells count >12,000/cu mm, <4,000/cu mm, or >10% immature (band) forms.

Individually these conditions are non-specific as any may be present for a variety of reasons unrelated to SIRS or sepsis. For example, a combination of abnormal lactate levels, temperature and hyperventilation may be present in liver dysfunction, diabetic ketoacidosis, or bronchiolitis each of which may, incidentally, exacerbate sepsis outcomes [10]–[12]. Conversely, these markers may even be absent in people who have sepsis/SIRS as is the case with 'SIRS-negative sepsis' diagnoses (i.e., sepsis with less than two SIRS criteria) [13]. Furthermore, only 60% of sepsis patients produce a positive blood culture test [14] and blood cultures determine causative microbes in merely 20-40% of patients with severe sepsis [3]. It can therefore be said the multitude of vital signs and test results are merely guidelines by which the clinician must make a subjective diagnosis and a single biomarker is clearly inadequate to make a reliable diagnosis. Moreover, minor variations between centre definitions of SIRS and measurement methods result in large variations in sepsis incidence suggesting either poor clinical acceptance of the consensus definitions, or the consensus definitions do not adequately describe the enormously complex pathological conditions of SIRS and sepsis [15], [16]. It follows that multivariate and machine learning techniques are well poised to provide reliable and objective tools for sepsis diagnostics [17], [18]. Each of the varied conditions and SIRS criteria mentioned above will result in changes to the infrared biochemical profile of serum from SIRS and sepsis patients, changes that can be rapidly detected using SIREs and trained machine learning algorithms. This has been attempted before by using standard clinical observations, laboratory test results, and demographics as input parameters to train and test predictive models, but only low sensitivities could be achieved [19]. In this chapter, three-way classifier models are built using Partial Least Squares Discriminant

Analysis (PLS-DA) and Random Forest (RF) to discriminate between control, SIRS and sepsis patient serum samples using spectral data acquired on SIREs. PCA and binary classification models are generated to investigate the specific spectral regions that contribute most variation in the dataset. Finally, an investigation is carried out to determine if the specific pathogen present in the sepsis patients can be determined using FTIR.

6.2. Materials and Methods

6.2.1. Serum Samples

Serum samples (n = 148) were acquired from the Defence Science & Technology Laboratory (Dstl), Porton Down, United Kingdom. Samples were collected from patients requiring elective high risk surgery (liver and renal surgery, lung resection, large bowel resection, gastrectomy, Whipple's procedure). Patients who were under 18 or over 80 years old, pregnant, and requiring immunosuppressors were excluded. Consent forms were signed by all patients involved at the medical centres they were admitted to (Liverpool Royal and University Hospital; University College Hospital, London; Bristol Royal Infirmary; Guy's & St Thomas' Hospital, London; Birmingham Queen Elizabeth Hospital; and University Hospital, Frankfurt). The project has been approved by Southampton & South West Hampshire Research Ethics Committee (A), 1st Floor, Regents Park Surgery, Shirley, Southampton, Hants, SO16 4RJ (Reference number: 06/Q1702/152). Blood samples were collected the day after surgery and if no complications developed were used in the control group (n =

46), otherwise, blood samples were collected again on the day of sepsis (n = 43) or SIRS (n = 59) diagnosis. Clotted blood samples (4ml) were centrifuged and serum was extracted then stored at -80°C until experiments were carried out. Patients were recruited from before 2016, so sepsis and SIRS diagnoses have been made according to former definitions [9], [20]:

1. Two or more of the four SIRS criteria.
2. Presence of an infection (sepsis only).

Sepsis, SIRS, and control patient's diagnoses made in the different medical centres involved were reviewed and confirmed by a clinical advisory panel. Patients undergoing high-risk surgery serve as a suitable control group as this provides patients experiencing sterile inflammation which is one of the reasons other biomarkers give low specificity [5]. Partially complete datasheets detailing a large variety of patient specific information have also been provided including ethnicity, gender, age, admission diagnosis, White Cell Count (WCC), CRP concentration, SOFA score, Antibiotics administrated, Glasgow coma score, and more.

6.2.2. Experimental Procedure

One SIRE was used per patient. 3µl of serum was pipetted into each of the three sample compartments of the SIRE. Samples were then allowed to dry for one hour in an incubator at 35°C. SIREs were then placed in a Spectrum2 (PerkinElmer, USA) with a custom automated slide indexing unit (see chapter 3, section 3.2.3) and Quest ATR accessory unit (Specac, UK) for analysis. Spectra were collected between 4000-450cm^{-1} range and averaged over 16 co-added scans at a resolution of 4cm^{-1} with

$1cm^{-1}$ data spacing. A background scan was acquired from the 'zero' position of the SIRE slide and automatically removed from sample spectra. SIRE sample compartments were analysed in triplicate equalling nine spectra per patient. In total, 1332 spectra were collected.

6.2.3. Analytical Procedure

A Matlab (MathWorks, USA) based General User Interface (GUI) developed in-house was generally used for exploratory analysis. An in-house R-studio based computer programme called PRFFECT was used to build all machine learned classification models [21]. Numerous multivariate and machine learning approaches were used as exploratory and classification techniques to analyse the serum samples. These were Principal Component Analysis (PCA), Partial Least Squares Discriminant Analysis (PLS-DA) and Random Forest (RF). PCA was intended to be a qualitative exploratory technique to visualise where the greatest variance in the dataset resided, so pre-processing techniques which already established were used to quickly explore the data. A cut-off value of 90% explained variance was taken when selecting the number of components to investigate. PLS-DA and RF were used to develop classification models to predict disease classes. Initially, investigations were carried out using PLS-DA and RF to develop three-way classification models (i.e., control vs sepsis vs SIRS) to deliver as close a resemblance to the real-world clinical scenario as possible. Later, a binary classification model (sepsis vs SIRS) was developed using RF only to elucidate the specific biochemical differences that could separate the two clinical conditions. Each technique used here necessitated slightly different spectral pre-processing methods. These methods have been represented as a flow diagram in

Figure 6.1. For the development of all classification models 70% of patient samples were used to train models, and 30% of the patient samples were used to test them. Statistical characteristics presented are the average and standard deviation of accuracy, sensitivity, specificity, negative predictive value (NPV), and positive predictive values (PPV) across 51 classification models. The reason for doing this is to ensure that test and training set patients are changed at each iteration reducing bias in the test set and providing results more representative of the wider patient population [22]. Thus, to demonstrate how robust the predictive models are, the mean and standard deviation of the outputs across many models are presented. Confusion matrices have also been presented to help illustrate performance of each classification regime. However, unlike the predictive value outputs, confusion matrix predictions cannot be averaged over all iterations as patient within the training and test set are effectively altered each time. Instead, a confusion matrix from only the first iteration has been presented.

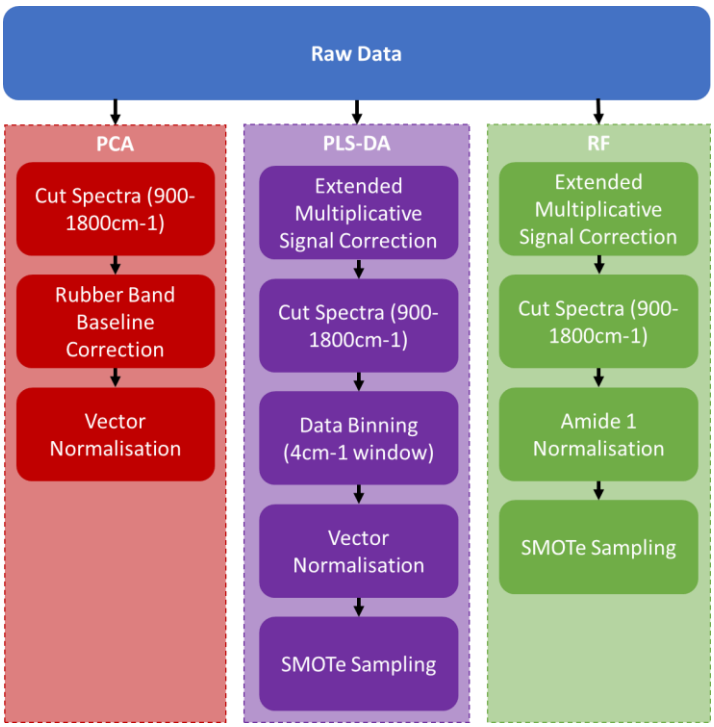


Figure 6.1. *Flowchart presenting consecutive pre-processing techniques used before initiating each analysis method.*

During the study numerous combinations of pre-processing techniques were attempted, but the methods presented here have been selected based on those that elicited the highest and most reliable sensitivity, specificity, NPV and PPV, and accuracy of predictive models. The most noteworthy difference is the use of a data binning step utilised in PLS-DA models to smooth spectra and reduce computation time for constructing predictive models. Extended Multiplicative Signal Correction (EMSC) served as an alternative baseline correction technique that produced superior predictive models [23].

6.3. Results and Discussion

6.3.1. Qualitative Observations

Average pre-processed spectra of all patients across the three disease classes have been presented in Figure 6.2. Upon visual observation it is apparent that noticeable changes in the Amide I and II peaks and the 1000-1100cm^{-1} region as well as subtler variations of the Amide III and 1100-1200cm^{-1} region between each group. The SIRS and sepsis waveforms are notably similar suggesting discriminant analysis between these groups may be difficult whereas discrimination of the control group can be expected to be fairly reliable. Spectral variations are likely due to an overall trend toward hyperglycaemia (abundance of glucose; 1000-1100cm^{-1}) and hypoalbuminemia (depletion of albumin; Amide I and II) in the sepsis and SIRS groups relative to the control group.

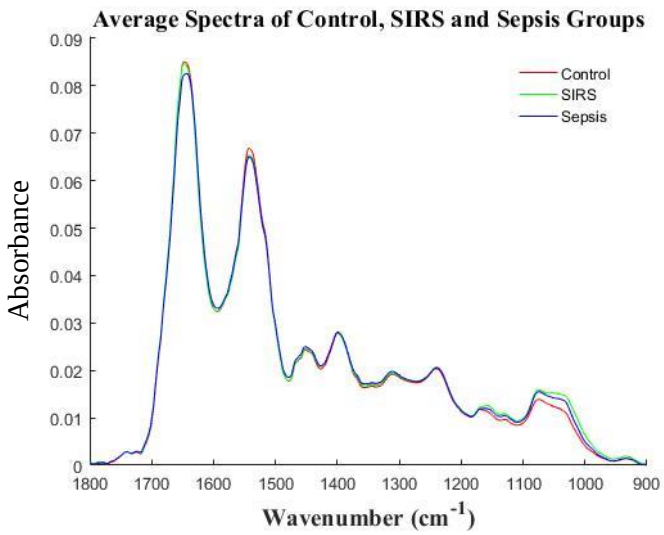


Figure 6.2. *Average finger-print region of pre-processed spectra of the serum from sepsis, SIRS, and control groups. SIRS and sepsis groups are remarkably similar indicating discriminate analysis between these classes may be difficult. The differences between the control group and SIRS/sepsis is more pronounced specifically at the Amide I, Amide II and 1000-1100cm^{-1}.*

Figure 6.3 shows the results of a PCA of all spectra in a 3D plot of PC 1, 2, and 3. The first five PCs explained 90.8% of the variance with PC4 and PC5 accounting for 3.5% and 3.2% of the total variance, respectively. PC4 and PC5 scores plot did not display any separation between groups. It was found that three SIRS and one sepsis patients (circled regions of Figure 6.3.(a)) were potential outliers. These patients have been removed in a second PCA so that separation for groups is more clearly visualised (Figure 6.3.(e)). Observing the difference in PC 1 loading vector before and after removing patient spectra indicates PC 1 variance is more dominated by glucose content with the inclusion of these patients (C-O related vibrations 1000-1100cm^{-1}, Figure 6.3.(b)), whereas evidence of variance due to lactate as well as glucose content can be observed in the PC 1 when these patients have been excluded (carbonyl related vibrations at 1650-1750cm^{-1}, Figure 6.3.(d)) [24]. Checking the average spectra of each patient confirms a significant contribution of spectral features at 1000-1100cm^{-1} (Chapter 8, Appendix VI). It should be noted that these patients have not been removed from subsequent analysis as they still represent valid variance within the dataset, but only removed here so that separation of the three classes can be more clearly illustrated by visual analysis of the 3D PCA plots. Loadings vectors of PC 2, and 3 have also been displayed (Figures 6.3.(e)-(h)). There is minimal difference observed in these loading directions whether the four patient outliers have been removed or not. Sharp peaks at C-O regions of loading 1, carbonyl regions and dips at the amide I and II regions in loading 2, and the dips at amide III regions of loading 3 suggest that for classes tending toward the positive scale of the PC1, PC2 and PC3 (i.e., sepsis and SIRS classes) have an increased C-O and carbonyl group presence and decreased amide I, II and III presence compared to classes tending toward negative PC1, PC2

and PC3 values (i.e., control samples). These qualitative observations are important because it is evidence that biochemical changes known to occur in the serum of sepsis and SIRS patients have corresponding patterns in the vibrational spectra of the serum. This lends credence to the validity of spectroscopy-based machine learning approaches in SIRS, sepsis, and control group classification by providing a link to spectral features and the physiological differences between groups.

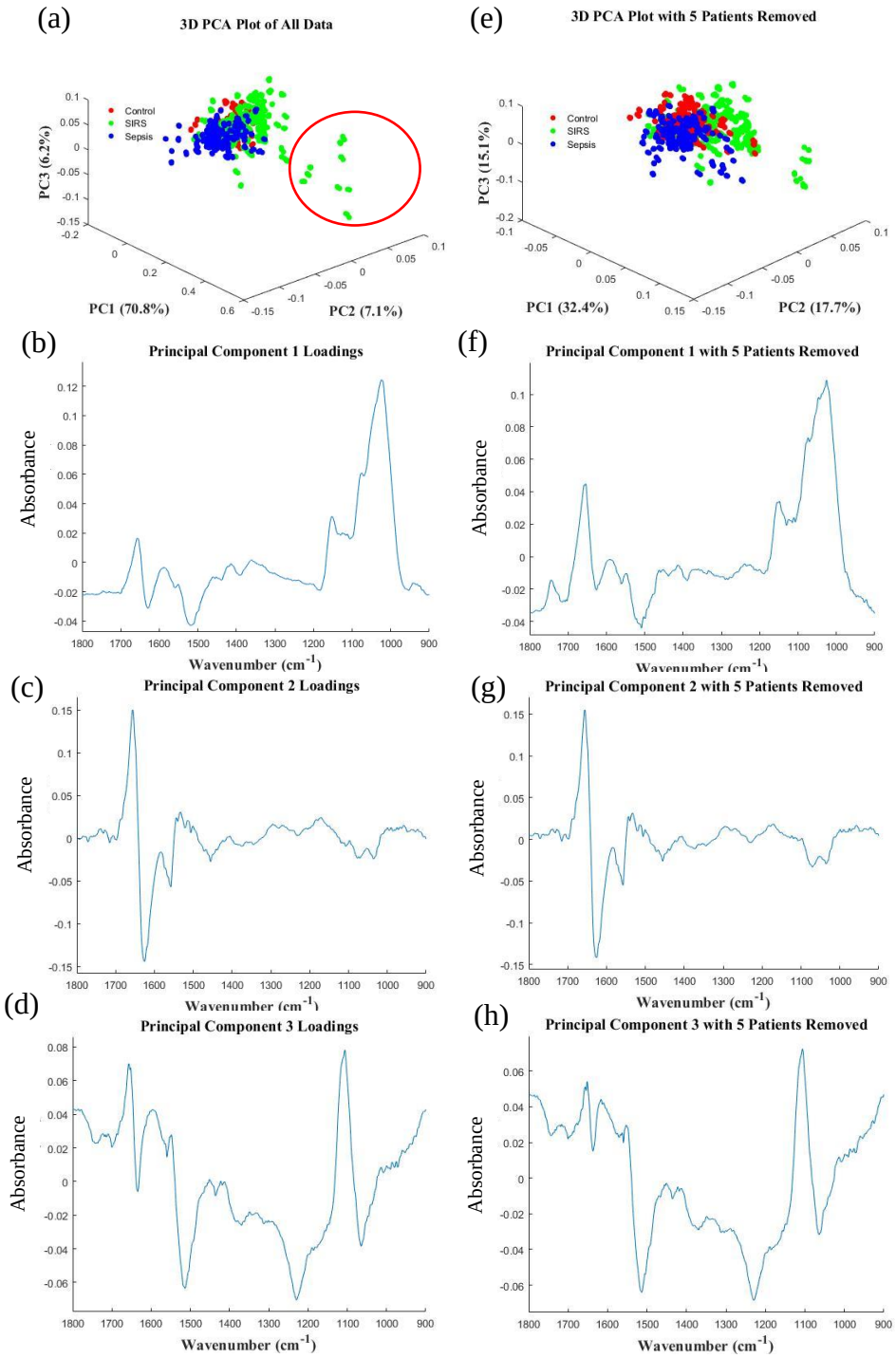


Figure 6.3. *(a-d) PCA plot with loading directions 1, 2, and 3 with all data collected in the experiment. Outliers in the SIRS group were identified as indicated by a red circle in (a). Loadings suggest variance in the dataset is owed to a combination of increased carbohydrate (1000-1200cm^{-1}, Loading 1) and decreased protein (Amide II and Amide II, Loading 2 and Amide III, Loading 3) related molecular species between groups. (e-h) PCA plot and loading diretions 1, 2, and 3 with four patient outliers removed. The removal of outliers has little impact on the loading plots with the exception of a greater degree of variance observed in the carbonyl related wavenumbers (1650-1750cm^{-1}) of loading 1.*

6.3.2. Discriminant Analysis

6.3.2.1. Partial Least Squares Three-Way Classifier

3D scatter plots of the PLS-DA scores of the first three PLS components (Figure 6.4.(a)) seems to exhibit better separation of the three classes than the previous PCA (Figure 6.3.(a)). Three SIRS patients and one sepsis patient were found to be particularly prominent in the component 1 and 2 directions (circled region in Figure 6.4.(a)). These same patients were also observed as outliers in the PCA (section 6.3.1). The first 10 components were selected for analysis as this was the minimum number of components required to minimise RMSE of predictive models (see chapter 2, section 2.4.2.2). Scores and loadings plots of PCs 1, 2, and 3 have been presented below (Figure 6.4). Loadings are complex, suggesting a wide variety of molecular species contribute to discrimination of classes as can be expected (Figure 6.4.(b), (c), (d)) although this could be due to aliasing of the data: a foreseeable ramification of data binning. Nevertheless, it can be tentatively prescribed that the major sources of variance by which classes can be discriminated is by concurrent hyperlactatemia ($1650cm^{-1}$) and hypoalbuminia ($1625cm^{-1}$ and $1575cm^{-1}$) according to features of component one loadings (Figure 6.4.(b)) and in agreement with previous qualitative observations. Loadings of components 2 and 3 are less salient and so vibrational modes or biochemical species cannot be as reliably assigned. Similarly, PC 4 through PC 10 score plots did not display any separation between groups and their respective loading plots are non-specific and do not highlight any particular spectral bands as discriminant factors. Statistical characteristics of the PLS-DA classifier model have been presented in Table 6.1 showing the accuracy, optimal number of PLS components, sensitivity,

specificity, NPV, and PPV of the test set of each class. The predictive capabilities of the model in correctly identifying control group samples are, while still good, lagging the SIRS and sepsis groups. This is reflected in the 3D PLS-DA score plot where the SIRS and sepsis groups are well discriminated and the control group is dispersed between. Specificity and NPV for each class is high and reliable although sensitivity and PPV is perhaps too low and too variable to be accepted by the clinical community [25], [26].

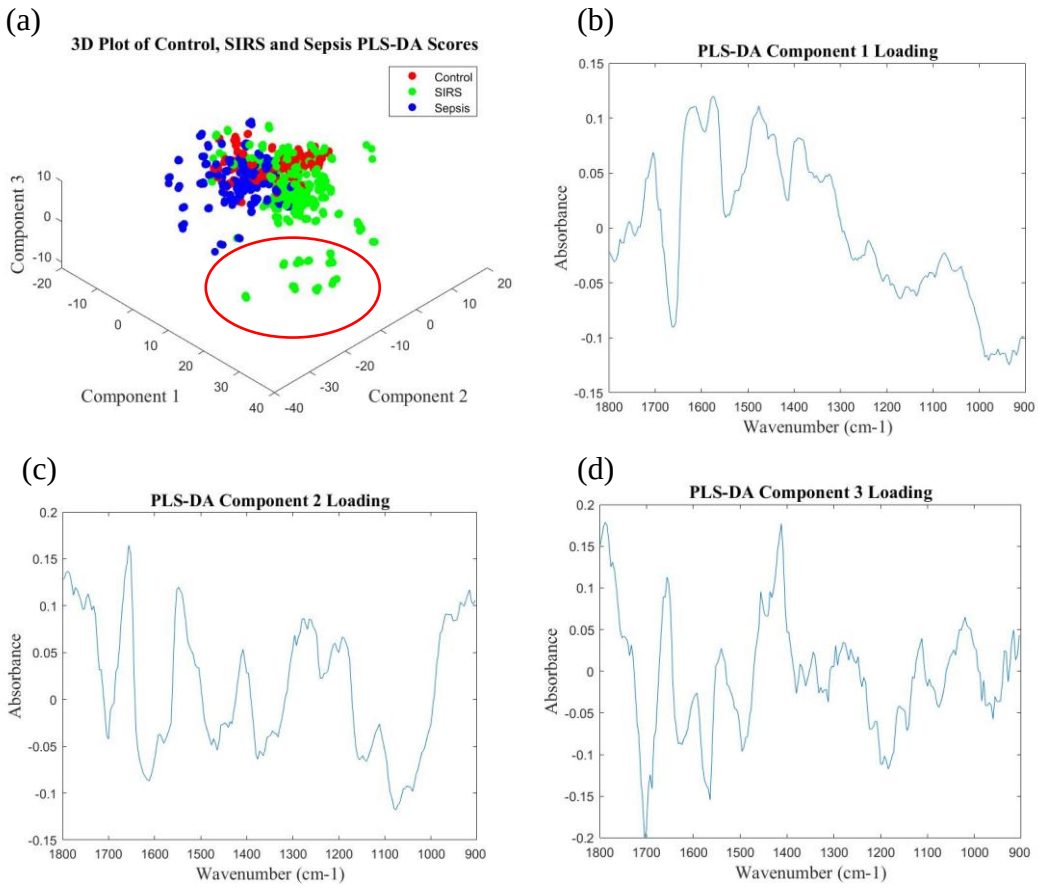


Figure 6.4.*(a) 3D plot of PLS-DA scores of SIRS, sepsis and control patients. Good separation is visually obvious between SIRS and sepsis groups. A red circle indiates two SIRS patient outliers in the dataset. (b-d) Loading vectors of components 1, 2, and 3 of the PLS-DA. It is difficult to associate biochemical IR fingerprints to the loadings, but sharp dips at 1575 and 1625cm-1 are akin to inverted amide peaks immediately adjacent to a carbonyl peak suggestive of protein and lactate related variance in the dataset.*

Table 6.1. *Statistical characteristics of the three-way classifier model constructed using the PLS-DA method. Quoted are the average and standard deviations of 51 models.*

PLS-DA Classifier Characteristics						
	Control		SIRS		Sepsis	
Sensitivity	0.6018	±0.1312	0.7647	±0.1046	0.8252	±0.1182
Specificity	0.8668	±0.0676	0.8667	±0.0700	0.8641	±0.0592
Neg. Pred. Val.	0.8312	±0.0491	0.8480	±0.0606	0.9278	±0.0455
Pos. Pred. Val.	0.6799	±0.1240	0.8029	±0.0925	0.7181	±0.0910
Accuracy		0.7316		±0.0625		
PLS Components			10			

Table 6.2. *Confusion matrix of the first iteration of three-way classifier model constructed using the PLS-DA method displaying observed and predicted diseased or non-diseased states of the test set patient samples.*

		Observed		
		Control	SIRS	Sepsis
Prediction	Control	8	1	0
	SIRS	2	14	1
	Sepsis	3	2	11

6.3.2.2. Random Forest Three-Way Classifier

Random Forest machine learning approaches yielded better prediction capabilities than PLS-DA as shown by the improved sensitivities, specificities, NPVs, PPVs, and accuracy presented in Table 6.3. Sensitivity and positive predictive value are good and specificity and NPVs are excellent across all classes. A strong ability to detect true negatives is an essential requirement for clinical acceptance due to the importance of correctly ruling out sepsis [27]. An importance plot can be acquired from RF that indicates which spectral regions are of greater importance for discrimination between classes (Figure 6.5). As expected, protein associated regions of the Amide bands are highly important for the prediction of classes. To be more precise, the shoulders of the Amide I band show significant contribution perhaps since

the data has been normalised to Amide I thereby eliminating variance at the centre of the amide I peak. Furthermore, only patients with significant drop in serum proteins will have detectable decrease of amide I band due to the saturation of the IR detector at the amide I band [28]. Another study have found that the Amide I band has little impact in the detection of cancer, but this may not be directly relevant to SIRS/sepsis/control classification using random forest [29]. The Amide II band has the greatest influence on classification as it may be sensitive to variations of pro- and anti-inflammatory protein species or changes in secondary protein structures. However, interpretation is difficult due to the multitude of biomolecule species known to fluctuate during SIRS and sepsis such as C-reactive protein, interleukin-6, procalcitonin and many more [30]. Wavenumbers between 900-1000cm^{-1} also play a significant part in the discrimination of classes possibly due to P-O and C-O stretching vibrations from nucleic acid and carbohydrate species such as deoxyribose, DNA, glucose or lactate. Upon closer inspection, the prominent peaks at roughly 927cm^{-1} and 960cm^{-1} likely correspond to vibrational modes of DNA and the C-O bond of deoxyribose respectively suggesting DNA fragments are in some way contributing to class discrimination in the RF model [24]. Circulation of exogenous DNA fragments from pathogenic species and endogenous cell-free DNA (cfDNA), which are both known to rise in sepsis, presents a plausible source for this variation [18], [31]–[33].

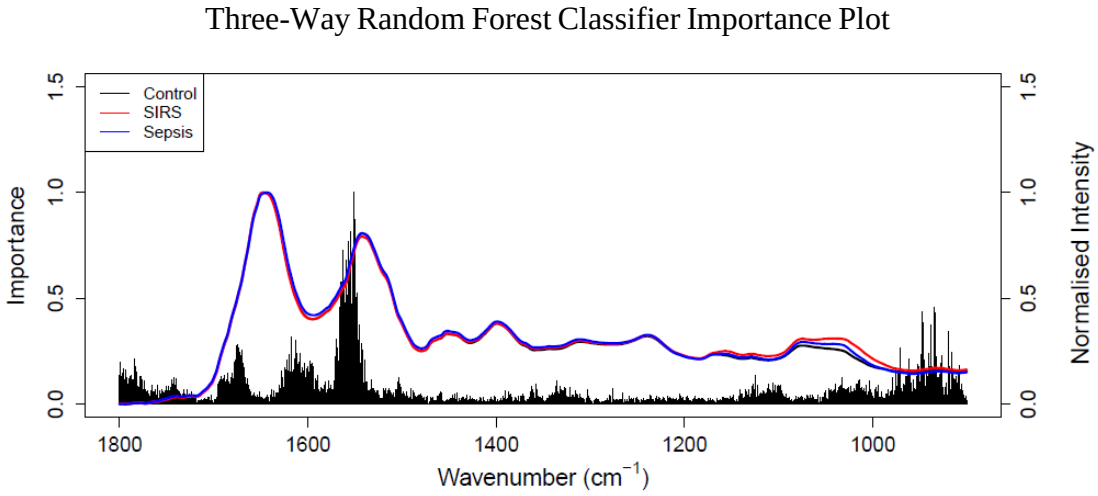


Figure 6.5. *Importance plot highlighting spectral regions that were instrumental in the construction of the three-way RF classification model. Amide I and II peaks contribute significantly likely due to fluctuations in pro- and anti-inflammatory cytokines. The 900-1000cm^{-1} wavenumber region indicates DNA related molecular species are also important in discriminating between SIRS, sepsis, and control classes.*

Table 6.3. *Statistical characteristics of the three-way RF classifier model constructed using the RF method. Quoted are the average and standard deviations of 51 models.*

Three-Way Random Forest Classifier Characteristics

	Control		SIRS		Sepsis	
Sensitivity	0.7753	±0.1007	0.8385	±0.0814	0.8415	±0.1083
Specificity	0.9101	±0.0543	0.9341	±0.0465	0.8895	±0.0501
Neg. Pred. Val.	0.9020	±0.0401	0.8975	±0.0464	0.9355	±0.0409
Pos. Pred. Val.	0.8056	±0.1002	0.9019	±0.0655	0.7608	±0.0812
Accuracy		0.8198		±0.0467		

Table 6.4. *Confusion matrix of the first iteration of three-way classifier model constructed using the RF method displaying observed and predicted diseased or non-diseased states of the test set patient samples.*

		Observed		
		Control	SIRS	Sepsis
Prediction	Control	11	1	1
	SIRS	0	16	1
	Sepsis	2	0	10

6.3.2.3. Random Forest Binary Classifier

The eminent clinical need is in rapid discriminative capabilities between sepsis and non-infectious SIRS due to the subtle distinction between the two syndromes, highly inconsistent presentation, and frequently contested clinical definitions. To examine the utility of SIRS with respect to this more specific clinical question and to elucidate the specific biochemical differences between serum of SIRS and sepsis groups the previous discriminatory techniques have been employed again to build binary RF classification models this time designating the SIRS patients as the control group and removing the elective high-risk surgery patients. Figure 6.6 and Table 6.5 display the importance plot and statistical characteristics of the random forest classifier model of the two disease state groups. Accuracy, positive predictive value, specificity, and sensitivity are all improved upon compared to three-way classifier investigations in the previous section as can be expected due to the decrease in complexity required to describe the datasets. NPV has remained roughly the same. In contrast to the three-way classifier model importance plot, variations in spectra between SIRS and sepsis groups lies decisively within the Amide II band and DNA (900-1000cm^{-1}) associated peaks of the fingerprint region while influences of wavenumbers with proximity to the Amide I band is further diminished. Again, it is difficult to associate specific biomolecular species to the differences in IR absorbance, but the emphasis on two distinct spectral peaks at *ca.* 945cm^{-1} and *ca.* 1585cm^{-1}) is a key indicator that profound differences in protein and DNA related biochemistry exists between the SIRS and sepsis groups [24]. Differences in protein concentrations could manifest due to host immunological response cytokines specific to infection such as procalcitonin (PCT),

interleukin-6 (IL-6), IL-8, IL-10, IL-18, and tumour necrosis factor-alpha (TNF-alpha) [34].

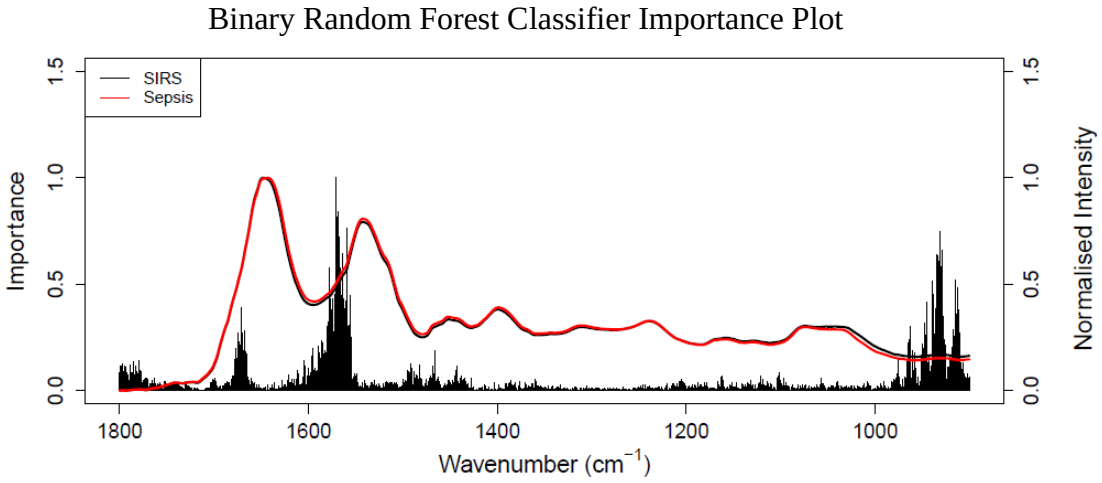


Figure 6.6. *Importance plot highlighting spectral regions that were instrumental in the construction of the binary RF classification model. Amide II peaks contribute significantly likely due to fluctuations in pro- and anti-inflammatory cytokines. The 900-1000cm^{-1} wavenumber region indicates DNA related molecular species are also important in discriminating between SIRS and sepsis classes.*

Table 6.5. *Statistical characteristics of the binary RF classifier model constructed using the RF method. Quoted are the average and standard deviations of 51 models.*

Binary Random Forest Classifier Characteristics

Sensitivity	0.8676	±0.1056
Specificity	0.9043	±0.0715
Neg. Pred. Val.	0.9112	±0.0837
Pos. Pred. Val.	0.8734	±0.0644
Accuracy	0.8891	±0.0553

Table 6.6. *Confusion matrix of the first iteration of binary classifier model constructed using the RF method displaying observed and predicted diseased or non-diseased states of the test set patient samples.*

		Reference	
		SIRS	Sepsis
Prediction	SIRS	16	3
	Sepsis	1	9

6.3.2.4. Evaluation of Predictive Capabilities

An excellent NPV (>98%) and test sensitivity (>98%) is widely specified as an essential characteristic of a sepsis test which the SIRE is capable of delivering on [18], [25]–[27]. This is because a false negative result may incur the inappropriate cessation of antibiotics and may pose great risk to patient welfare [18]. A false positive, on the other hand, is likely to lead to inappropriate administration of antibiotics which may introduce anti-microbial resistant pathogens but is not immediately threatening to the patient's life. Poor NPV may be one of the reasons there is no clinically accepted biomarker for the diagnoses of sepsis [3], although many have been proposed as potential candidates and some are widely used to guide clinical decisions with the caveat of low specificity such as lactate, CRP, and PCT [5], [35], [36]. The predictive capabilities presented here outperform predictive capabilities of all blood biomarkers commonly used in the clinic by virtue of the fact that this test does not require a concentration cut-off value to be selected which generally forces a trade-off between good sensitivity or good specificity [37]. Furthermore, the SIRE based test requires very low volumes of samples which is a crucial characteristic for the screening of sepsis in neonates where the disease is equally prevalent, but carries with it the added risk of causing life-long disorders [38]. It is estimated that as many as 60% of blood cultures are falsely negative if blood samples used are 0.5ml or less [39].

The specificity, sensitivity, negative predictive value, positive predictive value, and accuracy of common clinical indicators used to differentiate sepsis from sterile SIRS were compared to corresponding predictive characteristics of classifiers constructed from patient serum analysed using SIREs (Tables 6.3 and 6.5). Predictive values of lactate, C-reactive protein (CRP), procalcitonin (PCT), Interleuking-6 (IL-

6), presepsin, neutrophil CD64, lipopolysaccharide binding protein (LBP), and soluble triggering receptor expressed on myeloid cells-1 (sTREM-1) were derived from a multitude of studies and have been tabulated in Table 6.7 for comparison against predictive values for sepsis detection achieved by the three-way and binary classifiers developed using serum spectra acquired from SIREs. In studies where conflicting results were found (e.g., studies stating different diagnostic accuracies for patients in the Intensive Care Unit or Emergency Department patient groups) the results that most generously reflect on each respective biomarker has been selected for comparisons sake. Results published in systematic reviews and meta-analysis have been indicated by 'meta'. In one instance, positive predictive value, negative predictive value, and accuracy were not explicitly stated in the source material, but were instead calculated using true positive, false positive, true negative, and false negative values stated in the study [30]. These calculated values are indicated with an underline in Table 6.7.

Table 6.7. *Statistical characteristics of commonly used biomarkers for sepsis and the results achieved by the high throughput SIRE (**bold**). The best results found in the literature have been selected for comparison to the present study. '–' not available. '**meta**' pooled results drawn from meta-analysis studies. Underlined values have been calculated using data extracted from the respective study.*

Biomarker	Spec.	Sens.	PPV	NPV	Acc.	Reference
Lactate	28.18	83.36	-	-	-	[40]
	82.00	34.00	89.00	23.00	-	[41]
	74.30	56.70	20.80	93.50	-	[42]
CRP	65.60	61.10	32.00	85.90	-	[35]
	79.00	77.00	-	-	-	[43] **meta**
	67.00	75.00	76.62	60.68	69.35	[30] **meta**
	71.00	83.00	-	-	-	[44] **meta**
	46.15	84.30	84.00	42.80	-	[45]
	75.00	98.50	-	-	-	[46]
PCT	69.50	72.60	39.00	90.40	-	[35]
	79.00	78.00	-	-	-	[43] **meta**
	79.00	76.00	-	-	-	[44] **meta**
	78.00	79.00	78.93	68.41	74.15	[30] **meta**
	75.00	80.00	-	-	-	[47] **meta**
	76.56	63.64	-	-	-	[40]
	84.90	83.10	85.30	82.70	-	[48]
IL-6	62.00	68.00	32.50	87.80	-	[35]
	73.00	72.00	77.35	62.60	70.20	[30] **meta**
Presepsin	75.00	84.00	-	-	-	[43] **meta**
	77.00	84.00	83.37	70.93	78.61	[30] **meta**
	73.00	84.00	-	-	-	[47] **meta**
CD64	89.00	87.00	-	-	-	[44] **meta**
	75.00	85.00	96.00	38.00	82.08	[49]
	77.00	75.00	-	-	-	[36]
	79.00	91.00	-	-	-	[50] **meta**
	93.00	87.00	95.85	68.98	84.05	[30] **meta**
	95.20	84.40	94.90	85.30	-	[48]
LBP	70.00	62.00	59.45	69.34	64.88	[30] **meta**
sTREM-1	86.30	53.20	80.40	63.60	-	[48]
	78.00	78.00	83.20	63.27	74.97	[30] **meta**
SIRE (3-way)	**90.43**	**84.15**	**76.08**	**93.55**	**81.98**	
SIRE (binary)	**90.04**	**86.76**	**87.34**	**91.12**	**88.91**	

A number of studies cited in Table 6.7 assess predictive capabilities in binary sepsis vs non-infectious SIRS [40], [45], [46], [49] which is most comparable in design to binary SIRS/sepsis investigations in the present study. It is indeterminable from the literature that the tests listed above would be capable of discriminating between SIRS/sepsis and unrelated sterile inflammation. PCT, for instance, is known to poorly discriminate between sepsis and sterile inflammation and is therefore not used on patients who have undergone surgery, and CRP is frequently elevated in a variety of autoimmune disorders [5]. The SIRE based serum test demonstrated superior predictive capabilities compared to other biomarkers arguably with the exception of CD64 which consistently displays superior sensitivity and positive predictive value, but poorer specificity and negative predictive value. However, this comes with the caveat that two of the meta-analysis papers explicitly state that numerous CD64 studies suffered from low methodological quality and bias [30], [50]. Sepsis and SIRS are highly dynamic conditions that advances through hyper-inflammatory and immunosuppressive phases meaning that the selection of concentration cut-off values is inherently flawed since any given marker is both expressed and supressed to varying degrees along the progression of the syndrome [36]. Machine learning is not as affected by this provided patient samples across every stage of sepsis development are used to construct predictive models.

Binary classification has demonstrated that it is possible to achieve the desired PPV and specificity of above 85% for sepsis detection as specified in the literature [18], [26]. However, sensitivity and NPV achieved, while still high, may need to be improved upon to gain clinical acceptance. The precise pre-analytical steps taken, such as sample collection method, are not known for the present study, but it is possible that

mitigating these sources of error and optimising and standardising preanalytical factors SIRE test sensitivity and NPV can improve. For instance, levels of pathogen cfDNA in blood and urine are known to be influenced by preanalytical factors such as sample collection tube, processing delays, processing method, and sample volume [51]. Since pathogen cfDNA is believed to be an important factor in SIRE test discrimination, it is therefore likely that control of these factors would lead to more accurate sepsis detection (section 6.3.2.2). Lastly, the SIRE test presented here performs better than any other comparable laboratory biomarker test identified in Table 6.7 in terms of predictive capabilities, particularly NPV, and practical use due to simplicity of operation. This suggests that with further refinement this technique could be a top competitor in rapid sepsis testing.

6.3.3. Infectious Pathogen Determination

In order to prescribe effectual antibiotics and combat the rise of antimicrobial resistance through over usage of broad spectrum antibiotics it is desirable to know the specific pathogen(s) present during sepsis [52]. While section 6.3.2. determined that the presence of infection could be detected spectroscopically, possibly in part due to DNA fragments, the task of distinguishing the culprit is far more complex. Causative bacteria vary widely across all sepsis cases the world over [53]. The patient data supplied with the patient sample set identifies 11 different known pathogen strains and a further 8 unknown strains within the 43 sepsis samples used in the present study alone. In 9 cases patients were suffering from multiple concurrent infections. The different pathogen strains encountered in this study are presented in Table 6.8 along with the number of occurrences within the sample set. A PCA was carried out on the

four most prevalent bacteria species using spectra of patients who had been infected by only a single known pathogen (Figure 6.7). No clear separation between groups could be observed. The first most obvious reason for this is the small number of samples available to conduct this particular analysis. The second is that the pathogens themselves are not present in the serum as these will have been lost during centrifugation. Thus, analysis must depend on the detection of molecular species shed by the pathogen. It is likely these molecular species exist in low concentrations in serum. Moreover, the biochemicals shed would need to be specific to the causative pathogen and not just generic microbial by-products. Perhaps a more realistic and almost equally constructive goal would be the classification of causative microbial agents into their broad pathogen categories such as bacteria, fungi, or viruses [3], [18]. This added function would be of great benefit in narrowing down causative pathogen species and particularly for viral related sepsis incidences which currently rely on diagnosis by exclusion [2], [26].

Table 6.8. *List of all the pathogen species encountered and the frequency of occurrence.*

Pathogen	Frequency
Escherichia Coli (*E. Coli*)	7
Haemophilus Influenzae (*H. Inf*)	3
Candida species (*Cand.*)	3
Pseudomonas Aeruginosa (*P. Aeruginosa*)	3
Enterococcus Faecalis (*E. faecalis*)	2
Streptococcus species (*Strep.*)	2
Staphylococcus species (*Staph.*)	2
Clostridium species (*Clost.*)	1
Faeces CDT	1
Klebsiella species (*Kleb.*)	1
Bacteriodes Fragilis (*B. Fragilis*)	1
Unknown	8

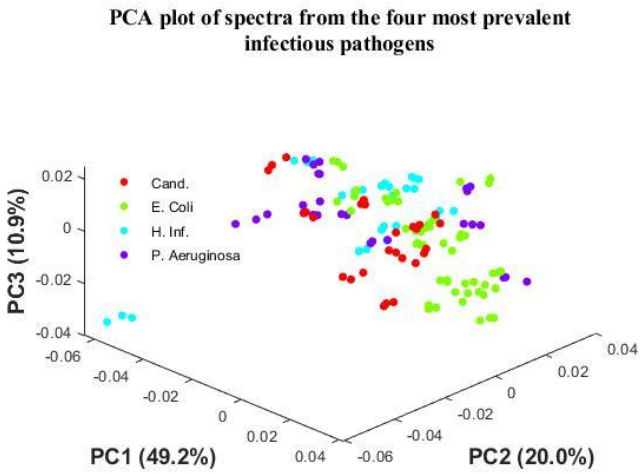


Figure 6.7. *PCA plot of spectra from the four most prevalent pathogen species encountered in the study. Separation between classes was not apparent.*

The silicon IREs may benefit from surface functionalisation as a means to either detect or rule out some of the more prominent sepsis culprits [54]. For instance, Mura *et al.* demonstrate the use of Single Side Polished silicon wafers coated in functionalised mesoporous titania films for the detection of E. Coli in water and food samples [55]. However, sepsis/SIRS/healthy disease state classification when carried out in conjunction with microbe detection may be impaired due to the underlying risk of Mie scattering caused by cells unless some corrective action is taken (e.g., Extended Multiplicative Signal Correction) [56].

6.3.4. Study Limitations

It is important to recognise some of the caveats which concern the validity of the present study to fully appreciate the potential of the SIRE as a sepsis diagnostic test. Firstly, samples analysed in this study were acquired on the day of sepsis diagnosis. To conclusively determine if the SIRE serum test can outperform the current status-quo the test must be able to accurately predict sepsis from serum samples

acquired in the days leading up to the time point a sepsis diagnosis can be made using current methods. Secondly, predictive models are built on the assumption that all the diagnoses made on the source patient samples are sound clinical assessments, although all diagnoses were reviewed and confirmed by a clinical advisory panel adding credence to the validity of the sample set. However, it has been observed that minor subjective differences in definitions heavily impacts apparent incidence of sepsis [15]. Lastly, investigations aimed at determining causative pathogens (Section 6.3.3) suffered from low sample size. Increased sample size would determine if specific pathogens could be delineated from blood serum alone, or at least if broad pathogen categories could be determined (e.g., viral, fungal, bacterial, etc.).

6.4. Conclusion

The crucial aspects in an effective sepsis diagnosis are speed and assurance in the result. A low-cost high-throughput ATR-FTIR based test could certainly provide rapid diagnosis of approximately two hours when factoring in the most time limiting factors in the process: sample collection and centrifugation. Although the test was not able to achieve NPV or sensitivity of >98% as specified in the literature [18], [26], the test is reliable and does possess good predictive capabilities (specificity and PPV >85%) comparable, or better, than existing blood biomarkers which have gained mainstream use in the clinic. This suggests the test could be commercially competitive and it is believed that with further refinement to minimise error sources, such as by utilising high purity FZ Si, hydrophilic surface treatments, and optimisation of pre-

analytical factors, test sensitivity and NPV can be improved. There is no cut-off value to choose from for the SIRE serum test which inherently gives an advantage over single biomarker tests as a cut-off value fails to capture the dynamic nature of sepsis. The SIRE serum test also requires very low sample volumes (minimum 9µl) which is of particular importance in the screening of neonates. This is a significant improvement from the notoriously slow (2-5 days) [18], [57] gold standard blood cultures that require large sample volumes to obtain reasonable sensitivity and specificity [58]. To realise the full potential of a SIRE serum test; future studies need to recruit larger patient cohorts to investigate specific or broad-based pathogen determination capabilities and explore the possibility of silicon surface functionalisation techniques to target specific pathogens.

6.5. References

[1] S. L. Fan, M. John Lee, and D. G. Remick, "Diagnosing Sepsis - The Role of Laboratory Medicine," *Physiol. Behav.*, vol. 460, no. 1, pp. 203–210, 2016, doi: 10.1016/j.physbeh.2017.03.040.

[2] M. I. El-Amir, M. A. El-Feky, D. A. A. Elwafa, and E. A. Abd-Elmawgood, "Rapid diagnosis of neonatal sepsis by PCR for detection of 16s rRNA gene, while blood culture and PCR results were similar in e.Coli-predominant EOS cases," *Infect. Drug Resist.*, vol. 12, pp. 2703–2710, 2019, doi: 10.2147/IDR.S213958.

[3] N. Schaub, R. Frei, and C. Mueller, "Addressing unmet clinical needs in the early diagnosis of sepsis," *Swiss Med. Wkly.*, vol. 141, no. July, pp. 1–4, 2011, doi: 10.4414/smw.2011.13244.

[4] M. Inada-Kim, B. Page, I. Maqsood, and C. Vincent, "Defining and measuring suspicion of sepsis: An analysis of routine data," *BMJ Open*, vol. 7, no. 6, pp. 1–7, 2017, doi: 10.1136/bmjopen-2016-014885.

[5] K. Reinhart, M. Bauer, N. C. Riedemann, and C. S. Hartog, "New approaches to sepsis: Molecular diagnostics and biomarkers," *Clin. Microbiol. Rev.*, vol. 25, no. 4, pp. 609–634, 2012, doi: 10.1128/CMR.00016-12.

[6] P. Comstedt, M. Storgaard, and A. T. Lassen, “The systemic inflammatory response syndrome (SIRS) in acutely hospitalised medical patients: A cohort study,” *Scand. J. Trauma. Resusc. Emerg. Med.*, vol. 17, no. 1, pp. 1–6, 2009, doi: 10.1186/1757-7241-17-67.

[7] A. Lever and I. Mackenzie, “Sepsis: Definition, epidemiology, and diagnosis,” *Br. Med. J.*, vol. 335, no. 7625, pp. 879–883, 2007, doi: 10.1136/bmj.39346.495880.AE.

[8] NICE, *Sepsis: recognition, assessment and early management*, no. 51. 2016.

[9] F. Gül, M. K. Arslantaş, İ. Cinel, and A. Kumar, “Changing definitions of sepsis,” *Turk Anesteziyoloji ve Reanimasyon Dern. Derg.*, vol. 45, no. 3, pp. 129–138, 2017, doi: 10.5152/TJAR.2017.93753.

[10] J. A. Antonow, K. Hansen, C. A. Mckinstry, and C. L. Byington, “Sepsis evaluations in hospitalized infants with bronchiolitis,” *Pediatr. Infect. Dis. J.*, vol. 17, no. 3, pp. 231–236, Mar. 1998, doi: 10.1097/00006454-199803000-00011.

[11] L. W. Andersen, J. Mackenhauer, J. C. Roberts, K. M. Berg, M. N. Cocchi, and M. W. Donnino, “Etiology and Therapeutic Approach to Elevated Lactate,” *Mayo Clin. Proc.*, vol. 88, no. 10, pp. 1127–1140, 2014, doi: 10.1016/j.mayocp.2013.06.012.Etiology.

[12] L. M. Frydrych, F. Fattahi, K. He, P. A. Ward, and M. J. Delano, “Diabetes and sepsis: Risk, recurrence, and ruination,” *Front. Endocrinol. (Lausanne).*, vol. 8, no. OCT, 2017, doi: 10.3389/fendo.2017.00271.

[13] B. G. Yipp and B. W. Winston, “Sepsis without SIRS is still sepsis,” *Ann. Transl. Med.*, vol. 3, no. 19, pp. 19–20, 2015, doi: 10.3978/j.issn.2305-5839.2015.11.13.

[14] J. L. Vincent *et al.*, “Sepsis in European intensive care units: Results of the SOAP study,” *Crit. Care Med.*, vol. 34, no. 2, pp. 344–353, 2006, doi: 10.1097/01.CCM.0000194725.48928.3A.

[15] P. M. C. K. Klouwenberg, D. S. David, M. J. M. Bonten, and O. L. Cremer, “Classification of sepsis, severe sepsis and septic shock: The impact of minor variations in data capture and definition of SIRS criteria,” *Intensive Care Med.*, vol. 38, no. 5, pp. 811–819, 2012, doi: 10.1007/s00134-012-2549-5.

[16] K. M. Kaukonen, M. Bailey, D. Pilcher, D. J. Cooper, and R. Bellomo, “Systemic inflammatory response syndrome criteria in defining severe sepsis,” *N. Engl. J. Med.*, vol. 372, no. 17, pp. 1629–1638, 2015, doi: 10.1056/NEJMoa1415236.

[17] S. E. Calvano, S. M. Coyle, K. S. Barbosa, P. S. Barie, and S. F. Lowry, “Multivariate analysis of 9 disease-associated variables for outcome prediction in patients with sepsis,” *Arch. Surg.*, vol. 133, no. 12, pp. 1347–1350, 1998, doi: 10.1001/archsurg.133.12.1347.

[18] M. Sinha, J. Jupe, H. Mack, T. P. Coleman, S. M. Lawrence, and I. Fraley,

"Emerging Technologies for Molecular Diagnosis of Sepsis," *Clin. Microbiol. Rev.*, vol. 31, no. 2, pp. 1–26, 2018.

[19] H. M. Giannini *et al.*, "A Machine Learning Algorithm to Predict Severe Sepsis and Septic Shock: Development, Implementation, and Impact on Clinical Practice," *Crit. Care Med.*, vol. 47, no. 11, pp. 1485–1492, 2019, doi: 10.1097/CCM.0000000000003891.

[20] P. E. Marik and A. M. Taeb, "SIRS, qSOFA and new sepsis definition," *J. Thorac. Dis.*, vol. 9, no. 4, pp. 943–945, 2017, doi: 10.21037/jtd.2017.03.125.

[21] B. R. Smith, M. J. Baker, and D. S. Palmer, "PRFFECT: A versatile tool for spectroscopists," *Chemom. Intell. Lab. Syst.*, vol. 172, no. October 2017, pp. 33–42, 2018, doi: 10.1016/j.chemolab.2017.10.024.

[22] H. J. Butler, B. R. Smith, R. Fritzsch, P. Radhakrishnan, D. S. Palmer, and M. J. Baker, "Optimised spectral pre-processing for discrimination of biofluids via ATR-FTIR spectroscopy," *Analyst*, vol. 143, no. 24, pp. 6121–6134, 2018, doi: 10.1039/c8an01384e.

[23] N. K. Afseth and A. Kohler, "Extended multiplicative signal correction in vibrational spectroscopy, a tutorial," *Chemom. Intell. Lab. Syst.*, vol. 117, pp. 92–99, 2012, doi: 10.1016/j.chemolab.2012.03.004.

[24] Z. Movasaghi, S. Rehman, and I. U. Rehman, "Fourier transform infrared (FTIR) spectroscopy of biological tissues," *Appl. Spectrosc. Rev.*, vol. 43, no. 5, pp. 134–179, 2008, doi: 10.1080/05704928.2016.1230863.

[25] L. Srinivasan and M. C. Harris, "New technologies for the rapid diagnosis of neonatal sepsis," *Curr. Opin. Pediatr.*, vol. 24, no. 2, pp. 165–171, 2012, doi: 10.1097/MOP.0b013e3283504df3.

[26] P. C. Ng and H. S. Lam, "Biomarkers for Late-Onset Neonatal Sepsis: Cytokines and Beyond," *Clin. Perinatol.*, vol. 37, no. 3, pp. 599–610, 2010, doi: 10.1016/j.clp.2010.05.005.

[27] T. G. Connell, M. Rele, D. Cowley, J. P. Buttery, and N. Curtis, "How reliable is a negative blood culture result? Volume of blood submitted for culture in routine practice in a children's hospital," *Pediatrics*, vol. 119, no. 5, pp. 891–896, 2007, doi: 10.1542/peds.2006-0440.

[28] F. Bonnier *et al.*, "Screening the low molecular weight fraction of human serum using ATR-IR spectroscopy," *J. Biophotonics*, vol. 9, no. 10, pp. 1085–1097, 2016, doi: 10.1002/jbio.201600015.

[29] C. L. Song, M. Z. Vardaki, R. D. Goldin, and S. G. Kazarian, "Fourier transform infrared spectroscopic imaging of colon tissues: evaluating the significance of amide I and C–H stretching bands in diagnostic applications with machine learning," *Anal. Bioanal. Chem.*, vol. 411, no. 26, pp. 6969–6981, 2019, doi: 10.1007/s00216-019-02069-6.

[30] Y. Liu, J. Hou, Q. Li, K. Chen, S. N. Wang, and J. Wang, "Biomarkers for diagnosis of sepsis in patients with systemic inflammatory response syndrome :

a systematic review and meta - analysis," *Springerplus*, 2016, doi: 10.1186/s40064-016-3591-5.

[31] Y. Long *et al.*, "Diagnosis of Sepsis with Cell-free DNA by Next-Generation Sequencing Technology in ICU Patients," *Arch. Med. Res.*, vol. 47, no. 5, pp. 365–371, 2016, doi: 10.1016/j.arcmed.2016.08.004.

[32] S. Hamaguchi *et al.*, "Origin of circulating free DNA in sepsis: Analysis of the CLP mouse model," *Mediators Inflamm.*, vol. 2015, 2015, doi: 10.1155/2015/614518.

[33] A. Rhodes and M. Cecconi, "Cell-free DNA and outcome in sepsis," *Crit. Care*, vol. 16, no. 6, 2012, doi: 10.1186/cc11508.

[34] M. Chaudhry, H., Zhou, J., Zhong, Y., Ali, M. M., McGuire, F., Nagarkatti, P. S., & Nagarkatti, "Role of Cytokines as a Double-edged Sword in Sepsis," vol. 27, no. 6, pp. 669–684, 2015.

[35] E. L. Tsalik *et al.*, "Discriminative value of inflammatory biomarkers for suspected sepsis," *J. Emerg. Med.*, 2012, doi: 10.1016/j.jemermed.2011.05.072.

[36] S. Y. Cho and J. H. Choi, "Biomarkers of Sepsis," *Crit. Rev. Clin. Lab. Sci.*, vol. 50, no. 1, pp. 23–36, 2013, doi: 10.3947/ic.2014.46.1.1.

[37] C. M. Florkowski, "Sensitivity , Specificity , Receiver-Operating Characteristic (ROC) Curves and Likelihood Ratios : Communicating the Performance of Diagnostic Tests," vol. 29, no. August, pp. 83–87, 2008.

[38] J. L. Wynn and R. A. Polin, "Progress in the management of neontal sepsis: the importance of a consensus definition," *Nat. Publ. Gr.*, vol. 83, no. 1–1, pp. 13–15, 2017, doi: 10.1038/pr.2017.224.

[39] M. P. Griffin and J. R. Moorman, "Toward the Early Diagnosis of Neonatal Sepsis and Sepsis-Like Illness Using Novel Heart Rate Analysis," *J. Am. Acad. Pediatr.*, vol. 107, no. 1, pp. 97–104, 2001, doi: 10.1542/peds.107.1.97.

[40] E. Keçe, E. Yaka, S. Yilmaz, N. Ö. Doğan, C. Alyeşil, and M. Pekdemir, "Comparison of diagnostic and prognostic utility of lactate and procalcitonin for sepsis in adult cancer patients presenting to emergency department with systemic inflammatory response syndrome," *Turkish J. Emerg. Med.*, vol. 16, no. 1, pp. 1–7, 2016, doi: 10.1016/j.tjem.2016.02.003.

[41] A. J. Singer *et al.*, "Diagnostic characteristics of a clinical screening tool in combination with measuring bedside lactate level in emergency department patients with suspected sepsis," *Acad. Emerg. Med.*, vol. 21, no. 8, pp. 853–857, 2014, doi: 10.1111/acem.12444.

[42] Y. J. Park *et al.*, "Serum lactate upon emergency department arrival as a predictor of 30-day in-hospital mortality in an unselected population," *PLoS One*, vol. 13, no. 1, pp. 1–14, 2018, doi: 10.1371/journal.pone.0190519.

[43] C. C. Wu *et al.*, "Comparison of diagnostic accuracy in sepsis between presepsin, procalcitonin, and C-reactive protein: a systematic review and meta-analysis," *Ann. Intensive Care*, vol. 7, no. 1, 2017, doi: 10.1186/s13613-017-

0316-z.

[44] C. F. Yeh, C. C. Wu, S. H. Liu, and K. F. Chen, “Comparison of the accuracy of neutrophil CD64, procalcitonin, and C-reactive protein for sepsis identification: a systematic review and meta-analysis,” *Ann. Intensive Care*, vol. 9, no. 1, 2019, doi: 10.1186/s13613-018-0479-2.

[45] S. Pradhan *et al.*, “The role of C-reactive protein as a diagnostic predictor of sepsis in a multidisciplinary Intensive Care Unit of a tertiary care center in Nepal,” *Indian J. Crit. Care Med.*, vol. 20, no. 7, pp. 417–420, Jul. 2016, doi: 10.4103/0972-5229.186226.

[46] P. Póvoa *et al.*, “C-reactive protein as an indicator of sepsis,” *Intensive Care Med.*, vol. 24, no. 10, pp. 1052–1056, 1998, doi: 10.1007/s001340050715.

[47] Y. Kondo, Y. Umemura, K. Hayashida, Y. Hara, M. Aihara, and K. Yamakawa, “Diagnostic value of procalcitonin and presepsin for sepsis in critically ill adult patients : a systematic review and meta-analysis,” *J. Intensive Care*, vol. 7, no. 22, pp. 1–13, 2019.

[48] S. Gibot *et al.*, “Combination biomarkers to diagnose sepsis in the critically ill patient,” *Am. J. Respir. Crit. Care Med.*, vol. 186, no. 1, pp. 65–71, 2012, doi: 10.1164/rccm.201201-0037OC.

[49] S. T. Dal Ponte *et al.*, “Diagnostic accuracy of CD64 for sepsis in emergency department,” *J. Glob. Infect. Dis.*, vol. 10, no. 2, pp. 42–46, Apr. 2018, doi: 10.4103/jgid.jgid_130_16.

[50] J. Cid, R. Aguinaco, R. Sánchez, G. García-Pardo, and A. Llorente, “Neutrophil CD64 expression as marker of bacterial infection: A systematic review and meta-analysis,” *J. Infect.*, vol. 60, no. 5, pp. 313–319, 2010, doi: 10.1016/j.jinf.2010.02.013.

[51] K. Murugesan *et al.*, “Investigation of Preanalytical Variables Impacting Pathogen Cell-Free DNA in Blood and Urine,” *J. Clin. Micorbiology*, vol. 57, no. 11, pp. 1–13, 2019.

[52] I. S. Pradipta *et al.*, “Antibiotic Resistance in Sepsis Patients : Evaluation and Recommendation of Antibiotic Use,” vol. 5, no. 6, 2013, doi: 10.4103/1947-2714.114165.

[53] M. P. Morgan *et al.*, “Sepsis patients with first and second-hit infections show different outcomes depending on the causative organism,” *Front. Microbiol.*, vol. 7, no. 207, pp. 1–9, 2016, doi: 10.3389/fmicb.2016.00207.

[54] S. Mura, G. Greppi, A. M. Roggio, L. Malfatti, and P. Innocenzi, “Polypeptide binding to mesostructured titania films,” *Microporous Mesoporous Mater.*, vol. 142, no. 1, pp. 1–6, 2011, doi: 10.1016/j.micromeso.2010.10.047.

[55] S. Mura *et al.*, “FTIR nanobiosensors for Escherichia coli detection,” *Beilstein J. Nanotechnol.*, vol. 3, no. 1, pp. 485–492, 2012, doi: 10.3762/bjnano.3.55.

[56] J. A. Kimber and S. G. Kazarian, “Spectroscopic imaging of biomaterials and biological systems with FTIR microscopy or with quantum cascade lasers,”

Anal. Bioanal. Chem., vol. 409, no. 25, pp. 5813–5820, 2017, doi: 10.1007/s00216-017-0574-5.

[57] P. E. Marik, “Don’t miss the diagnosis of sepsis!,” *Crit. Care*, vol. 18, no. 529, pp. 1–3, 2014, doi: 10.1186/s13054-014-0529-6.

[58] T. E. Sweeney, O. Liesenfeld, and L. May, “Diagnosis of bacterial sepsis: why are tests for bacteremia not sufficient?,” *Expert Rev. Mol. Diagn.*, vol. 19, no. 11, pp. 959–962, 2019, doi: 10.1080/14737159.2019.1660644.

Chapter 7: Concluding Remarks and Future Perspectives

7.1. Conclusion

ATR-FTIR spectroscopy is a clinically useful technique with well recognised diagnostic utility. Technical limitations have prevented the translation of ATR-FTIR spectroscopy to the clinical landscape. These limitations have been addressed in the present studies through the creation of a novel high-throughput ATR-FTIR serum diagnostics platform. This advancement in the technology has led to a prospective clinical validation study for the detection and rapid triage of brain cancers. This, together with the promising sepsis detection study results presented here, indicate the high-throughput SIREs can be a valuable clinical tool. High-throughput IR spectroscopy has been enabled through development of a simple, low cost, disposable microscope-sized silicon IRE slide called a SIRE was developed along with an accessory unit to interface the SIRE with IR spectrometers. Design specifics and rationale has been presented in chapter three. Reduction of IR beam pathlength through the SIRE successfully mitigated the optical limitations of silicon. V-grooves etched into the beam facing surface permitted a single internal reflection and use of thin (375µm) double-side polished Czochralski silicon reduced pathlength. V-groove width (250µm) and pitch (25µm) was optimised for spectral signal to noise ratio (136.7 SNR) and reproducibility (2.1 amide I RSD%, 1.5 amide II RSD%). It was confirmed that none of the design aspects, such as micron scale features that may scatter IR light, interfered with spectral quality. A design for manufacturability approach was employed in the design of SIREs whereby consideration has been given to the availability of raw materials, feasibility of scaled productions, ease of assembly, and unit cost of production. This entailed use of low cost, ubiquitous, and processable

materials including a silicon IRE substrate, PLA or ABS receptacle to protect the IRE crystal, and a laminated label with RH07 adhesive to restrain the components. Scalable microfabrication, injection moulding, label printing, and pick and place systems can be employed for the rapid low-cost mass production of SIRE slides. Similarly, an accessory unit was designed to be cheap and easy to manufacture and assemble but is not a consumable item like the SIRE so does not require the same degree of automated assembly.

SIRE spectrum reproducibility was optimised in chapter four by characterising intrinsic and extrinsic sources of variance and by defining pre-analytical sample preparation procedures. Batched procedures were established specifically to enable a high-throughput workflow. An incubator set to a stable temperature of 30°C with fan set to 60CFM was able to simultaneously prepare 48 samples for analysis in two minutes or less. The resultant workflow regime was capable of analysing 24 (or 72 if no instrument repeats are taken) samples within an 8-hour period compared to a rate of 8 samples within the same timeframe using a standard Di IRE. This equates to a 300% increase in sample processing rate. In a clinical setting, background scans could be acquired in advance while patient samples are being centrifuged to further improve time efficiency. Moreover, acquisition times could be shortened by reducing spectral resolution window and/or number of coadded scans. It was found that negligible variance existed between spectra of SIREs despite nonhomogeneous distribution of impurities across silicon wafers except for SIREs fabricated at wafer peripheries which exhibit elevated interstitial oxygen concentrations and impaired spectral quality. Consequently, a manufacturing yield of 86.67% per wafer has been estimated. Inter IRE variance is low across both the pure silicon spectra and serum sample spectra.

However, spectral reproducibility of samples may not be stable over the course of multiple days. Going forward, serum drying by lyophilisation may be the best approach to mitigating this problem. Lack of consistency in serum droplet spreading has clear detrimental impacts on sample spectrum reproducibility and quality, but standardised pipetting techniques can negate this. Furthermore, surface treatment using Piranha solution will salinize SIRE surface resulting in a highly wettable sampling area and consequently uniform sample coverage. It may be beneficial to explore how more stable surface treatments, such as plasma activation, might impact SIRE performance.

SIRE performance was characterised and compared to spectral quality and analytical capabilities of other IRE substrates. Chiefly among these was diamond as diamond is widely respected as a high-quality IRE substrate. Chapter five found SIREs detection and quantification limits comparable to a Di IRE, and in some instances better, depending on the target molecule even though spectra were collected from various SIREs in comparison to a single Di IRE. Moreover, SIRE spectra are less distorted by anomalous dispersion artefacts than Di IRE spectra. The intrinsic absorption profiles of SIREs and the Di IRE differ greatly alluding to a difference in sensitivity across the mid-IR spectrum and by extension to specific bonding environments. This was confirmed through clinically relevant serum concentration studies of three of the four biomolecular classes (proteins, carbohydrates, lipids) with which sections of the IR spectrum can be apportioned into. Glucose, triglycerides, and albumin were selected as they are archetypal serum carbohydrate, lipids, and protein, respectively. Lactate was also selected since it possesses both characteristic carbohydrate and lipid bonding environments. All analytes also have clinical

indications for sepsis. Silicon is a strong attenuator of light at 610cm^{-1} and 1107cm^{-1} which impairs detection and quantification of glucose in comparison to a Di IRE. Conversely, Di IRE attenuates IR light at 1640cm^{-1} and 3300cm^{-1} which impairs detection and quantification of albumin in comparison to SIREs. Di IRE and SIREs are comparable in the detection and quantification of triglycerides, but the Di IRE was better at detecting and quantifying lactate concentrations. Multivariate analysis (PLSR) was able to overcome the lack of SIRE sensitivity at localised wavenumbers producing excellent predictive capabilities at clinically relevant concentration ranges for all analytes under investigation including glucose. The evidence provided here suggests SIREs are capable of rapid point of care monitoring of the complex and dynamic biochemical shifts underpinning sepsis progression without need for reagents or complicated procedures.

A 148 patient study presented in chapter six confirmed that SIREs are clinically useful for their ability to rapidly discriminate between sepsis and closely related health conditions using only serum. The SIRE test possessed Negative predictive value (93.55%), positive predictive value (76.08%), sensitivity (84.15%), specificity (90.43%), and accuracy (81.98%) for the detection of sepsis demonstrating that the high-throughput ATR-FTIR serum spectroscopy platform developed herein can discriminate between sepsis from SIRS and control patients (healthy patients post-high risk surgery) returning rapid and reliable results. These predictive values perform well compared to common laboratory biomarker tests for sepsis, but do not quite satisfy the desirable predictive values as cited in the literature (NPV and sensitivity >98%, PPV and specificity >85%). In contrast to the gold standard blood culture method for diagnosing sepsis, this approach is rapid (<2 hour compared to 3-5 days to result) and

potentially less expensive. However, it is not possible to directly compare SIRE and blood culture costs at this early development stage as the true cost of a SIRE test may be influenced by a multitude of factors (e.g., cost of packaging, distribution, quality assurance, regulatory clearance, etc.) before reaching the clinical market. Nevertheless, these findings indicate there is potential for SIREs to fulfil an unmet clinical need whereby clinicians require an effective sepsis/SIRS triage tool. While it is envisaged that a SIRE sepsis test would ideally be performed in the emergency department to conserve time, there seems no reason the approach could not be used in other clinical settings, such as the hospital laboratory or the intensive care unit, for other clinical applications. Aspects of the workflow can be improved upon to reduce time to result. Specifically, it is feasible result turnaround could approach 1 hour with employment of rapid serum separation and sample dehydration techniques. Use of machine learning algorithms enable this method to address the notorious heterogeneous nature of sepsis by evolving predictive models as more data is fed into the system. The low sample volumes required (9µl) are especially attractive for the diagnosis of neonatal sepsis where issues with the blood volume required for blood cultures are well known. Identification of causative pathogens could not be demonstrated, but sample sizes were low as distinct pathogen species were too numerous within the sepsis patient sample set. In addition to meeting unmet clinical needs regarding timely sepsis diagnosis, the technology presented here has seen demonstrable potential to meet unmet clinical needs in the timely diagnosis of brain cancer through parallel studies. Promising results in two distinct diagnostic applications demonstrate the clinical utility of high-throughput ATR-FTIR spectroscopy as a general serum diagnostics platform.

In conclusion, the project undertaken aspired to create a clinically useful, low cost, rapid, high-throughput ATR-FTIR spectroscopic serum diagnostics platform. The developed technology presented herein meet this criteria and subsequent investigations substantiate the clinical utility through demonstration of analytical reliability, analyte quantification, and disease state classification. In particular, this technology has the potential to deliver a rapid and low-cost sepsis test thereby fulfilling an unmet clinical need. The work presented here represents a significant step toward the translation of IR spectroscopy to the clinical setting.

7.2. Future Directions

The results of these investigations demonstrate the reliability and advantages of a high-throughput ATR-FTIR serum spectroscopy platform and presents clear pathways for the translation of IR spectroscopy to the clinic. Imminent clinical trials by ClinSpec dx will establish the diagnostic capabilities of SIREs with respect to both predictive capabilities and suitability to clinical workflows. The claims made herein will be tested in the real-world. This will comprise multi-centre validation of pre- and post-analytical processes as well as test efficacy under intended use as evidence toward regulatory clearance for UK and EU markets. Clinical investigations will be registered with the Medicine and Healthcare products Regulatory Authority (MHRA) and the IVD must demonstrate compliance with the In Vitro Diagnostics Regulation (IVDR). This implies an ISO 13485 certified quality management system is in place and the device has been appropriately classified. As classification is dependant on associated

risk and intended use, SIRE serum tests may need to be classified and regulated differently depending on the clinical application they are to be used for. Classification definitions indicate that a brain cancer screening test would likely fall into class C as it is intended as a cancer screening tool. There is a chance a sepsis test may fall into class C or even the higher risk class D since failing to detect any transmissible agents could be life-threatening. However, the classification criterion is broad and may not be applicable as infectious agents are not yet being directly identified within the scope or the SIRE sepsis test.

Integration of user controls into a single General User Interface (GUI) whereby SIRE slide indexing and spectral acquisition are controlled from a monitor would be advantageous to help streamline the analysis procedure, but this would require coordination with every spectrometer manufacturer which would be challenging for a truly universal interface. However, universal spectrometer accessibility is not necessarily desirable for clinical applications as tests would be difficult to validate across multiple platforms. A GUI which returns the test results to the user in an accessible manner would also be necessary and more easily achievable. Determining what the clinician expects to see aside from a 'positive/negative' test result needs to be the first step here.

To enable storage of samples the longevity of the sample stability needs to be addressed. Samples may need to be dried in a different way, stored in a different way, or preserved in some other manner. Vacuum packing may be an option, albeit less practical. Alternatively, dehydration by lyophilisation may be an effective solution and could quite easily comply with a batched workflow process. Different silicon materials or surface treatments should be explored as this may further improve spectral

reproducibility. Hydrophilization by plasma activation may help enhance serum film homogeneity by establishing uniform surface conditions on the silicon and improving wettability. The use of dopant free FZ wafers could improve manufacturing yield by establishing crystal homogeneity. Preliminary cost estimates have deemed it feasible that this improvement would not violate the financial restrictions imposed on the SIRE slide device. Although a health economic assessment has helped clarify the financial constraints and cost effectiveness of SIRE testing with respect to brain cancer diagnostics, a separate health economic assessment needs to be carried out for the use of SIREs as a sepsis testing platform.

Sepsis is a dynamic condition that requires continuous monitoring of vital signs. Biomarker concentrations should also be continuously monitored, but this requires multiple tests, so the clinician is often limited to univariate non-specific biomolecule monitoring. Simultaneous determination of analyte concentration within serum by high-throughput ATR-FTIR needs to be demonstrated. This offers a clear advantage in sepsis monitoring as a clinician will be able to quickly determine a patient's status without need for ordering multiple tests. The three most commonly requested clinical tests to diagnose and manage sepsis patients are lactate, c-reactive protein, and procalcitonin. Ability to quantify the latter two within serum samples at clinically relevant concentration ranges using SIREs needs to be demonstrated before development of a simultaneous biomarker determination test. A simultaneous sepsis biomarker test should aim to at least quantify these three biomarkers. However, procalcitonin and c-reactive protein are present in low concentrations in serum so conventional analysis may not be sensitive enough for quantification. Surface functionalisation of the silicon sampling area to immobilise CRP and/or PCT may be

necessary to enhance sensitivity. Better yet, silicon surface functionalisation aimed at detection of causative sepsis pathogens would enable rapid administration of effectual antibiotics and undoubtedly improve patient mortality rates. This has the added benefit of working as a system whereby clinicians may rule out common pathogen culprits.

IR spectroscopy has clinical ability and is ready for translation into the clinical setting. It clearly has a multitude of applications within the clinic and is applicable to all clinical sample types (e.g., biofluids, tissue, and cells) and can aid in clinical decision making. Unmet clinical needs can be addressed effectively and inexpensively through IR spectroscopic methods. In particular, this thesis has illuminated brain cancer patients and sepsis patients as two key beneficiaries of SIRE serum testing in terms of the potential improvement to mortality rates and reduced economic burdens imposed on healthcare institutions. Theoretically, any proof-of-concept biofluid FTIR spectroscopy diagnostic could be performed on the SIREs presenting the possibility to address further unmet clinical needs in Alzheimer's disease, detection of other cancers (e.g., ovarian, breast, or bile duct), colitis, atherosclerosis, malaria, and more. The high-throughput ATR-FTIR spectroscopy method developed here presents a clear method to implement the diagnostic abilities of ATR-FTIR spectroscopy in the clinical workplace.

Chapter 8: Appendices

8.1. Appendix I – SIRE Design Iterations

Numerous SIRE design concepts were generated prior to the prototyping stage. These where overlooked either because they were not so easy to practically implement, or because they would be better suited to different applications.

8.1.1. Radial SIRE

Figures 8.1.(a) and (b) show a design where the SIRE sample compartments revolve around a spindle mounted upon the accessory module top plate, using radial movement instead of linear movement to position sample compartments above the IR aperture. The need to drill a hole through the SIRE itself would be challenging due to the brittleness of Si. However, this process may be possible to implement in a production environment.

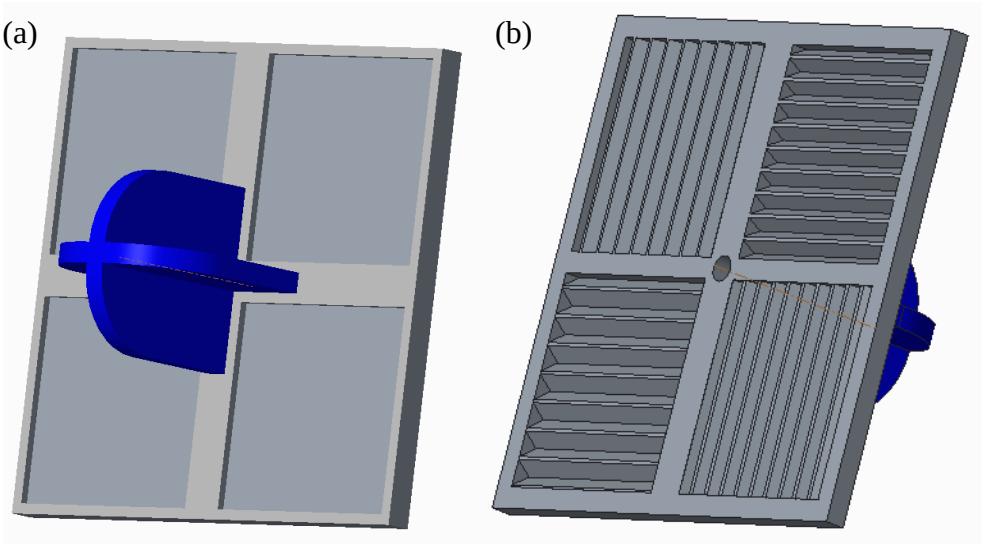


Figure 8.1. *Concept of a radial movement multi-compartmental SIRE. (a) Top view, the blue structure in the middle is a handle the user can grasp to move the device. (b) Bottom view, the device is placed onto a shaft through a hole drilled through the centre of the SIRE.*

8.1.2. Scaled SIRE production

Design concepts as shown in Figures 8.2 and 8.3 were a product of upscaling considerations. Figure 8.2 shows a device with individually diced SIREs. More devices may be manufactured (*ca.* ~45 more fully assembled multi-compartmental SIREs) by maximising the number of SIREs that may be fabricated from a single Si wafer. This has the obvious downside of having more discrete components, increasing the risk of defects and improper assembly and may require more complex fixation methods and eventually more complex automated assembly mechanisms when considering production scale up.

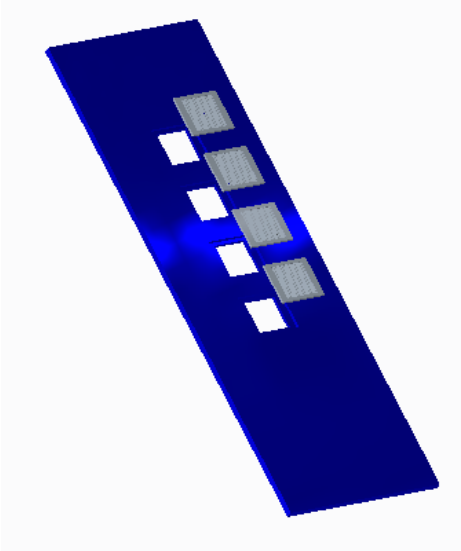


Figure 8.2. *Multi-compartmental slide concept with discretely diced SIREs.*

96-well microtiter plates are commonplace in clinical laboratories and are good for maximising sample throughput. A similar concept applied to SIRE slide design could understandably improve throughput capabilities whilst being compatible with clinical workflows (Figure 8.3). Adjacent wells need to be spaced 9mm from centre to centre to match the pitch on standard multichannel pipettes. This constraint determines the size of the SIRE component (Figure 8.3.(b)) which would result in large amounts

of waste material from the Silicon wafer. A simple solution to this is to construct 96-well plates with discrete SIRE components similar to those shown in Figure 8.2, though this would exacerbate assembly concerns raised using a discrete SIRE approach. In addition, more complicated translational XY movement is required to interface 96-well SIREs with an IR spectrometer.

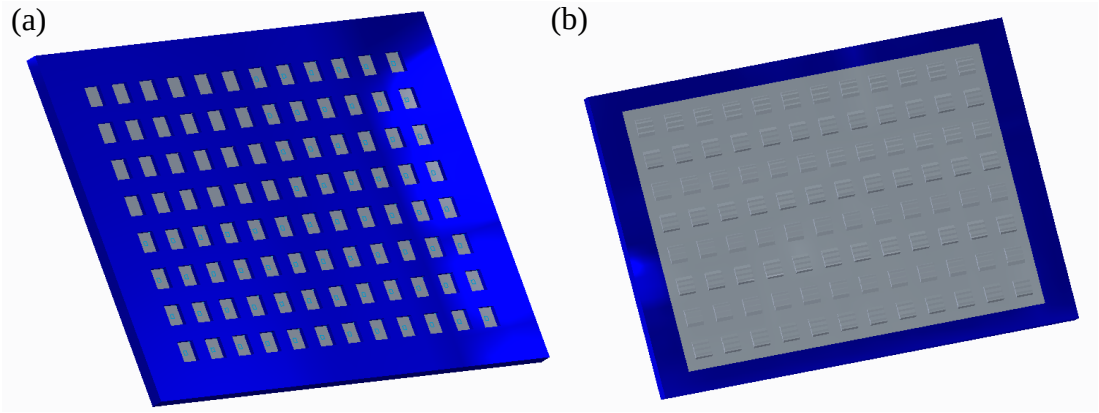


Figure 8.3. *Concept of a 96-compartment SIRE slide. (a) Top view. (b) Bottom view.*

8.1.3. Sample Tube SIRE Cap

ATR-FTIR is used in a variety of applications that require samples to be collected on site and preserved until they can be analysed in the laboratory such as in crime scene forensics or environmental monitoring studies. In some of these circumstances it is imperative that the sample has not been tampered with *en route* to analysis. A method to ensure this is to integrate a SIRE into the cap of an Eppendorf tube (Figures 8.4.(a) and (b)) so that the tube may then be sealed within the tube and analysed without ever having to reopen. This concept digresses from the focus of the present research and has not been developed further as it is not entirely suitable for

clinical applications where samples are generally collected in controlled procedures and environments, but it is none the less an interesting direction.

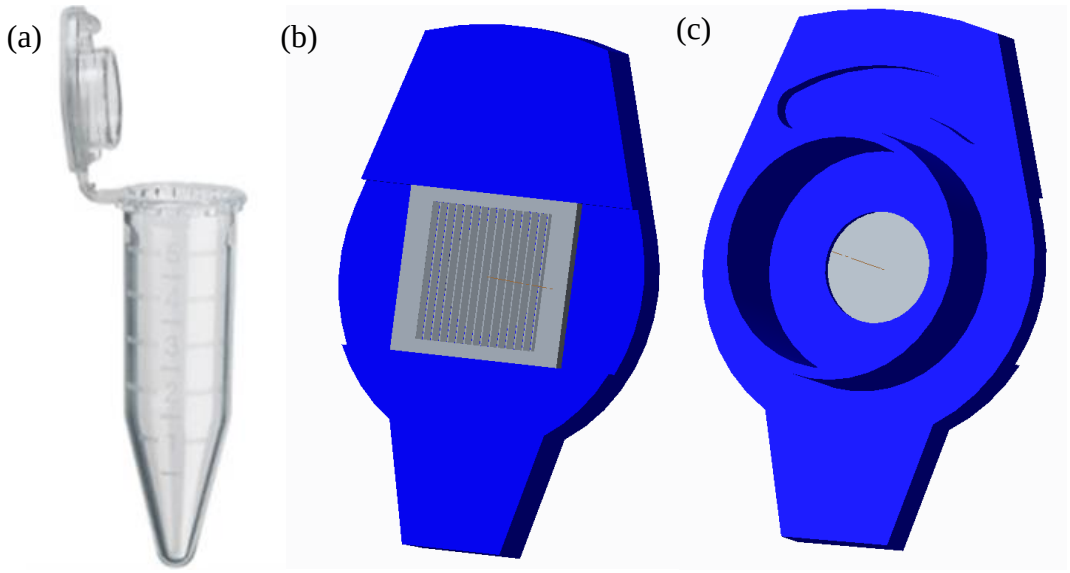


Figure 8.4. *Concept of a sample tube cap (blue) with a SIRE fixed within (grey) that allows samples to be stored within an analysed without ever opening the tube. (a) Eppendorf tube with cap open (a) top view of SIRE in Eppendorf tube cap. (b) bottom view.*

8.2. Appendix II – Accessory Module Design Iterations

An ATR accessory contains a system of highly engineered optics that deliver the beam to the IRE. Early designs aimed to replace only the metal IRE top-plate to take advantage of the existing optics within the PerkinElmer Spectrum2 spectrometer's accessory housing. This design was a basic manual linear translator orientated at 45° from to align with the internal optics of the Spectrum2 ATR accessory unit (Figures 8.5.(a) and (b)). The click-in-place system created a lot of mechanical

vibration which would throw the SIRE out of place making spectrum acquisition difficult.

(a)

(b)

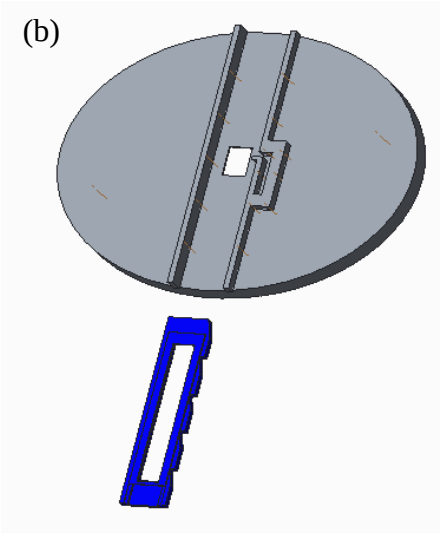


Figure 8.5. *Two early concepts for a Slide Indexing Unit to move the multi-compartmental SIRE above the IR beam aperture. (a) a rack and pinion is used to move the slide. (b) a click-in-place system is used to secure the slide in each position.*

Subsequent designs aimed to allow access to a wide variety of spectrometer instruments. Initially, efforts were directed at development of inhouse optics to deliver the optimum beam angle to SIREs. There was also a focus on automation, utilising motorised linear actuator mechanisms which permitted a higher level of control and reduced SIRE vibrations. Prior to this, the optimal incident beam angle was determined on a Cary 660 FTIR Agilent Technologies spectrometer with PIKE technologies VEEMAX variable angle specular reflection accessory. Dried HPS spectra was scanned on a SIRE and the angle of incidence adjusted from 30°-80° in 2° intervals (Figure 8.6). The highest signal intensity was achieved with an incident angle of 30°.

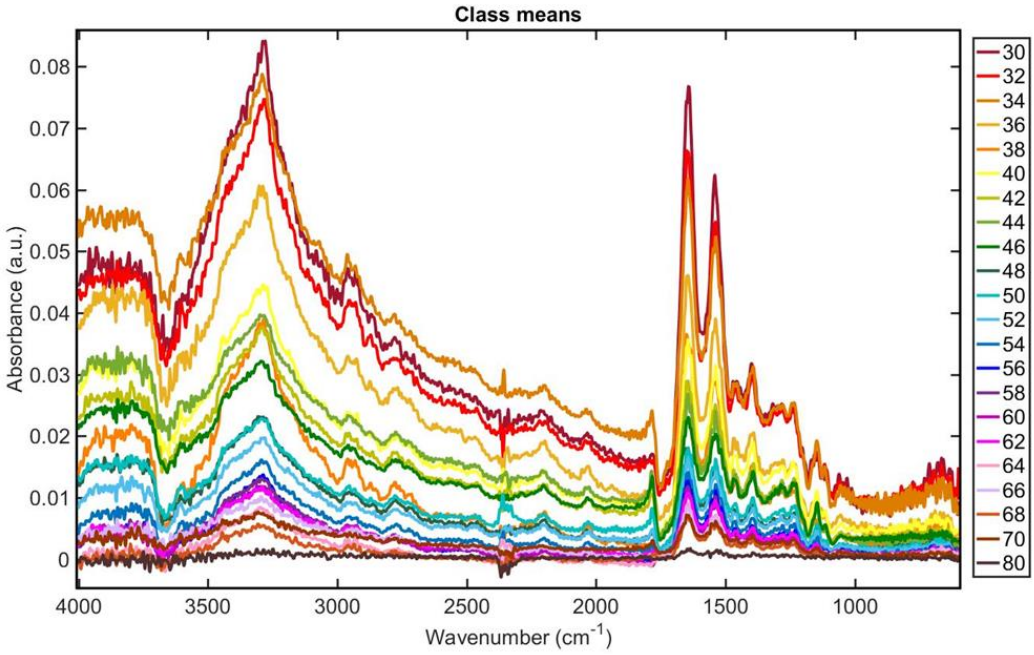
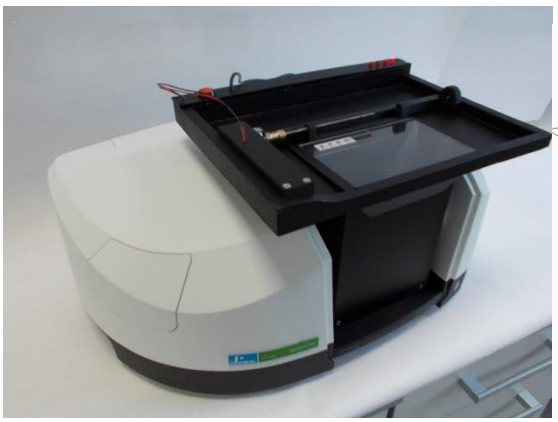


Figure 8.6. *HPS spectra acquired from a SIRE at angles of incidence varying from 30°-80°.*

The first automated accessory unit working prototype used an Arduino Nano (Arduino, Italy) and a motor driver were used to actuate a Maxon A-max stepper motor. This spun a lead screw that in turn moved a plastic carriage the SIRE rested within (Figures 8.7.(a) and (b)). The user inputs commands via push button which directs the linear actuator to index each successive SIRE compartment, positioning them above the IR beam aperture. The accessory housing was 3D printed in Aluminium using the Direct Metal Laser Sintering (DMLS) technique. Contained within were six flat gold mirrors directing the beam to the SIRE at an angle of 30°.

(a)



(b)

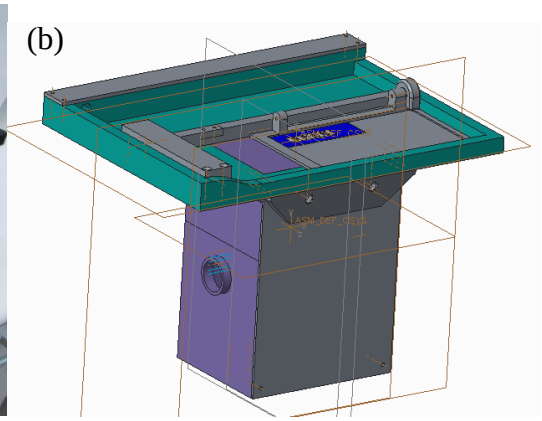


Figure 8.7. *Images of the first automated slide indexing unit. (a) attached within a PerkinElmer Spectrum2 spectrometer. (b) CAD model of the fully assembled design.*

The Specac Quest ATR unit features adjustable optics which resolved issues whereby different spectrometer manufacturers engineer different IR beam focal points into their products. The decision was thus made to develop a slide indexing unit that could exploited the optics of the Quest ATR (Figure 8.8). The initial slide indexer design used a rack and pinion to move the slide within a carriage. The unit was constructed entirely from PLA plastic which flexed while actuating resulting in considerable unwanted movement and vibration of the SIRE. Furthermore, the open-ended design of the carriage resulting in frequent dislodging of the SIRE. These downfalls were rectified in the final slide indexing unit design which utilised stiff anodised aluminium components, a closed top carriage, and slide guides to stabilise the drive shaft.

(a) (b)

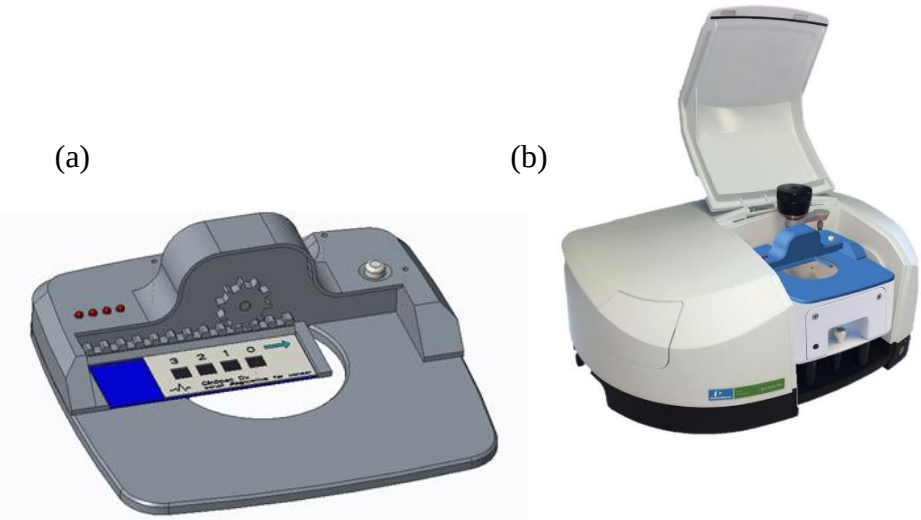


Figure 8.8. *Former slide indexing unit design intended to compliment the Quest ATR universal accessory module. (a) CAD drawing of second slide indexing unit prototype with rack and pinion mechanism revealed. (b) Slide indexing unit mounted on a Quest ATR in a Spectrum2 spectrometer.*

8.3. Appendix III – Schematics of Accessory Module Electronics

Figure 8.9 presents a schematic diagram of the electronic circuitry used to control the slide indexing unit of the accessory unit developed in this project.

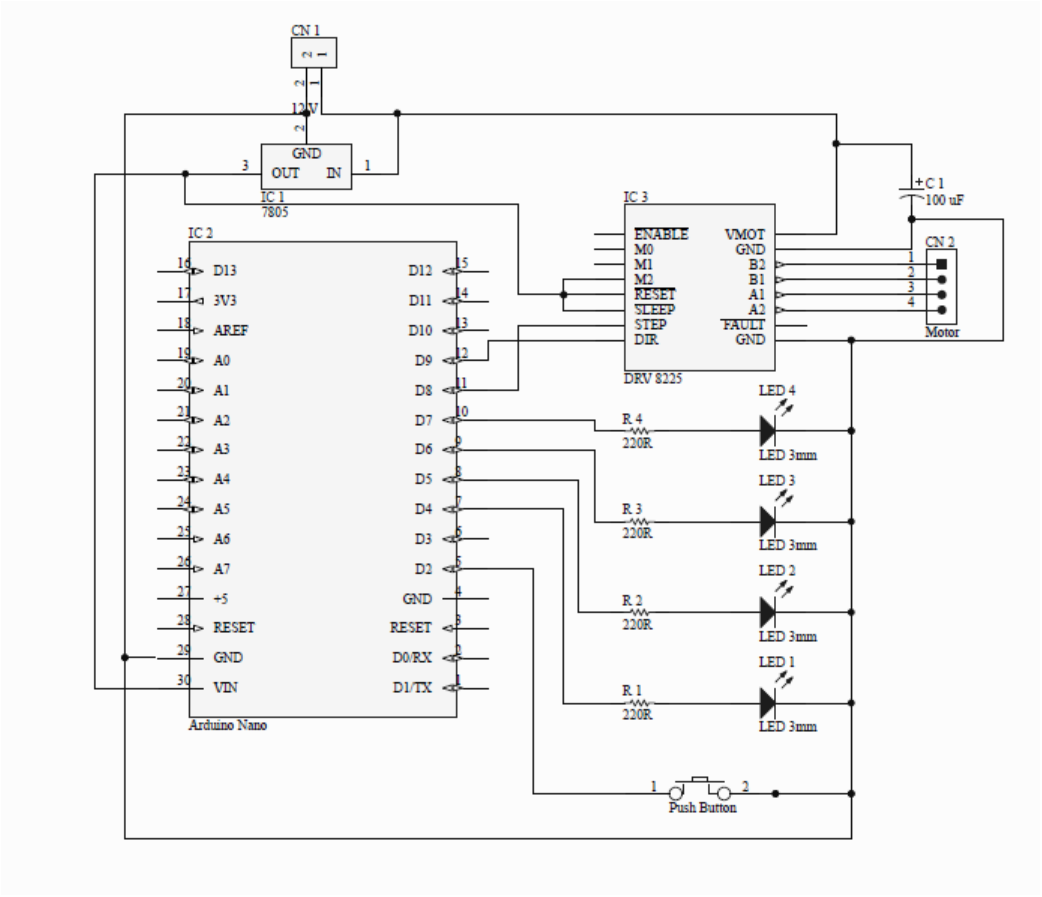


Figure 8.9. *Schematic of the electrical circuit used to control the accessory unit slide indexer.*

8.4. Appendix IV – Serum Drying Study Experimental Set-up

An Arduino Mega 2560 microcontroller (Arduino, Italy) was used to facilitate the control systems necessary to heat the SIREs at stable temperatures. To prevent any denaturation of proteins in the serum, the samples were kept strictly below 37°C. Two 12V adhesive heating elements (RS components, UK) were bound to either side of the ATR crystal along with a 10kΩ (at 25°C) thermistor (RS components, UK) for each heating element. The thermistors were arranged as a voltage divider by connecting one end of the thermistor to a ground pin of the Arduino Mega, the 10kΩ resistor connected

to the 5V supply voltage pin, and the remaining thermistor and 10k resistor connectors connected to the A1 pin via a third wire and a breadboard. The result is an output voltage interfacing with the analogue to digital converter of the Arduino Mega. This digitised value was used to calculate the resistance of the thermistor (Equation 8.1) and hence the temperature of the thermistors (Equation 8.2).

$$R_T = \frac{10{,}000}{1{,}023 / (ADC_{Value} - 1)}$$

Equation 8.1. *Digitised voltage divider equation.*

$$\frac{1}{Temperature(°K)} = a + b\left(l_n\left(\frac{R_T}{R_{25}}\right)\right) + c\left(l_n\left(\frac{R_T}{R_{25}}\right)\right)^2 + d\left(l_n\left(\frac{R_T}{R_{25}}\right)\right)^3$$

Equation 8.2. *Thermistor temperature equation.*

Where a, b, c, and d = constants only valid between a temperature range of 0°C to 50°C.

A negative feedback control system was programmed into the Arduino Mega to maintain plate temperatures of either 30°C or 35°C. This was achieved using Pulse Width Modulation (PWM) to control the states of two TIP120 transistors to permit the flow of current to the heating elements. The use of two transistors allowed independent heater activation and hence finer temperature control. Both a mercury thermometer and a TMP102 temperature sensor breakout board (Cool Components, England) were used to verify the control system was working properly. A small 5-volt DC Sunon® fan was mounted above the ATR crystal using 4 3D printed PLA legs. The fans speed was controlled by changing the input voltage from the Arduino Mega to 3.3V and 5V to produce slow (~5CFM) and fast (~9CFM) flow rates respectively.

8.5. Appendix V – SIRS Patient Glucose Interference

Figure 8.10 shows spectra of four potential outliers identified during PCA of sepsis, SIRS, and control groups. A spectrum of human pooled serum (HPS) has been added as a visual reference. The patient spectra show elevated absorbance in the 1000-1100cm^{-1} spectral region which is indicative of increased carbohydrate content, for instance from lactate or glucose. All except patient G1259 presented elevated blood lactate levels (Table 8.1).

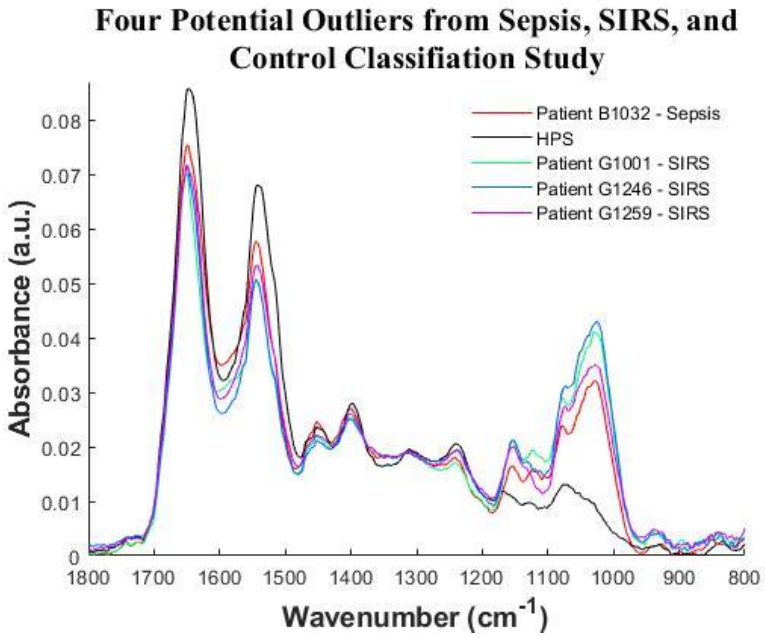


Figure 8.10. *Spectra of four potential outliers and an HPS spectrum as a visual reference. The outlier spectra exhibit high absorbance in the 1000-1100 region compared to the HPS spectrum.*

Table 8.2. *Lactate levels of four potential outliers and the normal range of blood lactate concentration in a healthy individual.*

Patient Identifier	Blood Lactate (mmol/L)
B1032	9.1
G1246	1.9
G1001	14.0
G1259	1.1
Normal range	0.5-1.3