



UNIVERSIDADE
ESTADUAL DE LONDRINA

MAURÍCIO MOREIRA

**DESENVOLVIMENTO E TESTE DE UM FOTÔMETRO PARA
QUANTIFICAÇÃO SIMULTÂNEA DE ETANOL E METANOL
EM GASOLINAS UTILIZANDO COMPRIMENTOS DE ONDA
DISCRETOS NO INFRAVERMELHO PRÓXIMO**

Londrina
2020

MAURÍCIO MOREIRA

**DESENVOLVIMENTO E TESTE DE UM FOTÔMETRO PARA
QUANTIFICAÇÃO SIMULTÂNEA DE ETANOL E METANOL
EM GASOLINAS UTILIZANDO COMPRIMENTOS DE ONDA
DISCRETOS NO INFRAVERMELHO PRÓXIMO**

Tese apresentada ao Programa de Doutorado Associado em Engenharia Elétrica da Universidade Estadual de Londrina (UEL) e da Universidade Tecnológica Federal do Paraná campus Cornélio Procópio (UTFPR-CP) como requisito parcial para a obtenção do título de Doutor em Engenharia Elétrica.

Orientador: Prof. Dr. José Alexandre de França

Londrina
2020

Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática do Sistema de Bibliotecas da UEL

M838 MOREIRA, MAURÍCIO.
Desenvolvimento e Teste de um Fotômetro para Quantificação Simultânea de Etanol e Metanol em Gasolinas Utilizando Comprimentos de Onda Discretos no Infravermelho Próximo / MAURÍCIO MOREIRA. - Londrina, 2020.
171 f.

Orientador: José Alexandre de França.
Tese (Doutorado em Engenharia Elétrica) - Universidade Estadual de Londrina, Centro de Tecnologia e Urbanismo, Programa de Pós-Graduação em Engenharia Elétrica, 2020.
Inclui bibliografia.

1. Espectroscopia NIR - Tese. 2. Calibração - Tese. 3. Álcool - Tese. 4. Variabilidade amostral - Tese. I. França, José Alexandre de. II. Universidade Estadual de Londrina. Centro de Tecnologia e Urbanismo. Programa de Pós-Graduação em Engenharia Elétrica. III. Título.

CDU 62

MAURÍCIO MOREIRA

**DESENVOLVIMENTO E TESTE DE UM FOTÔMETRO PARA
QUANTIFICAÇÃO SIMULTÂNEA DE ETANOL E METANOL EM
GASOLINAS UTILIZANDO COMPRIMENTOS DE ONDA DISCRETOS
NO INFRAVERMELHO PRÓXIMO**

Tese apresentada ao Programa de Doutorado Associado em Engenharia Elétrica da Universidade Estadual de Londrina (UEL) e da Universidade Tecnológica Federal do Paraná campus Cornélio Procópio (UTFPR-CP) como requisito parcial para a obtenção do título de Doutor em Engenharia Elétrica.

BANCA EXAMINADORA

Orientador: Prof. Dr. José Alexandre de França
Universidade Estadual de Londrina – UEL

Prof. Dr. Ivo Milton Raimundo Junior
Universidade Estadual de Campinas –
UNICAMP

Prof. Dr. Flavio Henrique Vasconcelos
Universidade Federal de Minas Gerais – UEMG

Prof. Dr. Francisco Granziera Júnior
Universidade Estadual de Londrina – UEL

Prof.a Dr.a Maria Bernadete de Moraes França
Universidade Estadual de Londrina - UEL

Londrina, 24 de agosto de 2020.

*A minha amada esposa Mariana e a minha filha
Marina.*

AGRADECIMENTOS

Agradeço primeiramente a Deus;

Aos meus familiares, especialmente à minha esposa Mariana Angela Rossaneis Moreira, à minha mãe Ana Rissi Moreira e ao meu pai Marcos Moreira (*in memoriam*);

Ao professor José Alexandre de França pela orientação e pela oportunidade oferecida para a realização deste trabalho;

Aos demais integrantes do LA2I, especialmente à professora Maria Bernadete de Moraes França, à Ana Lúcia de Souza Madureira Felício e à Ana Carla Ranucci Carneiro, pelo auxílio na realização dos experimentos, pelo compartilhamento das idéias e dos conhecimentos em todos os momentos;

Ao coordenador de curso, professor Rogério Marcos da Silva, e aos demais professores da Coordenação de Engenharia Elétrica (COELT-AP); bem como à Diretoria de Pesquisa e Pós-Graduação (DIRPPG-AP), da Universidade Tecnológica Federal do Paraná - campus Apucarana, pelo apoio oferecido para a realização deste trabalho;

À CAPES, pela bolsa de estudos;

E, por fim, aos demais professores do Programa de Doutorado Associado em Engenharia Elétrica UEL/UTFPR-CP e a todos que direta ou indiretamente, com sua contribuição, possibilitaram a realização deste trabalho.

MOREIRA, Maurício. **Desenvolvimento e Teste de um Fotômetro para Quantificação Simultânea de Etanol e Metanol em Gasolinas Utilizando Comprimentos de Onda Discretos no Infravermelho Próximo**. 2020. Tese (Doutorado em Engenharia Elétrica) – Universidade Estadual de Londrina, Londrina, 2020.

RESUMO

A concentração de um constituinte em uma mistura pode ser determinada medindo-se parâmetros ópticos a partir de uma amostra, em diferentes comprimentos de onda. Em seguida, um modelo matemático multivariado pode ser utilizado para relacionar a concentração de interesse com os dados ópticos medidos. Normalmente, equipamentos caros e com sistemas ópticos complexos são requeridos para medir os dados ópticos com elevada resolução espectral e em uma faixa ampla do espectro. Por outro lado, quanto menor o número de variáveis espectrais necessários para compor um modelo satisfatório, mais simples e barato se torna o sistema optoeletrônico do instrumento. Este trabalho apresenta o projeto, desenvolvimento e teste de um fotômetro baseado em LEDs e em filtros ópticos de interferência para quantificar, simultaneamente, as concentrações de etanol e metanol em gasolinas comerciais, em misturas ternárias, por meio de comprimentos de onda discretos no infravermelho próximo. Realiza-se uma análise crítica de diferentes abordagens identificadas na literatura para calcular as variáveis preditoras do modelo a partir da medição e do processamento das informações ópticas. Devido às semelhanças físico-químicas entre o metanol e o etanol, os espectros de absorbâncias dos dois álcoois são muito semelhantes, o que dificulta a sua diferenciação. Também, a composição da gasolina possui uma variabilidade natural, o que corresponde a uma fonte de variância nos sinais analíticos que não está associada à mudança nas concentrações de interesse. Por isso, para que os modelos matemáticos de medição sejam viáveis na prática, realizou-se um estudo espectral para determinar os comprimentos de onda mais adequados para compor os modelos, considerando a variabilidade natural da gasolina. Para avaliar as performances de predição dos modelos, desenvolveu-se um método de validação independente que emprega um grupo de amostras derivadas do conjunto de calibração e outro grupo de amostras com matrizes independentes do grupo de calibração, a fim de se obter uma estimativa da real capacidade preditiva futura do instrumento. O instrumento construído é composto por seis LEDs infravermelhos e oito filtros ópticos de interferência. Resultados experimentais mostraram que os modelos desenvolvidos por Regressão Linear Múltipla (MLR) com sete variáveis espectrais permitem quantificar as concentrações dos álcoois com incertezas menores do que 3 % do volume.

Palavras-chave: Espectroscopia NIR. Calibração. Álcool. Variabilidade amostral. Matriz de amostra.

MOREIRA, Maurício. **Development and Testing of a Photometer for Simultaneous Quantification of Ethanol and Methanol in Gasolines Using Discrete Wavelengths in Near Infrared.** 2020. Thesis (Doctorate in Electrical Engineering) – State University of Londrina, Londrina, 2020.

ABSTRACT

The concentration of a constituent in a mixture can be determined by measuring optical parameters of a sample, at different wavelengths. Then, a multivariate mathematical model can be used to relate the concentration of interest to the measured optical data. Typically, expensive equipment with complex optical systems are required to measure optical data with high spectral resolution and over a wide spectral range. On the other hand, the smaller the number of spectral variables needed to compose a satisfactory model, the simpler and cheaper becomes the optoelectronic system of the instrument. This work presents the design, development and testing of a photometer based on LEDs and optical interference filters to simultaneously quantify the concentrations of ethanol and methanol in commercial gasolines, in ternary mixtures, using discrete wavelengths in the near infrared. A critical analysis of different approaches, identified in the literature, to calculate the predictive variables of the model from the measurement and processing of the optical information is done. Due to the physico-chemical similarities between methanol and ethanol, the absorbance spectra of the two alcohols are very similar, which makes their discrimination difficult. Also, the gasoline composition has a natural variability, which is a source of variance in the analytical signals that is not associated with variations in the concentrations of interest. Therefore, in order to make the mathematical measurement models relevant in practice, a spectral study was carried out to determine the most appropriate wavelengths to compose the models, considering the natural variability of gasoline. To evaluate the prediction performances of the models, an independent validation method was developed that employs a set of samples derived from the calibration set and another set of samples with matrices independent from the calibration set, in order to obtain an estimate of the real future predictive ability of the instrument. The built instrument consists of six infrared LEDs and eight optical interference filters. Experimental results showed that Multiple Linear Regression (MLR) models with seven spectral variables make it possible to quantify the concentrations of the alcohols with uncertainties lower than 3 % in volume.

Keywords: NIR spectroscopy. Calibration. Alcohol. Sample variability. Sample matrix.

LISTA DE FIGURAS

1.1	Espectros de absorbâncias do etanol e do metanol na faixa de 900 a 1.650 nm obtidos por um espectrofotômetro de bancada.	17
2.1	Algumas técnicas de instrumentação para obtenção do espectro de absorbâncias: (a) transmissão; (b) reflexão total atenuada; (c) reflexão difusa; e (d) reflexão especular. Adaptado de Sun (2009).	22
2.2	Espectros de 84 amostras de gasolina contendo diferentes concentrações de etanol: (a) sem pré-processamento; (b) após correção da linha de base; (c) após processamento por SNV; (b) e primeira derivada.	25
2.3	Diagrama de blocos do fotômetro desenvolvido por Moreira (2013).	31
2.4	Circuito de detecção e condicionamento de sinal óptico do protótipo desenvolvido por Moreira (2013).	31
2.5	Representação dos dados espectrais de uma substância genérica medidos a partir de diferentes tipos de instrumentos.	33
3.1	Diagrama representativo da primeira iteração do método de validação cruzada utilizado para avaliar cada modelo matemático de medição.	52
3.2	Espectros de emissão normalizados dos LEDs e espectros de transmitância dos filtros de interferência.	59
3.3	(a) Representação do sistema óptico adotado; (b) desenho das principais peças que compõem a estrutura de sustentação do protótipo; (c) desenho de parte da estrutura do protótipo montada.	60
4.1	Curvas espectrais das 102 amostras de calibração e validação, sem pré-processamento espectral, as quais foram obtidos a partir das absorbâncias calculadas por: (a) método B; (b) método D; (c) método E; e (d) método F.	63
4.2	Relação entre concentrações especificadas e preditas, considerando o modelo de MLR com correção da linha de base: (a) utilizando o método B, que utiliza como referência uma medição com uma cubeta vazia no início do ciclo de análises; e (b) utilizando o método E, que calcula a absorbância em relação a uma amostra de referência.	66
4.3	Curvas espectrais pré-processadas de 119 amostras de gasolina contendo diferentes concentrações de etanol, obtidas: (a) pelo protótipo, por meio do Método B de cálculo de absorbâncias; (b) pelo protótipo, por meio do Método E de cálculo das absorbâncias; (c) pelo espectrofotômetro de bancada, para a faixa espectral de 900 a 1.650 nm.	70

4.4	Resultados após calibração utilizando amostras exclusivamente do primeiro conjunto de dados, com pontos espectrais obtidos pelo Método B: (a) relação entre concentrações especificadas pelo método de referência e concentrações estimadas pelo modelo; (b) resíduos de calibração e erros de predição; (c) histogramas dos resíduos de calibração e dos erros de predição.	72
4.5	Resultados após calibração utilizando amostras distribuídas entre os quatro conjuntos, com pontos espectrais obtidos pelo Método B: (a) relação entre concentrações especificadas pelo método de referência e concentrações estimadas pelo modelo; (b) resíduos de calibração e erros de predição; (c) histogramas dos resíduos de calibração e dos erros de predição.	73
4.6	Resultados após calibração utilizando amostras exclusivamente do primeiro conjunto de dados, com pontos espectrais obtidos pelo Método E: (a) relação entre concentrações especificadas pelo método de referência e concentrações estimadas pelo modelo; (b) resíduos de calibração e erros de predição; (c) histogramas dos resíduos de calibração e dos erros de predição.	76
4.7	Resultados após calibração utilizando amostras distribuídas entre os quatro conjuntos, com pontos espectrais obtidos pelo Método E: (a) relação entre concentrações especificadas pelo método de referência e concentrações estimadas pelo modelo; (b) resíduos de calibração e erros de predição; (c) histogramas dos resíduos de calibração e dos erros de predição.	77
4.8	Resultados obtidos pelo espectrofotômetro de bancada após calibração utilizando amostras distribuídas entre os quatro conjuntos: (a) relação entre concentrações especificadas pelo método de referência e concentrações estimadas pelo modelo; (b) resíduos de calibração e erros de predição; (c) histogramas dos resíduos de calibração e dos erros de predição.	79
4.9	(a) Espectros originais das 170 amostras, com resolução de 8 cm^{-1} ; (b) espectros originais com número de variáveis reduzido para 37 pontos espectrais coincidentes com os comprimentos de onda dos filtros disponíveis comercialmente; (c) Espectros dos álcoois puros.	81
4.10	Estatísticas RMSEV e RMSEP conjuntas para ambos os modelos matemáticos de medição, em função do número de variáveis espectrais selecionadas.	83
4.11	Concentrações especificadas versus concentrações preditas para a primeira iteração do <i>algoritmo de seleção de variáveis</i> , para modelos MLR com sete variáveis espectrais: (a) para o etanol; (b) para o metanol.	84
4.12	Fotos do protótipo desenvolvido: (a) roda de filtros ópticos; (b) parte do sistema óptico em fase de montagem; (c) placa de acionamento dos LEDs infravermelhos; (d) placa do circuito de detecção acoplada à placa de desenvolvimento do microcontrolador; (e) visão interna do protótipo em fase final de montagem; (f) protótipo finalizado em operação.	86
4.13	Circuito de acionamento dos LEDs infravermelhos.	87

4.14	Diagrama esquemático do circuito de transimpedância e supressor de resposta contínua.	91
4.15	Circuito de demodulação síncrona e amplificação.	92
4.16	(a) Espectros das 237 amostras, obtidos pelo espectrofotômetro; (b) espectros obtidos pelo espectrofotômetro após pré-processamento e redução para as oito variáveis espectrais do protótipo; (c) Espectros obtidos pelo protótipo.	95
4.17	Concentrações especificadas versus concentrações preditas pelos modelos MLR com sete variáveis espectrais para concentrações de etanol ((a) e (b)), metanol ((c) e (d)) e total de álcoois ((e) e (f)). Os dados em (a), (c) e (e) são do protótipo; e os dados em (b), (d) e (f) são do espectrofotômetro.	97
C1	Diagrama esquemático da placa do circuito de detecção e condicionamento de sinais - folha 1.	127
C2	Diagrama esquemático da placa do circuito de detecção e condicionamento de sinais - folha 2.	128
C3	Diagrama esquemático da placa do circuito de detecção e condicionamento de sinais - folha 3.	129
C4	Diagrama esquemático da placa de acionamento dos LEDs infravermelhos - folha 1.	130
C5	Diagrama esquemático da placa de acionamento dos LEDs infravermelhos - folha 2.	131
C6	Diagrama esquemático da placa de acionamento dos LEDs infravermelhos - folha 3.	132
C7	Diagrama esquemático da placa de acionamento dos LEDs infravermelhos - folha 3.	133

LISTA DE TABELAS

3.1	Comprimentos de onda centrais e larguras de banda à meia altura (FWHM) dos 37 filtros de interferência ópticos disponíveis comercialmente.	50
4.1	Comparação entre os resultados para os sete diferentes métodos de cálculo de absorbâncias, considerando calibração multivariada por MLR sem pré-processamento espectral.	64
4.2	Comparação dos resultados para os sete diferentes métodos de cálculo de absorbâncias, considerando calibração multivariada por MLR com correção da linha de base (<i>offset</i>).	64
4.3	Concentrações de referência do conjunto de validação e concentrações previstas para os sete métodos de cálculo das absorbâncias, considerando calibração multivariada por MLR com correção da linha de base (<i>offset</i>). A medição de cada amostra foi realizada uma única vez e empregada em todos os métodos.	65
4.4	Vantagens e desvantagens dos sete métodos estudados, considerando suas características e questões de utilização prática para instrumentos de diferentes tipos construtivos e diferentes aplicações. Os métodos são apresentados em ordem crescente de RMSEC.	68
4.5	Comprimentos de onda mais informativos selecionados pelo <i>algoritmo de seleção de variáveis</i>	82

SUMÁRIO

1	INTRODUÇÃO	15
1.1	ORGANIZAÇÃO DO TRABALHO	20
2	FUNDAMENTAÇÃO TEÓRICA	21
2.1	PRINCÍPIOS BÁSICOS DA ESPECTROSCOPIA DO INFRAVERMELHO	21
2.2	MÉTODOS DE CALIBRAÇÃO MULTIVARIADA E PRÉ-PROCESSAMENTOS ESPECTRAIS	23
2.3	MEDIÇÃO DOS DADOS ÓPTICOS E MÉTODOS DE CÁLCULO DAS ABSORBÂNCIAS	25
2.3.1	Método A	27
2.3.2	Método B	27
2.3.3	Método C	28
2.3.4	Método D	28
2.3.5	Método E	28
2.3.6	Método F	29
2.3.7	Método G	29
2.3.8	Fotômetro baseado em LEDs empregado nos experimentos preliminares	30
2.4	SELEÇÃO DE COMPRIMENTOS DE ONDA E RESOLUÇÃO ESPECTRAL	32
2.5	ESPECTROSCOPIA NIR NA ANÁLISE DE GASOLINAS COMERCIAIS	35
3	PROCEDIMENTOS EXPERIMENTAIS	41
3.1	ESTUDO DA MEDIÇÃO E MODELAGEM MATEMÁTICA DOS SINAIS ÓPTICOS PARA OBTENÇÃO DAS VARIÁVEIS PREDITORAS	42
3.1.1	Preparação das amostras e medições	43
3.1.2	Pré-processamento de dados e calibração	44
3.2	AValiação DA INFLUÊNCIA DA VARIABILIDADE DE MATRIZ	45
3.3	ESTUDO ESPECTRAL DAS MISTURAS TERNÁRIAS PARA SELEÇÃO DE VARIÁVEIS	47
3.3.1	Determinação dos comprimentos de onda analiticamente importantes	47
3.3.1.1	Preparação das amostras	47
3.3.1.2	Aquisição dos espectros das amostras	48
3.3.1.3	Pré-processamento dos dados espectrais	48
3.3.1.4	Procedimento de seleção das variáveis espectrais	49
3.3.2	Avaliação dos modelos candidatos	56
3.4	PROJETO DO PROTÓTIPO E TESTES FINAIS	58

4	RESULTADOS EXPERIMENTAIS	62
4.1	COMPARAÇÃO ENTRE OS MÉTODOS DE CÁLCULO DAS ABSORBÂNCIAS	62
4.2	ANÁLISE DA INFLUÊNCIA DA VARIABILIDADE DE MATRIZ	69
4.2.1	Avaliação da influência da seleção de amostras utilizando o Método B	71
4.2.2	Avaliação da influência da seleção de amostras utilizando o Método E	75
4.2.3	Comparação com o espectrofotômetro de bancada	78
4.3	RESULTADOS DO ESTUDO ESPECTRAL PARA SELEÇÃO DE VARIÁVEIS .	80
4.4	MONTAGEM DO PROTÓTIPO E IMPLEMENTAÇÃO DO MODELO DE SETE VARIÁVEIS ESPECTRAIS	85
4.4.1	Circuito de acionamento dos LEDs	87
4.4.2	Circuito de detecção e condicionamento de sinais	89
4.4.3	Circuitos auxiliares	92
4.4.4	Lógica de programação e medição dos dados ópticos	92
4.5	TESTE DO PROTÓTIPO E COMPARAÇÃO COM O ESPECTROFOTÔMETRO	93
5	CONCLUSÃO	100
	APÊNDICE A - Artigo Publicado na Revista IEEE Transactions on Instrumentation and Measurement	102
	APÊNDICE B - Artigo Publicado na Revista Measurement Science and Technology	113
	APÊNDICE C - Diagramas Esquemáticos das Placas Eletrônicas Desenvolvidas	126
	ANEXO A - Folhas de Dados dos LEDs Infravermelhos e do Fotodetector	134
	ANEXO B - Espetros de Transmitância dos Filtros de Interferência Ópticos	151
	ANEXO C - Folhas de Dados dos Principais Componentes Eletrônicos Utilizados	159
	REFERÊNCIAS	168

1 INTRODUÇÃO

A quantificação de substâncias por meio da espectroscopia do infravermelho próximo (NIR - *Near-Infrared*) se baseia em observar a interação da radiação eletromagnética com a matéria. Basicamente, este método analítico consiste em posicionar uma amostra no caminho óptico de um feixe luminoso e medir as intensidades de radiação transmitidas ou refletidas por ela, em diferentes comprimentos de onda do NIR. Esta informação óptica é dependente da estrutura molar da substância sob análise. Assim, os dados espectrais medidos podem ser empregados na construção de uma função de medição multivariada que permita estimar a concentração de um constituinte de uma mistura em interesse.

O modelo matemático normalmente utilizado para relacionar a concentração de um analito de interesse com os dados ópticos é uma função linear multivariada da forma

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \epsilon, \quad (1.1)$$

em que y é a concentração do analito; β_0, \dots, β_k são os coeficientes de regressão do modelo; x_1, \dots, x_k são as variáveis preditoras, e correspondem às absorvâncias calculadas a partir das intensidades de luz medidas nos k comprimentos de onda $\lambda_1, \dots, \lambda_k$, respectivamente; e ϵ é o termo do erro aleatório. Este modelo multivariado é baseado na Lei de Beer-Lambert (Burns e Ciurczak, 2008). Do ponto de vista matemático, a determinação de uma curva de calibração na forma (1.1) a partir de um conjunto de amostras com concentrações pré-determinadas é um problema relativamente simples. Apesar disso, construir um modelo matemático de medição que tenha uma performance de predição satisfatória na prática não é um problema trivial.

Um primeiro aspecto que deve ser considerado se refere ao número de variáveis que compõem o modelo. Em muitas aplicações, existem compostos químicos que apresentam bandas de absorção muito estreitas, que só são observáveis quando a resolução espectral é elevada. Também, nas situações em que existe sobreposição de bandas de absorção por diferentes compostos em uma mistura, dados com elevada resolução podem ser requeridos para que seja possível eliminar a influência de uma substância interferente. Nesse sentido, modelos com ótima capacidade preditiva podem ser construídos a partir milhares de variáveis. Evidentemente, quanto maior a resolução e a amplitude da faixa espectral dos dados disponíveis, maiores são as possibilidades de análise e processamento destes dados para a obtenção da informação química desejada. Entretanto, para que seja possível medir os dados ópticos em uma ampla faixa do espectro e com elevada resolução espectral, equipamentos caros e com sistemas ópticos complexos são requeridos, tais como os espectrofotômetros de bancada.

Por outro lado, dependendo das características físico-químicas da mistura a ser analisada, é possível construir modelos satisfatórios com apenas algumas unidades de variáveis discretas (isto é, componentes espectrais de banda estreita centrados em diferentes regiões do espectro). A medida que um menor número de comprimentos de onda seja necessário para

a determinação analítica em uma aplicação específica, torna-se possível construir instrumentos com sistemas optoeletrônicos mais simples, robustos e opticamente estáveis, resultando em soluções mais baratas, leves e portáteis.

Diante do exposto, torna-se evidente que o projeto dos sistemas óptico e de instrumentação eletrônica de um instrumento dedicado a uma aplicação específica deve ser baseado nos requisitos ópticos do modelo analítico. Isto é, a escolha da fonte de luz; do tipo de fotodetector; do dispositivo ou método utilizado para separar os comprimentos de onda; entre outras especificações, devem possibilitar a obtenção de dados ópticos analiticamente importantes para a quantificação da substância desejada. Por sua vez, para definir estes requisitos ópticos, faz-se necessário estudar as características espectrais das amostras, considerando as diversas fontes de variância espectral existentes. Por isso, um número de amostras suficiente que cubra a variabilidade esperada em uma situação prática de utilização do instrumento é um requisito indispensável.

Neste trabalho, realizou-se um estudo para avaliar a viabilidade prática de se construir modelos matemáticos de medição na forma (1.1) para quantificar, simultaneamente, as concentrações de etanol e metanol em gasolinas comerciais utilizando comprimentos de onda discretos no NIR. Para implementar estes modelos, o estudo avalia a construção de um fotômetro baseado em LEDs e filtros de interferência ópticos, que permitem obter informações espectrais discretas com uma larguras de banda estreitas, da ordem de 10 a 20 nm. Para tornar esta solução viável, o estudo investiga qual o menor número de comprimentos de onda possível que possibilite quantificar os álcoois com incertezas dentro de limites aceitáveis, da ordem 2 a 3 % do volume (v/v).

A principal aplicação deste instrumento será para detectar adulterações em gasolinas comerciais. No Brasil, as gasolinas disponibilizadas aos consumidores finais possuem etanol em sua composição, em concentração de 27 % v/v fixada pela legislação vigente. Por sua vez, a adição de metanol é proibida, sendo que o teor máximo admissível é de 0,5 % do volume. Assim, a adição de etanol em quantidades acima da permitida, ou a contaminação por metanol, são dois tipos de adulterações possíveis. Neste caso, o instrumento proposto poderá ser utilizado em campo como uma ferramenta de triagem, para auxiliar no controle de qualidade de combustíveis de forma rápida, reduzindo a necessidade de análises mais complexas em laboratório. Além desta, outras utilizações para este tipo de equipamento são possíveis, como por exemplo o controle de injeção de combustíveis em motores.

Alguns autores já propuseram instrumentos portáteis baseados no NIR para quantificar álcoois na gasolina, tais como Harris et al. (1993); Chagas (2006) e Lutz et al. (2014). Apesar disso, estes trabalhos consideram apenas a quantificação de etanol ou metanol separadamente, em misturas binárias com gasolina. Contudo, para a aplicação estudada, é necessário quantificar os dois álcoois simultaneamente em misturas ternárias de etanol/metanol/gasolina. A detecção simultânea de etanol e metanol em gasolinas também já foi estudada por Correia et al. (2018); Ouyang e Liu (2013) e Fernandes et al. (2008), porém empregando muitas variáveis espectrais. Fernandes et al. (2008) e Ouyang e Liu (2013) empregaram espectrofotômetros

de bancada, enquanto Correia et al. (2018) utilizaram um mini-espectrômetro comercial de baixa resolução, com 128 elementos sensores. Já quantificação simultânea dos dois álcoois em misturas ternárias, por meio de comprimentos de onda discretos no NIR, ainda não foi estudada. Uma discussão mais detalhada dos trabalhos citados relacionados à quantificação de álcoois na gasolina é feita no Capítulo 2, na Seção 2.5.

Um dos principais desafios do trabalho consiste em selecionar as variáveis espectrais analiticamente relevantes que permitam diferenciar os dois álcoois. O etanol (C_2H_6O) e o metanol (CH_4O) são caracterizados por ligações O-H. Por sua vez, a gasolina é uma mistura de hidrocarbonetos voláteis cujos principais componentes são parafinas de cadeia ramificada, ciclo-parafinas e aromáticos (Harris et al., 1993). Os espectros de absorção dos álcoois na região do NIR são substancialmente diferentes dos da gasolina, o que torna esta região espectral particularmente apropriada para este tipo de medição. Contudo, a diferenciação entre etanol e metanol não é simples, conforme apontam Neto et al. (2014); Correia et al. (2018) e Ouyang e Liu (2013). As características físico-químicas dos dois álcoois são muito semelhantes. Consequentemente, também são muito semelhantes os espectros de absorbâncias dos dois compostos na região do NIR, os quais possuem bandas de absorção coincidentes, como mostra os espectros da Figura 1.1. Dessa forma, pode-se considerar que os dois analitos em interesse são mutuamente interferentes. Por isso, determinar um conjunto de comprimentos de onda discretos que possibilite diferenciar e quantificar os dois álcoois em misturas ternárias, com incertezas dentro de limites aceitáveis, não é simples.

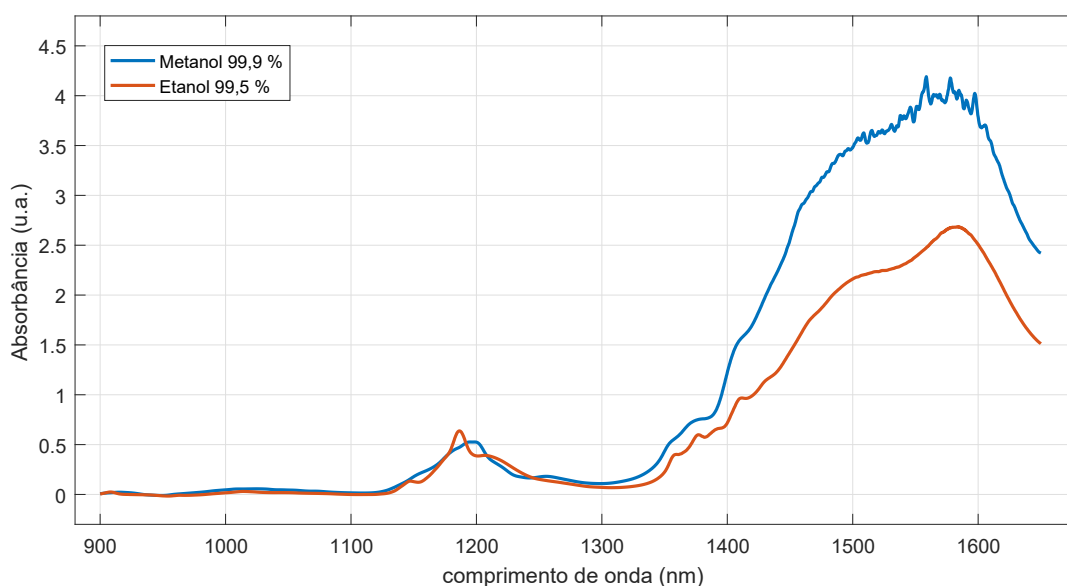


Figura 1.1: Espectros de absorbâncias do etanol e do metanol na faixa de 900 a 1.650 nm obtidos por um espectrofotômetro de bancada.

Fonte: o autor.

Outro aspecto a ser considerado na determinação das variáveis espectrais se refere à chamada variabilidade de matriz. Normalmente, a substância a ser quantificada é um componente de uma mistura contendo diversos outros compostos químicos (matriz de amostra),

cuja composição possui uma variabilidade natural. Esta corresponde a uma fonte de variância nos sinais analíticos que não está associada à mudança na concentração em interesse. A variabilidade de matriz é muito evidente na análise de amostras de origem biológica, tais como alimentos. Entretanto, mesmo em aplicações que envolvem substâncias sintéticas processadas em laboratório (como na própria análise de combustíveis, por exemplo), a variabilidade de matriz pode influenciar significativamente na determinação do modelo matemático de medição.

Mais especificamente, a composição química da gasolina pode variar consideravelmente devido a diversos fatores, tais como: origem do óleo utilizado na produção; diferenças nos processos de fabricação em cada refinaria; contaminações devido ao transporte e ao armazenamento em cada estabelecimento comercial; além da possível presença de adulterantes. Uma evidência da influência de matriz nos espectros da gasolina é o trabalho de Balabin et al. (2010), que analisa a variabilidade espectral para classificar gasolinas provenientes de diferentes refinarias. O estudo de Fernandes et al. (2008) mostra que a influência de matriz no resultado da calibração para a quantificação de álcoois na gasolina é significativa, mesmo quando o modelo é construído a partir de dados com alta resolução espectral. Além dessa variabilidade natural, deve-se considerar que diferentes tipos de gasolinas (comum e aditivada) apresentam composições distintas. Evidentemente, a influência de matriz pode ser muito mais significativa para um instrumento com poucos comprimentos de onda discretos.

Naturalmente, um modelo quimiométrico deve ser composto por variáveis espectrais que apresentem a maior sensibilidade possível às variações nas concentrações de cada analito. Isto pode ser obtido simplesmente selecionando-se comprimentos de onda em regiões características de absorção pelos componentes em interesse presentes na amostra. Por outro lado, para que a função de medição produza bons resultados na prática, é preciso que ela tenha a menor sensibilidade possível às variações de matriz. Por isso, a seleção dos comprimentos de onda mais adequados para compor os modelos deve considerar a variabilidade amostral em uma condição real de utilização futura do instrumento. Isto normalmente requer um estudo espectral utilizando muitas amostras. Isto porque, na prática, uma curva de calibração é obtida por um método de regressão por mínimos quadrados. A análise de regressão deve empregar dados de amostras que cubram toda a faixa de concentrações em interesse, para todos os analitos, e para diferentes matrizes.

A importância de se considerar a variabilidade amostral no desenvolvimento de um modelo matemático de medição já é bem conhecida na literatura. Apesar disso, os estudos já desenvolvidos por Chagas (2006); Harris et al. (1993); Correia et al. (2018); Fernandes et al. (2008) e Ouyang e Liu (2013), que quantificam álcoois em gasolina, não consideram ou não avaliam adequadamente a variabilidade de matriz real. Consequentemente, a aplicabilidade dos modelos desenvolvidos por estes autores pode ser muito limitada na prática. Por isso, este trabalho propõe um novo método para determinar os comprimentos de onda mais adequados e para avaliar os modelos de calibração resultantes, com o objetivo de obter uma estimativa mais real da capacidade preditiva futura destes modelos. O estudo empregou, no total, 547 amostras de gasolinas dos tipos comum e aditivadas, provenientes de diferentes distribuidores, o que

corresponde a um número significativamente superior a todos os trabalhos relativos ao tema identificados na literatura. Para avaliar os modelos de calibração, desenvolveu-se um método de validação que emprega dois conjuntos de amostras independentes: um conjunto de validação contendo amostras derivadas do conjunto de calibração, cujas matrizes estão representadas nos dados utilizados para construir os modelos; e um segundo conjunto de amostras com matrizes novas, distintas daquelas utilizadas na calibração. Dessa forma, este método fornece uma melhor estimativa da real capacidade preditiva futura dos modelos.

Por fim, um terceiro fator que pode influenciar na capacidade preditiva de um modelo quimiométrico está associado à medição dos dados ópticos em si, e ao processamento destes dados para gerar as variáveis predictoras. A transformação dos dados ópticos em absorbâncias é calculada em função da intensidade de luz medida no fotodetector após interagir com a amostra (I_S) e de uma intensidade de luz medida em uma condição experimental de referência (I_0), que define o zero de absorbâncias. Contudo, de acordo com a descrição dos procedimentos experimentais encontrados em diversos trabalhos na literatura (Lutz et al., 2014; Visconti et al., 2017; Gaião et al., 2008; Guchardi e Pasquini, 2001; Isaac e Na, 2016; Chagas, 2006; Kittipanyangam et al., 2016; Visconti et al., 2017; Ribeiro et al., 2016; Gentilin et al., 2016), verifica-se que existem diferentes abordagens para esta medição de referência e, conseqüentemente, diferentes métodos de cálculo das absorbâncias. Além disso, comumente, técnicas de pré-processamento são empregadas para preparar os dados antes de serem aplicados ao modelo matemático. Isto porque, normalmente, a maior parte da variância na intensidade de luz medida em um dado comprimento de onda está associada a fatores físicos, como o espalhamento da luz pelas partículas da amostra, e não à absorção pelos analitos. Dessa forma, em muitas situações, só é possível interpretar e extrair alguma informação útil dos espectros quando algum método de pré-processamento de dados é aplicado (Burns e Ciurczak, 2008). Por isso, em geral, as variáveis predictoras em (1.1) são absorbâncias transformadas. Assim, para um mesmo conjunto de comprimentos de onda, diferentes modelos podem ser construídos a partir de variáveis predictoras distintas, os quais terão diferentes performances preditivas.

Diante do exposto, neste trabalho, criou-se um procedimento experimental que permite avaliar e comparar empiricamente sete métodos de cálculo de absorbâncias identificados na literatura, a partir de uma única medição de um conjunto de amostras de gasolinas, ou seja, utilizando a mesma amostra estatística. Isto é importante porque elimina a influência de outras variáveis na comparação entre os métodos. Os sete métodos são detalhados no Capítulo 2, na Seção 2.3. O estudo analisa a aplicabilidade e a influência de cada método nos modelos para predição dos álcoois, e também analisa as implicações para outros tipos de aplicações. Este tipo de comparação também não foi identificado na literatura e é uma contribuição deste trabalho. Por fim, são discutidos também conceitos sobre a modelagem matemática do fotodetector e do circuito de instrumentação responsável pela medição dos dados ópticos, os quais geralmente não são considerados por outros autores. Dessa forma, apesar de tratar de uma aplicação específica, o presente trabalho apresenta um estudo abrangente e que aborda conceitos aplicáveis também à análise de outros tipos substâncias.

1.1 ORGANIZAÇÃO DO TRABALHO

Além desta introdução, o trabalho contém outros quatro capítulos, os quais tratam dos assuntos descritos a seguir.

No Capítulo 2, é apresentada a fundamentação teórica sobre o tema. São abordados assuntos gerais importantes para o bom entendimento do trabalho, tais como: conceitos sobre a medição e o processamento dos dados ópticos; métodos de calibração multivariada; métodos de seleção de variáveis espectrais; conceitos sobre dispositivos e sistemas ópticos; e uma breve revisão bibliográfica sobre métodos analíticos baseados no NIR para a análise de gasolinas.

O Capítulo 3 descreve os materiais e métodos utilizados para a solução do problema desta tese. São descritos os experimentos realizados em cada etapa da pesquisa e as decisões de projeto adotadas para a determinação dos modelos de calibração finais. Também, apresenta-se o projeto e a construção do protótipo que implementa os modelos de calibração adotados.

O Capítulo 4, por sua vez, apresenta e discute os resultados dos experimentos realizados. Além disso, as performances de predição dos modelos de calibração finais implementados são estimadas e comparadas com os modelos gerados por dados de um espectrofotômetro de bancada.

Conclusões e comentários finais são apresentados no Capítulo 5. Também, são sugeridos trabalhos futuros relacionados ao tema.

Por fim, após o Capítulo 5, são incluídos três apêndices e três anexos. Nos Apêndices A e B, apresenta-se dois artigos publicados pelo autor durante o desenvolvimento deste trabalho e que são relacionados a ele. O apêndice C contém os diagramas esquemáticos das placas eletrônicas desenvolvidas. Por sua vez, os anexos A a C apresentam as folhas de dados dos principais dispositivos ópticos e eletrônicos utilizados.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo, apresenta-se um resumo dos principais conceitos relacionados ao problema de calibração utilizando comprimentos de onda discretos no NIR. Os conceitos descritos aqui fornecem uma dimensão do número de variáveis envolvidas e permitem uma melhor compreensão do problema, e servem como base para as discussões dos capítulos posteriores.

2.1 PRINCÍPIOS BÁSICOS DA ESPECTROSCOPIA DO INFRAVERMELHO

A espectroscopia NIR é um tipo de espectroscopia vibracional que emprega fótons na região espectral entre a luz visível e o infravermelho médio (MIR), compreendida entre os comprimentos de onda 750 e 2.500 nm. Esta região espectral é caracterizada por fótons com energia entre $7,96 \times 10^{-20}$ e $2,65 \times 10^{-19} J$. Este nível de energia é suficiente para promover moléculas de seu estado vibracional fundamental para estados vibracionais excitados, porém é menor do que o necessário para a excitação de elétrons na grande maioria das moléculas (Chagas, 2006). Dessa forma, esta técnica é uma importante ferramenta para a análise de compostos com ligações C-H, N-H, S-H ou O-H de forma rápida, não-destrutiva e não-invasiva (Pasquini, 2003).

A quantificação de substâncias por meio da espectroscopia NIR se baseia em observar a interação da radiação eletromagnética com a matéria. A energia da radiação eletromagnética pode ser absorvida por uma molécula, a qual é convertida em energia mecânica associada ao movimento dos átomos, levando-a a um estado vibracional excitado de maior energia. Para isto, a energia do fóton, que é dependente do comprimento de onda da radiação, deve ser igual à diferença entre dois níveis de energia. Assim, apenas radiações com certos comprimentos de onda podem excitar os níveis vibracionais de uma determinada espécie química: aquelas cuja energia corresponda à diferença de energia entre dois níveis vibracionais adjacentes ou de seus sobretons, bem como combinações de duas ou mais vibrações. Por isso, é possível obter informações qualitativas ou quantitativas de uma espécie química observando-se seus sinais ópticos em diferentes regiões do NIR. Isto é feito posicionando-se uma amostra no caminho óptico de um feixe de luz e medindo-se a intensidade de luz transmitida ou refletida por ela, nos diferentes comprimentos de onda. A partir dessa medição, pode-se obter o espectro de absorbâncias, o qual é dependente da constituição molar da amostra em estudo.

Existem diferentes técnicas instrumentais para a obtenção dos dados espectrais, algumas das quais estão representadas na Figura 2.1. A transmissão é o método mais simples, em que a luz atravessa a amostra e é coletada no lado oposto ao da incidência, como mostra a Figura 2.1(a). Este método permite obter os dados espectrais com alta relação sinal-ruído (SNR), sendo empregado na análise de amostras líquidas e gasosas. Na reflexão, ao contrário da transmissão, mede-se a luz refletida pela amostra. Na reflexão total atenuada, mostrada na Figura 2.1(b), é utilizado o princípio da reflexão interna total da radiação eletromagnética

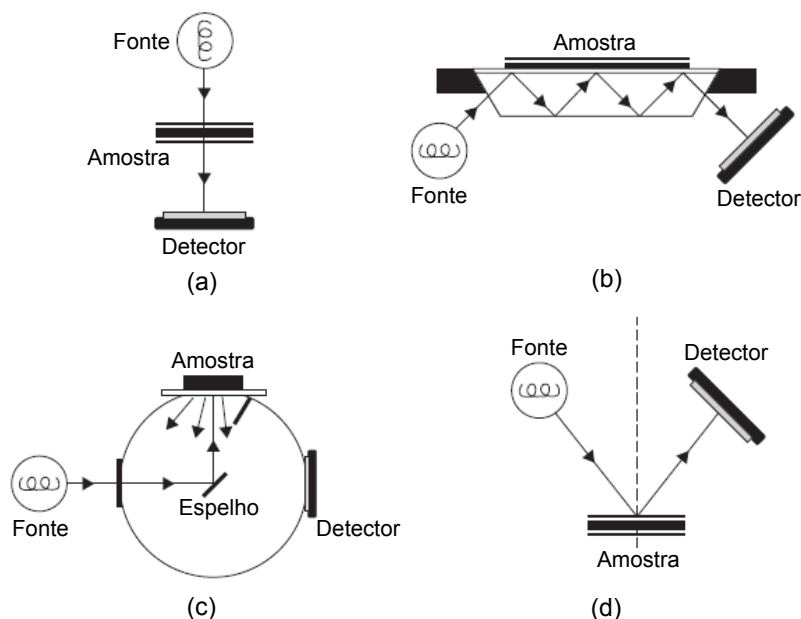


Figura 2.1: Algumas técnicas de instrumentação para obtenção do espectro de absorvâncias: (a) transmissão; (b) reflexão total atenuada; (c) reflexão difusa; e (d) reflexão especular. Adaptado de Sun (2009).

incidente, e o feixe de luz incide mais de uma vez na amostra. Já a reflexão difusa corresponde ao caso em que a luz é refletida pela superfície da amostra em uma direção aleatória, devido à irregularidades em sua superfície; este método, representado na Figura 2.1(c), é mais indicado para materiais sólidos e pós, sendo bastante empregado na análise de alimentos. Por fim, a reflexão especular, mostrada na Figura 2.1(d), é indicada para amostras com superfície regular, sendo utilizada principalmente na análise de polímeros.

Dependendo do estado físico (sólido, líquido ou gasoso) e das características físico-químicas da amostra em estudo, uma determinada técnica de instrumentação pode ser mais adequada e levar a melhores resultados. De qualquer forma, estas várias técnicas de instrumentação podem ser classificadas em dois grupos: aquelas que medem a luz que é transmitida pela amostra (baseadas em transmitância); e aquelas que medem a luz refletida pela amostra (baseadas na reflectância). A partir dessa medição, pode-se determinar a absorvância (A), que é definida pela relação entre a intensidade de luz incidente na amostra (I_0) e a intensidade da luz que chega ao detector após interagir com a amostra (I_S), dada por

$$A = \log \left(\frac{I_0}{I_S} \right). \quad (2.1)$$

Neste caso, a razão I_0/I_S é o inverso da transmitância, para uma técnica baseada em transmissão da luz; ou, equivalentemente, o inverso da reflectância, para uma técnica baseada em reflexão. A relação (2.1) considera uma luz incidente monocromática. Se esta relação for aplicada para diferentes comprimentos de onda, obtém-se o espectro de absorvâncias da amostra.

Técnicas quantitativas baseadas na espectroscopia de absorção tem o objetivo

de associar a concentração de um analito de interesse com a intensidade de luz absorvida por ele. Nesse sentido, a grande maioria das técnicas de calibração de instrumentos espectroscópicos visam estabelecer uma relação linear entre a concentração do analito e as absorvâncias, baseada na Lei de Beer-Lambert (Qiang e Chu, 2012; Haxha e Jhoja, 2016; Kittipanyangam et al., 2016; Yadav et al., 2015; Burns e Ciurczak, 2008). Isto porque, de acordo com esta lei, a contribuição de um absorvedor para o aumento da absorvância, em um dado comprimento de onda, é proporcional à concentração do absorvedor presente na amostra. Por isto, a transformação dos dados ópticos dada em (2.1) é a mais frequentemente utilizada.

Uma vez que dados espectrais de uma espécie química em estudo estejam disponíveis, técnicas matemáticas e estatísticas podem ser utilizadas para relacioná-los com a informação química desejada, isto é, a concentração de um analito de interesse. Nesse sentido, a performance do modelo analítico está associada a diferentes fatores: às características construtivas do sistema optoeletrônico de instrumentação, que é responsável pela medição dos dados ópticos; às técnicas matemáticas para processamento desta informação óptica; e ao processo de preparação e seleção de amostras representativas a partir das quais os dados ópticos são medidos. Assim, nas próximas seções, são apresentados e discutidos diferentes aspectos relacionados à calibração de instrumentos baseados no NIR relevantes para o desenvolvimento deste trabalho.

2.2 MÉTODOS DE CALIBRAÇÃO MULTIVARIADA E PRÉ-PROCESSAMENTOS ESPECTRAIS

A partir da medição das absorvâncias para um conjunto de amostras com concentrações de referência pré-determinadas, diversos métodos podem ser utilizados para estimar os coeficientes de regressão de um modelo linear. Entre eles, pode-se destacar: regressão linear múltipla (MLR - *Multiple Linear Regression*); regressão por mínimos quadrados parciais (PLS - *Partial Least Squares*); regressão de componentes principais (PCR - *Principal Components Regression*); e suas variações. A técnica mais adequada a cada problema depende, entre outros fatores, da dimensionalidade dos dados, isto é, do número de amostras e do número de variáveis (comprimentos de onda) disponíveis. Alguns métodos, como o PLS e o PCR, são frequentemente empregados quando existem muitas variáveis espectrais e poucas amostras. Estes métodos empregam procedimentos matemáticos para decompor os conjuntos de dados originais das amostras em componentes ortogonais cujas combinações lineares aproximam os dados originais para um determinado grau de exatidão desejado (Burns e Ciurczak, 2008). Dessa forma, a análise de regressão para obtenção dos coeficientes de regressão não é feita no domínio dos dados ópticos medidos diretamente, mas no domínio dos componentes ortogonais. Por outro lado, o método MLR corresponde à abordagem padrão, em que a análise de regressão é feita empregando os dados ópticos originais. Este método só é aplicável quando o número de amostras é superior ao número de variáveis espectrais. Assim, tendo em vista a construção de modelos com poucas variáveis discretas, o método MLR é utilizado neste trabalho.

Uma questão determinante para a qualidade dos modelos é a existência de outros fatores não associados às concentrações dos analitos em interesse, os quais também causam variações nos valores de absorbâncias medidos. Além do fator químico (absorção da luz pelos diferentes constituintes da amostra), a interação da radiação eletromagnética com a amostra envolve também fatores físicos, tais como a reflexão e o espalhamento da luz promovidos pelos constituintes da amostra. Assim, todos estes fatores em conjunto determinam o valor absoluto da intensidade luminosa que chega ao detector. Também, conforme discutido por Yadav et al. (2015), dependendo das características físico-químicas da solução a ser analisada, a maior parte da variância na intensidade de luz medida pode estar relacionada aos fatores físicos, especialmente ao espalhamento da luz, e não à absorção pelos analitos.

Diante do exposto, técnicas de pré-processamento de dados normalmente são necessárias para reduzir a influência dos fatores físicos nos dados espectrais originais e melhorar a aderência dos dados ao modelo (1.1). Entre as técnicas de pré-processamento normalmente empregadas, pode-se citar: correção de linha de base (*baseline correction*); correção multiplicativa do espalhamento de luz (MSC - *Multiplicative Scatter Correction*); variação normal padrão (SNV - *Standard Normal Variate*); derivadas; filtragens digitais por algoritmo Savitz-Golay; e suas variações. A Figura 2.2 ilustra a aplicação de três métodos de pré-processamento distintos em espectros de gasolina comum. Na Figura 2.2(a), são apresentados os espectros originais, sem pré-processamento, de 84 amostras de gasolina contendo diferentes concentrações de etanol entre 25% e 40% do volume. Pode-se observar que existe um espalhamento vertical das curvas espectrais, provavelmente associado aos efeitos físicos. Por sua vez, as Figuras 2.2(b), 2.2(c) e 2.2(d) mostram os espectros obtidos após pré-processamentos por correção da linha de base, SNV e primeira derivada, respectivamente.

É importante observar que alguns dos métodos de pré-processamento mais comumente empregados, como por exemplo as derivadas, só são aplicáveis ou geram bons resultados quando informações ópticas com elevada resolução estão disponíveis. Entretanto, este tipo de pré-processamento não seria adequado para um instrumento baseado em comprimentos de onda discretos. Dessa forma, a seleção de variáveis deve ser considerada em conjunto com o método de pré-processamento espectral escolhido. A grande variedade de técnicas de pré-processamento de dados espectrais e de análise de funções multivariadas indica que a calibração de um instrumento na área da espectroscopia NIR não é uma tarefa trivial.

Por fim, um terceiro fator que também é determinante para a qualidade da curva de calibração está relacionado com a determinação das concentrações de referência das amostras de calibração. O método de análise por NIR caracteriza-se por ser um método secundário, de forma que a incerteza obtida após a calibração do instrumento fica limitada à incerteza do método primário de referência. Uma discussão sobre o método de referência utilizado para quantificar os álcoois na gasolina é feita na Seção 2.5.

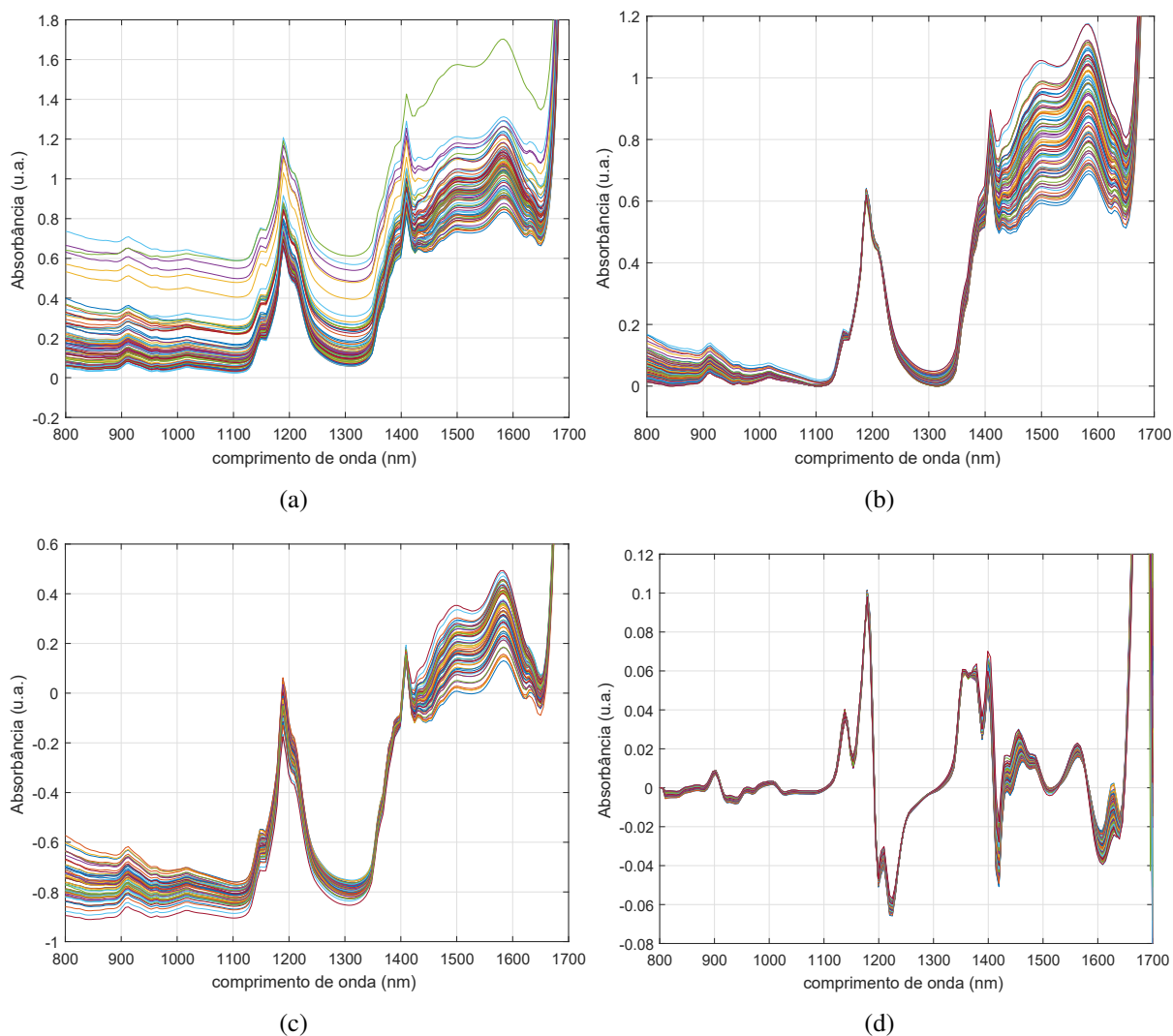


Figura 2.2: Espectros de 84 amostras de gasolina contendo diferentes concentrações de etanol: (a) sem pré-processamento; (b) após correção da linha de base; (c) após processamento por SNV; (d) e primeira derivada.

Fonte: o autor.

2.3 MEDIÇÃO DOS DADOS ÓPTICOS E MÉTODOS DE CÁLCULO DAS ABSORBÂNCIAS

Evidentemente, uma boa performance preditiva de um modelo depende da qualidade dos dados de entrada. Neste sentido, dois aspectos em particular são discutidos neste trabalho: a modelagem matemática do sistema eletrônico de instrumentação, responsável pela medição dos dados ópticos brutos; e os procedimentos experimentais utilizados para medir esses dados ópticos e posteriormente calcular os valores de absorvâncias.

No que se refere à medição dos sinais ópticos, diferentes tecnologias de fotodetectores estão disponíveis no mercado, tais como os fotoresistivos, fotoelétricos, ou piroelétricos. Entre os detectores mais comumente utilizados em construções alternativas identificadas na literatura, estão os fotodiodos baseados em semicondutores InGaAs, os fotoresistores de PbS e os detectores piroelétricos. Tais detectores podem ser encontrados como elementos sensores

pontuais ou em *arrays* lineares, geralmente com 128 ou 256 elementos sensores.

De forma geral, as características de cada tipo de sensor e sua aplicabilidade em diferentes tipos de aplicações são amplamente estudadas na literatura. Também, avaliações experimentais de características como sensibilidade, relação sinal ruído, estabilidade ao longo do tempo e influência da temperatura nos circuitos empregados por diferentes autores são amplamente discutidas (Haxha e Jhoja, 2016; Chagas, 2006; Ribeiro et al., 2016; Moreira et al., 2016). Apesar disso, o modelo matemático do detector em conjunto com o circuito eletrônico de instrumentação, a fim de relacionar a saída do circuito eletrônico com a intensidade de luz que chega ao detector, normalmente não é considerado. Naturalmente, a informação óptica que se deseja medir está relacionada com a resposta do circuito eletrônico, e deve ser sensível às variações na concentração do analito de interesse. Contudo, dependendo do tipo de detector e do circuito de instrumentação utilizados, esta relação entre a informação óptica e a resposta eletrônica pode ser não linear, ou pode existir um *offset* que deve ser considerado no cálculo das absorvâncias. Um exemplo de circuito não linear foi utilizado por Chagas (2006) para a análise de combustíveis, em que fotoresistores de PbS foram conectados em ponte de Wheatstone. Por sua vez, um exemplo de circuito de condicionamento eletrônico que introduz *offset* foi proposto por Moreira (2013). Este circuito é apresentado na Seção 2.3.8. Neste caso, se os valores de absorvância forem determinados a partir da resposta eletrônica desconsiderando-se a relação com os dados ópticos, isto pode aumentar o desvios dos dados em relação ao modelo matemático linear.

Por sua vez, além da medição dos dados ópticos em si, a determinação das absorvâncias depende do método utilizado para apresentar as amostras ao instrumento durante a coleta desses dados ópticos. De acordo com (2.1), a transformação dos dados ópticos em absorvâncias envolve a definição de uma condição experimental de referência em que o zero de absorvâncias é determinado. Isto é necessário porque, na prática, o termo I_0 , referente à intensidade de luz que incide na amostra, corresponde a uma intensidade de luz que também é desconhecida e que precisa ser medida. De acordo com a descrição dos procedimentos experimentais encontrados em diversos trabalhos na literatura, pode-se verificar que existem diferentes abordagens para a medição desta referência: as que utilizam um porta amostras vazio no caminho óptico entre a fonte de luz e o detector; as que utilizam uma amostra de referência; as que utilizam o espaço vazio; e ainda as que desconsideram a necessidade dessa medição de referência. Naturalmente, estas diferentes abordagens devem gerar valores de absorvâncias distintos e podem, assim, originar modelos de calibração com performances significativamente diferentes. Por isso, neste trabalho, sete diferentes métodos para determinação das absorvâncias identificados na literatura são estudados e comparados. Estes métodos são descritos nas subseções a seguir. Para permitir uma distinção clara entre eles, pode-se explicitar o cálculo das absorvâncias para diferentes amostras em termos dos diferentes comprimentos de onda. Assim, pode-se denotar por $A^j(\lambda)$ a absorvância em um dado comprimento de onda λ para uma dada

amostra de índice j , e pode-se reescrever a equação (2.1) como

$$A^j(\lambda) = \log \left(\frac{I_0(\lambda)}{I_S^j(\lambda)} \right), \quad (2.2)$$

em que $I_0(\lambda)$ é a intensidade de luz com comprimento de onda λ incidente na amostra; $I_S^j(\lambda)$ é a intensidade de luz transmitida ou refletida pela j -ésima amostra, de mesmo comprimento de onda λ , que chega ao detector; $j = 1, \dots, n$; $\lambda = \lambda_1, \dots, \lambda_k$.

2.3.1 Método A

Lutz et al. (2014) desenvolveram um espectrômetro NIR portátil para quantificação de etanol em gasolina que utiliza cubetas como porta amostras. De acordo com a descrição dos seus procedimentos experimentais, para cada amostra, seis medições foram realizadas e o espectro médio foi considerado. Além disso, uma medição de referência era tomada a cada seis medições. Os autores não especificaram em que condições estas medições de referência eram realizadas, mas acredita-se que tenham sido realizadas a partir de uma cubeta vazia no caminho óptico da luz, que é a forma mais comumente empregada em análises com espectrofotômetros de bancada. Também, Visconti et al. (2017) desenvolveram um protótipo para análise de gases ou líquidos, que pode funcionar nos modos de feixe simples ou de feixe duplo. No sistema de feixe simples, uma medição de referência é realizada com a cela de gás vazia antes da análise da amostra gasosa. Com base nesses dois exemplos, pode-se definir um método em que uma medição de referência é realizada para cada amostra, e as absorvâncias são calculadas de acordo com

$$A_A^j(\lambda) = \log \left(\frac{I_0^j(\lambda)}{I_S^j(\lambda)} \right), \quad (2.3)$$

em que $I_0^j(\lambda)$ corresponde às intensidades de luz medidas com uma cubeta vazia, as quais estão associadas à j -ésima amostra.

2.3.2 Método B

Gaião et al. (2008) desenvolveram um fotômetro portátil baseado em LEDs para análise qualitativa de gasolina, destinado a indicar adulterações por solventes e querosene. Os autores utilizaram como porta amostras uma cubeta de quartzo. Nas medições realizadas pelos autores, 100% de transmitância (ou zero de absorvância) também foi determinado a partir do porta amostras vazio. Contudo, entende-se pela descrição dos experimentos que esta referência foi obtida apenas uma vez no início do ciclo de análises. Assim, o método de cálculo realizado por Gaião et al. (2008), diferentemente do Método A, pode ser representado por

$$A_B^j(\lambda) = \log \left(\frac{I_0'(\lambda)}{I_S^j(\lambda)} \right), \quad (2.4)$$

onde $I_0'(\lambda)$ corresponde a uma única medição das intensidades de luz realizadas para uma cubeta

vazia, no início do ciclo de análises.

2.3.3 Método C

Guchardi e Pasquini (2001), por sua vez, desenvolveram um espectrômetro NIR que foi testado e avaliado para determinação de água em etanol combustível. O protótipo proposto por eles pode funcionar nos modos de feixe simples ou de feixe duplo. De acordo com os autores, no sistema de feixe simples, a medida de referência foi tomada uma vez quando a cubeta foi removida do caminho óptico, e as absorvâncias foram calculadas por

$$A_C^j(\lambda) = \log \left(\frac{I_0''(\lambda)}{I_S^j(\lambda)} \right), \quad (2.5)$$

em que $I_0''(\lambda)$ corresponde a uma medição das intensidades de luz que atingem o detector sem qualquer obstáculo. Este método pode ser considerado análogo ao utilizado por Isaac e Na (2016), baseado em reflectância, em que $I_0(\lambda)$ é a luz refletida por um material de referência, tal como disco cerâmico.

2.3.4 Método D

Chagas (2006), desenvolveu e avaliou diferentes protótipos de fotômetros, tanto de feixe simples como de feixe duplo, para análise de combustíveis. Para o sistema de feixe duplo, o autor empregou dois detectores, sendo o primeiro chamado detector de medida, que foi associado a um filtro óptico de comprimento de onda 1.480 nm (comprimento de onda correspondente a uma banda de absorção da substância sob teste); e o segundo, um detector de referência associado a um filtro em 1.300 nm (região espectral sem informação química). Então, a absorvância foi calculada a partir da razão entre as tensões geradas por ambos os detectores. Na prática, isto corresponde a um terceiro método diferente em que um comprimento de onda $\lambda_{ref} \in \{\lambda_1, \dots, \lambda_k\}$, onde não há informação química, é definido como referência para o cálculo das absorvâncias, isto é,

$$A_D^j(\lambda) = \log \left(\frac{I_S^j(\lambda_{ref})}{I_S^j(\lambda)} \right). \quad (2.6)$$

Isto pode ser entendido como uma medida da magnitude da luz atenuada em cada comprimento de onda em relação ao comprimento de onda de referência.

2.3.5 Método E

Nos testes para quantificação de água em álcool combustível utilizando o fotômetro de feixe simples, Chagas (2006) descreveu o cálculo da transmitância como sendo a razão entre as intensidades medidas para a amostra em interesse e aquelas medidas para uma amostra de referência, a qual continha, naquele caso, uma concentração específica do analito (94% de álcool). Isto é similar aos métodos utilizados por Kittipanyangam et al. (2016), que definem

$I_0(\lambda)$ como as intensidades de luz transmitidas por um solvente; Visconti et al. (2017), que considera um gás de referência, quando utiliza um sistema de feixe duplo; ou quando a referência é a matriz em que a substância a ser determinada está inserida. Nestes casos, uma absorbância relativa é calculada, a qual é dada por

$$A_E^j(\lambda) = \log \left(\frac{I_{ref}(\lambda)}{I_S^j(\lambda)} \right), \quad (2.7)$$

onde $I_{ref}(\lambda)$ corresponde a uma medição das intensidades de luz que chegam ao detector após interagir com uma solução definida como referência.

2.3.6 Método F

Uma abordagem mais simples foi utilizada por Ribeiro et al. (2016) e Gentilin et al. (2016), os quais desenvolveram protótipos baseados na reflexão da luz. Estes autores consideraram apenas que a intensidade de luz refletida por uma amostra é diretamente proporcional à reflectância. A partir desta consideração, pode-se assumir a reflectância (ou, analogamente, a transmitância) como sendo o valor absoluto da intensidade de luz medida para cada amostra, o que leva a outro método de cálculo das absorbâncias:

$$A_F^j(\lambda) = \log \left(\frac{1}{I_S^j(\lambda)} \right). \quad (2.8)$$

Assim, este método não considera nenhuma medição de referência.

2.3.7 Método G

Por fim, pode-se propor outro método que é similar ao definido em (2.3). Este também consiste em obter uma medição de referência com uma cubeta vazia para cada amostra. A única diferença é que, neste método, a mesma cubeta utilizada para determinar $I_0(\lambda)$ é utilizada na medição subsequente da amostra em interesse, sendo que a cubeta deve ser mantida imóvel no caminho óptico enquanto é preenchida com a amostra. Desta forma, este método de cálculo das absorbâncias pode ser escrito como

$$A_G^j(\lambda) = \log \left(\frac{I_{imovel}^j(\lambda)}{I_S^j(\lambda)} \right), \quad (2.9)$$

onde $I_{imovel}^j(\lambda)$ corresponde às intensidades de luz medidas para uma cubeta vazia, a qual é mantida imóvel no caminho óptico para a medida subsequente da j -ésima amostra.

Pode-se observar que os métodos A, B, C e G são baseados na medição direta ou indireta da radiação incidente na amostra, uma vez que $I_0(\lambda)$ é a intensidade de radiação transmitida ou refletida pelo ar ou por um material de referência (cubeta vazia pra transmitância, disco cerâmico para reflectância). Por sua vez, os métodos D e E são baseados na medição de

absorbâncias relativas a um comprimento de onda de referência (método D) ou a uma amostra de referência (método E), enquanto o método F é independente de uma medição de referência.

É evidente que, para qualquer tipo de aplicação, a calibração de um instrumento a partir de cada um destes métodos levará a diferentes resultados, o que, neste caso, é independente da exatidão do sistema de instrumentação. Neste trabalho, a influência de cada um desses métodos na performance de predição dos modelos foi avaliada empiricamente por meio do experimento descrito na Seção 3.1. Neste experimento, a medição dos dados ópticos foi realizada utilizando um fotômetro baseado em LEDs desenvolvido por Moreira (2013). Também, os dados deste experimento foram utilizados para avaliar experimentalmente a influência da variabilidade de matriz na gasolina. O protótipo utilizado é apresentado a seguir.

2.3.8 Fotômetro baseado em LEDs empregado nos experimentos preliminares

O diagrama de blocos da Figura 2.3 apresenta a estrutura do protótipo desenvolvido por Moreira (2013), o qual foi empregado na aquisição de dados ópticos no primeiro experimento desta tese. Trata-se de um fotômetro que utiliza três LEDs como fontes de radiação, centrados em comprimentos de onda distintos, e emprega análise por transmitância. Este protótipo foi desenvolvido inicialmente para a quantificação de água adicionada ao leite, em sistema de fluxo. Neste caso, foi necessário adaptá-lo para a aplicação estudada neste trabalho. Para isto, a estrutura mecânica do protótipo foi modificada para que fosse possível utilizar cubetas móveis ao invés de uma célula de fluxo fixa. Além desta, as únicas modificações significativas foram a seleção de novos LEDs, com comprimentos de onda diferentes do sistema original, e a realização de ajustes no circuito de condicionamento de sinais.

Na Figura 2.4, é apresentado o circuito de amplificação e condicionamento implementado no protótipo utilizado. Uma descrição detalhada do funcionamento deste circuito pode ser encontrada em Moreira (2013) e Moreira et al. (2016). Com base nas características de funcionamento do circuito, as tensões de saída do protótipo podem ser relacionadas com as intensidades de luz incidentes no detector como segue.

Neste protótipo, foi utilizado como detector um fotodiodo de InGaAs (PT511, da *Roithner Lasertechnik*), o qual foi conectado a um circuito de transimpedância que correspondente ao primeiro estágio do circuito da Figura 2.4. Neste modo de funcionamento, o fotodiodo opera em curto circuito. Conseqüentemente, a fotocorrente gerada é linearmente relacionada com a potência luminosa incidente, e vale a relação

$$I_r(\lambda) = I_p(\lambda)/r(\lambda), \quad (2.10)$$

onde $I_p(\lambda)$ é a fotocorrente gerada pelo fotodiodo para uma luz incidente de intensidade $I_r(\lambda)$ e comprimento de onda λ ; e $r(\lambda)$ é a responsividade do fotodiodo no comprimento de onda λ . Deve-se notar que $I_r(\lambda)$ pode corresponder tanto à luz que chega ao detector após interagir com a amostra, $I_S^j(\lambda)$, quanto à luz incidente na condição de referência, $I_0(\lambda)$.

Por sua vez, conforme descrito por Moreira (2013), o circuito da Figura 2.4

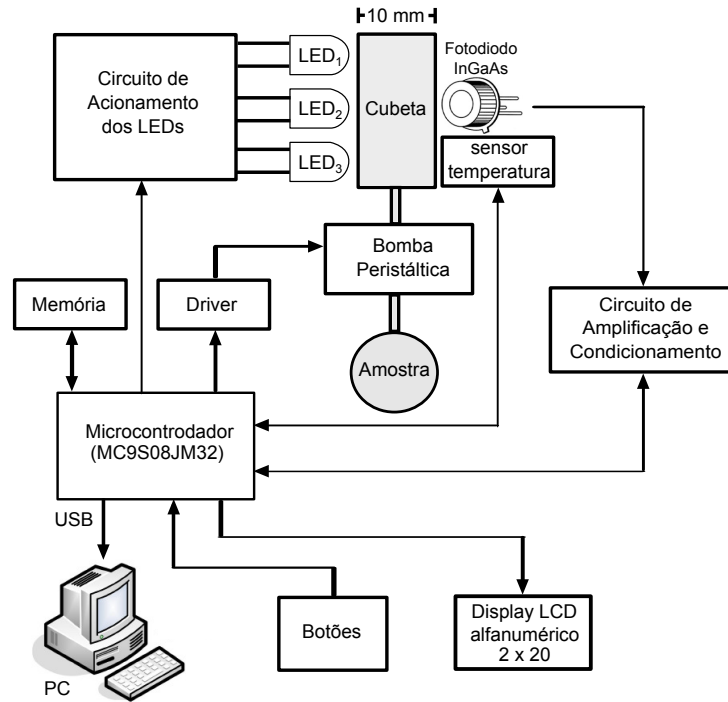


Figura 2.3: Diagrama de blocos do fotômetro desenvolvido por Moreira (2013).

produz uma tensão de saída $V(\lambda)$ que é dada por

$$V(\lambda) = c_1(\lambda)I_p(\lambda) - c_2(\lambda), \lambda \in \{\lambda_1, \lambda_2, \lambda_3\}, \quad (2.11)$$

onde $c_1(\lambda)$ e $c_2(\lambda)$ são constantes que dependem das resistências e tensões de referência do circuito. Neste trabalho, os valores de $c_1(\lambda)$ e $c_2(\lambda)$ foram pré-determinadas a partir dos ajustes descritos na Seção 3.1.

A partir de (2.11), observa-se que as fotocorrentes podem ser obtidas a partir da medição das tensões de saída. Por sua vez, relacionando (2.10) com (2.3), (2.4), (2.5), (2.7) e (2.9), pode-se verificar que as intensidades de luz podem ser substituídas pelas fotocorrentes

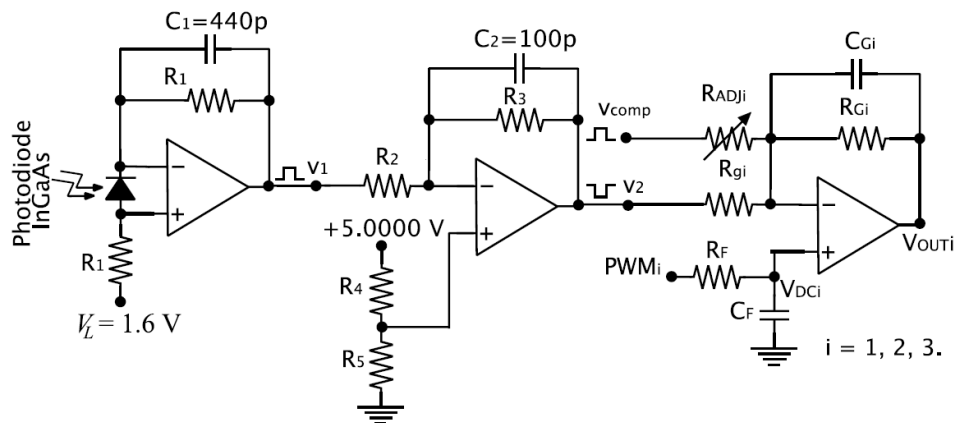


Figura 2.4: Circuito de detecção e condicionamento de sinal óptico do protótipo desenvolvido por Moreira (2013).

medidas no cálculo das absorbâncias, as quais se tornam independentes da responsividade do sensor. Isto é válido para os métodos descritos na Seção 2.3, com exceção dos métodos D e F. Por outro lado, no método D, relacionando-se (2.10) e (2.6), obtém-se

$$A_D^j(\lambda) = \log \left(\frac{r(\lambda)}{r(\lambda_{ref})} \right) + \log \left(\frac{I_{ps}^j(\lambda_{ref})}{I_{ps}^j(\lambda)} \right), \quad (2.12)$$

onde $I_{ps}^j(\lambda)$ é a fotocorrente gerada quando a luz de comprimento de onda λ atinge o detector após interagir com a j -ésima amostra, e λ_{ref} é o comprimento de onda utilizado como referência para este método. Por sua vez, no método F, relacionando-se (2.10) e (2.8), obtém-se

$$A_F^j(\lambda) = \log(r(\lambda)) + \log \left(\frac{1}{I_{ps}^j(\lambda)} \right). \quad (2.13)$$

As equações (2.12) e (2.13) mostram que as absorbâncias, quando calculadas pelos métodos D e F, tornam-se dependentes da responsividade do sensor, que é uma função de λ e varia com a temperatura, por exemplo. Assim, se este termo for ignorado, ao se medir uma fotocorrente pequena em um dado λ , pode-se interpretar que existe uma grande absorção associada àquele comprimento de onda. Na realidade, poderia ser apenas que a resposta do sensor naquele comprimento de onda fosse menos intensa. Apesar disso, neste estudo, os termos dependentes da responsividade em ambos os métodos foram ignorados para fins de comparação entre todos os métodos nos testes realizados. Isto é razoável uma vez que os experimentos, foram realizados em um intervalo de tempo relativamente curto e em um ambiente com temperatura controlada, de forma que a responsividade em cada λ pode ser considerada constante durante a execução do experimento. Dessa forma, nos testes realizados com este protótipo, as absorbâncias em todos os sete métodos avaliados foram calculadas por meio das equações (2.3) a (2.9), em que as intensidades de luz foram substituídas pelas fotocorrentes geradas pelo fotodiodo.

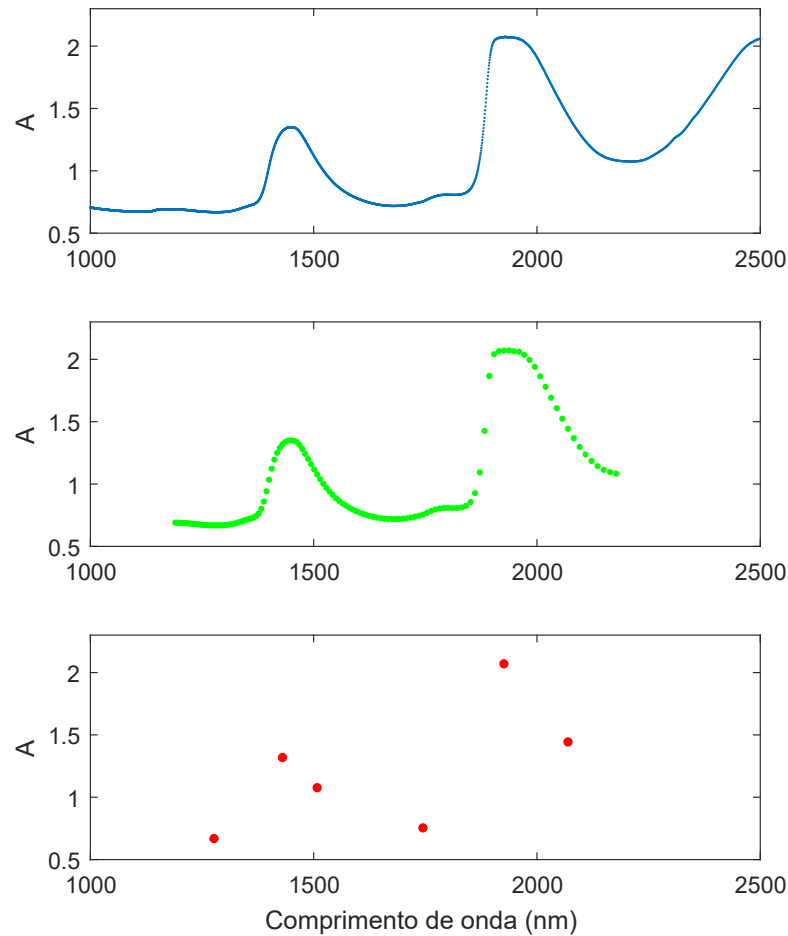
Na seção a seguir, são apresentados algumas tecnologias existentes que permitem a separação e medição dos diferentes comprimentos de onda da luz, que consistem na base da análise espectral.

2.4 SELEÇÃO DE COMPRIMENTOS DE ONDA E RESOLUÇÃO ESPECTRAL

Diversos tipos de instrumentos analíticos baseados na espectroscopia NIR tem sido desenvolvidos, os quais diferem entre si, entre outros fatores, pela tecnologia óptica utilizada para separar e permitir a observação dos diferentes comprimentos de onda. A Figura 2.5 mostra uma representação dos dados espectrais de uma substância genérica, quando medidos a partir de diferentes tipos de instrumentos, os quais são discutidos a seguir.

Os instrumentos mais completos, sem dúvida, são os espectrofotômetros de bancada, baseados em interferômetros e transformada de Fourier, que permitem aquisição de espectros em uma ampla faixa do NIR, com alta resolução espectral, alta relação sinal-ruído,

Figura 2.5: Representação dos dados espectrais de uma substância genérica medidos a partir de diferentes tipos de instrumentos.



Fonte: o autor.

elevada reprodutibilidade dos comprimentos de onda, entre outras características, e são genéricos para os mais diferentes tipos de aplicações. A primeira curva da Figura 2.5 representa o espectro de uma substância genérica medido a partir de um espectrofotômetro. Apesar do elevado desempenho oferecido por tais equipamentos, as principais desvantagens são o alto custo e limitação de uso em laboratório. Construções alternativas, por sua vez, pode ser realizadas empregando instrumentos dispersivos, como grades de difração, ou filtros ópticos lineares variáveis (LVOF - *Linear Variable Optical Filter*) aliadados a *arrays* de detectores; ou ainda por meio de filtros opto-acústicos sintonizáveis (AOTF - *Acousto Optic Tunable Filter*). Todos esses dispositivos permitem varredura de uma ampla faixa do espectro eletromagnético, fornecendo informações sobre múltiplos comprimentos de onda, mas em geral com menor resolução em relação a um espectrofotômetro de bancada, como representado na segunda curva da Figura 2.5. Estes tipos de soluções apresentam vantagens como maior robustez para uso em campo, e também permitem construções genéricas para diversas aplicações. Apesar disso, este tipo de solução ainda é relativamente cara para uso em larga escala.

Outra categoria de instrumentos são aqueles baseados em filtros de interferência óptica e/ou diodos emissores de luz (LED). Estes dispositivos possibilitam a construção de instrumentos que fornecem informações espectrais em comprimentos de onda discretos de bandas estreitas, centrados em diferentes regiões espectrais de interesse, com bom custo-benefício. Isto é representado no terceiro gráfico da Figura 2.5. Este tipo de solução facilita a construção de instrumentos portáteis e dedicados para aplicações específicas, em que um pequeno número de comprimentos de onda seja requerido para realizar uma determinação analítica. Além do custo, tais sistemas possuem vantagens como maior facilidade na montagem (menor dificuldade em alinhamentos ópticos) e maior estabilidade ao longo do tempo, o que pode aumentar a robustez para uso em campo. Nesse sentido, muitos trabalhos foram desenvolvidos propondo instrumentos com as mais variadas configurações ópticas (Visconti et al., 2017; Gaião et al., 2005; Gentilin, 2012; Lima, 2012; Fonseca e Raimundo, 2007; Nascimento, 2008; Ribeiro et al., 2016; Moreira, 2013), para diferentes tipos de aplicações. Neste trabalho, avalia-se a construção de um protótipo que emprega LEDs como fontes de luz e filtros de interferência para seleção dos comprimentos de onda de interesse.

Estudos já foram realizados para a análise de combustíveis a partir de espectrofotômetros de bancada, que possibilitaram a coleta de dados espectrais com cerca de cerca de mil pontos e com resolução da ordem 2 nm (Fernandes et al., 2008; Ouyang e Liu, 2013). Outros autores, como Correia et al. (2018) e Lutz et al. (2014), propuseram estudos a partir da aquisição de 128 pontos espectrais, com resoluções de 6 a 8 nm. Por fim, estudos com combustíveis também foram realizados a partir de instrumentos com dois comprimentos de ondas discretos (Harris et al., 1993), e até mesmo por meio de um único comprimento de onda (Chagas, 2006; Gaião et al., 2008; Nascimento, 2008).

A seleção dos comprimentos de onda mais adequados para compor um modelo matemático de medição normalmente é feita a partir de um estudo espectral empregando dados medidos por um espectrofotômetro de bancada, em uma ampla faixa espectral e com alta resolução. A partir dos dados dos espectros completos, dois tipos de métodos de seleção de variáveis são comumente empregados: métodos de seleção de variáveis individuais, os quais selecionam conjuntos de variáveis discretas relacionando-as com a concentração em interesse; e métodos de seleção de variáveis em grupos, os quais estudam a associação entre a concentração em interesse com intervalos de comprimentos de onda (Lin et al., 2016). Entre os métodos que empregam seleção de variáveis discretas individuais, pode-se citar: UVE (*Uninformative Variable Elimination*); SA (*Simulated Annealing*); CARS (*Competitive Adaptive Reweighted Sampling*); VISSA (*Variable Iterative Space Shrinkage Approach*); BOSS (*Bootstrapping Soft Shrinkage*), SPA (*Successive Projections Algorithm*), entre outros. Já entre os métodos de seleção de variáveis baseados em agrupamento de variáveis, pode-se citar: iPLS (*Interval Partial Least Squares*); FOSS (*Fisher optimal subspace shrinkage*) e iSPA (*Interval Successive Projections Algorithm*), entre diversas outras variações. Estes métodos exploram o conceito de que comprimentos de onda vizinhos, normalmente, são altamente correlacionados e carregam informações similares. Assim, este tipo de técnica consiste em agrupar variáveis consecutivas, de

forma que variáveis altamente correlacionadas tendem a ser selecionadas ou deixadas de fora do modelo juntas (Lin et al., 2016).

A grande quantidade de métodos de seleção de variáveis existentes demonstra que este tema já é amplamente discutido na literatura, para muitos tipos de aplicações (Lin et al., 2016; Rui et al., 2017; Marois et al., 2018). Contudo, normalmente, o intuito desta seleção de variáveis é reduzir a dimensão dos dados disponíveis, com o objetivo de otimizar a performance preditiva do modelo. Isto é, utilizar um número de variáveis menor do que a quantidade disponível, eliminando-se comprimentos de onda não informativos, pode reduzir, em muitos casos, o chamado ruído espectral, resultando em modelos com melhores performances de predição (Rui et al., 2017). Também, neste tipo de abordagem, a seleção de variáveis pode ser realizada a partir dos espectros já pré-processados, incluindo técnicas como as derivadas.

Diferentemente, este trabalho emprega uma outra abordagem da seleção de variáveis. Ela é utilizada como uma ferramenta para especificar e otimizar a configuração óptica do instrumento. Isto é, a escolha das variáveis deve considerar que, na prática, dados de um número já restrito de comprimentos de onda estarão disponíveis para serem pré-processados e empregados na construção do modelo. Nesse sentido, técnicas de pré-processamento como as derivadas não seriam aplicáveis. Por isso, o estudo espectral deve partir dos dados originais medidos pelo espectrofotômetro, sem qualquer transformação de variáveis. Neste trabalho, apenas o pré-processamento por correção de linha de base é empregado, e é proposto um método de seleção de variáveis baseado no algoritmo SPA.

Na próxima seção, é apresentada uma revisão bibliográfica sobre os trabalhos existentes na literatura relacionados à análise de gasolinas. São analisadas as características e os procedimentos experimentais utilizados pelos autores, relacionado-os com os diversos conceitos discutidos até aqui.

2.5 ESPECTROSCOPIA NIR NA ANÁLISE DE GASOLINAS COMERCIAIS

A gasolina automotiva é um tipo de combustível derivado do petróleo, que corresponde a uma mistura complexa de hidrocarbonetos contendo de oito a doze átomos de carbono. Em muitos países, metanol e etanol anidros são adicionados às gasolinas, resultando em um combustível com menor emissão de poluentes. Também, com o objetivo de aumentar a octanagem, aditivos são frequentemente adicionados à gasolina (Fernandes et al., 2008). Mais especificamente, as gasolinas automotivas comercializadas no Brasil são classificadas entre comum (que pode ser com ou sem aditivos) e premium (que normalmente possui aditivos). Ambos os tipos são obtidos a partir da gasolina A (sem etanol anidro), produzida pelas refinarias, por formuladores, centrais petroquímicas ou, ainda, importada por empresas autorizadas pela Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP). As gasolinas recebem adição de etanol anidro pelos distribuidores autorizados, em quantidade oficialmente estabelecida, e posteriormente são disponibilizadas aos consumidores finais. Por sua vez, a adição de metanol é proibida pela legislação vigente, caracterizando-se como um contaminante, o qual possui

teor máximo admissível de 0,5 % do volume. Atualmente, o teor obrigatório de etanol anidro é fixado pela Portaria MAPA Nº 75 de 05/03/2015, do Ministério da Agricultura, Pecuária e Abastecimento, em 27% para a gasolina comum e 25 % para a gasolina premium.

Para avaliar a qualidade dos combustíveis, a Resolução ANP nº 807 de 23 de janeiro de 2020 estabelece que a medição do teor de etanol deve ser realizada mediante o emprego dos métodos de ensaio definidos pelas normas ABNT NBR 13992: "Gasolina automotiva - Determinação do teor de álcool etílico anidro combustível (AEAC)" e ASTM D5501: "*Determination of Ethanol and Methanol Content in Fuels Containing Greater than 20% Ethanol by Gas Chromatography*". Por sua vez, a ANP estabelece como método referência para a determinação de metanol em gasolinas a cromatografia gasosa, por meio dos ensaios definidos nas normas ABNT NBR 16041: "Etanol combustível - Determinação dos teores de metanol e etanol por cromatografia gasosa".

Como já discutido, a diferenciação entre etanol e metanol não é simples. O próprio método de referência ABNT NBR 13992 (ABNT, 1997), por exemplo, não permite detectar a adulteração de gasolina por metanol, uma vez que não faz distinção entre os dois compostos. Este método de referência é um teste simples de extração em que a quantificação do álcool é baseada no aumento de volume da fase aquosa. Por meio desse método, a adulteração é avaliada apenas em termos do volume total de álcool presente na amostra. Também, a exatidão do método ABNT NBR 13992 é limitada por fatores como: resolução das provetas; interpretação livre da leitura da escala das provetas e dos arredondamentos; e paralaxe da medição. Por isso, as concentrações medidas por este método podem apresentar incertezas superiores a 1 % em volume (v/v) (Zucchini et al., 2008; Fernandes et al., 2008; ABNT, 1997).

Por outro lado, o método de referência alternativo baseado em cromatografia gasosa permitiria determinar as concentrações de referência com uma incerteza muito menor do que o método ABNT NBR 13992. A cromatografia gasosa foi empregada como método de referência nos trabalhos de Ouyang e Liu (2013) e Lutz et al. (2014). Entretanto, trata-se de um método complexo, que requer: um equipamento e outros recursos específicos para a análise das amostras; conhecimento técnico especializado para a operação e modelagem de dados coletados; além de um extenso trabalho prévio para estabelecer e validar o método analítico. Em contraste, método ABNT NBR 13992 emprega materiais simples e o resultado é obtido em apenas quinze minutos. Por esses motivos, o método ABNT NBR 13992 foi empregado em diversos trabalhos descritos na literatura para determinar as concentrações de referência para a calibração. Este método também foi empregado neste trabalho.

Evidentemente, a incerteza das concentrações medidas pelo método secundário baseado do NIR fica limitada à incerteza dada pelo método de referência. Também, o tipo de instrumento proposto neste trabalho não tem a função de substituir o método de referência. Ao invés disso, busca-se um instrumento portátil para a triagem de adulterações em campo, reduzindo o tempo de medição e a necessidade de análises mais complexas em laboratório. Por isso, considerando todas as variáveis já discutidas e as características do instrumento proposto (baseado em comprimentos de onda discretos, de baixo custo), é aceitável uma incerteza na

quantificação dos álcoois da ordem 2 a 3 % v/v.

Por meio do NIR, a quantificação de metanol ou etanol na gasolina, separadamente, já foi estudada por alguns autores. Chagas (2006), por exemplo, avaliou a construção de um fotômetro portátil baseado em filtros ópticos para medir o teor de água no álcool combustível e o teor de etanol na gasolina. Para determinar os comprimentos de onda mais adequados para ambas as aplicações, o autor realizou análises espectrais por meio de um espectrofotômetro de bancada e selecionou um único comprimento de onda (filtro óptico de banda estreita em 1.480 nm), região espectral característica de absorção pelas ligações OH existentes tanto da água quanto do álcool. Esta seleção foi feita exclusivamente com base na sensibilidade do sinal analítico medido neste comprimento de onda. O autor optou por realizar calibração com base em um modelo matemático linear univariado baseada na Lei de Beer, e reportou valores médios para a incerteza na quantificação do etanol da ordem de 3 % v/v, tendo como referência o método padrão ABNT NBR 13992. A utilização de apenas um comprimento de onda proposta por Chagas (2006) provavelmente não é adequada em termos práticos devido à variabilidade amostral e aos diversos outros fatores já discutidos. Neste caso, este tipo de abordagem provavelmente gerará erros de predição ainda maiores na prática, o que compromete a aplicabilidade do instrumento. Também, de acordo com o autor, o protótipo foi projetado para que fosse possível utilizá-lo para ambas as aplicações (determinação de etanol na gasolina e de água no etanol combustível). Uma abordagem semelhante que emprega método univariado foi proposto por Nascimento (2008), porém para análise qualitativa da adulteração de gasolinas por adição de solventes e querosene. Estas abordagens demonstram que utilizar apenas um comprimento de onda não permite diferenciar dois compostos distintos. Analogamente, este tipo de abordagem univariada também não permitiria diferenciar metanol e etanol e, portanto, não é adequada para o problema estudado neste trabalho.

Por sua vez, Harris et al. (1993) propôs uma técnica para analisar misturas de combustíveis internamente em veículos. A técnica é baseado em transmitância e destina-se a detectar concentrações de álcool (etanol ou metanol) em misturas com gasolinas. A estrutura óptica do instrumento proposto emprega filtros de interferência em dois comprimentos de onda discretos. De acordo com o autor, os dois comprimentos de onda são escolhidos tal que, em um deles, os álcoois sejam fortemente absorvedores enquanto a gasolina seja transparente; e, no outro, tanto os álcoois quanto a gasolina devem apresentar pequena absorção. O instrumento foi proposto para medir concentrações de álcool de até 85%, e o autor aponta como adequados comprimentos de onda nas regiões de 1.500 e 1.300 nm, respectivamente. Evidentemente, este tipo de solução também não permite a diferenciação entre os dois álcoois. Também, a descrição do procedimento proposto para seleção dos comprimentos de ondas demonstra-se extremamente superficial, tendo em vista as diversas questões discutidas neste trabalho. Por fim, a ampla faixa de concentrações considerada permite inferir que esta metodologia provavelmente apresenta erros de quantificação elevados e incompatíveis com a proposta estudada.

Já Lutz et al. (2014) desenvolveram um espectrômetro NIR portátil para determinar etanol em gasolina. O instrumento é baseado em um array de diodos de 128 pixels,

utiliza como elemento dispersivo um LVOF e permite aquisição de dados na região de 1128 a 2161 nm, com resolução de 8 nm. Os autores utilizaram seis amostras adquiridas em postos de combustíveis locais, as quais foram analisadas previamente por cromatografia gasosa e espectrometria de massa para caracterização inicial quanto à presença de etanol e metanol. Os autores preparam 36 amostras com concentrações de etanol entre 0 e 10 % (w/w). A curva de calibração foi obtida por PLS com pré-processamento por correção da linha de base, e o valor de RMSEP (raiz quadrada do erro quadrático médio avaliado para as amostras de teste) reportado foi de 0,20%. O trabalho realizado por Lutz et al. (2014) considerou variabilidade de matriz, e o erro obtido a partir da aquisição de 128 pontos espectrais foi satisfatório. Contudo, também avaliou-se apenas a quantificação de etanol em misturas binárias. Além disso, o número de amostras utilizado pode ser considerado pequeno para cobrir todas as fontes de variância em uma aplicação prática. Este, assim como os demais estudos citados, não permite quantificar o metanol em gasolinas tipo C pois, neste caso, tem-se uma mistura ternária, o que requer a distinção entre os dois compostos. Por sua vez, a detecção simultânea de etanol e metanol em gasolinas foi estudada por Correia et al. (2018), Fernandes et al. (2008) e por Ouyang e Liu (2013).

Correia et al. (2018) avaliaram a aplicabilidade de um instrumento analítico portátil no controle de qualidade de combustíveis. Os autores utilizaram um mini espectrofotômetro comercial (MicroNIR Pro 1700, da Viavi Solutions Inc.), o qual utiliza como fonte de radiação duas lâmpadas de tungstênio; um filtro linear variável como elemento dispersivo; e um *array* de fotodiodos de 128 pixels como detector (resolução de 6,2 nm para a faixa 908-1676 nm). Os autores avaliaram diferentes tipos de aplicações, entre elas a quantificação de metanol e etanol em gasolina tipo C. Os autores utilizaram misturas ternárias e planejaram o experimento a fim de cobrir a faixa de concentrações entre 25 e 75% para o etanol, e de 0 a 50% (w/w) para o metanol. Contudo, foram utilizadas para calibração e validação apenas 16 amostras. Os autores não esclarecem se as amostras foram preparadas a partir de gasolina tipo A ou de gasolina tipo C, nem qual o método utilizado para determinação das concentrações de referência neste caso. De qualquer maneira, o número de amostras utilizado sugere que não houve variação de matriz. Na análise multivariada de dados, foi utilizada como pré-processamento espectral a primeira derivada pelo algoritmo Savitz-Golay, e a calibração por PLS. Devido ao número limitado de amostras, foi empregado apenas um método de validação cruzada, e os erros médios reportados foram de 3,81% e 1,80% (w/w) na quantificação de etanol e metanol, respectivamente. Os autores apontam como variáveis importantes para a quantificação de etanol e metanol em gasolina a região espectral entre 1.300 e 1.600 nm, a qual é atribuída ao segundo sobretom do grupo funcional OH presentes em ambos os álcoois. Apesar disso, nenhuma análise a respeito das características espectrais que permitem diferenciar os dois compostos é feita. O trabalho de Correia et al. (2018) demonstra claramente que a faixa do NIR é sensível às variações nas concentrações dos dois álcoois, e sugere que é possível diferenciar e quantificar os dois compostos a partir de 128 pontos espectrais na faixa de 908 a 1676 nm. Contudo, devido ao pequeno número de amostras utilizado e à não consideração da influência da variabilidade

amostral, torna-se difícil avaliar os parâmetros de adequação do modelo analítico bem como sua aplicabilidade prática. Também, pode-se destacar que a faixa de concentrações empregada para ambos os álcoois foi muito extensa. Uma redução desta faixa poderia contribuir para uma melhor performance do método analítico.

Por sua vez, Ouyang e Liu (2013) estudaram a quantificação de etanol e metanol em gasolina a partir de dados coletados de um espectrofotômetro FT-NIR na faixa de 12.000 a 4.000 cm^{-1} (833 a 2.500 nm), com resolução de 8 cm^{-1} . Os autores utilizaram um conjunto de 64 amostras, sendo 32 misturas de gasolina A com metanol, e outras 32 misturas de gasolina A com etanol, ambos com concentrações na faixa entre 1,2 e 24% v/v. A análise de dados realizada empregou pré-processamentos por correção de linha de base e primeira derivada por Savitz-Golay. Na calibração, os autores utilizaram o método PLS para uma faixa mais restrita, de 1100 a 1680 nm; e também avaliaram a aplicação dos algoritmos genético e SPA (*Successive projections algorithm*) para a seleção de variáveis. Os valores de RMSEP reportados para a quantificação de metanol e etanol foram de 0,163% e 0,235%, utilizando 19 e 7 variáveis, respectivamente. A partir da análise realizada com o espectro completo, os autores indicam a possibilidade de classificar corretamente gasolinas com etanol ou com metanol. Apesar dos valores de RMSEP serem baixos, observa-se que apenas misturas binárias foram consideradas, o que não atende às necessidades do problema estudado neste trabalho. Também, a seleção de variáveis foi realizada a partir dos espectros completos e já pré-processados, o que não elimina a necessidade de aquisição de todo o espectro. Por fim, o estudo também não avaliou a influência de matriz.

Por fim, Fernandes et al. (2008) também utilizam um espectrômetro FT-NIR de bancada, com aquisição de espectros na região de 14.000 a 4.000 cm^{-1} (714 a 2.500 nm) com resolução de 8 cm^{-1} (cerca de 1250 pontos espectrais). Neste estudo, os autores utilizaram 80 amostras ternárias para calibração e 40 para validação, as quais foram preparadas a partir de gasolina tipo A, pela adição de ambos os álcoois com concentrações finais na faixa de 0 % a 30 % v/v. Os autores avaliaram diferentes regiões espectrais e selecionaram a região entre 1105 e 1682 nm, a partir da qual construíram modelos de calibração por PLS com no máximo 4 variáveis latentes. Os valores reportados de RMSECV (raiz quadrada do erro quadrático médio de validação cruzada) foram 0,297 % e 0,459 % v/v para metanol e etanol, respectivamente, quando foi utilizado pré-processamento por primeira derivada. Já para o pré-processamento por correção da linha de base, os valores de RMSECV cresceram para 0,573 % e 0,814 %, respectivamente; e, finalmente, atingiram 0,650 % e 0,787 % quando nenhum pré-processamento foi aplicado. Para o conjunto de validação independente, foram reportados valores de RMSEP de 0,28 % v/v para a quantificação de etanol; e de 0,32 % v/v para a quantificação de metanol. Em outra etapa, os autores investigaram a influência da composição da gasolina na determinação dos álcoois. Para isto, a composição inicial da gasolina A foi modificada em laboratório pela adição de diferentes hidrocarbonetos normalmente encontrados na gasolina, em diferentes concentrações. A partir deste estudo, os autores concluíram que variações nas concentrações dos hidrocarbonetos isooctano, o-xileno, m-xileno, p-xileno e tolueno podem afetar significa-

tivamente a determinação de etanol e metanol, podendo levar a erros de predição de até 2 %, e indicam a necessidade de se utilizar amostras representativas da variabilidade total da composição da gasolina durante a construção do modelo matemático. Por fim, os autores também avaliaram a determinação de álcool em 15 gasolinas tipo C comerciais diferentes, e os resultados foram comparados com as concentrações de referência dadas pelo método padrão ABNT NBR 13992. Neste caso, os autores reportam valores de RMSEP de 0,83 % na determinação do etanol. Os resultados apresentados por Fernandes et al. (2008) mostram a importância do pré-processamento espectral, e indicaram que a primeira derivada pode melhorar significativamente a aderência dos pontos ao modelo. Apesar disso, conforme discutido na Seção 2.2, este tipo de pré-processamento pode não ser aplicável, no caso de um instrumento com poucos comprimentos de onda discretos; ou pode levar a resultados ruins, caso a resolução espectral seja pequena. Também, os autores concluíram que resultados significativamente diferentes podem ser obtidos a partir da seleção de diferentes regiões espectrais. De acordo com os autores, a seleção de regiões adequadas pode, ao mesmo tempo, otimizar a distinção entre etanol e metanol, ao incluir regiões que apresentem absorção pelo metil (-CH₃, característico do etanol) e metileno (-CH₂, comum a ambos os álcoois), e minimizar a dependência da composição da gasolina, ao eliminar regiões com maior interferência dos hidrocarbonetos. Dessa forma, fica evidente a importância de se determinar os comprimentos de onda para a construção e validação do protótipo desejado a partir de amostras reais. Isto porque o erro obtido analisando-se amostras comerciais de gasolina tipo C (RMSEP = 0,83%) foi significativamente superior àquele obtido na validação com amostras preparadas em laboratório a partir da gasolina tipo A (RMSEP = 0,28%). Estes resultados evidenciam que a construção de um protótipo de baixo custo não é um problema simples, quando consideradas as condições reais de operação esperadas para o instrumento.

Após a análise dos trabalhos citados, conclui-se que a quantificação de qualquer álcool baseada em apenas um ou dois comprimentos de onda discretos não é adequada em termos práticos. Por sua vez, os resultados obtidos por Lutz et al. (2014) indicam que a utilização de um tipo construtivo com até 128 pontos espectrais pode levar a resultados satisfatórios na quantificação simultânea de etanol e metanol proposta. Assim, este trabalho investiga qual o menor conjunto de comprimentos de onda que produz modelos de calibração com performances preditivas satisfatória. No próximo capítulo, são descritos os materiais e métodos utilizados para a solução do problema estudado.

3 PROCEDIMENTOS EXPERIMENTAIS

Para desenvolver os modelos de calibração para estimar as concentrações de etanol e metanol na gasolina, empregou-se os procedimentos introduzidos a seguir.

Inicialmente, comparou-se experimentalmente os sete métodos de cálculo de absorvâncias descritos na Seção 2.3. O objetivo deste experimento foi analisar e comparar a aplicabilidade, bem como as vantagens e as desvantagens de cada método, a fim de determinar qual deles é o mais adequado para gerar as variáveis preditoras dos modelos. Neste experimento, foram empregadas amostras provenientes de um único posto de combustíveis. Analisou-se um modelo de duas variáveis, para a predição de etanol, construído a partir de dados ópticos medidos pelo protótipo desenvolvido por Moreira (2013), descrito na Seção 2.3.8.

Em seguida, analisou-se também os dados deste primeiro experimento quanto à metodologia de preparação e seleção das amostras de calibração. Conforme já discutido, a capacidade preditiva futura de um modelo também é influenciada pela qualidade dos dados de entrada, o que é independente da medição e do processamento de dados ópticos em si. O objetivo foi avaliar se existe variação de matriz significativa entre diferentes amostras de um único posto de combustíveis, a qual influencie na capacidade preditiva do modelo de duas variáveis. Os resultados deste experimento foram responsáveis por guiar a preparação e a seleção de amostras nos experimentos subsequentes, bem como para definir a estratégia de validação dos modelos finais.

Posteriormente, para fundamentar o projeto do sistema óptico do protótipo, foi necessário determinar os comprimentos de onda analiticamente mais importantes para a quantificação dos dois álcoois. O objetivo foi selecionar as variáveis espectrais que geram modelos com menores erros na predição de ambos os álcoois em misturas ternárias. Para isto, realizou-se um estudo espectral utilizando amostras provenientes de diferentes fornecedores. O estudo foi realizado a partir de dados ópticos com alta resolução espectral, medidos por um espectrofotômetro de bancada, contendo informações de 2.194 comprimentos de onda na faixa de 900 a 1.650 nm. Então, seguindo a metodologia de Paiva et al. (2019), estes dados foram reduzidos para 37 comprimentos de onda candidatos, que correspondem a comprimentos de onda de filtros ópticos de interferência disponíveis comercialmente. Estes correspondem aos comprimentos de onda implementáveis na prática na construção do protótipo. Os dados dos espectros reduzidos foram então utilizados para identificar as dez variáveis espectrais mais informativas e para construir diferentes modelos candidatos, contendo entre três e dez variáveis preditoras.

Por fim, a partir do estudo espectral empregando dados do espectrofotômetro, projetou-se um protótipo baseado em LEDs e filtros ópticos de interferência para implementar modelos com sete variáveis. A performance preditiva dos modelos construídos a partir dos dados do protótipo foram então comparados com os mesmos modelos construídos a partir dos

dados do espectrofotômetro. Cada uma dessas etapas são detalhadas nas seções a seguir.

3.1 ESTUDO DA MEDIÇÃO E MODELAGEM MATEMÁTICA DOS SINAIS ÓPTICOS PARA OBTENÇÃO DAS VARIÁVEIS PREDITORAS

Para comparar experimentalmente os sete métodos de cálculo de absorbâncias, optou-se por limitar as variáveis do problema de forma a facilitar a análise e interpretação dos resultados. O experimento foi realizado utilizando amostras binárias de gasolinas diluídas apenas com etanol. Assim, nesta etapa, a adulteração por metanol não foi considerada. Também, o experimento foi realizado a partir de gasolinas comerciais adquiridas de um único posto de combustíveis, de forma que a influência da variabilidade de matriz fosse limitada. Assim, considerando esse universo amostral restrito, um modelo de duas variáveis espectrais, selecionadas simplesmente pelas características de absorção do analito de interesse, é suficiente para este estudo. Neste caso, empregou-se neste estudo o protótipo baseado em LEDs de Moreira (2013), que permite medir as informações ópticas de até três comprimentos de onda. Evidentemente, como discutido no Capítulo 1, esta abordagem para determinação dos comprimentos de onda não seria adequada em termos práticos para a quantificação dos álcoois desejada, devido aos problemas de variabilidade de matriz e outros problemas já discutidos. Apesar disso, tendo em vista o universo restrito de amostras considerado, essa metodologia permite avaliar e comparar a influência de cada método em si, sem considerar necessariamente a qualidade dos resultados ou a aplicabilidade prática desta solução.

Para a realização das medições, foi necessário ajustar o protótipo de Moreira (2013), o que foi feito como segue. Primeiramente, determinou-se novos comprimentos de onda dos LEDs para permitir a quantificação do etanol. Para isto, após uma revisão de literatura (Harris et al., 1993; Chagas, 2006) e com base no espectro de absorbâncias do etanol, apresentado no Capítulo 1, foram selecionados três LEDs com máximos de emissão nos seguintes comprimentos de onda: $\lambda_1 = 970$ nm, $\lambda_2 = 1.450$ nm, and $\lambda_3 = 1.530$ nm. Os últimos dois estão compreendidos em uma região espectral em que há absorção pelo etanol. Por outro lado, o comprimento de onda de 970 nm corresponde a uma região espectral onde não há informação química: a resposta do sistema neste comprimento de onda é utilizada para pré-processamento espectral; também, a informação óptica neste comprimento de onda é utilizada como referência para o cálculo de absorbâncias pelo método D (isto é, $\lambda_{ref} = \lambda_1$).

Após alguns testes preliminares utilizando gasolinas tipo C comerciais e gasolinas 40% de etanol, as resistências do circuito da Figura 2.4 foram ajustadas. Os ajustes foram realizados de forma a se obter níveis de tensão adequados na saída do circuito, para que não ocorresse saturação nas medições de referência sem cubetas, em que as intensidades de luz são maiores; e, de outro lado, para que fosse possível observar uma variação significativa para diferentes concentrações dentro da faixa de medição.

3.1.1 Preparação das amostras e medições

Para permitir uma comparação adequada entre os sete métodos de cálculo de absorvâncias, era necessário que todos eles fossem aplicados às mesmas amostras; analisadas com as mesmas cubetas; e sob mesmas condições de medição. Assim, adotou-se um procedimento experimental com o objetivo de se obter uma única medição para cada amostra, a qual poderia ser utilizada em todos os métodos avaliados. Este procedimento é descrito a seguir.

Primeiro, foram preparadas amostras pela adição de diferentes quantidade de etanol absoluto (99,5%) a gasolinas tipo C disponíveis comercialmente. Foram utilizadas gasolinas adquiridas em um único posto de combustíveis da região de Londrina, Paraná, em quatro dias diferentes, em intervalos de três a quatro dias entre as aquisições. As concentrações iniciais de etanol nas gasolinas comerciais foram medidas por meio do método de referência ABNT NBR 13992. As medições de referência nos quatro dias resultaram, em ordem cronológica, nos valores 25%, 24%, 28% e 28%. Então, o experimento foi projetado para a faixa de concentrações de etanol entre 24 e 40% do volume, que engloba possíveis valores reais de adulteração de gasolina por adição de etanol. Assim, para cada um dos quatro dias, as amostras iniciais foram diluídas de forma a cobrir toda a faixa de medição desejada, e as concentrações de referência obtidas após as diluições foram calculadas de acordo com a quantidade de etanol adicionada. Dessa forma, no total, foram obtidas 119 amostras com concentrações de referência entre 24,198% e 40,578% v/v, distribuídas em quatro conjuntos de dados que diferem entre si exclusivamente pela matriz.

O procedimento de medição foi realizado como segue. No início do ciclo de análises de cada conjunto de amostras, as tensões de saída do circuito foram medidas para uma cubeta vazia no caminho óptico, e também sem cubeta. Após a transformação das tensões em fotocorrentes, estas correspondem aos valores de referência para os métodos B e C, respectivamente. Então, para cada amostra individual, o seguinte procedimento foi adotado:

- uma medição foi realizada com uma cubeta vazia e anotada como referência para o método A;
- a mesma cubeta vazia foi retirada e inserida novamente no compartimento de medição, e uma nova medição foi realizada (que corresponde à referência para o método G);
- finalmente, a amostra foi inserida cuidadosamente na cubeta sem que ela fosse removida ou movimentada de sua posição no caminho óptico, e uma terceira medição foi anotada (esta é a medição da amostra em si, a qual é comum para todos os sete métodos).

Os valores de referência para o método D foram obtidos a partir das tensões para $\lambda_1 = 970$ nm nas medições com a amostra. Por sua vez, para o método E, devido à indisponibilidade de gasolina tipo A (sem etanol), foi utilizada como referência os valores medidos para as gasolinas comerciais sem adição de etanol, ou seja, as amostras com menores concentrações de etanol. Por fim, o método F é independente de uma medição de referência, como descrito anteriormente.

Após a realização das medições, a partir das tensões medidas e transformadas em fotocorrentes por (2.11), os diferentes métodos de cálculo de absorbâncias apresentados na Seção 2.3 foram aplicados, os quais originaram os pontos espectrais de absorbâncias.

Neste estudo, optou-se por determinar os modelos de calibração e realizar uma validação independente, a fim de se estimar a performance futura destes modelos. Então, para as calibrações, foram utilizadas 104 amostras, as quais foram selecionadas aleatoriamente a partir de todos os quatro conjuntos de dados. Desse conjunto inicial, 17 amostras foram descartadas durante o processo de calibração por caracterizarem-se como *outliers* ou corresponderem a erros grosseiros de medição. Estatisticamente, a remoção de *outliers* do modelo é válida desde que não exceda 20% do número total de amostras selecionadas para o desenvolvimento do modelo. Posteriormente, para o conjunto de validação externa, foram utilizadas outras 15 amostras independentes. Este número se enquadra no requerimento mínimo para desenvolvimento de um modelo quimiométrico, segundo o qual o número de amostras de validação deve ser de pelo menos 10% do conjunto de calibração. Todas as amostras medidas pelo protótipo também foram analisadas, ao mesmo tempo, em um espectrofotômetro de bancada.

3.1.2 Pré-processamento de dados e calibração

Para avaliar o efeito do pré-processamento espectral, foram avaliados modelos de regressão empregando os pontos espectrais antes e após o pré-processamento de dados, e os resultados obtidos foram comparados. Seguindo o procedimento utilizado por Lutz et al. (2014), utilizou-se uma correção por deslocamento da linha de base, que consiste em aplicar um deslocamento vertical em cada espectro a fim de igualar as linhas de base. Outros métodos de pré-processamento não foram avaliados pois informações de apenas três comprimentos de onda estão disponíveis. Após o pré-processamento, foram obtidas as absorbâncias corrigidas, as quais foram então aplicadas a um modelo MLR de duas variáveis (1.450 nm e 1.530 nm).

Para todos os métodos de cálculo de absorbâncias estudados, foram utilizadas como variáveis independentes as absorbâncias com e sem pré-processamento de dados, e então diferentes coeficientes de regressão foram determinados em cada caso. Na sequência, as concentrações de etanol estimadas pelas sete diferentes curvas de calibração foram determinadas e utilizadas para calcular as estatísticas RMSEC (*Root Mean Square Errors of Calibration*), RMSEP (*Root Mean Square Errors of Prediction*), dadas por

$$\text{RMSEC} = \sqrt{\frac{1}{N-3} \sum_{i=1}^N (y_i - \hat{y}_i)^2}, \quad (3.1)$$

$$\text{RMSEP} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}, \quad (3.2)$$

$$(3.3)$$

onde y_i e \hat{y}_i denotam os valores especificados e preditos pelo modelo avaliado, respectivamente,

e N é o número de amostras do conjunto avaliado. Também, foram calculados os coeficientes de determinação (R^2) e as estatísticas RPD (*residual prediction deviations*), dada por

$$\text{RPD} = \frac{\sigma_{ref}}{\text{RMSEP}}, \quad (3.4)$$

onde σ_{ref} denota o desvio padrão das concentrações de referência do conjunto de teste. Toda a análise de dados foi realizada por meio do *software* Matlab[®] 2016.

Os resultados deste experimento mostraram que o Método E resultou em modelo com melhor performance preditiva na quantificação de etanol em gasolina. Contudo, avaliou-se que este método seria inviável na prática, para a aplicação proposta, porque as gasolinas de referência necessárias para a medição de $I_0(\lambda)$ não estão disponíveis. Por sua vez, verificou-se que a diferença entre os demais métodos não foi tão significativa em termos práticos, para esta aplicação. Também, avaliou-se que resultados semelhantes provavelmente seriam obtidos na predição de metanol. Por isso, para permitir uma comparação mais adequada do protótipo desenvolvido com o espectrofotômetro de bancada e considerando suas vantagens práticas, o Método B foi escolhido para os experimentos subsequentes. Os resultados detalhados deste experimento são apresentados no Capítulo 4, na Seção 4.1.

3.2 AVALIAÇÃO DA INFLUÊNCIA DA VARIABILIDADE DE MATRIZ

Para avaliar a influência desses dois fatores na quantificação de álcool em gasolina e guiar a preparação e a seleção de amostras nos experimentos subsequentes, os mesmos dados utilizados no experimento anterior foram analisados para diferentes formas de seleção de amostras. O objetivo desta análise foi avaliar se existe variação de matriz significativa para amostras de um único posto de combustíveis, a qual influencie na capacidade preditiva do modelo de duas variáveis.

As amostras utilizadas no experimento foram preparadas a partir de quatro gasolinas tipo C distintas adquiridas em um único posto de combustíveis. Neste caso, foram obtidos quatro conjuntos de dados que diferem entre si exclusivamente pela matriz. Assim, a variabilidade esperada se refere principalmente às diferenças na composição química de diferentes lotes de gasolinas daquele fornecedor. Nesse sentido, este experimento não analisou as diferentes fontes possíveis de variabilidade. Ao invés disso, limitou-se o espaço amostral para permitir a interpretação dos resultados em termos de uma única fonte de variabilidade pré-estabelecida. Assim, neste experimento, os dados referentes a cada conjunto de amostras foram agrupados e considerados individualmente no processo de calibração, e os erros de predição obtidos foram comparados. Isto foi feito também considerando diferentes métodos de cálculo de absorvâncias, para avaliar esta influência em cada método. Neste caso, foram analisados os resultados para os métodos que apresentaram maior e menor valores de RMSEC entre os sete métodos estudados (Métodos E e B, respectivamente).

Mais especificamente, neste experimento, considerou-se dois procedimentos

de seleção dos dados de calibração:

- Procedimento de Seleção 1: foram selecionadas dados de 33 amostras de calibração exclusivamente do primeiro conjunto (aquelas obtidas a partir da diluição da gasolina coletada no primeiro dia), enquanto outras 42 amostras de teste foram utilizadas para validação, das quais 29 correspondem a amostras do segundo conjunto, e 13 são do terceiro conjunto;
- Procedimento de Seleção 2: os dados de calibração foram selecionadas aleatoriamente a partir dos quatro conjuntos de amostras. Este procedimento de seleção foi o mesmo empregado no experimento descrito na Seção 3.1. Neste caso, a calibração foi realizada a partir de 104 amostras para a calibração (das quais 17 foram posteriormente descartadas); e 15 outras amostras de teste foram utilizadas para validação.

As curvas de calibração foram determinadas considerando estes dois procedimentos. Posteriormente, comparou-se as concentrações estimadas por cada modelo graficamente e em termos dos parâmetros RMSEC e RMSEP. No Procedimento de Seleção 1, optou-se por utilizar amostras de teste apenas de dois outros conjuntos para facilitar a distinção entre eles e a interpretação dos resultados. Por fim, os resultados também foram analisados para diferentes métodos de cálculo de absorvâncias.

Por sua vez, além da influência da matriz, os erros de estimação obtidos em cada caso também estão relacionados com a exatidão do método de referência ABNT NBR 13992, o qual foi utilizado para especificar as concentrações de etanol iniciais de cada gasolina antes das diluições. Assim, para avaliar os modelos de calibração eliminando a influência do método de referência, as mesmas amostras também foram analisadas em um espectrofotômetro de bancada (FT-IF/NIR Frontier, da *PerkinElmer*). Dessa forma, comparou-se os erros de predição obtidos pelo protótipo com aqueles obtidos pelo espectrofotômetro NIR de bancada. A performance dos modelos de calibração foram então avaliadas em termos da diferença entre os parâmetros RMSEC e RMSEP obtidos em ambos os instrumentos, tendo em vista que a influência do método primário, neste caso, é a mesma para ambos. Nesta comparação, apenas o segundo procedimento de seleção de amostras foi considerado.

Para o espectrofotômetro de bancada, a calibração foi realizada considerando três regiões espectrais com larguras de 100 nm em torno dos comprimentos de onda centrais dos LEDs empregados no protótipo. Esta largura foi escolhida para que fosse similar às bandas de emissão dos LEDs. Assim, a calibração do espectrofotômetro foi realizada apenas com o conteúdo espectral também disponível no protótipo. A diferença, contudo, é que os dados espectrais foram medidos pelo espectrofotômetro com resolução de 5 nm. Consequentemente, a calibração daquele instrumento empregou cerca de 60 variáveis, diferentemente dos três variáveis disponíveis no protótipo. As curvas espectrais obtidas por ambos os instrumentos também são apresentadas e comparadas. Por fim, no espectrofotômetro, a calibração foi realizada por PLS com seis variáveis latentes. Em ambos os instrumentos, as amostras foram analisadas em cubetas de acrílico com comprimento óptico de 10 mm.

Os resultados deste experimento revelaram que existe variabilidade de matriz muito significativa mesmo para gasolinas adquiridas de um mesmo posto revendedor, provenientes de um único distribuidor. Além disso, os resultados mostraram que a influência de matriz na qualidade do modelo é mais significativa do que a exatidão nas medições dos dados ópticos em si pelo sistema de instrumentação. Isto evidenciou a necessidade de empregar, nas próximas etapas do trabalho, uma quantidade de amostras suficiente para cobrir uma variabilidade amostral mais representativa de uma situação prática de utilização do instrumento. Por fim, estes resultados também guiaram a metodologia de validação do novo protótipo, a fim de proporcionar uma estimativa da real capacidade preditiva futura dos modelos. Os resultados deste experimento são detalhados no Capítulo 4, na Seção 4.2.

3.3 ESTUDO ESPECTRAL DAS MISTURAS TERNÁRIAS PARA SELEÇÃO DE VARIÁVEIS

Para fundamentar o projeto do sistema óptico do protótipo, realizou-se um estudo espectral em amostras ternárias de gasolina comercial/etanol/metanol, com o objetivo de identificar um conjunto de comprimentos de onda discretos que possibilitem a quantificação simultânea dos dois álcoois. Este estudo avaliou as performances de predição de diferentes modelos de calibração candidatos, construídos a partir de diferentes números de variáveis espectrais. O método utilizado para selecionar os comprimentos de onda mais informativos e para avaliar os modelos de calibração candidatos são apresentados nas subseções a seguir.

3.3.1 Determinação dos comprimentos de onda analiticamente importantes

A primeira parte do estudo espectral consistiu na investigação dos comprimentos de onda mais informativos para a determinação simultânea dos dois álcoois em amostras ternárias. Nesta etapa, um espectrofotômetro de bancada foi utilizado para medir os espectros das amostras com alta resolução espectral. Os dados coletados foram então analisados por meio de um algoritmo de seleção de variáveis. O algoritmo implementado identifica as variáveis espectrais que geram modelos com melhores performances de predição, para um dado um número de variáveis desejado. Assim, nas subseções 3.3.1.1 a 3.3.1.4 a seguir, são apresentados os procedimentos experimentais utilizados para a preparação das amostras, aquisição dos espectros e seleção de variáveis para compor os modelos de calibração candidatos.

3.3.1.1 Preparação das amostras

Foram adquiridas gasolinas tipo C de nove postos revendedores de combustíveis da cidade de Londrina, Paraná, sendo seis delas não-aditivadas e outras três aditivadas, provenientes de quatro distribuidores diferentes. A concentração inicial de álcool em cada gasolina foi determinada pelo método de referência ABNT NBR 13992. Cinco delas apresentaram teor de álcool igual a 27 % v/v, valor correspondente à concentração esperada de etanol, conforme a legislação vigente. Já os resultados dos testes para as outras quatro gasolinas apontaram

concentrações iniciais de álcool iguais a 25% v/v. Conforme descrito anteriormente na Seção 2.5, o método de referência ABNT NBR 13992 não permite distinguir o metanol do etanol. Contudo, é muito pouco provável que qualquer das gasolinas utilizadas nesta etapa do experimento tenha sido adulterada com metanol, pois as concentrações medidas não ultrapassaram os percentuais de etanol adicionados obrigatoriamente pelos distribuidores. Dessa forma, considerando a indisponibilidade de outro método de referência com menor incerteza, foi assumido neste trabalho que os percentuais medidos se referem especificamente ao etanol. Dessa forma, as concentrações iniciais de metanol foram estabelecidas como zero.

Em seguida, as nove gasolinas distintas foram diluídas a partir da adição de diferentes volumes de metanol (com grau de pureza 99,95%) e etanol (com grau de pureza 99,5%). As diluições foram realizadas com o auxílio de duas micropipetas (Kasvi, modelos K1-1000PB e K1-5000PB). Foram obtidas 88 misturas binárias (sendo 43 gasolina comercial/etanol e 45 gasolina comercial/metanol) e 79 misturas ternárias (gasolina comercial/metanol/etanol). Assim, as concentrações de referência foram calculadas a partir dos volumes dos álcoois adicionados. As concentrações finais obtidas estão na faixa de 21,9 % a 44,9 % v/v, para o etanol, e de 0,0 % a 12,3 % v/v, para o metanol. No total, as gasolinas iniciais e suas diluições resultaram em 176 amostras, as quais foram divididas em nove grupos. Cada grupo consiste em uma gasolina inicial e suas diluições, de forma que não existe variação de matriz entre amostras pertencente a um mesmo grupo. Neste caso, elas se diferenciam entre si apenas pelas concentrações dos analitos em interesse.

3.3.1.2 Aquisição dos espectros das amostras

Os espectros das 176 amostras foram medidos por meio de um espectrofotômetro de bancada (FT-IF/NIR Frontier, da PerkinElmer), na faixa espectral $11.111,1 \text{ cm}^{-1}$ a $6.060,6 \text{ cm}^{-1}$ (900 a 1.650 nm), com resolução de 8 cm^{-1} , empregando 32 acumulações por amostra. As amostras foram analisadas em cubetas descartáveis de acrílico com comprimento óptico de 10 mm. Estas configurações resultaram em espectros com 2.194 variáveis cada.

3.3.1.3 Pré-processamento dos dados espectrais

Inicialmente, os dados espectrais de cada conjunto de amostras foram analisados separadamente, com o objetivo de se identificar possíveis *outliers* (decorrentes, por exemplo, de erros no manuseio das amostras durante as diluições ou na medição dos espectros). Foram construídos modelos de calibração para as duas variáveis de interesse (concentrações de etanol e metanol), por meio do método de regressão por mínimos quadrados parciais (PLS). Então, os espectros completos de cada grupo de amostras foram utilizados para a determinação dos parâmetros dos modelos. Isto foi feito porque, como não existe variação de matriz entre amostras de um mesmo conjunto, maior é a correlação entre as absorvâncias e as concentrações em interesse, de forma que os *outliers* se tornam mais evidentes. Como resultado, foram identificados no total 6 *outliers*. As amostras correspondentes foram então eliminadas, resultando em

170 amostras úteis para a sequência do experimento.

Em seguida, os espectros foram submetidos a um processo de redução de variáveis, baseado no estudo proposto por Paiva et al. (2019). Isto foi feito com o objetivo de limitar as variáveis espectrais candidatas àquelas que pudessem efetivamente ser empregadas na construção do protótipo. Neste trabalho, avalia-se a construção de um protótipo cujo sistema óptico emprega LEDs como fonte de luz e filtros de interferência para a separação dos comprimentos de onda. Neste caso, identificou-se os comprimentos de onda dos filtros disponíveis comercialmente. Verificou-se que existiam comercialmente filtros passa-banda (da Knight Optical Ltd.) centrados em 37 comprimentos de onda distintos na faixa estudada. A Tabela 3.1 apresenta os comprimentos de onda nominais dos filtros disponíveis comercialmente, juntamente com os valores de largura de banda à meia altura (FWHM - *Full Width at Half Maximum*) especificados pelo fabricante. Assim, 2.194 variáveis espectrais originais foram reduzidas para 37 variáveis candidatas a comporem o modelo matemático de medição final. A intensidade da absorbância em cada ponto do espectro reduzido foi obtida pela média aritmética das absorbâncias do espectro original na faixa espectral próxima ao comprimento de onda central de cada filtro de interferência, limitada pela largura de banda FWHM do espectro de transmitância do respectivo filtro.

Após a redução de variáveis, os espectros resultantes passaram por um pré-processamento igual ao utilizado por Lutz et al. (2014). Foi empregada uma correção por deslocamento da linha de base, que consiste em aplicar um deslocamento vertical em cada espectro a fim de igualar as absorbâncias em um comprimento de onda de referência. Após uma investigação utilizando os dados das 170 amostras e o método de regressão por PLS, definiu-se o comprimento de onda 1.040 nm como referência para o pré-processamento espectral, uma vez que este resultou em modelos com menores erros de predição. Dessa forma, o comprimento de onda 1.040 nm corresponde à primeira variável selecionada para a construção do protótipo, a qual é utilizada como base para o pré-processamento espectral. Os dados espectrais pré-processados das demais variáveis compõem uma matriz de observações X de dimensão 170×36 . Estes dados foram empregados no procedimento de seleção de variáveis descrito na próxima seção.

3.3.1.4 Procedimento de seleção das variáveis espectrais

A próxima etapa consistiu em investigar os comprimentos de onda mais informativos para a quantificação simultânea dos álcoois. Também, foi necessário avaliar o número de comprimentos de onda necessários para produzir resultados de predição aceitáveis em termos práticos. Os procedimentos adotados e os critérios de avaliação utilizados são apresentados a seguir.

De forma genérica, seja $X = (x_1, x_2, \dots, x_k)$ a matriz de dimensão $n \times k$ que contém os dados espectrais (matriz de observações) para k variáveis (comprimentos de onda) candidatas, onde x_j , $j = 1, \dots, k$, denota a j -ésima coluna de X , e n é o número total de amostras. Por sua vez, sejam y_1 e y_2 duas matrizes de dimensão $n \times 1$ que correspondem

Tabela 3.1: Comprimentos de onda centrais e larguras de banda à meia altura (FWHM) dos 37 filtros de interferência ópticos disponíveis comercialmente.

Filtro	1	2	3	4	5	6	7	8	9	10
Comp. onda central (nm)	900	920	930	940	950	960	970	980	990	1000
FWHM (nm)	10	10	10	10	10	10	10	10	10	15
Filtro	11	12	13	14	15	16	17	18	19	20
Comp. onda central (nm)	1020	1040	1050	1064	1080	1140	1180	1200	1220	1240
FWHM (nm)	15	15	15	16	15	17,5	11	11	11	11,5
Filtro	21	22	23	24	25	26	27	28	29	30
Comp. onda central (nm)	1260	1280	1300	1320	1360	1400	1420	1440	1460	1480
FWHM (nm)	12	12	12	12,5	13	14	14	14	14	15
Filtro	31	32	33	34	35	36	37			
comp. onda central (nm)	1500	1520	1540	1580	1600	1620	1640			
FWHM (nm)	15	15	15	16	13	13	16			

às variáveis de respostas das duas propriedades de interesse (neste estudo, as concentrações de etanol e de metanol).

Deseja-se selecionar um conjunto $S^* = \{s_1^*, s_2^*, \dots, s_z^*\} \subseteq \{1, \dots, k\}$ correspondente aos índices de z variáveis (colunas de X), $z \leq k$, e utilizá-las como preditores para estimar, simultaneamente, as duas variáveis de resposta em interesse, por meio de dois modelos de regressão linear múltipla (MLR) da forma

$$\mathbf{y}_1 = \mathbf{x}_0\beta_{01} + \sum_{j \in S^*} \mathbf{x}_j\beta_{j1} + \boldsymbol{\epsilon}_1, \quad (3.5)$$

$$\mathbf{y}_2 = \mathbf{x}_0\beta_{02} + \sum_{j \in S^*} \mathbf{x}_j\beta_{j2} + \boldsymbol{\epsilon}_2, \quad (3.6)$$

onde: \mathbf{x}_0 é um vetor $n \times 1$ com todos os elementos iguais a 1; $\{\beta_{j1}, \forall j \in S^*\}$ e $\{\beta_{j2}, \forall j \in S^*\}$, são os coeficientes de regressão de cada modelo; e $\boldsymbol{\epsilon}_1$ e $\boldsymbol{\epsilon}_2$ denotam os vetores de erros.

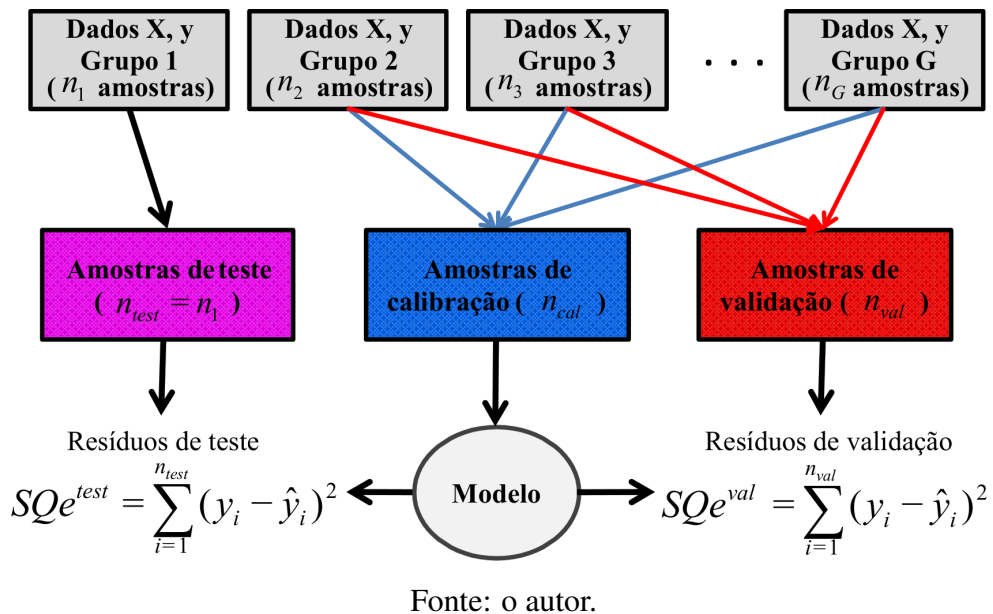
O problema de seleção de variáveis consiste em determinar quais são os comprimentos de onda mais informativos sobre a variável de resposta, isto é, aquelas que produzem modelos com melhor performance de predição por meio de (3.5) e (3.6). Naturalmente, a quantidade de variáveis espectrais selecionadas gera modelos com performances de predição distintas. É esperado que os erros de predição sejam menores a medida que mais variáveis sejam incluídas no modelo. Por outro lado, quanto menor o número de comprimentos de necessários, mais simples e mais barata se torna a construção do instrumento. Por isso, este estudo avaliou possíveis soluções a partir de diferentes números de variáveis selecionadas, comparando as performances de predição dos dois analitos em cada caso.

De forma geral, as variáveis mais informativas são aquelas que geram um modelo com maior sensibilidade às concentrações dos analitos em interesse. Contudo, conforme já discutido, um dos principais fatores que devem ser considerados na seleção de variáveis é a variabilidade natural na composição da substância a ser analisada. O conjunto ótimo de variáveis, que geram menores erros de predição, provavelmente são distintos quando se considera amostras de diferentes matrizes. Por isso, é importante considerar amostras com a maior variabilidade de matriz possível, de forma a se obter uma boa amostragem da variabilidade real esperada no uso de rotina do instrumento. Apesar disso, na prática, um número limitado de grupos estará disponível para a calibração de qualquer instrumento. Por isso, faz-se necessário estimar as performances de predição dos modelos para amostras distintas daquelas empregadas no momento da calibração. Isto é uma medida da capacidade de predição futura do modelo.

Diante do exposto, as amostras foram separadas em grupos, conforme descrito na seção 3.3.1.1. Seja G o número de grupos de amostras distintos, e sejam n_1, n_2, \dots, n_G as quantidades de amostras pertencentes a cada grupo. O procedimento de seleção de variáveis proposto emprega um método de validação cruzada, baseado no método denominado *leave-one-out*. Trata-se de um processo iterativo em que, para cada iteração, um dos grupos é separado como grupo de teste, enquanto as amostras dos demais $G - 1$ grupos são divididas entre calibração e validação. A primeira iteração do método proposto está representada na Figura 3.1. Este método de validação fornece uma medida mais realista da capacidade de predição futura de um determinado modelo, pois duas formas de avaliação são realizadas. A primeira delas utiliza amostras do grupo de validação, cujas características espectrais estão representadas no grupo de calibração. Já a segunda emprega amostras do grupo de teste, cujas matrizes são totalmente distintas daquelas representadas na calibração, o que ocorre em uma situação real de utilização do instrumento.

O *algoritmo de seleção de variáveis* desenvolvido e implementado neste trabalho é definido a seguir.

Figura 3.1: Diagrama representativo da primeira iteração do método de validação cruzada utilizado para avaliar cada modelo matemático de medição.



Algoritmo de seleção de variáveis

Passo 0: dados de entrada: dados espectrais, \mathbf{X} ($n \times k$); dados das variáveis de resposta, \mathbf{y}_1 ($n \times 1$) e \mathbf{y}_2 ($n \times 1$); número de grupos de amostras, G ; número de amostras pertencentes a cada grupo de amostras, $\{n_1, n_2, \dots, n_G\}$; e número de variáveis espectrais a serem selecionadas, z . Seja $g = 1$.

Passo 1: seja $n_{test} = n_g$. Selecione as linhas de \mathbf{X} , de \mathbf{y}_1 e de \mathbf{y}_2 correspondentes às n_g amostras pertencentes ao grupo g como conjunto de teste, e armazene em \mathbf{X}^{test} ($n_{test} \times k$), \mathbf{y}_1^{test} ($n_{test} \times 1$) e \mathbf{y}_2^{test} ($n_{test} \times 1$), respectivamente.

Passo 2: armazene os dados das demais amostras não pertencentes ao conjunto de teste em \mathbf{X}^R ($n_R \times k$), \mathbf{y}_1^R ($n_R \times 1$) e \mathbf{y}_2^R ($n_R \times 1$), onde $n_R = n - n_{test}$.

Passo 3: execute o algoritmo SPXY para dividir as amostras não pertencentes ao conjunto de teste entre n_{cal} amostras para o conjunto de calibração e n_{val} amostras para o conjunto de validação, com $n_{cal} \geq z + 1$, isto é, obtenha as matrizes \mathbf{X}^{cal} ($n_{cal} \times k$), \mathbf{y}_1^{cal} ($n_{cal} \times 1$) e \mathbf{y}_2^{cal} ($n_{cal} \times 1$); e \mathbf{X}^{val} ($n_{val} \times k$), \mathbf{y}_1^{val} ($n_{val} \times 1$) e \mathbf{y}_2^{val} ($n_{val} \times 1$).

Passo 4: execute o algoritmo SPA-MLR para obter o subconjunto $S = \{s_1, s_2, \dots, s_z\} \subseteq \{1, \dots, k\}$ dos índices das z variáveis mais informativas considerando os dados atuais. Em seguida, seja $S_g = S$.

Passo 5: seja $g = g + 1$. Se $g \leq G$, retorne ao Passo 1;

Passo 6: seja $S^* = \{s_1^*, s_2^*, \dots, s_z^*\} \subseteq \{1, \dots, k\}$ os z índices pertencentes aos subconjuntos S_1, S_2, \dots, S_G que ocorram com maior frequência.

Fim: os índices das variáveis espectrais resultantes são $S^* = \{s_1^*, s_2^*, \dots, s_z^*\}$.

Analisando-se o *algoritmo de seleção de variáveis*, pode-se observar que as n amostras disponíveis são divididas em três conjuntos mutuamente exclusivos: um conjunto de teste; um conjunto de calibração, com número de amostras maior ou igual a $z + 1$, o qual é empregado na seleção de variáveis e na estimação dos coeficientes de regressão do modelo MLR resultante; e um grupo de validação. É sabido que a separação entre amostras de calibração e validação é um aspecto muito importante no desenvolvimento de um modelo quantitativo, a qual deve ser feita por um método apropriado. Diferentes métodos são comumente utilizados para este fim, entre os quais pode-se citar o método de seleção aleatória (RS - *Random Sampling*), o método Kennard-Stone (KS), e o algoritmo de partição do conjunto de amostras baseado na distâncias conjuntas x-y (SPXY - *Sample set portioning based on joint X-Y distances*). Neste trabalho, o algoritmo SPXY foi utilizado.

O algoritmo SPXY avalia as amostras disponíveis em termos das distâncias euclidianas dos vetores formados no espaço das variáveis espectrais (X) e no espaço das composições em interesse (Y). Este método se baseia no conceito de que amostras próximas entre si são espectralmente semelhantes, enquanto amostras distantes entre si possuem maiores diferenças espectrais. Então, o algoritmo SPXY visa selecionar um conjunto de amostras de calibração seguindo uma distribuição uniforme em ambos os espaços. Isto é feito como segue.

Dado um par de amostras (p, q) , define-se por $d_x(p, q)$ e $d_y(p, q)$ as distâncias euclidianas no espaço espectral e no espaço das composições em interesse, respectivamente. Então, para atribuir a mesma importância para as distribuições das amostras em ambos os espaços, as distâncias euclidianas $d_x(p, q)$ e $d_y(p, q)$ são divididas pelas respectivas máximas distâncias em todo o conjunto de amostras disponível, e uma distância normalizada $d_{xy}(p, q)$ é calculada. O algoritmo inicia selecionando um par de amostras (p, q) para o qual $d_{xy}(p, q)$ é máximo. Em seguida, o algoritmo executa um processo iterativo a fim de selecionar novas amostras até que o número n_{cal} desejado seja alcançado. Neste trabalho, adotou-se n_{cal} como 70% do número de amostras disponíveis, enquanto 30% são utilizadas para validação. Este algoritmo é detalhado a seguir.

Algoritmo SPXY

Passo 0: dados de entrada: matriz de dados espectrais, \mathbf{X}^R ($n_R \times k$); dados das variáveis de resposta, \mathbf{y}_1^R ($n_R \times 1$) e \mathbf{y}_2^R ($n_R \times 1$); e número de amostras a serem selecionadas para a calibração, n_{cal} .

Passo 1: para todo par de amostras (p, q) , determine as distâncias euclidianas $d_x(p, q)$ e $d_y(p, q)$ entre os vetores linhas formados no espaço das variáveis espectrais \mathbf{X}^R e no espaço das variáveis de resposta \mathbf{y}_1^R e \mathbf{y}_2^R , respectivamente, dadas por

$$d_x(p, q) = \sqrt{\sum_{j=1}^k (x_{pj}^R - x_{qj}^R)^2}, p, q \in [1, n_R], \quad (3.7)$$

$$d_y(p, q) = \sqrt{(y_{1p}^R - y_{1q}^R)^2 + (y_{2p}^R - y_{2q}^R)^2}, p, q \in [1, n_R], \quad (3.8)$$

onde x_{pj}^R e x_{qj}^R correspondem aos dados da j -ésima variável espectral para as amostras p e q , respectivamente; y_{1p}^R e y_{1q}^R correspondem aos valores das linhas p e q de y_1^R , respectivamente; e y_{2p}^R e y_{2q}^R correspondem aos valores das linhas p e q de y_2^R , respectivamente.

Passo 2: para todo par de amostras (p, q) , determine a distância euclidiana normalizada $d_{xy}(p, q)$, dada por

$$d_{xy}(p, q) = \frac{d_x(p, q)}{\max_{p, q \in [1, n]} d_x(p, q)} + \frac{d_y(p, q)}{\max_{p, q \in [1, n]} d_y(p, q)}. \quad (3.9)$$

Passo 3: selecione, para o grupo de calibração, um par de amostras (p, q) para o qual $d_{xy}(p, q)$ é máximo; seja $V = \{p, q\}$ o conjunto de índices de amostras já selecionadas para a calibração; e seja $cont = 2$, o número de amostras já selecionadas.

Passo 4: para cada uma das $(n_R - cont)$ amostras que ainda não foram selecionadas, determine a mínima distância d_{xy} entre ela e todas as amostras já selecionadas. Em seguida, selecione, para o grupo de calibração, a amostra para a qual a mínima distância seja máxima; e inclua o índice da amostra selecionada no conjunto V .

Passo 5: faça $cont = cont + 1$. Se $cont < n_{cal}$, retorne ao Passo 4.

Passo 6: Selecione as linhas de \mathbf{X}^R , \mathbf{y}_1^R e \mathbf{y}_2^R referentes às amostras selecionadas para o grupo de calibração, isto é, seja $\mathbf{X}^{cal} = \{\mathbf{x}_i^R, \forall i \in V\}$; seja $\mathbf{y}_1^{cal} = \{y_{1i}^R, \forall i \in V\}$; e seja $\mathbf{y}_2^{cal} = \{y_{2i}^R, \forall i \in V\}$, onde \mathbf{x}_i^R , y_{1i}^R e y_{2i}^R denotam as i -ésimas linhas de \mathbf{X}^R , \mathbf{y}_1^R e \mathbf{y}_2^R , respectivamente.

Passo 7: Selecione, para o grupo de validação, as linhas de \mathbf{X}^R , \mathbf{y}_1^R e \mathbf{y}_2^R referentes às amostras que não foram selecionadas para o grupo de calibração, isto é, seja $\mathbf{X}^{val} = \{\mathbf{x}_i^R, \forall i \notin V\}$; seja $\mathbf{y}_1^{val} = \{y_{1i}^R, \forall i \notin V\}$; e seja $\mathbf{y}_2^{val} = \{y_{2i}^R, \forall i \notin V\}$, onde \mathbf{x}_i^R , y_{1i}^R e y_{2i}^R denotam as i -ésimas linhas de \mathbf{X}^R , \mathbf{y}_1^R e \mathbf{y}_2^R , respectivamente.

Fin: as matrizes contendo os dados de calibração são \mathbf{X}^{cal} , \mathbf{y}_1^{cal} , \mathbf{y}_2^{cal} ; e as matrizes contendo os dados de validação são \mathbf{X}^{val} , \mathbf{y}_1^{val} e \mathbf{y}_2^{val} .

Uma vez que os conjuntos de calibração, validação e teste tenham sido definidos, o *algoritmo de seleção de variáveis* executa a subrotina denotada por *algoritmo SPA-MLR*. Esta subrotina é responsável por selecionar as variáveis mais informativas para compor os modelos MLR, e avaliar os modelos a partir das amostras de validação e teste.

Neste trabalho, o método empregado para selecionar as variáveis é o Algoritmo das Projeções Sucessivas (SPA - *Successive Projections Algorithm*). Trata-se de uma técnica de seleção que minimiza os efeitos da colinearidade entre variáveis (Galvão et al., 2007). No algoritmo implementado neste trabalho, subconjuntos candidatos de variáveis são construídos de acordo com uma sequência de operações de projeções envolvendo as colunas da matriz \mathbf{X}^{cal} . Estes subconjuntos candidatos são avaliados de acordo com a performance dos modelos MLR resultantes na predição das amostras de validação. O *algoritmo SPA-MLR* é descrito a seguir.

Algoritmo SPA-MLR

Passo 0: dados de entrada: dados espectrais de calibração e validação, \mathbf{X}^{cal} ($n_{cal} \times k$) e \mathbf{X}^{val} ($n_{val} \times k$); dados das variáveis de resposta, \mathbf{y}_1^{cal} ($n_{cal} \times 1$), \mathbf{y}_1^{val} ($n_{val} \times 1$), \mathbf{y}_2^{cal} ($n_{cal} \times 1$) e \mathbf{y}_2^{val} ($n_{val} \times 1$); número de variáveis a serem selecionadas, z . Seja $iter = 1$ (primeira de k iterações).

Passo 1: selecione a variável $iter$ como a primeira variável para o modelo, isto é, faça $\lambda_1 = iter$; e seja $i = 1$.

Passo 2: seja \mathbf{w}_j a j -ésima coluna de \mathbf{X}^{cal} , $j = 1, \dots, k$.

Passo 3: seja C_{iter} o conjunto de índices das variáveis que já foram selecionadas na iteração atual, isto é, $C_{iter} = \{\lambda_1, \dots, \lambda_i\}$.

Passo 4: calcule a projeção de \mathbf{w}_j no subespaço ortogonal à coluna referente à última variável selecionada, \mathbf{w}_{λ_i} , como

$$P\mathbf{w}_j = \mathbf{w}_j - (\mathbf{w}_j^T \mathbf{w}_{\lambda_i}) \mathbf{w}_{\lambda_i} (\mathbf{w}_{\lambda_i}^T \mathbf{w}_{\lambda_i})^{-1} \quad (3.10)$$

para todo $\{j : 1 \leq j \leq k; j \notin C_{iter}\}$, onde P é o operador projeção, e o superescrito T denota a transposta.

Passo 5: selecione a próxima variável, aquela cuja projeção calculada no passo anterior tenha maior módulo; isto é, faça $\lambda_{i+1} = \arg(\max |\mathbf{P}\mathbf{w}_j|; 1 \leq j \leq k; j \notin C_{iter})$;

Passo 6: faça $\mathbf{w}_j = \mathbf{P}\mathbf{w}_j$, para todo $j \notin C_{iter}$; $1 \leq j \leq k$.

Passo 7: faça $i = i + 1$. Se $i < z$, volte ao Passo 3.

Passo 8: as variáveis espectrais selecionadas na iteração atual são $C_{iter} = \{\lambda_1, \dots, \lambda_z\}$. Faça $iter = iter + 1$. Se $iter < k$, volte ao Passo 1.

Passo 9: $\{C_1, C_2, \dots, C_k\}$ são os subconjuntos candidatos, cada um contendo os índices de z variáveis espectrais. Para todo subconjunto candidato C_i ($i = 1, \dots, k$), faça:

Passo 9.1: construa modelos de calibração MLR para estimar as variáveis de resposta, na forma

$$\mathbf{y}_1 = \mathbf{x}_0 \beta_{01} + \sum_{j \in C_i} \mathbf{x}_j \beta_{j1} + \boldsymbol{\epsilon}_1, \quad (3.11)$$

$$\mathbf{y}_2 = \mathbf{x}_0 \beta_{02} + \sum_{j \in C_i} \mathbf{x}_j \beta_{j2} + \boldsymbol{\epsilon}_2, \quad (3.12)$$

onde: \mathbf{x}_0 é um vetor $n \times 1$ com todos os elementos iguais a 1; \mathbf{x}_j denota a coluna referente à j -ésima variável espectral; $\{\beta_{01}, \beta_{11}, \dots, \beta_{z1}\}$ e $\{\beta_{02}, \beta_{12}, \dots, \beta_{z2}\}$ são os coeficientes de regressão de cada modelo; e $\boldsymbol{\epsilon}$ denotam os vetores de erros.

Passo 9.2: utilize os dados do grupo de calibração (\mathbf{X}^{cal} , \mathbf{y}_1^{cal} e \mathbf{y}_2^{cal}) para estimar os coeficientes de regressão dos modelos, e utilize os modelos para estimar as concentrações a partir dos dados dos grupos de validação (\mathbf{X}^{val}).

Passo 9.3: calcule a estatística $RMSEV_i$ para o subconjunto candidato, isto é, faça

$$RMSEV_i = \sqrt{\frac{1}{n_{val}} \sum_{i=1}^{n_{val}} [(y_{1_i}^{val} - \hat{y}_{1_i}^{val})^2 + (y_{2_i}^{val} - \hat{y}_{2_i}^{val})^2]}, \quad (3.13)$$

onde: $y_{1_i}^{val}$ e $y_{2_i}^{val}$ correspondem aos valores especificados para a amostra de índice i do grupo de validação; e $\hat{y}_{1_i}^{val}$ e $\hat{y}_{2_i}^{val}$ correspondem aos valores preditos pelos modelos para a amostra de índice i do grupo de validação.

Passo 10: Selecione o subconjunto candidato que originou a mínima estatística $RMSEV$, isto é, determine $idx = \arg(\min RMSEV_i; i = \{1, \dots, k\})$; e seja $S = C_{idx}$.

Fim: O subconjunto de índices das z variáveis mais informativas para os dados atuais é S .

No *algoritmo SPA-MLR*, a seleção dos comprimentos de onda e a construção dos modelos candidatos são feitas utilizando as amostras de calibração, enquanto as amostras de validação são utilizadas para avaliar a qualidade destes modelos. Para um dado número z de variáveis, é esperado que os comprimentos de onda mais informativos para estimar o etanol sejam diferentes dos comprimentos de onda ótimos para estimar o metanol. Por isso, o problema consiste em selecionar z comprimentos de onda que maximizem a performance de predição de ambos os modelos simultaneamente. Neste caso, para avaliar a performance de predição conjunta dos dois modelos e guiar a seleção de variáveis, foi utilizada a estatística raiz quadrada do erro quadrático médio para o grupo de validação ($RMSEV$), definida em (3.13).

Ao final da execução do *algoritmo de seleção de variáveis*, os índices das z variáveis espectrais finais (S^*) são determinados. Estas correspondem às variáveis selecionadas com maior frequência em todas as iterações do processo de validação cruzada. Este critério foi adotado porque as variáveis que geram melhores performances de predição provavelmente são distintas em cada iteração, quando diferentes subconjuntos de calibração são considerados. Então, dois modelos candidatos contendo z variáveis espectrais são construídos por meio de (3.5) e (3.6). Em seguida, o processo de validação cruzada é repetido, com o objetivo de avaliar as performances de predição destes modelos candidatos. Nesta etapa, tanto as amostras de validação como as de teste são empregadas. Isto é descrito a seguir.

3.3.2 Avaliação dos modelos candidatos

Para avaliar as performances de predição dos modelos candidatos de z variáveis, foi utilizado o algoritmo descrito a seguir. Este algoritmo executa o mesmo processo de validação cruzada utilizado anteriormente na seleção de variáveis, aplicado agora aos modelos já definidos.

Algoritmo de avaliação dos modelos candidatos

Passo 0: dados de entrada: dados espectrais, \mathbf{X} ($n \times k$); dados das variáveis de resposta, \mathbf{y}_1 ($n \times 1$) e \mathbf{y}_2 ($n \times 1$); número de grupos de amostras, G ; número de amostras pertencentes a cada grupo de amostras, $\{n_1, n_2, \dots, n_G\}$; e conjunto de índices das variáveis espectrais selecionadas, $S^* = \{s_1^*, s_2^*, \dots, s_z^*\}$. Seja $g = 1$; $SQe^{val} = 0$ e $SQe^{test} = 0$; $N_V = 0$ e $N_T = 0$.

Passo 1: seja $n_{test} = n_g$. Selecione as linhas de \mathbf{X} , de \mathbf{y}_1 e de \mathbf{y}_2 correspondentes às n_g amostras pertencentes ao grupo g como grupo de teste, e armazene em \mathbf{X}^{test} ($n_{test} \times k$), \mathbf{y}_1^{test} ($n_{test} \times 1$) e \mathbf{y}_2^{test} ($n_{test} \times 1$), respectivamente.

Passo 2: armazene os dados das demais amostras não pertencentes ao grupo de teste em \mathbf{X}^R ($n_R \times k$), \mathbf{y}_1^R ($n_R \times 1$) e \mathbf{y}_2^R ($n_R \times 1$), onde $n_R = n - n_{test}$.

Passo 3: execute o algoritmo SPXY para dividir as amostras não pertencentes ao grupo de teste entre n_{cal} amostras para o grupo de calibração e n_{val} amostras para o grupo de validação, com $n_{cal} \geq z + 1$, isto é, obtenha as matrizes \mathbf{X}^{cal} ($n_{cal} \times k$), \mathbf{y}_1^{cal} ($n_{cal} \times 1$) e \mathbf{y}_2^{cal} ($n_{cal} \times 1$); e \mathbf{X}^{val} ($n_{val} \times k$), \mathbf{y}_1^{val} ($n_{val} \times 1$) e \mathbf{y}_2^{val} ($n_{val} \times 1$).

Passo 4: construa modelos de calibração MLR para estimar cada variável de resposta, utilizando as variáveis selecionadas, na forma

$$\mathbf{y}_1 = \mathbf{x}_0\beta_{01} + \sum_{j \in S^*} \mathbf{x}_j\beta_{j1} + \boldsymbol{\epsilon}_1, \quad (3.14)$$

$$\mathbf{y}_2 = \mathbf{x}_0\beta_{02} + \sum_{j \in S^*} \mathbf{x}_j\beta_{j2} + \boldsymbol{\epsilon}_2, \quad (3.15)$$

onde: \mathbf{x}_0 é um vetor $n \times 1$ com todos os elementos iguais a 1; \mathbf{x}_j denota a coluna referente à j -ésima variável espectral, para todas as colunas selecionadas $j \in S^*$; $\{\beta_{01}, \beta_{11}, \dots, \beta_{z1}\}$ e $\{\beta_{02}, \beta_{12}, \dots, \beta_{z2}\}$ são os coeficientes de regressão de cada modelo; e $\boldsymbol{\epsilon}$ denotam os vetores de erros.

Passo 5: utilize os dados do grupo de calibração (\mathbf{X}^{cal} , \mathbf{y}_1^{cal} e \mathbf{y}_2^{cal}) para estimar os coeficientes de regressão dos modelos, e utilize os modelos para estimar as concentrações a partir dos dados dos grupos de validação (\mathbf{X}^{val}) e de teste (\mathbf{X}^{test}).

Passo 6: atualize a soma dos quadrados dos erros de estimação para os grupos de validação (SQe^{val}) e de teste (SQe^{test}), isto é, faça

$$SQe^{val} = SQe^{val} + \sum_{i=1}^{n_{val}} [(y_{1_i}^{val} - \hat{y}_{1_i}^{val})^2 + (y_{2_i}^{val} - \hat{y}_{2_i}^{val})^2], \quad (3.16)$$

$$SQe^{test} = SQe^{test} + \sum_{i=1}^{n_{test}} [(y_{1_i}^{test} - \hat{y}_{1_i}^{test})^2 + (y_{2_i}^{test} - \hat{y}_{2_i}^{test})^2], \quad (3.17)$$

onde: onde $y_{1_i}^{val}$ e $y_{1_i}^{test}$ correspondem aos valores especificados para as amostras de índice i dos grupos de validação e de teste, respectivamente; e $\hat{y}_{1_i}^{val}$ e $\hat{y}_{1_i}^{test}$ correspondem aos

valores preditos pelo modelo para as amostras de índice i dos grupos de validação e de teste, respectivamente.

Passo 7: atualize o total de amostras avaliadas (N_V e N_T), isto é, faça

$$N_V = N_V + n_{val}, \quad (3.18)$$

$$N_T = N_T + n_{test}. \quad (3.19)$$

Passo 8: faça $g = g + 1$. Se $g \leq G$, retorne ao Passo 1;

Passo 9: determine os valores das estatísticas RMSEV e RMSEP, isto é, faça

$$\text{RMSEV} = \sqrt{\frac{SQe^{val}}{N_V}}, \quad (3.20)$$

$$\text{RMSEP} = \sqrt{\frac{SQe^{test}}{N_T}}, \quad (3.21)$$

Fim: as estatísticas de performances de predição dos modelos são RMSEV e RMSEP.

Neste estudo espectral, foram avaliadas as performances de predição de diferentes modelos candidatos contendo entre três e dez variáveis. Dessa forma, as estatísticas RMSEV e RMSEP foram utilizadas como base para a tomada de decisão sobre o número e a especificação das dos comprimentos de onda a serem empregados na construção do protótipo. Toda a análise de dados foi realizada por meio do *software* Matlab[®] 2016. Os resultados deste estudo são apresentados no Capítulo 4, na Seção 4.3. Baseado nos valores obtidos para a estatística RMSEV, decidiu-se por construir um protótipo para implementar e avaliar modelos de sete variáveis. Os procedimentos utilizados para projetar o sistema optoeletrônico do protótipo são descritos a seguir.

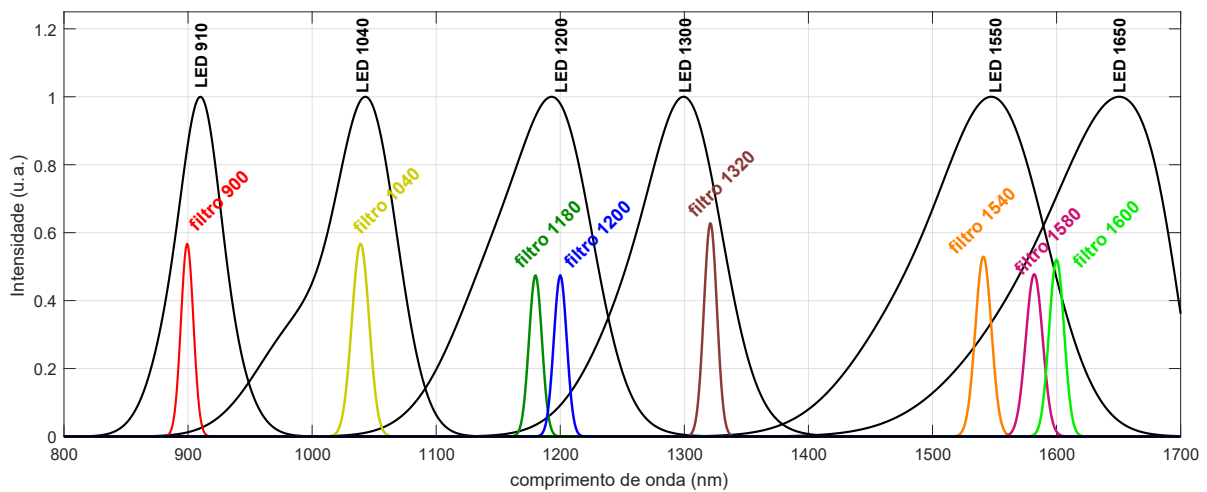
3.4 PROJETO DO PROTÓTIPO E TESTES FINAIS

Após a seleção de variáveis, projetou-se o sistema óptico do protótipo a fim de possibilitar a aquisição dos dados espectrais requeridos para a construção dos modelos de calibração adotados. Para implementar os modelos de sete variáveis, foram necessários oito filtros de interferência ópticos (da *Knight Optical*) centrados nos seguintes comprimentos de onda, dados pelo resultado da seleção de variáveis: 900, 1.040, 1.180, 1.200, 1.320, 1.540, 1.580 e 1.600 nm. As informações ópticas medidas no comprimento de onda 1.040 nm são utilizadas para o pré-processamento espectral, enquanto as informações dos outros sete comprimentos de onda originam as variáveis preditoras dos modelos. Os filtros empregados são do tipo passa-banda, com picos de transmitância entre 46% e 56%, larguras de banda à meia altura entre 10 e 16 nm, e diâmetro de 12,5 mm, com exceção do filtro de 1.600 que possui diâmetro de 25 mm.

Por sua vez, como fonte de luz, foram utilizados LEDs infravermelhos cujos espectros de emissão abrangem as faixas de comprimentos de onda de interesse. Assim,

considerada a disponibilidade, foram selecionados seis LEDs (da *Roithner Lasertechnik*) com espectros de emissão centrados nos seguintes comprimentos de onda: 910, 1.040, 1.200, 1.300, 1.550 e 1.650 nm. Os LEDs selecionados possuem larguras de banda à meia altura (FWHM) entre 46 nm e 120 nm, ângulos de emissão entre 15° e 30°, e suportam correntes contínuas máximas entre 40 mA e 100 mA. Mais especificações técnicas dos LEDs utilizados podem ser obtidas no Anexo A. Os espectros de emissão normalizados dos LEDs, juntamente com os espectros de transmitância dos filtros de interferência, são apresentados na Figura 3.2.

Figura 3.2: Espectros de emissão normalizados dos LEDs e espectros de transmitância dos filtros de interferência.



Fonte: o autor.

Após alguns testes preliminares utilizando diferentes configurações ópticas, adotou-se a configuração mostrada na Figura 3.3(a). Nesta, os LEDs são posicionados de forma que a radiação emitida é direcionada diretamente para o filtro óptico, e a radiação filtrada incide na cubeta (caminho óptico de 10 mm) que contém a amostra a ser analisada. Por sua vez, a luz transmitida pela amostra passa por uma lente condensadora esférica plano-convexa (da *Edmund Optics*) com diâmetro de 25 mm e distância focal de 14 mm, a qual direciona a luz para o fotodetector. O fotodetector escolhido foi o fotodiodo de InGaAs PT511, da *Roithner Lasertechnik*, o qual possui encapsulamento TO-46, diâmetro da área sensível de 300 μm e faixa de trabalho entre 850 e 1.650 nm, o que abrange os comprimentos de onda de interesse. Para permitir a seleção dos diferentes comprimentos de onda, os filtros são posicionados em uma estrutura circular, a qual é girada por meio de um motor de passo. Dessa forma, radiações com diferentes comprimentos de onda podem incidir na amostra selecionando-se o filtro desejado e ativando-se um ou mais LEDs cuja emissão seja significativa no comprimento de onda a ser filtrado. Também, a montagem da estrutura do sistema óptico possibilita que a distância entre o fotodetector e a lente seja ajustável, de forma a se obter uma resposta significativa em todos os comprimentos de onda.

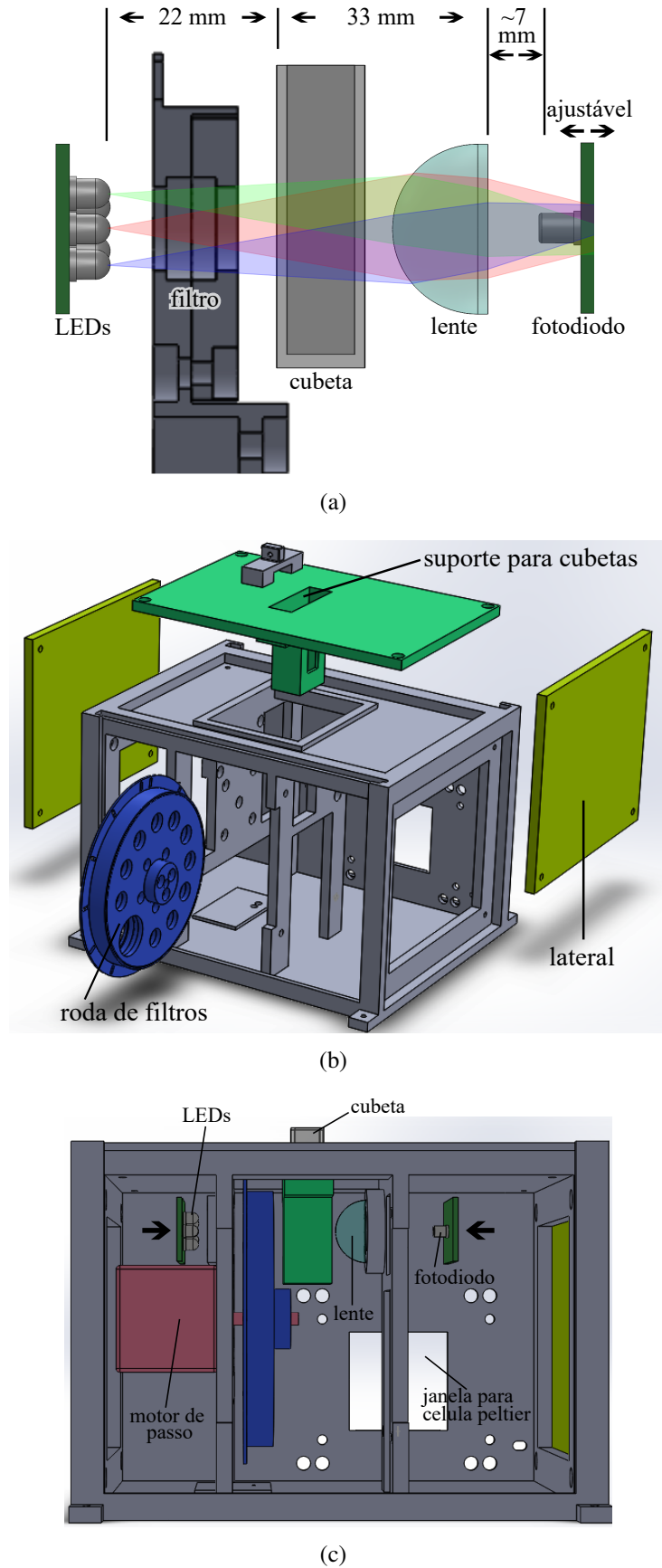


Figura 3.3: (a) Representação do sistema óptico adotado; (b) desenho das principais peças que compõem a estrutura de sustentação do protótipo; (c) desenho de parte da estrutura do protótipo montada.

Fonte: o autor.

Para implementar o sistema óptico da Figura 3.3(a), projetou-se a estrutura mostrada nas Figuras 3.3(b) e 3.3(c). Esta estrutura foi projetada por meio de um *software* de desenho assistido por computador (CAD) e fabricada por meio de uma impressora 3D em material PLA. A Figura 3.3(b) mostra os desenhos das principais peças que constituem a estrutura do protótipo. Por sua vez, a Figura 3.3(c) mostra parte da estrutura montada. Para possibilitar o correto alinhamento de cada filtro de interferência no sistema óptico da Figura 3.3(a), a extremidade da roda de filtros possui marcações e foi projetada para ser utilizada em conjunto com um sensor óptico de barreira. Além de possibilitar o alinhamento e o distanciamento adequados dos dispositivos ópticos, a estrutura projetada cria um encapsulamento fechado para os dispositivos ópticos e para as placas dos circuitos eletrônicos. Isto torna o sistema menos suscetível a influências ópticas externas. Por fim, a estrutura foi dimensionada ainda para possibilitar o controle da sua temperatura interna por meio de uma pastilha termoelétrica de efeito Peltier. As dimensões externas desta estrutura são $20 \times 18 \times 14$ cm.

Posteriormente, procedeu-se ao projeto do sistema eletrônico de instrumentação, responsável pela medição dos dados ópticos nos comprimentos de onda de interesse. Por fim, o protótipo foi montado, calibrado e testado a partir de um novo conjunto de amostras, e os resultados obtidos foram comparados com os resultados do espectrofotômetro de bancada. O projeto e a implementação do protótipo são apresentados no próximo capítulo, na Seção 4.4, e os resultados dos testes com as amostras de gasolinas são apresentados e discutidos na Seção 4.5.

4 RESULTADOS EXPERIMENTAIS

4.1 COMPARAÇÃO ENTRE OS MÉTODOS DE CÁLCULO DAS ABSORBÂNCIAS

A Figura 4.1 mostra as curvas espectrais obtidas para as 102 amostras de calibração e validação, obtidas a partir dos diferentes métodos de cálculo das absorvâncias, sem nenhum tipo de pré-processamento espectral. São apresentadas apenas as curvas referentes aos métodos B, D, E e F. Os métodos A, C e G resultaram em curvas espectrais de formato similar àquelas da Figura 4.1(a), já que em todos esses métodos as referências são medidas com cubetas vazias ou sem cubeta; por isso, as demais curvas não são apresentadas.

De acordo com os dados da Figura 4.1, a determinação das absorvâncias a partir dos diferentes métodos avaliados originou curvas espectrais com formatos distintos, embora a informação química (isto é, as concentrações de etanol e dos demais componentes das amostras) fossem as mesmas em todos os métodos. Esta análise poderia ser expandida para um sistema com mais pontos espectrais. Assim, deve-se notar que a aplicação de técnicas de pré-processamento espectral e a interpretação das bandas espectrais em comparação com outros trabalhos na literatura devem ser feitas de forma cuidadosa, dependendo do método empregado. Por exemplo, pode-se observar que na Figura 4.1(a) os pontos espectrais com menores valores de absorvância, que definem a linha de base do espectro, estão localizados no comprimento de onda 970 nm, enquanto nas Figuras 4.1(b) e 4.1(d), os menores valores de absorvâncias são encontrados no comprimento de onda 1.450 nm.

A Figura 4.1 também mostra que houve variação nos valores de absorvância, para diferentes amostras, em todos os comprimentos de onda. A única exceção é o comprimento de onda 970 nm mostrado na Figura 4.1(b), que é utilizado como referência para o Método D e, conseqüentemente, origina valores de absorvância iguais a zero para todas as amostras. Isto se deve às variações nas intensidades de luz medidas, cujos valores absolutos não representam por si só uma resposta às variações na concentração do etanol. Isto porque, como discutido anteriormente, o etanol não absorve luz na faixa de 970 nm; assim, a variância observada naquele comprimento de onda reflete a influência dos fatores físicos que modificam a luz transmitida por cada amostra. Ao invés disso, pode-se entender que a informação química está na relação entre as intensidades em diferentes comprimentos de onda, ou seja, no formato do espectro de cada amostra. Neste sentido, pode-se observar nas Figuras 4.1(a) e 4.1(c) que o método E apresentou uma maior variância em λ_2 e λ_3 em relação a λ_1 , quando comparado com o método B, por exemplo. Isto é porque o método E utiliza uma amostra de gasolina na medição da referência $I_0(\lambda)$, de forma que a influência da matriz é reduzido; assim, o efeito da absorção do etanol na banda que compreende os comprimentos de onda 1.450 nm e 1.530 nm é intensificado. Esta característica do método E teve influência significativa nos resultados da calibração, os quais são mostrados a seguir.

A Tabela 4.1 apresenta os resultados obtidos após calibração utilizando os

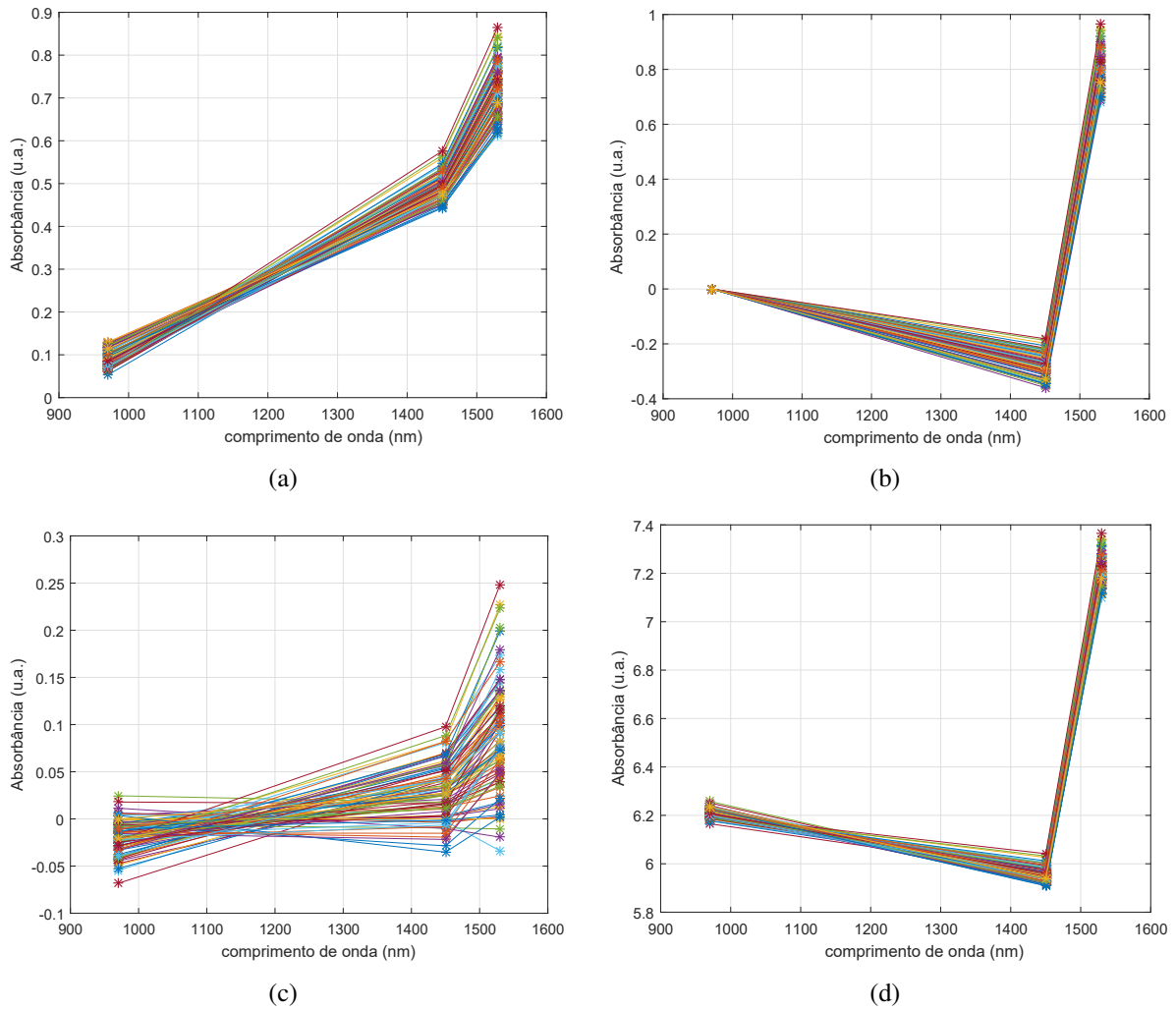


Figura 4.1: Curvas espectrais das 102 amostras de calibração e validação, sem pré-processamento espectral, as quais foram obtidos a partir das absorvâncias calculadas por: (a) método B; (b) método D; (c) método E; e (d) método F.

pontos espectrais mostrados na Figura 4.1, sem pré-processamento espectral. Pode-se observar nesta tabela que o método E alcançou os melhores resultados, menores valores de RMSEC (1,17%) e RMSEP (1,12%); maiores coeficientes de determinação ($R^2 = 0,9164$ para calibração e $R^2 = 0,9359$ para validação); e maior RPD (3,90). Ao mesmo tempo, os maiores valores de RMSEC e RMSEP foram obtidos para os métodos B e A, respectivamente. Também, o método A teve o menor RPD (2,70), juntamente com o método G. Por sua vez, a Tabela 4.2 apresenta os resultados obtidos após aplicação da correção da linha de base. É possível observar que, novamente, os menores erros, os maiores coeficientes de determinação e maiores desvios de previsão residual foram obtidos pelo método E, enquanto os métodos A, B e G alcançaram os piores resultados. Também, é possível verificar que, com exceção dos parâmetros RMSEC nos métodos E e F e seus correspondentes coeficientes de determinação, todos os demais parâmetros apresentaram resultados iguais ou melhores em relação aos dados da Tabela 4.1. Isto indica que o pré-processamento espectral melhorou a aderência dos pontos espectrais ao modelo de regressão linear. Pode-se observar, contudo, que a diferença nos valores de RMSEP

Tabela 4.1: Comparação entre os resultados para os sete diferentes métodos de cálculo de absorvâncias, considerando calibração multivariada por MLR sem pré-processamento espectral.

Método	R^2 Calibração	R^2 Validação	RMSEC (%)	RMSEP (%)	RPD
A	0,8295	0,8697	1,66	1,61	2,70
B	0,8225	0,8757	1,70	1,56	2,79
C	0,8405	0,8832	1,61	1,51	2,88
D	0,8406	0,8893	1,60	1,49	2,92
E	0,9164	0,9359	1,17	1,12	3,90
F	0,8508	0,8859	1,56	1,49	2,91
G	0,8358	0,8711	1,63	1,61	2,70

Tabela 4.2: Comparação dos resultados para os sete diferentes métodos de cálculo de absorvâncias, considerando calibração multivariada por MLR com correção da linha de base (*offset*).

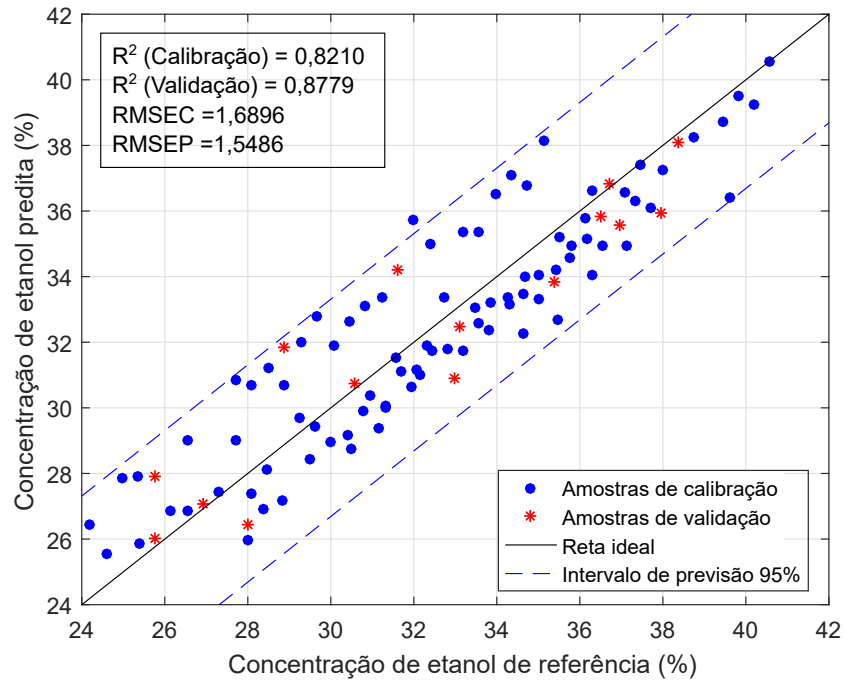
Método	R^2 Calibração	R^2 Validação	RMSEC (%)	RMSEP (%)	RPD
A	0,8295	0,8698	1,65	1,61	2,70
B	0,8210	0,8779	1,69	1,55	2,80
C	0,8359	0,8856	1,60	1,50	2,90
D	0,8406	0,8893	1,60	1,49	2,92
E	0,9152	0,9406	1,17	1,08	4,04
F	0,8487	0,8897	1,56	1,48	2,95
G	0,8358	0,8735	1,62	1,60	2,72

e RMSEC foi insignificante em termos práticos. Isto indica que os efeitos físicos do espalhamento da luz nos dados medidos por meio deste protótipo foram pouco significativos. Apesar disso, deve-se considerar que a variabilidade de matriz esperada na prática é muito maior do que a representada nas amostras utilizadas neste experimento, o que pode justificar a necessidade de pré-processamentos. Dessa forma, isto também foi avaliado nos experimentos subsequentes. Além disso, se os conceitos aqui discutidos forem expandidos para outras aplicações, em análises de outros tipos de substâncias heterogêneas, com matrizes mais complexas e/ou com maior variabilidade amostral, o efeito do espalhamento da luz poderá ser mais intenso. Neste caso, acredita-se que a melhora nos resultados causada pelos pré-processamentos espectrais será mais evidente. Finalmente, pode-se observar em ambas as tabelas que todos os métodos apresentaram valores de RPD maiores ou iguais a 2,70, o que indica boas performances preditivas de todos os modelos.

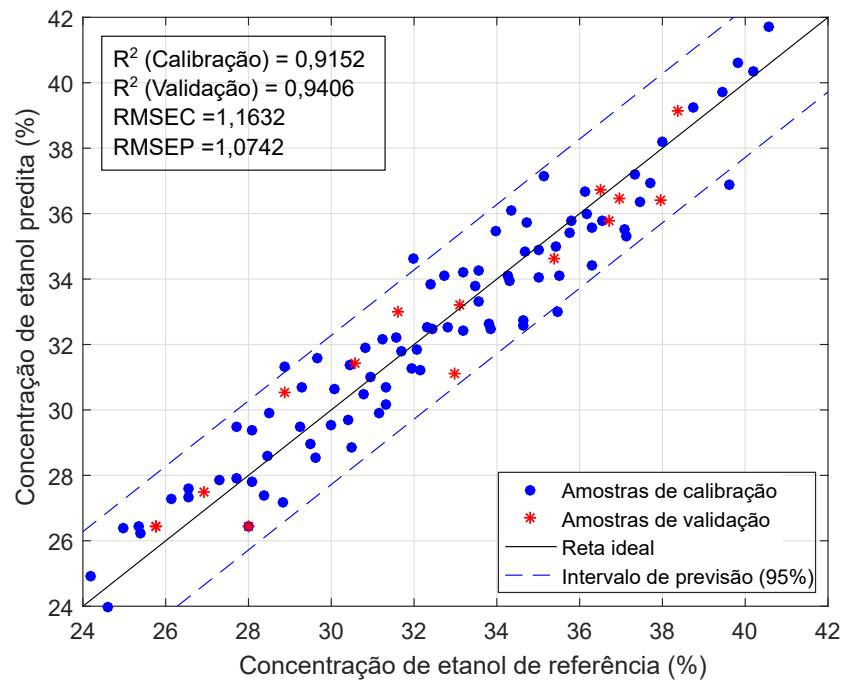
Analisando-se agora os dados da Tabela 4.2, pode-se observar que os maiores valores de RMSEP e RMSEC foram obtidos para aqueles métodos que utilizam medidas de referência com cubetas vazias (métodos A, B e G). Isto mostra que o material das cubetas também representa uma fonte de variabilidade nos sinais ópticos. Entre estes três métodos, os melhores resultados foram obtidos quando uma medida de referência foi tomada para cada

Tabela 4.3: Concentrações de referência do conjunto de validação e concentrações previstas para os sete métodos de cálculo das absorvâncias, considerando calibração multivariada por MLR com correção da linha de base (*offset*). A medição de cada amostra foi realizada uma única vez e empregada em todos os métodos.

Amostra	Concentração de referência (%)	Método A (%)	Método B (%)	Método C (%)	Método D (%)	Método E (%)	Método F (%)	Método G (%)
1	25,773	26,062	26,028	26,013	26,759	26,431	26,480	26,277
2	26,928	27,155	27,063	27,099	27,766	27,512	27,581	27,058
3	33,093	32,481	32,488	32,661	33,193	33,185	33,234	32,450
4	36,948	35,754	35,592	35,698	35,830	36,438	36,335	35,790
5	25,756	27,545	27,889	27,596	27,408	26,431	27,282	27,483
6	31,615	34,012	34,191	34,046	33,745	33,021	33,840	33,899
7	28,879	31,656	31,833	31,649	31,253	30,554	31,401	31,689
8	36,692	36,368	36,827	36,655	36,756	35,783	36,501	36,708
9	30,592	31,202	30,749	30,901	30,204	31,411	30,697	31,274
10	38,361	38,487	38,113	38,288	38,271	39,121	38,221	37,978
11	36,510	35,833	35,857	36,190	36,142	36,749	36,067	35,841
12	35,402	34,040	33,819	33,982	33,798	34,626	33,835	34,076
13	28,000	26,330	26,428	26,255	26,021	26,431	25,986	26,320
14	32,982	30,729	30,918	30,793	30,674	31,131	30,604	30,769
15	37,958	34,885	35,956	35,801	35,715	36,407	35,709	34,896



(a)



(b)

Figura 4.2: Relação entre concentrações especificadas e previstas, considerando o modelo de MLR com correção da linha de base: (a) utilizando o método B, que utiliza como referência uma medição com uma cubeta vazia no início do ciclo de análises; e (b) utilizando o método E, que calcula a absorbância em relação a uma amostra de referência.

amostra, mantendo-se a cubeta imóvel no caminho óptico para medição de referência e subsequente medição da amostra (método G), uma vez que isto reduz a variabilidade causada pela cubeta. Resultados intermediários foram obtidos pelo método C, que utiliza uma única medição sem cubeta; e pelo método D, em que as absorbâncias são relativas ao comprimento de onda de referência λ_1 . Por sua vez, os melhores resultados foram obtidos pelos métodos E (em que as absorbâncias são relativas a uma amostra de referência) e F (que é independente de uma medição de referência).

Na Tabela 4.3, são apresentadas numericamente as concentrações especificadas e as preditas pelos sete métodos avaliadas para as 15 amostras do conjunto de validação. Por sua vez, nas Figuras 4.2(a) e 4.2(b) são apresentadas as relações entre as concentrações especificadas e aquelas estimadas pelo protótipo a partir dos métodos B e E, respectivamente, os quais apresentaram maior e menor valores de RMSEC.

Uma comparação entre as Figuras 4.2(a) e 4.2(b) permite avaliar graficamente a diferença entre os dois métodos na estimação das concentrações, para ambos os grupos de calibração e de validação. Pode-se observar que a variância dos pontos em torno da reta ideal é menor para o Método E, o que indica uma maior exatidão do protótipo quando calibrado a partir deste método. É importante reafirmar que a coleta de dados das amostra foi realizada uma única vez, os quais foram utilizados para todos os métodos, de forma que as diferenças entre eles são dadas exclusivamente pela definição de diferentes referências. Também, deve-se notar que os valores de RMSEC e RMSEP apresentados correspondem aos desvios padrão amostrais dos erros de quantificação, que correspondem a estimativas pontuais dos desvios padrão populacionais em cada método. Mesmo assim, neste experimento, deve-se considerar que os sete métodos originaram sete estimativas pontuais diferentes para a mesma amostra estatística (isto é, para um único conjunto de medições que foram tomadas uma única vez, e não para diferentes conjuntos de observações). Assim, a comparação entre os valores de RMSEC e RMSEP para os diferentes métodos é válida. Nesta comparação, foram levadas em consideração as incertezas na estimação destes parâmetros com dois dígitos significativos, como recomendado por JGCM (2008). Por isso, os valores de RMSEC e RMSEP foram arredondados para cima (JGCM, 2008) e apresentados com dois dígitos decimais de forma a serem consistentes com suas incertezas.

Além dos resultados numéricos apresentados, outras diferenças podem ser inferidas pela análise das características e implicações práticas da utilização de cada método. Assim, algumas vantagens e desvantagens de cada método são resumidos na Tabela 4.4. Nesta tabela, os métodos são listados em ordem decrescente de performance, de acordo com o valor de RMSEC. A partir das informações da Tabela 4.4 e dos resultados experimentais apresentados anteriormente, conclui-se que o método que se mostrou mais apropriado na quantificação de etanol em gasolinas pelo protótipo foi o Método E. Acredita-se que resultados semelhantes seriam obtidos também para a quantificação de metanol. Este método também pode ser avaliado para otimizar a calibração para outros tipos de substâncias em diferentes aplicações. Contudo, dependendo da aplicação, a indisponibilidade de amostras de referência para a medição de $I_0(\lambda)$ pode tornar este método impraticável. Também, apesar de ter alcançado o segundo

Tabela 4.4: Vantagens e desvantagens dos sete métodos estudados, considerando suas características e questões de utilização prática para instrumentos de diferentes tipos construtivos e diferentes aplicações. Os métodos são apresentados em ordem crescente de RMSEC.

	Vantagens	Desvantagens
Método E	Requer apenas uma medição de referência a cada novo ciclo de análises.	Amostras ou soluções de referência devem estar disponíveis a cada novo ciclo de análises; Podem ser inadequado para sistemas que utilizam células de fluxo fixas.
Método F	É independente de uma medição de referência.	Provavelmente é mais dependente de fatores como a estabilidade na emissão da luz e do alinhamento optomecânico; Também é inadequado para sistemas com células de fluxo fixas; É dependente da responsividade do sensor.
Método D	Não requer uma medição de referência, uma vez que a referência é obtida pela medição da própria amostra.	Depende de uma escolha cuidadosa de λ_{ref} ; assim, requer um conhecimento aprofundado das características espectrais do tipo de amostra a ser analisada; Também é dependente da responsividade do sensor.
Método C	Requer apenas uma medição de referência a cada novo ciclo de análises. Medição de referência não é influenciada pelo material da cubeta.	Uma vez que o equipamento tenha sido calibrado, pode requerer uma nova calibração caso seja necessário utilizar um tipo diferente de cubeta; Ineficaz para sistemas com células de fluxo fixas.
Método G	Variabilidade nos dados ópticos causada pelo material da cubeta é minimizada, uma vez que a referência e a amostra são medidas com uma mesma cubeta a qual é mantida imóvel entre as duas medições.	Requer uma medição de referência para cada nova amostra; Requer grande cuidado no manuseio do instrumento, uma vez que a amostra deve ser inserida na cubeta sem removê-la do caminho óptico.
Método A	Maior facilidade no manuseio das amostras em comparação com o método G, já que permite o movimento da cubeta entre a medição da referência e a medição da amostra; Variabilidade causada pelo material da cubeta também é reduzida.	Também requer uma nova medição de referência para cada amostra.
Método B	Requer apenas uma medição de referência a cada novo ciclo de análises.	Pode também levar a resultados ruins em sistemas com célula de fluxo fixas devido, por exemplo, ao acúmulo de resíduos no porta amostras entre análises subsequentes.

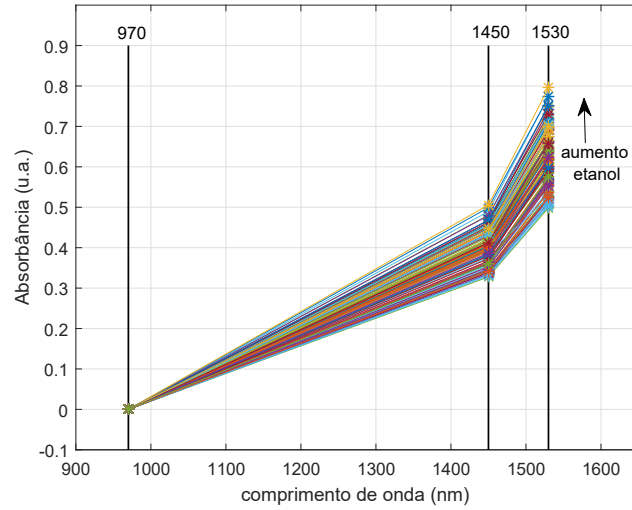
melhor desempenho neste experimento, o método F tem várias desvantagens que podem levar à necessidade de calibrações com maior frequência, uma vez que a resposta do instrumento pode ser afetada ao longo do tempo. Por sua vez, o método C demonstrou-se apropriado para esta aplicação desde que o tipo de cubeta a ser utilizado tenha sempre as mesmas características ópticas. Finalmente, a análise apresentada sugere que, em um sistema com célula de fluxo fixa, tal como o proposto inicialmente por Moreira (2013), o método G pode levar a melhores resultados do que o método B, uma vez que a medição da referência (célula vazia) a cada nova análise deve reduzir os efeitos da acumulação de resíduos entre amostras subsequentes. Neste sentido, métodos E e F são provavelmente inadequados para este tipo de construção, uma vez que estes métodos usam apenas uma medição de referência no início do ciclo de análises; enquanto o método C seria impraticável.

Tendo em vista todos os aspectos discutidos, conclui-se que a escolha do método mais apropriado depende das características do tipo de solução a ser analisada em uma aplicação específica, e pode ser utilizada como uma ferramenta para otimizar a resposta do modelo matemático de medição. Apesar disso, considerando a quantificação dos álcoois estudada neste trabalho, e tendo em vista uma possível utilização para o controle de qualidade da gasolina, pode-se considerar que a diferença entre os métodos não foi tão significativa. Assim, para permitir uma comparação mais adequada do protótipo desenvolvido com o espectrofotômetro de bancada, o Método B foi escolhido para os experimentos subsequentes.

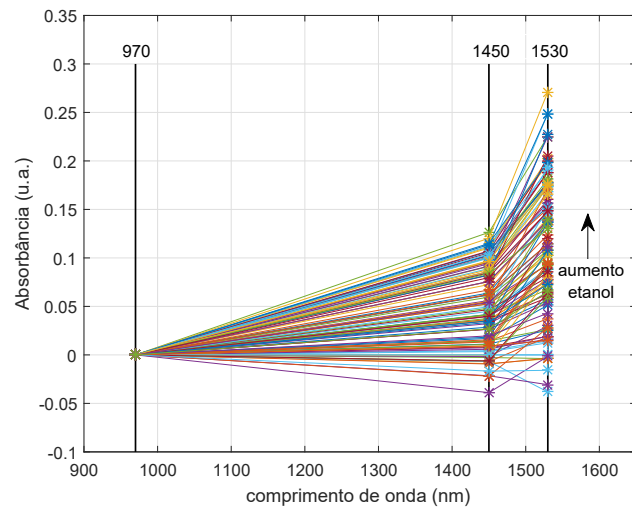
4.2 ANÁLISE DA INFLUÊNCIA DA VARIABILIDADE DE MATRIZ

A Figura 4.3 apresenta as curvas espectrais obtidas pelo protótipo para os métodos de cálculo de absorvâncias B e E, e também aquelas obtidas pelo espectrofotômetro, para as 102 amostras utilizadas, após o pré-processamento espectral. Estes dados espectrais foram utilizados para construir os modelos MLR de duas variáveis consideradas neste experimento. Nos três gráficos apresentados, os dados espectrais foram pré-processados tendo como base o comprimento de onda 970 nm. Também, para comparação, os comprimentos de onda correspondente aos valores nominais dos três LEDs do protótipo são indicados.

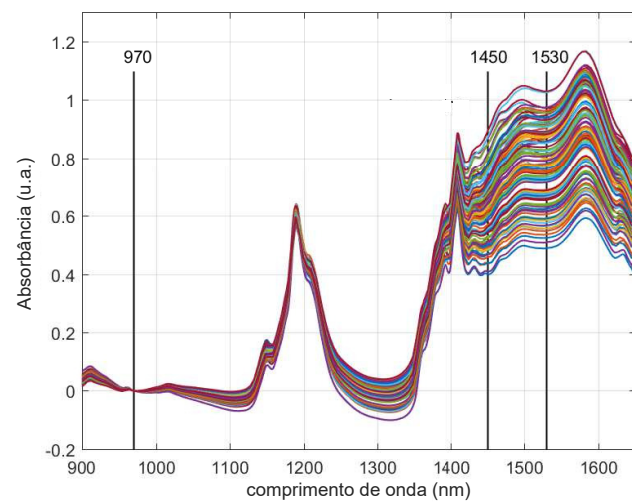
Pode-se observar na Figura 4.3 que as características de variação das absorvâncias obtidas pelo protótipo em ambos os métodos é consistente com aquela observada no espectrofotômetro, apesar de a amplitude de variação em 1.450 nm e 1.530 nm serem relativamente menores para o protótipo. Isto revela uma menor sensibilidade do protótipo em relação ao equipamento comercial, que pode estar associada com a relativa simplicidade do sistema óptico utilizado: devido à ausência de lentes, filtros ou outros dispositivos no sistema óptico para colimar a luz, a quantidade dos feixes de luz que interagem com a amostra e chegam ao detector para compor a resposta observada provavelmente é menor. Também, deve-se notar que os LEDs tem uma banda de emissão relativamente larga (cerca de 100 nm). Então, diferentemente do espectrofotômetro, os sinais que chegam ao sensor no protótipo correspondem na verdade às contribuições de comprimentos de onda em uma faixa espectral em torno do valor nominal



(a)



(b)



(c)

Figura 4.3: Curvas espectrais pré-processadas de 119 amostras de gasolina contendo diferentes concentrações de etanol, obtidas: (a) pelo protótipo, por meio do Método B de cálculo de absorbâncias; (b) pelo protótipo, por meio do Método E de cálculo das absorbâncias; (c) pelo espectrofotômetro de bancada, para a faixa espectral de 900 a 1.650 nm.

de cada LED. Apesar dessas observações, os gráficos mostram que, desconsiderando-se uma maior influência de matriz, esta solução baseada em LEDs permite analisar substâncias que apresentem bandas de absorção relativamente largas e bem definidas, mesmo sem a utilização de lentes, filtros ou outros tipos de elementos monocromadores.

4.2.1 Avaliação da influência da seleção de amostras utilizando o Método B

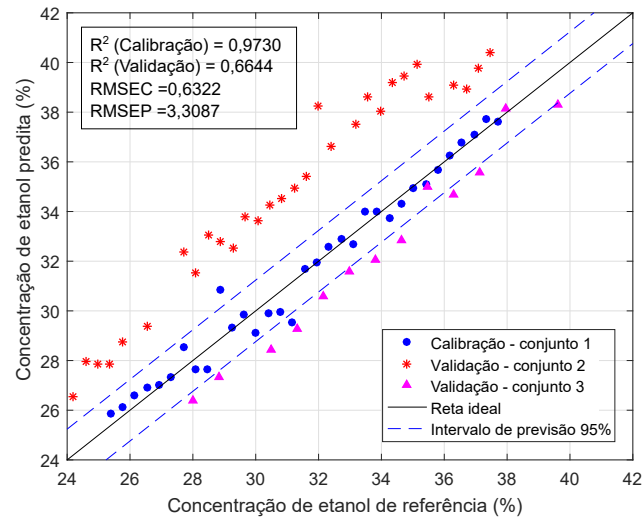
A partir dos pontos espectrais mostrados na Figura 4.3(a), as quais foram calculadas pelo Método B, determinou-se as curvas de calibração para os dois procedimentos de seleção de amostras descritos na Seção 3.2. No primeiro procedimento, em que a calibração foi realizada a partir de dados exclusivamente do primeiro conjunto de amostras, obteve-se a curva de calibração

$$\hat{y} = -7,8297 + 96,6083x_2 + 4,0831x_3, \quad (4.1)$$

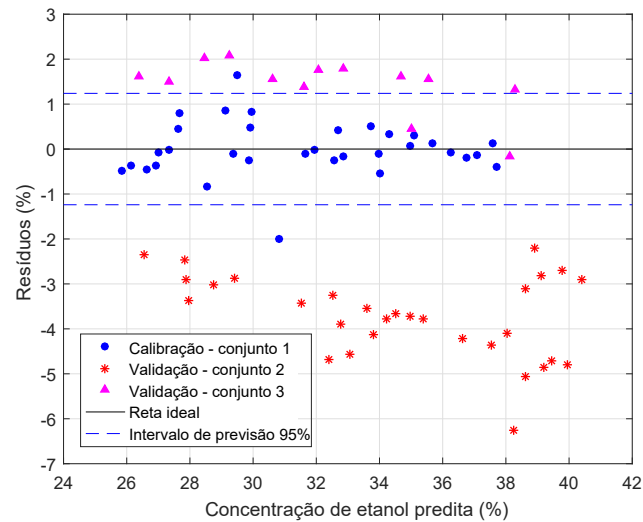
em que x_2 e x_3 são as absorbâncias pré-processadas nos comprimentos de onda 1.450 nm e 1.530 nm, respectivamente; e \hat{y} é a concentração de etanol estimada pelo modelo. Os resultados obtidos utilizando (4.1) são mostrados nas Figuras 4.4(a), 4.4(b) e 4.4(c).

A Figura 4.4(a) relaciona as concentrações de etanol de referência com aquelas estimadas pelo protótipo de acordo com (4.1). Nesta figura, os pontos estimados pelo modelo para as amostras de calibração (conjunto 1) são apresentados, os quais foram utilizados para determinar o valor de RMSEC. Este parâmetro corresponde ao desvio padrão amostral, que é uma estimativa pontual do desvio padrão populacional. Adicionalmente, o intervalo de probabilidades de 95% é representado, o qual é definido por $\pm 1,96 \times \text{RMSEC}$, considerando que o erro esperado tem uma distribuição de probabilidades gaussianas. Este intervalo indica que, se a curva de calibração obtida for uma boa estimativa da curva de regressão verdadeira, e se as amostras de calibração representam a população adequadamente, então espera-se que 95% das novas observações esteja dentro deste intervalo. Por fim, a Figura 4.4(a) também mostra os pontos estimados pelo modelo para as amostras de teste (conjuntos 2 e 3), as quais foram utilizadas para determinar o parâmetro RMSEP. Por sua vez, e apresentado na Figura 4.4(b) o gráfico dos erros de calibração e validação. Por último, na Figura 4.4(c), são apresentados histogramas dos erros mostrados na Figura 4.4(b), de forma a estimar a distribuição de frequências dos erros observados. Também, é apresentada a função densidade de probabilidades normal com média zero e desvio padrão igual a RMSEC, que corresponde aos parâmetros estimados pelo modelo, juntamente com o intervalo de probabilidades de 95%.

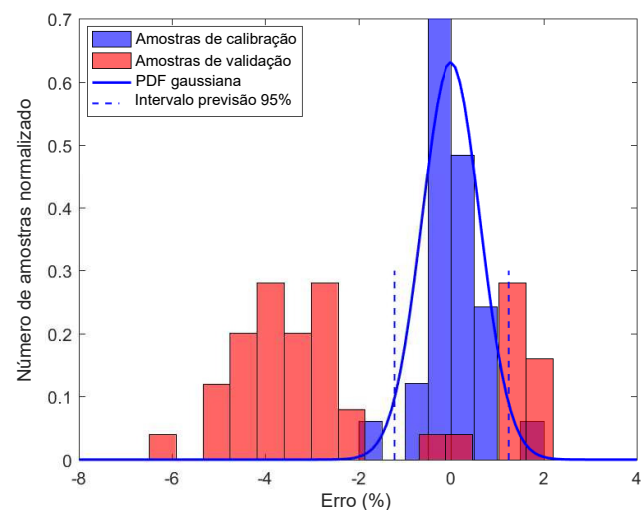
A análise da Figura 4.4(a) indica que o modelo ajusta bem os dados referentes às amostras de calibração, levando a um valor relativamente pequeno de RMSEC (0,67%) e a um coeficiente de determinação $R^2 = 0,9730$. Por outro lado, as amostras de validação estão mais distantes da reta referente à função identidade, resultando em RMSEP=3,31% e em um R^2 de apenas 0,6644. As Figuras 4.4(a) e 4.4(b) mostram que existe uma tendência clara associada a cada um dos três conjuntos de dados. Pode-se observar que os erros para as amostras de calibração estão dispersos em torno de zero, enquanto os erros são todos negativos



(a)

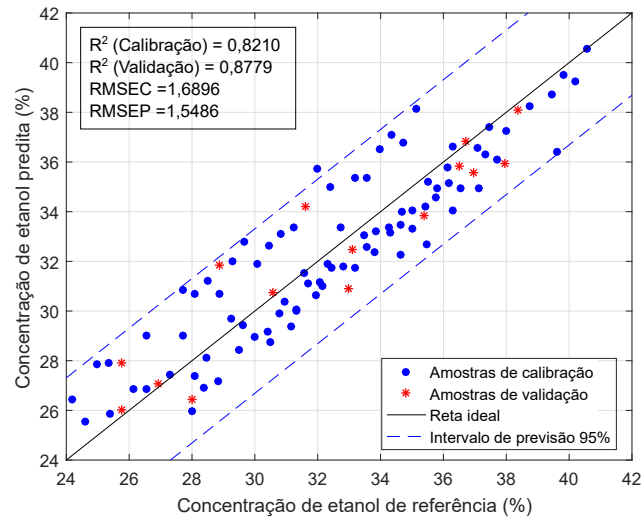


(b)

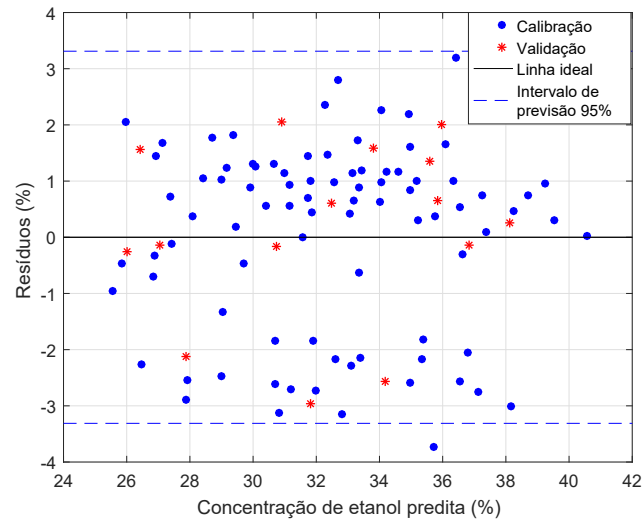


(c)

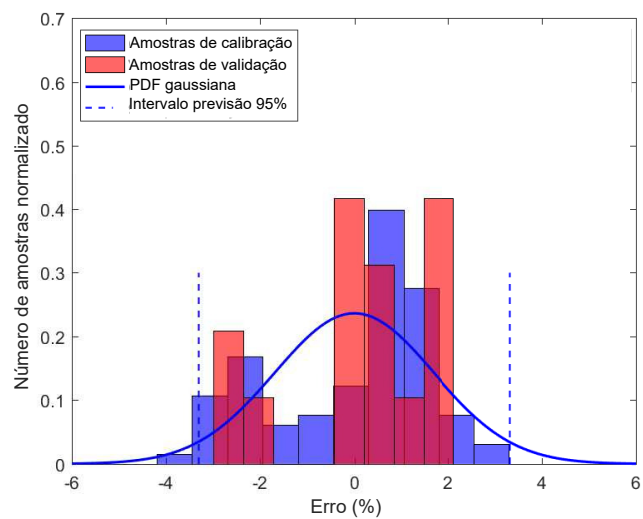
Figura 4.4: Resultados após calibração utilizando amostras exclusivamente do primeiro conjunto de dados, com pontos espectrais obtidos pelo Método B: (a) relação entre concentrações especificadas pelo método de referência e concentrações estimadas pelo modelo; (b) resíduos de calibração e erros de predição; (c) histogramas dos resíduos de calibração e dos erros de predição.



(a)



(b)



(c)

Figura 4.5: Resultados após calibração utilizando amostras distribuídas entre os quatro conjuntos, com pontos espectrais obtidos pelo Método B: (a) relação entre concentrações especificadas pelo método de referência e concentrações estimadas pelo modelo; (b) resíduos de calibração e erros de predição; (c) histogramas dos resíduos de calibração e dos erros de predição.

para as amostras de validação do segundo conjunto, e predominantemente positivos para as amostras de validação do conjunto 3. Também, os erros de predição do conjunto 2 foram maiores do que aqueles do conjunto 3. Isto pode indicar que a matriz referente ao conjunto 2 possui características espectrais mais distintas daquelas do grupo de calibração (conjunto 1). Soma-se a isto a possibilidade de que o erro na quantificação inicial da gasolina que originou o conjunto 2 foi maior em relação à gasolina inicial do conjunto 3. Finalmente, pode-se observar na Figura 4.4(c) que os erros de calibração apresentam uma distribuição que se assemelha à gaussiana, que é uma das premissas da análise de regressão. Contudo, os erros associados às amostras de teste estão distantes do intervalo de probabilidades de 95% estimado, o que indica que a curva de regressão (4.1), bem como a distribuição de probabilidades dos erros estimada para a população, são inadequadas. A partir desses resultados, pode-se também inferir que a variabilidade amostral, associada à incerteza do método de referência, tem um impacto mais significativo na variância dos erros de estimação do que a influência dos erros instrumentais na medição dos sinais ópticos. Assim, a análise das Figuras 4.4(a), 4.4(b) e 4.4(c) indica que os erros de predição em amostras futuras poderia ser muito elevados se as amostras selecionadas para a calibração não forem representativas da variabilidade populacional.

Diferentemente, no segundo procedimento de seleção de amostras, as amostras de calibração e validação foram selecionadas de forma distribuída entre os quatro conjuntos de dados. Neste caso, a função de medição obtida foi

$$\hat{y} = 1,9174 - 27,5043x_2 + 66,0505x_3, \quad (4.2)$$

e os resultados obtidos a partir dessa curva de calibração são mostrados nas Figuras 4.5(a), 4.5(b) e 4.5(c). Observa-se nestas figuras que o espalhamento dos pontos das amostras de calibração em torno da reta ideal é maior do que aquela observada anteriormente, conforme indica também o aumento no valor de RMSEC (1,69% v/v) e a diminuição no coeficiente de determinação ($R^2 = 0,8210$). Consequentemente, pode-se também notar que a região de probabilidades de 95% é mais larga. Por outro lado, agora os erros de predição para as amostras de validação estão distribuídas em torno de zero, e todos os pontos estão dentro da região de $\pm 1,96 \times \text{RMSEC}$, resultando em $\text{RMSEP} = 1,55\% \text{ v/v}$. A função densidade de probabilidade dos erros, mostrada na Figura 4.5(c), é mais larga e caracteriza melhor a variabilidade populacional. Consequentemente, conclui-se que este segundo procedimento de seleção de amostras para a calibração é mais apropriado do que o anterior. Neste caso, a variabilidade da população é melhor representada na calibração, e os erros de predição obtidos estão consistentes com aqueles estimados na calibração. Evidentemente, este fator já era conhecido da literatura, e uma diferença nos resultados para os dois casos já era esperada. Entretanto, a partir desse experimento, tem-se uma percepção quantitativa da influência deste fator na análise de gasolina, bem como da importância de se utilizar um método primário de referência com menor incerteza.

Para relacionar o efeito da seleção de amostras com os métodos de cálculo de absorvâncias, ambos os procedimentos descritos nesta subseção foram repetidos, para as

mesmas amostras, desta vez utilizando as absorvâncias calculadas a partir do Método E. Os resultados obtidos são apresentados a seguir.

4.2.2 Avaliação da influência da seleção de amostras utilizando o Método E

As curvas de calibração apresentadas na subseção 4.2.1 foram recalculadas, desta vez utilizando os pontos espectrais mostrados na Figura 4.3(b). No primeiro caso, utilizando apenas amostras do primeiro conjunto para a calibração, a função de medição obtida foi

$$\hat{y} = 25,9561 - 24,1222x_2 + 87,6033x_3, \quad (4.3)$$

e os resultados obtidos a partir de (4.3) são mostrados nas Figuras 4.6(a), 4.6(b) e 4.6(c). Por sua vez, para a calibração utilizando amostras distribuídas entre os quatro conjuntos, a função de medição obtida foi

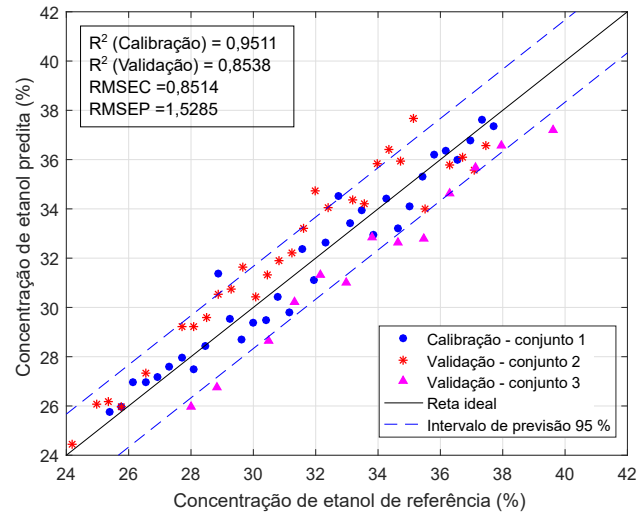
$$\hat{y} = 26,4313 - 27,8454x_2 + 68,5890x_3, \quad (4.4)$$

e os resultados obtidos a partir de (4.4) são mostrados nas Figuras 4.7(a), 4.7(b) e 4.7(c).

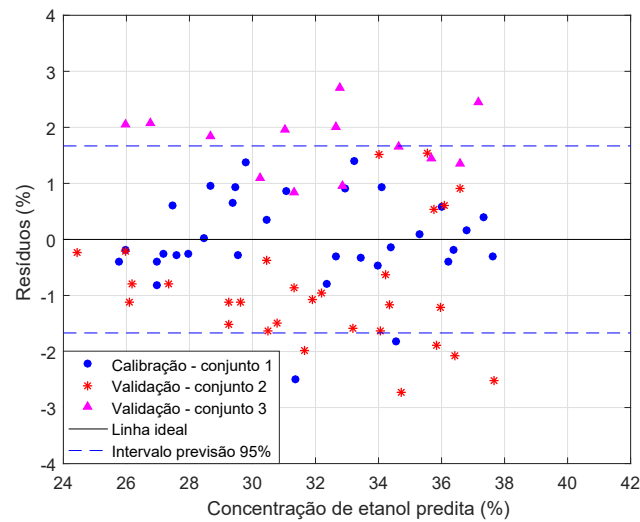
Comparando-se as Figuras 4.6 e 4.7, observa-se que a seleção de amostras de calibração de forma distribuída resultou, novamente, em menores erros de predição (RMSEP=1,08 % v/v, $R^2=0,9406$) em relação à calibração com amostras de um único conjunto (RMSEP=1,53 % v/v, $R^2=0,8538$), como esperado. As mesmas observações anteriores podem ser feitas com respeito à distribuição observada dos erros.

Por sua vez, comparando os resultados apresentados para diferentes métodos de cálculo de absorvâncias (métodos B e E), pode-se notar que as diferenças observadas nos erros de predição entre os diferentes procedimentos de seleção de amostras foi menor para o método E (0,45 % v/v) em relação ao método B (1,76 % v/v). Isto evidencia a existência de diferenças na composição das gasolinas utilizadas: como o Método E emprega a medição de $I_0(\lambda)$ a partir de uma amostra de gasolina, que contém a matriz, o seu efeito na variância dos dados entre amostras de diferentes conjuntos é menor do que no Método B, em que a medição de $I_0(\lambda)$ é feita para uma cubeta vazia.

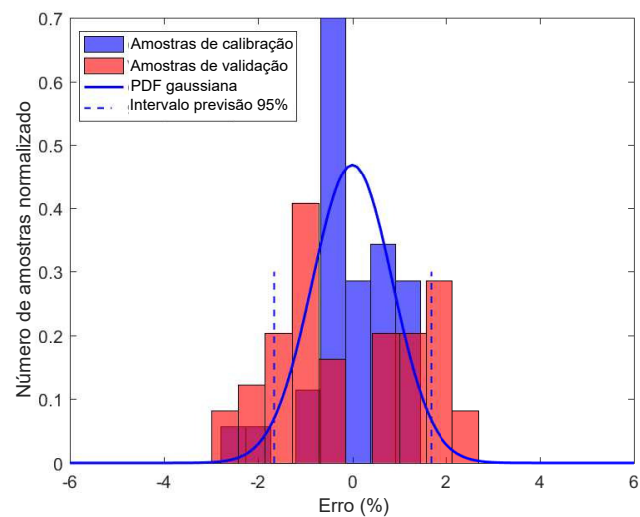
Por fim, comparando-se as equações de calibração obtidas nos quatro casos testados, pode-se observar que (4.2) e (4.4) apresentam coeficientes parciais de regressão estimados ($\hat{\beta}_2$ e $\hat{\beta}_3$) relativamente próximos. Isto indica que ambos os métodos de cálculo de absorvâncias resultaram em curvas de calibração com sensibilidades similares às variações nas absorvâncias em 1.450 nm e em 1.530 nm, quando as amostras de calibração foram selecionadas de forma distribuída. Em contraste, ao utilizar amostras de um único conjunto, os mesmos métodos levaram a curvas de calibração ((4.1) e (4.3)) com coeficientes estimados $\hat{\beta}_2$ e $\hat{\beta}_3$ totalmente distintos. Este é um indicativo de que esta forma de seleção de amostras gera instabilidade na estimação dos coeficientes de regressão. Deve-se registrar, porém, que nesta comparação, o número de amostras utilizado para calibração nos dois procedimentos de seleção



(a)

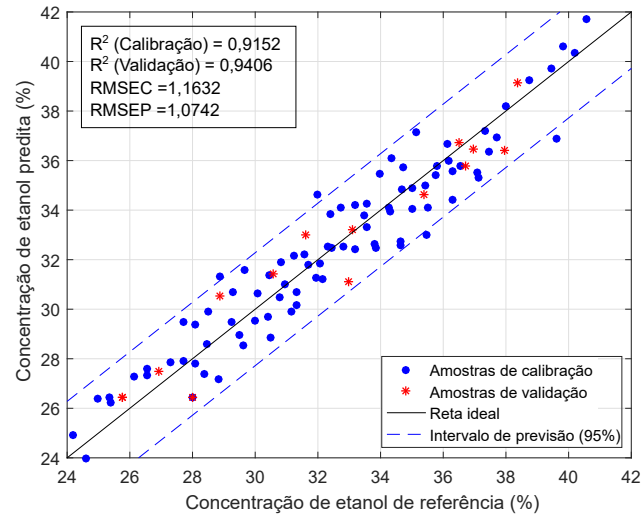


(b)

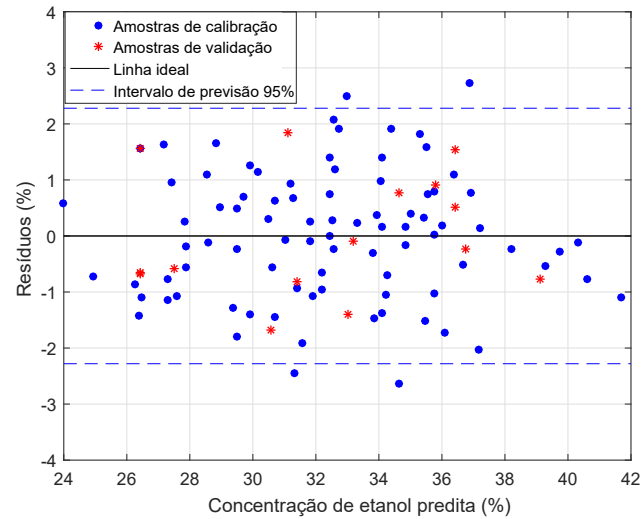


(c)

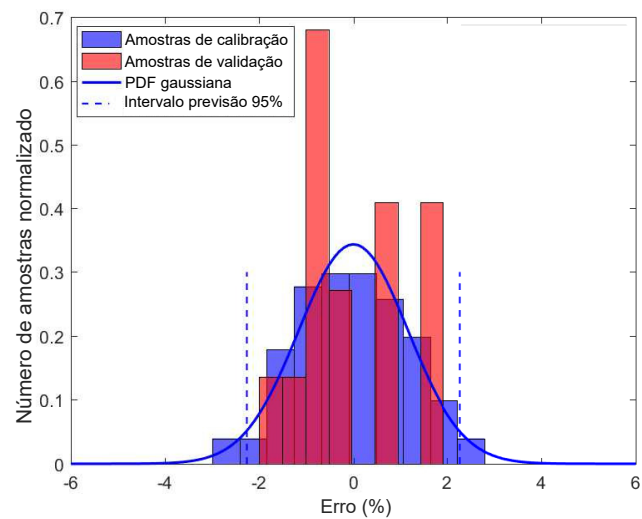
Figura 4.6: Resultados após calibração utilizando amostras exclusivamente do primeiro conjunto de dados, com pontos espectrais obtidos pelo Método E: (a) relação entre concentrações especificadas pelo método de referência e concentrações estimadas pelo modelo; (b) resíduos de calibração e erros de predição; (c) histogramas dos resíduos de calibração e dos erros de predição.



(a)



(b)



(c)

Figura 4.7: Resultados após calibração utilizando amostras distribuídas entre os quatro conjuntos, com pontos espectrais obtidos pelo Método E: (a) relação entre concentrações especificadas pelo método de referência e concentrações estimadas pelo modelo; (b) resíduos de calibração e erros de predição; (c) histogramas dos resíduos de calibração e dos erros de predição.

foi diferente.

Por fim, é importante lembrar que as gasolinas utilizadas nos quatro dias foram adquiridas em um mesmo posto de combustíveis. Assim, é esperado que a variabilidade amostral e sua influência, na prática, sejam ainda mais significativas quando considerados fornecedores distintos, ou para outros tipos de aplicações. Os resultados discutidos nas Seções 4.2.1 e 4.2.2 indicam a importância do planejamento no processo de preparação e seleção de amostras em laboratório, as quais devem ser feitas com distribuição uniforme em relação tanto à faixa de concentrações do analito de interesse, quanto às demais fontes de variância. Isto significa que qualquer amostra deve ter a mesma probabilidade de ser selecionada. Caso contrário, será mais provável que apenas uma porção relativamente pequena da população seja selecionada. Neste caso, muitas amostras seriam espectralmente similares, o que poderia inserir uma tendência na curva de calibração.

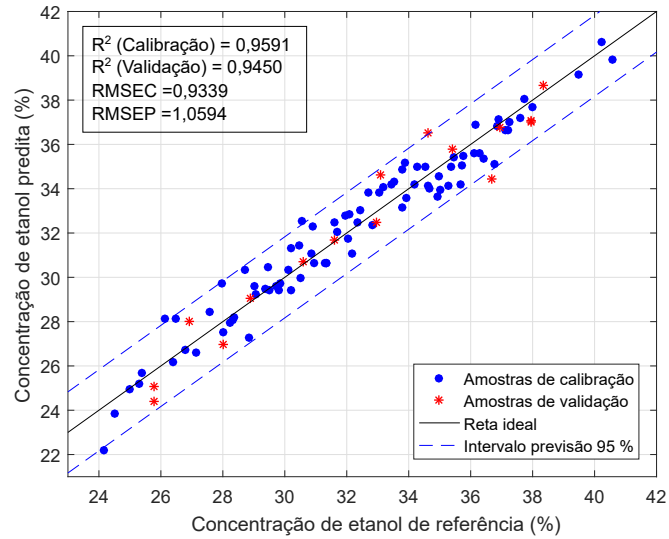
4.2.3 Comparação com o espectrofotômetro de bancada

Os resultados obtidos pelo protótipo foram comparados com aqueles fornecidos pelo espectrofotômetro de bancada, em termos dos erros de estimação e dos coeficientes de determinação obtidos. Nesta comparação, apenas o procedimento de seleção de amostras distribuídas foi considerado.

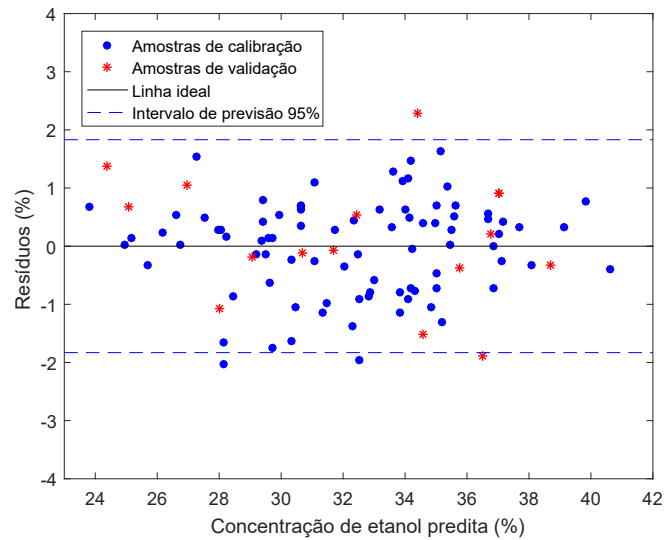
A Figura 4.8 mostra os resultados obtidos após a calibração do espectrofotômetro de bancada. Observa-se na Figura 4.8(a) que os pontos estão menos espalhados em torno da reta ideal em comparação com o protótipo, como também indicado pelo menor RMSEC (0,94 % v/v). Os coeficientes de determinação obtidos para as amostras de calibração e validação foram, respectivamente, 0,9591 e 0,9450. Por sua vez, ambos os erros de calibração e validação estão bem distribuídos em torno de zero, como mostrado na Figura 4.8(b), e consistentes com a distribuição gaussiana estimada, como mostra a Figura 4.8(c).

Comparando-se os resultados do equipamento comercial (Figura 4.8(a)) com aqueles do protótipo para o método E (Figura 4.7(a)), pode-se verificar que a diferença entre os valores de RMSEC correspondem a apenas 0,23% v/v de etanol, e as diferenças entre os RMSEP foi de cerca de 0,02% v/v de etanol. Assim, estima-se que, em média, a diferença entre os erros médios nas medições realizadas pelo protótipo e pelo espectrofotômetro comercial correspondam a menos de 0,3% v/v na concentração de etanol. Deve-se considerar, porém, que a referência $I_0(\lambda)$ foi medida no espectrofotômetro comercial para uma cubeta vazia, diferentemente do protótipo no Método E. Nesse sentido, pode-se comparar mais adequadamente os resultados do equipamento comercial com aqueles mostrados na Figura 4.5(a), obtidos a partir dos dados calculados pelo Método B. Neste caso, pode-se verificar que a diferença foi de 0,76 % v/v para as amostras de calibração, e de 0,49 % v/v para as amostras de validação. Assim, estima-se que o erro médio na quantificação do etanol pelo protótipo, em módulo, foi maior do que o espectrofotômetro comercial em menos de 0,8% v/v de etanol.

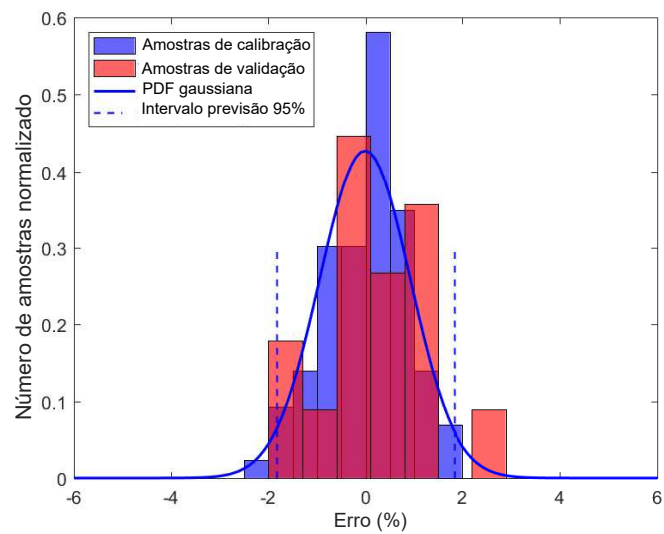
A partir dos resultados obtidos neste experimento, conclui-se que a prepara-



(a)



(b)



(c)

Figura 4.8: Resultados obtidos pelo espectrofotômetro de bancada após calibração utilizando amostras distribuídas entre os quatro conjuntos: (a) relação entre concentrações especificadas pelo método de referência e concentrações estimadas pelo modelo; (b) resíduos de calibração e erros de predição; (c) histogramas dos resíduos de calibração e dos erros de predição.

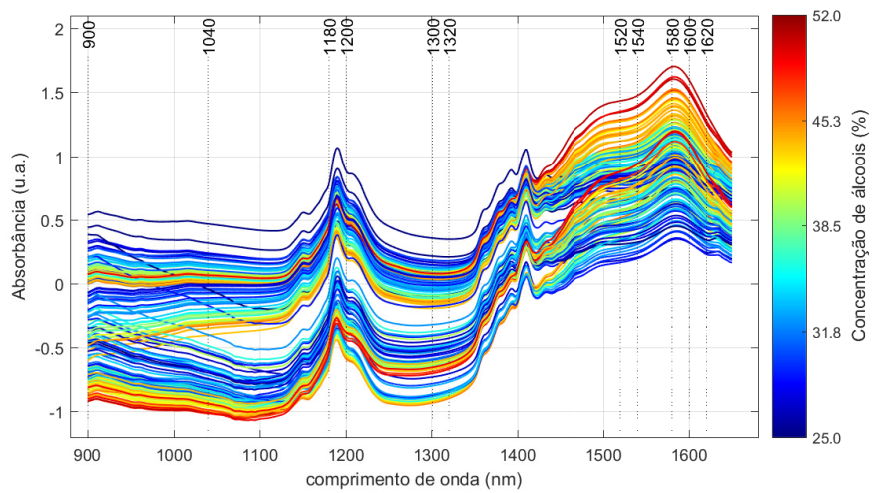
ção e seleção de amostras; e o método primário para determinação das concentrações de referência são fatores que tem um papel importante no processo de calibração e, conseqüentemente, na capacidade de predição do modelo analítico. Observa-se que a influência desses fatores pode ser mais significativa do que a exatidão nas medições dos dados ópticos em si pelo sistema de instrumentação. Provavelmente, resultados similares seriam obtidos para amostras diluídas com metanol. A relativa proximidade dos resultados obtidos por ambos os instrumentos mostra que este tipo de solução é promissora. Contudo, os resultados revelam que, mesmo para gasolinas adquiridas de um mesmo posto revendedor, provenientes de um mesmo distribuidor, existe uma variabilidade de matriz muito significativa. Isto evidencia a necessidade de empregar, nas próximas etapas do trabalho, uma quantidade de amostras suficiente para cobrir uma variabilidade amostral mais representativa de uma situação prática de utilização do instrumento. Por isso, nos experimentos subsequentes, foram empregadas gasolinas aditivadas e não-aditivadas, provenientes de seis distribuidores e adquiridas em dezoito postos revendedores distintos. Isto cobre uma variabilidade amostral muito superior à utilizada em todos os trabalhos identificados na literatura, relativos ao tema.

Na próxima seção, são apresentados os resultados do estudo espectral que foi realizado para identificar as variáveis analiticamente mais importantes para a quantificação simultânea dos dois álcoois em misturas ternárias.

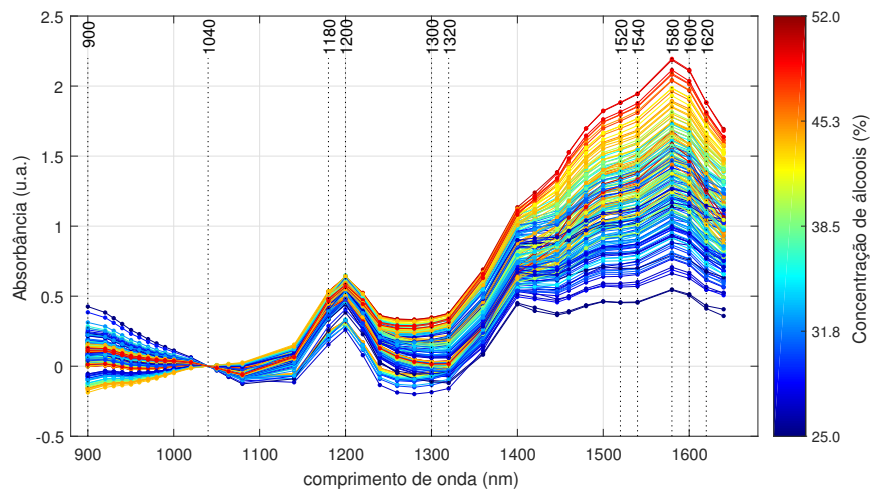
4.3 RESULTADOS DO ESTUDO ESPECTRAL PARA SELEÇÃO DE VARIÁVEIS

A Figura 4.9 apresenta os espectros das 170 amostras utilizadas na etapa de seleção de variáveis, obtidos por meio do espectrofotômetro de bancada. Na Figura 4.9(a), são apresentados os espectros originais, com 2.194 pontos espectrais. Por sua vez, a Figura 4.9(b) apresenta os espectros reduzidos após o pré-processamento espectral descrito na Seção 3.3.1.3. Nesta figura, cada espectro possui 37 variáveis espectrais correspondentes aos comprimentos de onda centrais dos filtros de interferência ópticos disponíveis comercialmente. Também, é apresentada uma escala de cores em função da concentração total dos álcoois (soma das concentrações de etanol e metanol). Por fim, a Figura 4.9(c) apresenta os espectros dos álcoois puros.

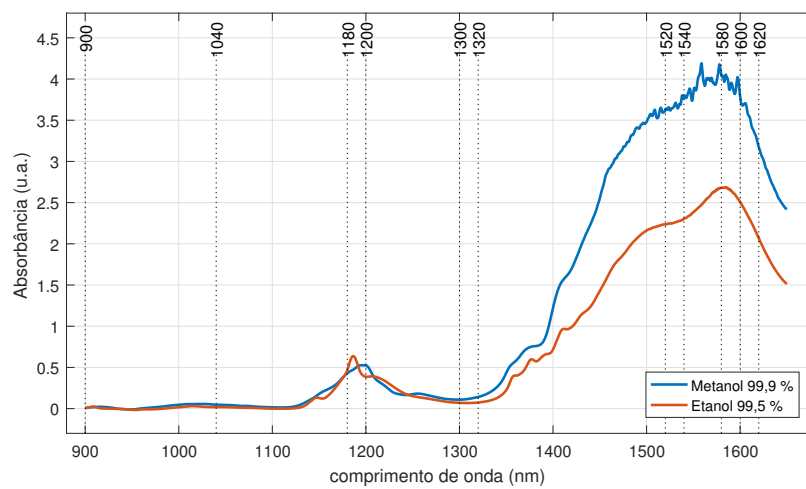
Após a aquisição dos espectros, os dados mostrados na Figura 4.9(b) foram utilizados para a seleção de variáveis descrita na Seção 3.3.1.4. Na Tabela 4.5, são apresentados os comprimentos de onda selecionados pelo *algoritmo de seleção de variáveis*, para valores de k entre três e dez. Em cada linha da tabela, os comprimentos de onda são ordenados de acordo com a frequência de seleção em todas as iterações do processo de validação cruzada. Os dados da tabela mostram que, a medida que se aumenta o número desejado de variáveis, uma nova variável é acrescentada, mantendo aqueles comprimentos de onda que já haviam sido selecionados anteriormente. Este novo comprimento de onda é destacado em negrito. Isto mostra que a seleção de variáveis é estável, isto é, aumentar o valor de k não resulta na seleção de novas variáveis de forma desordenada. As dez variáveis mais informativas selecionadas pelo



(a)



(b)



(c)

Figura 4.9: (a) Espectros originais das 170 amostras, com resolução de 8 cm^{-1} ; (b) espectros originais com número de variáveis reduzido para 37 pontos espectrais coincidentes com os comprimentos de onda dos filtros disponíveis comercialmente; (c) Espectros dos álcoois puros.

algoritmo, juntamente com o comprimento de onda utilizado para a correção de linha de base, também estão destacadas pelas linhas verticais nos espectros das Figuras 4.9(a), 4.9(b) e 4.9(c).

É interessante observar que os três primeiros comprimentos de onda selecionados (1580 nm, 1200nm e 1180nm) coincidem com regiões espectrais onde existem diferenças visíveis nos espectros dos álcoois. Pode-se observar na Figura 4.9(c) que existe um pico de absorbância no espectro do etanol próximo a 1180 nm, e outro no espectro do metanol em 1200 nm. Por sua vez, o comprimento de onda 1580 nm coincide com a região espectral de maiores absorbâncias dentro da faixa estudada, a qual está associada ao primeiro sobretom do estiramento das ligações O-H dos dois álcoois (Fernandes et al., 2008). Estas observações evidenciam que não seria viável utilizar somente LEDs centrados em diferentes comprimentos de onda, como no protótipo de Moreira (2013), utilizado na primeira etapa do estudo, e nos protótipos de Chagas (2006) e Nascimento (2008). A determinação analítica dos dois álcoois em misturas ternárias requer informações sobre regiões espectrais de banda estreita que permitam diferenciar os dois compostos. Neste caso, as bandas de emissão dos LEDs são relativamente largas.

Após a identificação dos comprimentos de onda mais informativos, foi realizada a avaliação dos modelos construídos a partir dos comprimentos de onda selecionados, descrita na Seção 3.3.2. A Figura 4.10 mostra os valores das estatísticas RMSEV e RMSEP, definidas em (3.20) e (3.21), obtidas em função do número de comprimentos de onda selecionados.

A Figura 4.10 mostra que a estatística RMSEV é decrescente em função do número de comprimentos de onda selecionados. Isto indica que a performance conjunta dos modelos na predição das amostras de validação melhoram à medida que novos comprimentos de onda são acrescentados. De fato, os comprimentos de onda são selecionados pelo algoritmo desenvolvido de forma a minimizar os valores de RMSEV resultantes, conforme descrito na Seção 3.3.1.4. Apesar disso, observa-se uma tendência de estabilização para valores de k acima de 6 ou 7 variáveis. Isto indica que utilizar um número excessivo de comprimentos de onda

Tabela 4.5: Comprimentos de onda mais informativos selecionados pelo *algoritmo de seleção de variáveis*.

Número de variáveis (k)	Variáveis espectrais selecionadas (nm)									
3	1580	1200	1180							
4	1580	1600	1200	1180						
5	1580	1600	1200	1180	900					
6	1580	1600	1200	1180	900	1540				
7	1600	1200	900	1540	1180	1580	1320			
8	1600	1200	1540	1580	900	1320	1180	1520		
9	1600	1200	1540	1580	900	1320	1520	1300	1180	
10	1600	1200	1540	1580	900	1520	1180	1320	1620	1300

tornaria o instrumento mais caro e complexo, sem necessariamente contribuir para melhorar a qualidade dos modelos de forma significativa.

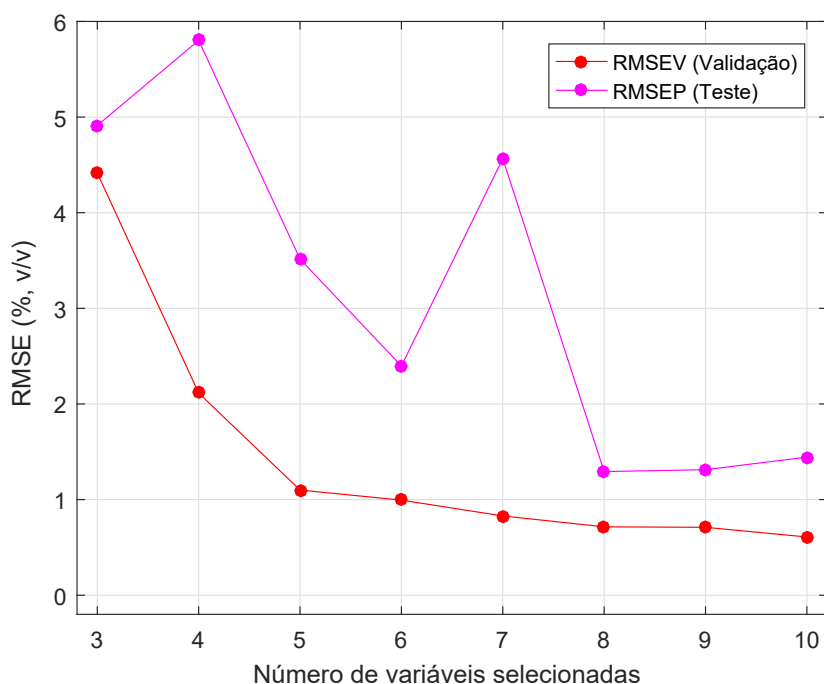
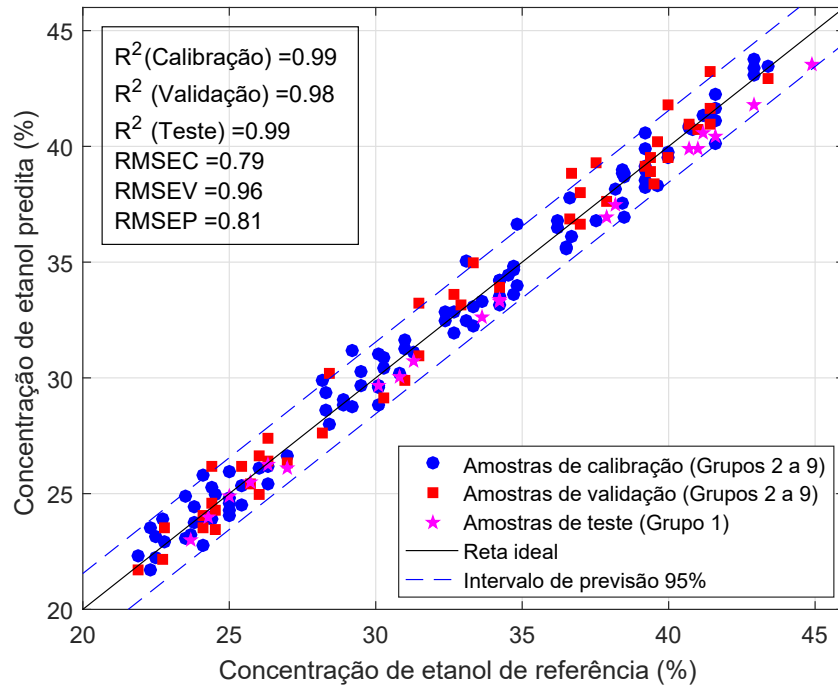


Figura 4.10: Estatísticas RMSEV e RMSEP conjuntas para ambos os modelos matemáticos de medição, em função do número de variáveis espectrais selecionadas.

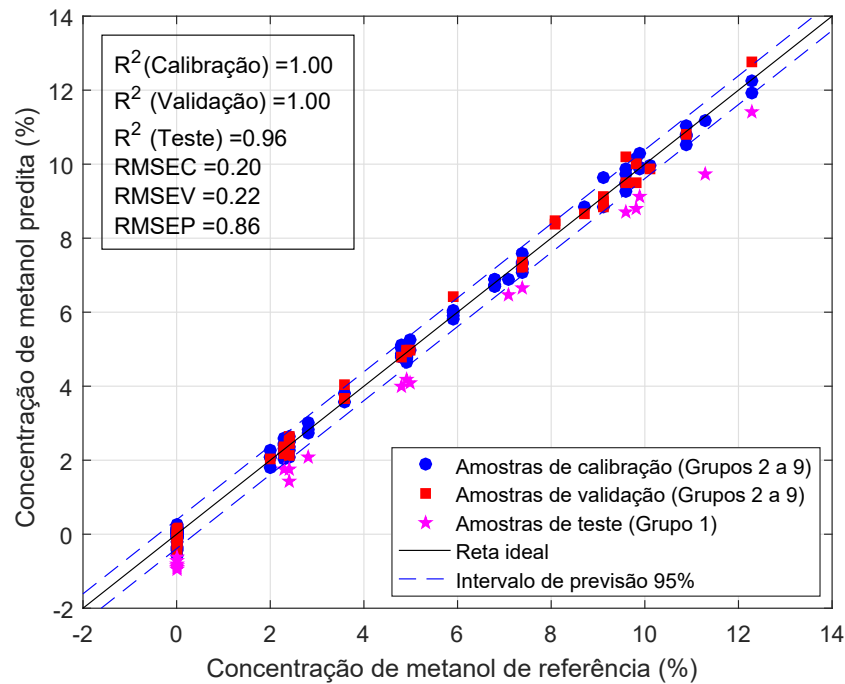
Por outro lado, quando os modelos são avaliados na predição das amostras de teste, que possuem matrizes distintas daquelas empregadas na calibração, a relação entre a estatística RMSEP e o número de variáveis não é tão evidente. Isto ocorre porque, a cada iteração do processo de validação cruzada, as gasolinas de um determinado grupo são separadas como conjunto de teste, enquanto as amostras dos demais grupos são empregadas na determinação dos coeficientes de regressão dos modelos. Dessa forma, as características espectrais de um conjunto de teste em particular podem ser significativamente distintas em relação aos demais grupos. Dessa forma, uma nova variável pode contribuir negativamente na predição deste grupo de teste em particular, elevando o valor da estatística RMSEP.

A Figura 4.11 mostra, como exemplo, os resultados de uma única iteração do *algoritmo de avaliação dos modelos candidatos*, para $k = 7$ variáveis. Nesta iteração, as amostras do Grupo 1 são utilizadas como grupo de teste, enquanto as amostras dos Grupos 2 a 9 são empregadas na calibração e validação dos modelos. As Figuras 4.11(b) e 4.11(a) mostram os gráficos de concentrações especificadas versus concentrações preditas pelos modelos MLR para estimação do metanol e do etanol, respectivamente. Para estes grupos de amostras especificamente, observa-se nos três modelos que as amostras de calibração e validação se distribuem de maneira aproximadamente uniforme em torno da reta ideal. Por outro lado, as amostras de teste apresentam uma tendência de deslocamento, especialmente para o metanol.

Apesar do exposto, é possível observar na Figura 4.10 uma tendência geral



(a)



(b)

Figura 4.11: Concentrações especificadas versus concentrações preditas para a primeira iteração do algoritmo de seleção de variáveis, para modelos MLR com sete variáveis espectrais: (a) para o etanol; (b) para o metanol.

de queda também da estatística RMSEP em função de k . Acredita-se que, a medida que uma maior variedade de gasolinas é empregada na calibração, mais clara deve se tornar a tendência de redução da estatística RMSEP em função do número de variáveis.

Os resultados apresentados comprovam que existem diferenças espectrais sig-

nificativas entre gasolinas, e que diferentes grupos de calibração podem originar modelos com coeficientes de regressão significativamente distintos. Isto reforça a necessidade de se utilizar a maior variedade possível de gasolinas na calibração. Os resultados demonstram que melhorar a qualidade das amostras é mais significativo para a qualidade do modelo do que aumentar o número de variáveis. Por fim, os resultados da Figura 4.10 também mostram que os erros de estimação para amostras futuras foram sempre maiores do que os obtidos para as amostras de validação, para qualquer número de variáveis. Isto mostra que é importante avaliar os modelos finais a partir de amostras com matrizes espectralmente distintas daquelas utilizadas na calibração. Dessa forma, é possível obter uma estimativa mais real da performance dos modelos, diferentemente do que foi realizado por autores como Correia et al. (2018); Lutz et al. (2014) e Ouyang e Liu (2013).

A partir dos resultados apresentados, optou-se por utilizar um conjunto de variáveis que geram valores de RMSEV inferiores a 1 % v/v para a construção do protótipo. Neste caso, selecionou-se o modelo com sete variáveis espectrais (RMSEV= 0,83 % v/v) para ser posteriormente implementado no protótipo. Consequentemente, o protótipo desenvolvido emprega oito filtros de interferência ópticos. Além dos dois modelos MLR para estimar as concentrações de etanol e metanol, avaliou-se um terceiro modelo para estimar a concentração total de ambos os álcoois, a partir dos mesmos comprimentos de onda. A construção do protótipo é apresentada no próximo capítulo. Os testes realizados para testar o protótipo e os modelos matemáticos de medição resultantes são apresentados no Capítulo 4.5.

4.4 MONTAGEM DO PROTÓTIPO E IMPLEMENTAÇÃO DO MODELO DE SETE VARIÁVEIS ESPECTRAIS

A Figura 4.12 apresenta fotos do protótipo implementado. A Figura 4.12(a) mostra a roda de sustentação dos filtros impressa em 3D, juntamente com os oito filtros de interferência. Por sua vez, a Figura 4.12(b) mostra o protótipo em fase de montagem, onde é possível observar em detalhe os LED e a lente condensadora. Os circuitos eletrônicos do protótipo foram implementados em três placas eletrônicas. Estes circuitos são descritos nas Seções 4.4.1 a 4.4.3. A primeira placa, mostrada na Figura 4.12(c) contempla o circuito de acionamento dos LEDs. Já a placa mostrada na Figura 4.12(d) contém o circuito de detecção e condicionamento de sinais e o circuito auxiliar para acionamento do motor de passo. Também, esta placa foi projetada para ser acoplada a uma placa de desenvolvimento comercial, que contém o microcontrolador. A placa de desenvolvimento comercial utilizada foi a *STM32F4 Discovery*, da *STMicroelectronics*, a qual é baseada no microcontrolador STM32F407VG. Por fim, também foi desenvolvida uma terceira placa para o circuito auxiliar de controle de temperatura, a qual é fixada externamente à estrutura do protótipo. As duas primeiras placas projetadas possuem circuitos de alimentação com reguladores de tensão próprios, bem como possuem circuitos conversores analógicos-digitais para digitalização dos sinais de interesse. Dessa forma, excetuando-se as linhas de alimentação, as conexões entre todas as placas e as interligações

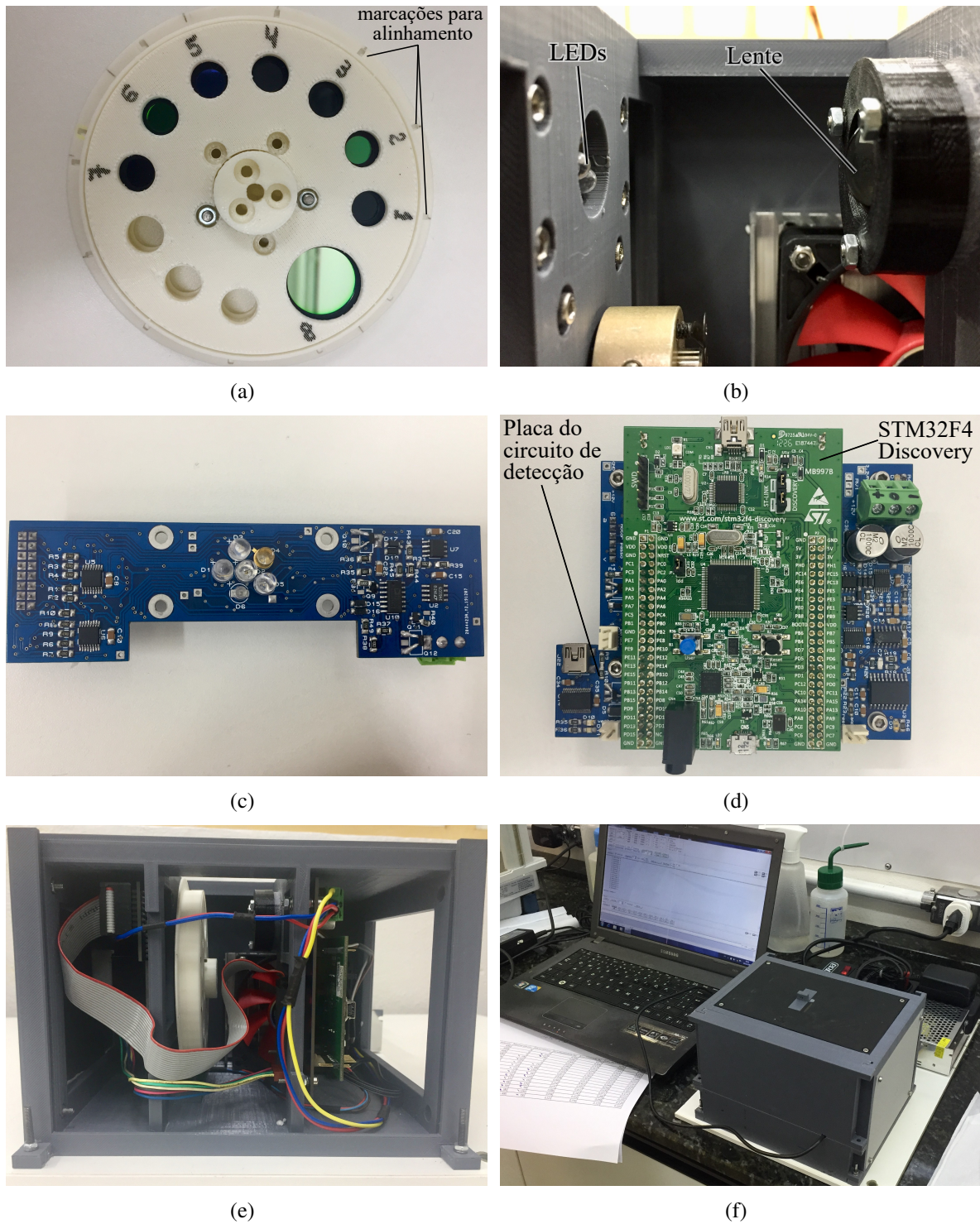


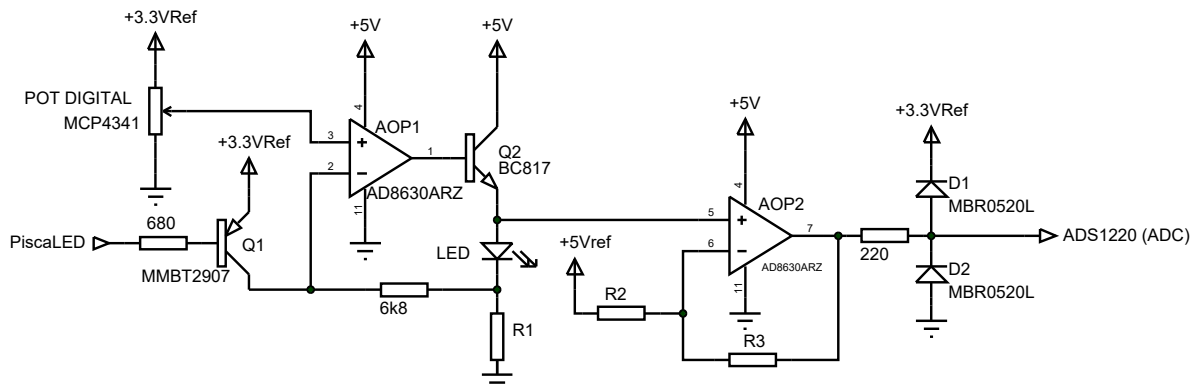
Figura 4.12: Fotos do protótipo desenvolvido: (a) roda de filtros ópticos; (b) parte do sistema óptico em fase de montagem; (c) placa de acionamento dos LEDs infravermelhos; (d) placa do circuito de deteção acoplada à placa de desenvolvimento do microcontrolador; (e) visão interna do protótipo em fase final de montagem; (f) protótipo finalizado em operação.

com o microcontrolador empregam apenas sinais digitais. A Figura 4.12(e) mostra uma visão interna do protótipo e as conexões elétricas entre as placas. Por fim, a Figura 4.12(f) mostra o protótipo final montado em operação. A interface de comunicação com o usuário é feita por meio de comunicação USB com um computador.

4.4.1 Circuito de acionamento dos LEDs

A Figura 4.13 apresenta o circuito utilizado para acionamento dos LED infravermelhos, o qual foi baseado no circuito de Ribeiro et al. (2016). Neste circuito, *PiscaLED* é um sinal digital pulsado gerado pelo microcontrolador, que tem a função de ligar o LED correspondente, quando em nível lógico alto, ou desligar o LED, quando em nível lógico baixo. Uma vez que *PiscaLED* esteja em nível lógico alto, o circuito ligado ao LED se comporta como uma fonte de corrente controlada por tensão. Neste caso, a corrente do LED é dada pela tensão na porta não-inversora do amplificador operacional AOP1, dividida pela resistência R_1 . Por outro lado, quando o sinal *PiscaLED* está em nível lógico baixo, o transistor Q2 está cortado e a corrente do LED é igual a zero. Dessa forma, o sinal *PiscaLED* pode ser utilizado para modular o sinal óptico produzido por cada LED em diferentes frequências e razões cíclicas.

Figura 4.13: Circuito de acionamento dos LEDs infravermelhos.



Fonte: o autor.

A tensão na porta inversora de AOP1 pode ser ajustada digitalmente, por meio de um potenciômetro digital, o MCP4341 da *Microchip*. Trata-se de circuito integrado com comunicação SPI, quatro canais e resolução de 7 bits, o que possibilita ajustar 129 níveis de tensão distintos. O protótipo implementado possui seis circuitos iguais ao figura, sendo um para cada LED. Contudo, um único canal do MCP4341 é responsável por controlar a tensão aplicada aos seis circuitos simultaneamente. Dessa forma, as correntes de trabalho desejadas para cada LED individualmente são estabelecidas definindo-se uma tensão de trabalho e ajustando-se os resistores R_1 de cada circuito. Por outro lado, o circuito permite aumentar ou diminuir as intensidades das correntes dos seis LEDs simultaneamente, controlando-se a posição do potenciômetro digital. Esta característica do circuito é importante pelo que segue.

Estabelecer uma corrente constante em um LED não resulta na emissão de uma potência luminosa constante por ele. Conforme observado por Pinheiro (2018) e Moreira (2013), após a energização do circuito, existe um decaimento da potência luminosa emitida pelo LED, que pode ser prontamente observada no fotodetector. Este decaimento ocorre devido ao

auto-aquecimento do LED, o qual eleva a temperatura de junção e, como consequência, altera a característica corrente \times tensão dele (Hegedüs et al., 2017; Chia-Hung Wang et al., 2010). A medida que a temperatura de junção aumenta, a tensão nos terminais do LED diminui, resultando em uma diminuição na potência elétrica dissipada, tendo em vista que a corrente é mantida constante. Consequentemente, isto gera também uma redução da potência luminosa emitida. Por esse motivo, quando o instrumento é energizado, normalmente é necessário aguardar um tempo de inicialização significativamente elevado até a estabilização térmica dos LEDs, que pode ser superior a uma hora.

Diante do exposto, o controle do potenciômetro digital permite criar uma lógica de controle em malha fechada para acelerar o processo de estabilização térmica dos LEDs. Contudo, neste caso, é necessário determinar a potência dissipada no LED para utilizá-la como realimentação. Por isso, o circuito da Figura 4.13 também permite medir a potência elétrica dissipada em cada LED. Uma vez que o LED esteja aceso, com uma corrente constante, então a tensão no terminal catodo do LED é constante. Assim, pode-se observar a variação da potência elétrica dissipada no LED ao longo do tempo medindo-se a tensão no terminal anodo do LED. Para isto, o circuito amplificador formado pelo AOP2 e pelos resistores R2 e R3, no circuito da Figura 4.13, é responsável por adequar o nível da tensão amostrada no terminal anodo do LED, quando aceso, para a faixa 0 a 3,3 V. Esta tensão é então digitalizada, por meio de um conversor analógico-digital delta-sigma de 24 bits (ADS1220, da *Texas Instruments*). A taxa de amostragem foi de 1.000 amostras por segundo. Por fim, o amplificador operacional utilizado foi o AD8630ARZ, da *Analog Devices*. Trata-se de um amplificador operacional quádruplo, com alimentação assimétrica de 2,7 a 5 V, *rail-to-tail*, CMRR de 130 dB e *drift* de $0,002 \mu\text{V}/^\circ\text{C}$.

A lógica de controle responsável por acelerar a estabilização térmica dos LEDs foi implementada como segue. Primeiramente, um ensaio é realizado em malha aberta para identificar a tensão de estabilização de cada LED, para a corrente de trabalho desejada e para modulação com sinal quadrado (com razão cíclica de 50 %). Posteriormente, a tensão de estabilização é utilizada como *setpoint* para um controlador PID, e um sinal *PiscaLED* de modulação PMW é utilizado como sinal de controle. Após cada energização do circuito, ajusta-se o potenciômetro digital para aumentar as correntes dos LEDs em relação à corrente de trabalho desejada, dentro dos limites máximos especificados para cada LED. Isto é feito para aumentar a dissipação de potência, acelerando assim o processo de aquecimento do LED. O controle atua então ajustando a razão cíclica do sinal *PiscaLED* a fim de estabilizar a tensão do LED no ponto de operação desejado. Por fim, quando os seis LEDs atingem a estabilidade térmica, as correntes de trabalho são restabelecidas e o circuito passa a operar em malha aberta, com razões cíclicas de 50%. O controlador implementado atua apenas na inicialização do sistema porque o circuito de detecção implementado, apresentado na Seção 4.4.2, requer que os sinais *PiscaLED* tenham razões cíclicas constantes. Por meio deste método, foi possível reduzir significativamente o tempo de inicialização do protótipo para cerca de dez minutos.

Após alguns testes preliminares, os valores dos componentes no circuito da Figura 4.13 foram estabelecidos como segue. Na condição de trabalho, a tensão na porta in-

versora de AOP1 foi ajustada por meio do potenciômetro digital em 2,423 V. O resistor R1 foi ajustado em 47Ω , para os LEDs de 910, 1200, 1300 e 1550 nm; em 180Ω , para o LED de 1.040 nm; e em 39Ω , para o LED de 1.650 nm. Estes valores resultam em correntes de aproximadamente 52 mA, 13,5 mA e 62 mA, respectivamente, para cada caso mencionado anteriormente. Por sua vez, durante a inicialização, o potenciômetro digital é ajustado para gerar uma tensão de 3,094 V, de forma que as correntes são, aproximadamente, 66 mA, 17 mA e 79 mA, respectivamente. Dessa forma, as correntes de trabalho dos seis LEDs estão próximas às correntes recomendadas pelo fabricante, e as correntes de inicialização estão dentro dos limites máximos estabelecidos.

Devido ao problema já discutido da estabilidade térmica dos LEDs, é importante que o microcontrolador mantenha o acionamento de todos os LEDs constantemente durante a operação do equipamento. Contudo, o circuito de detecção utilizado é síncrono: ele requer um sinal de entrada (denotado por *RefSinc*) que deve ser de mesma fase e frequência do sinal de modulação *PiscaLED* do LED que se deseja medir. Este circuito é apresentado na próxima seção. Por esse motivo, a modulação dos LEDs pelo microcontrolador é feita utilizando duas frequências distintas: 20 Hz e 10 Hz. A primeira corresponde à frequência do sinal *RefSinc*. Dessa forma, durante uma varredura espectral, cada LED em interesse é pulsado por vez na frequência de 20 Hz, enquanto os demais LEDs são pulsados em 10 Hz. Assim, apenas o sinal referente ao LED em interesse é detectado. Foram escolhidas duas frequências relativamente baixas com o objetivo de tornar menos significativos os efeitos do chaveamento e assim melhorar a estabilidade do sinal detectado. O diagrama esquemático completo do circuito de acionamento dos LEDs é apresentado no Apêndice C. Mais informações sobre os componentes utilizados estão disponíveis nas folhas de dados do Anexo C.

4.4.2 Circuito de detecção e condicionamento de sinais

O circuito de detecção e condicionamento de sinais utilizado neste trabalho é semelhante ao utilizado por Pinheiro et al. (2019). Este circuito, mostrado nas Figuras 4.14 e 4.15, tem a função de gerar uma tensão de saída proporcional à intensidade de luz que incide no fotodiodo proveniente do LED em interesse, após interagir com a amostra. Ou seja, o circuito deve detectar o sinal luminoso que tenha mesma frequência e fase do sinal *RefSinc*. Isto é feito como segue.

O circuito da Figura 4.14 é um circuito de transimpedância com supressor de resposta contínua. O circuito converte a fotocorrente (I_p) gerada pelo sinal óptico incidente no fotodiodo em um sinal de tensão. Entretanto, a resposta gerada pela luz pulsada do LED, a resposta estática da luz ambiente, bem como a corrente de escuro do fotodiodo, resultam em um sinal de tensão com componente contínua. Neste caso, o circuito de realimentação destacado na Figura 4.14 é responsável por suprimir a resposta de baixa frequência, mantendo a linearidade do sinal (Johnson, 2003). O circuito de realimentação ativo é um filtro passa-baixas

com frequência de corte

$$f_{c1} = \frac{1}{2\pi R_2 C_2}. \quad (4.5)$$

Por sua vez, suprimindo-se o capacitor C_1 , o circuito da Figura 4.14 como um todo possuirá resposta passa-altas, com frequência de corte

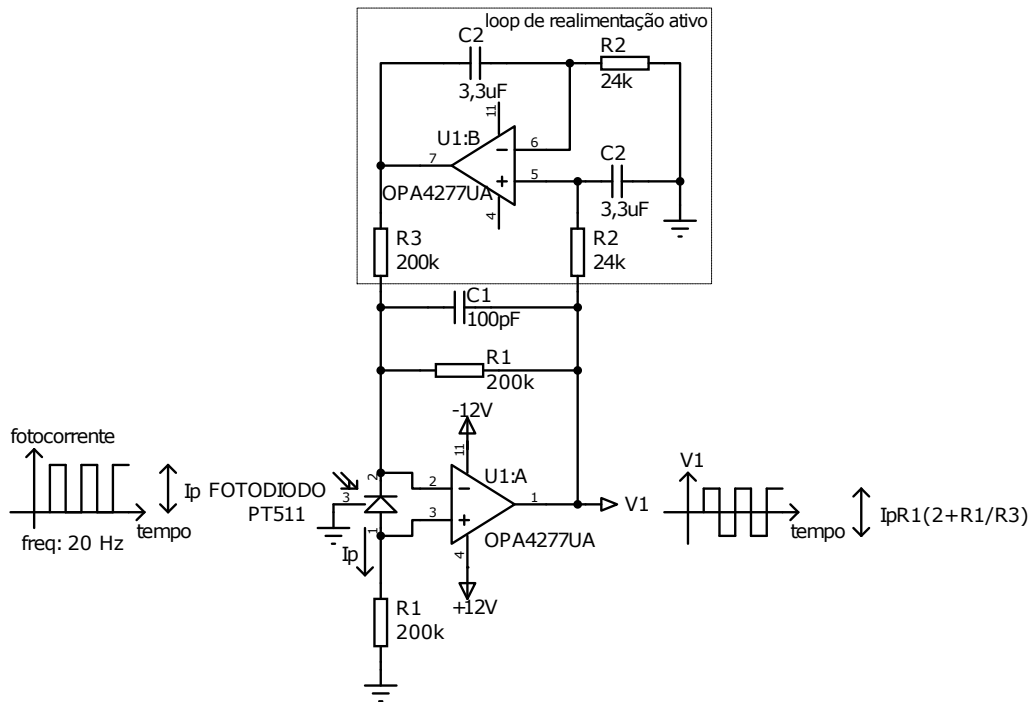
$$f_{c2} = \frac{R_1}{2\pi R_2 C_2 R_3}. \quad (4.6)$$

Em frequências muito maiores do que f_{c2} , a transimpedância é igual a $V_1/I_p = R_1 \left(2 + \frac{R_1}{R_3}\right)$. Assim, para os valores adotados indicados na Figura 4.14, tem-se uma frequência de corte f_{c2} de aproximadamente 2 Hz, e uma transimpedância $V_1/I_p = 3R_1 = 6 \times 10^5$ V/A. Por fim, o capacitor C_1 tem a função de reduzir a transimpedância em altas frequências (Johnson, 2003), o que contribui para a estabilidade da tensão de saída do amplificador circuito e evita oscilações nos componentes de altas frequências. Neste caso, C_1 atua como filtro para frequências acima de $f_{c3} = 1/(2\pi R_1 C_1)$. No circuito implementado no protótipo, a frequência de corte f_{c3} foi de aproximadamente 8 kHz. A Figura 4.14 exemplifica a forma de onda esperada para a saída deste estágio do circuito de detecção, para um único LED pulsando em 20 Hz. Neste circuito, o amplificador utilizado foi o OPA4277UA, da *Texas Instruments*, o qual possui alimentação simétrica de até ± 18 V, corrente de polarização máxima de 1 nA, CMRR de 140 dB e *drift* de $\pm 0,1 \mu\text{V}/^\circ\text{C}$.

A resposta gerada pelo primeiro estágio da Figura 4.14 resulta das componentes de fotocorrentes geradas por todos os LEDs, que piscam simultaneamente, bem como de possíveis componentes de luz do ambiente. Dessa forma, o circuito da Figura 4.15 tem a função de filtrar os componentes indesejados do sinal, detectar e amplificar apenas o sinal referente ao LED que se deseja medir (pulsado na frequência de 20 Hz). Para isto, é utilizado um amplificador *lock-in*. Este tipo de amplificador consiste basicamente de um demodulador síncrono seguido de um filtro passa-baixas. Um amplificador *lock-in* permite detectar sinais muito pequenos na presença de elevados ruídos descorrelacionados (da ordem de 100 dB), quando a frequência e a fase do sinal desejado são conhecidos. Por isso, o circuito da Figura 4.15 utiliza como entrada o sinal de referência *RefSinc*, o qual deve estar em fase com o sinal de modulação *PiscaLED* do LED que se deseja medir, ambos gerados pelo microcontrolador.

Para a implementação do amplificador *lock-in*, foi utilizado o circuito integrado AD630, da *Analog Devices*. Nesta configuração, este dispositivo atua como um demodulador síncrono balanceado. O funcionamento deste dispositivo, nesta configuração, pode ser descrito como segue. O CI AD630 amplifica o sinal de entrada (V_1) com dois ganhos fixos de ± 2 . Isto é, ocorre uma alternância entre dois amplificadores com configurações inversora e não-inversora, a qual é determinada pelo nível lógico do sinal *RefSinc*. Quando *RefSinc* está em nível lógico alto, o sinal V_1 é multiplicado +2; e quando *RefSinc* está em nível lógico baixo, o sinal V_1 é multiplicado por -2. Uma vez que *RefSinc* está em fase com o componente do sinal óptico proveniente do LED de interesse, então esta operação resulta no sinal de interesse

Figura 4.14: Diagrama esquemático do circuito de transimpedância e supressor de resposta contínua.



Fonte: o autor.

demodulado em V_2 . Contudo, os outros componentes indesejados do sinal também são amplificados para compor o sinal resultante V_2 . Apesar disso, estes componentes resultam em média zero. Dessa forma, o sinal combinado V_2 é aplicado a um filtro passa-baixas com frequência de corte próxima a zero, o que resulta no valor médio de V_2 . Dessa forma, o sinal resultante V_3 é proporcional à intensidade de luz que chega ao detector, proveniente do LED em interesse.

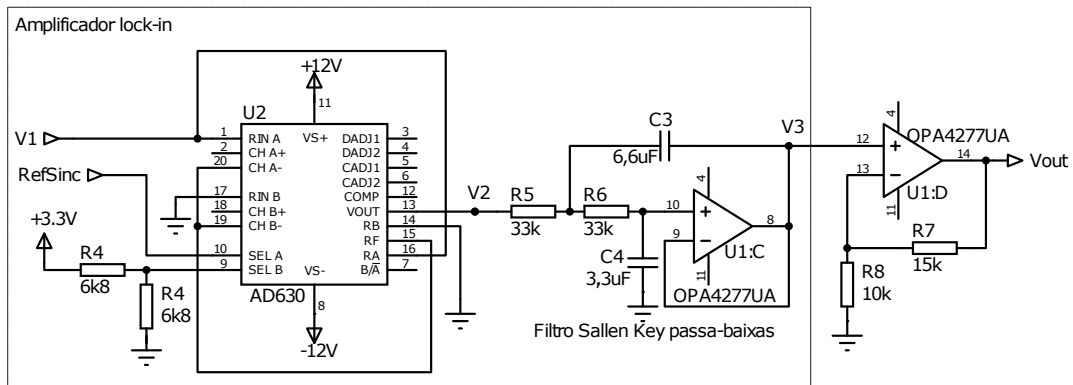
O filtro passa-baixas implementado possui topologia *Sallen Key*, com frequência de corte dada por

$$f_{c3} = \frac{1}{2\pi\sqrt{R_5 R_6 C_3 C_4}}, \quad (4.7)$$

o qual foi dimensionado para uma frequência de corte de aproximadamente 1 Hz. O sinal de saída do filtro V_3 é então amplificado, com ganho de 2,5 ajustado empiricamente, e resulta na tensão de saída V_{OUT} , a qual é finalmente digitalizada por meio de outro conversor analógico-digital ADS1220. No Apêndice C, está disponível o diagrama esquemático completo do circuito de detecção implementado.

Os componentes do circuito foram ajustados empiricamente para produzir maior variação possível, sem saturar nenhum estágio para a máxima condição de luz (até dois LEDs ativos, piscando em 20 Hz simultaneamente, e cubeta vazia).

Figura 4.15: Circuito de demodulação síncrona e amplificação.



Fonte: o autor.

4.4.3 Circuitos auxiliares

Foram implementados também dois circuitos auxiliares. O primeiro deles é responsável pela medição e controle da temperatura interna do protótipo. O controle da temperatura interna do sistema é importante para melhorar a estabilidade térmica dos LEDs, conforme já discutido, bem como dos demais componentes do sistema eletrônico. Também, as características espectrais dos componentes, tais como os espectros de emissão dos LEDs e de transmissão dos filtros, são afetados pela temperatura.

Assim, para a medição da temperatura, foi utilizado o circuito integrado LM92, da *Texas Instruments*. Este circuito possui comunicação I2C *Inter-Integrated Circuit* (I2C) e fornece uma saída digital, com resolução de $0,0625^{\circ}\text{C}$, com exatidão de $\pm 0,5^{\circ}\text{C}$ para a faixa de 10 a 50°C . Como atuador, foi utilizada uma pastilha termoelétrica de efeito Peltier, a qual é montada no sistema em conjunto com dissipadores de calor e ventoinhas. A pastilha peltier é ligada por meio de uma ponte H e modulada por PWM.

Já o segundo circuito auxiliar é responsável por acionar o motor de passo, utilizado para rotacionar a roda de filtros, e para controlar a sua posição. O motor de passo utilizado foi o modelo *Vexta PX244-03A*. Este é um motor de passo de duas fases, $1,8^{\circ}$ por passo, tensão de 12V e corrente de $0,4\text{A}$. Para controlar a posição da roda de filtros e, assim, possibilitar o alinhamento de cada filtro com o sistema óptico da Figura 3.3(a), foi utilizada uma chave óptica de barreira. Os diagramas esquemáticos detalhados dos dois circuitos auxiliares também estão disponíveis no Apêndice C.

4.4.4 Lógica de programação e medição dos dados ópticos

A lógica de funcionamento do protótipo para a medição dos dados espectrais foi feita como segue. Para cada amostra a ser analisada, são realizadas oito varreduras espectrais. Isto é, os dados ópticos correspondentes aos oito filtros são medidos sequencialmente, e são realizadas oito repetições. Os resultados das oito repetições são acumulados para gerar uma

média final. Isto é feito para reduzir a influência do posicionamento dos filtros no sistema óptico e, assim, melhorar a repetitividade das medições. Para esse número de varreduras, o tempo de cada análise é de aproximadamente 2,5 minutos. Ao final de cada análise, o protótipo fornece as tensões de saída do circuito de detecção médias digitalizadas para os oito filtros. Para o circuito de instrumentação implementado, a intensidade de luz em interesse é linearmente relacionada com a tensão de saída do circuito, de forma que as absorbâncias podem ser calculadas a partir das tensões de saída por meio da relação

$$A(\lambda) = \log \left(\frac{V_{OUT_0}(\lambda)}{V_{OUT_S}(\lambda)} \right), \quad (4.8)$$

em que $V_{OUT_0}(\lambda)$ denota as tensões de saída (que são proporcionais às intensidades de luz) de referência, medidas para uma cubeta vazia; e $V_{OUT_S}(\lambda)$ denota as tensões de saída medidas com a amostra. Este corresponde ao Método B definido na Seção 2.3. Também, um controlador de temperatura foi implementado a fim de manter a temperatura interna do protótipo em 25 °C. Durante todos os testes do instrumento, verificou-se que a variação máxima de temperatura obtida não excedeu 0,5 °C.

Na próxima seção, apresenta-se o procedimento utilizado para testar o protótipo com as amostras de gasolina. Também, os resultados obtidos pelo protótipo são apresentados e comparados com os resultados obtidos pelo espectrofotômetro de bancada.

4.5 TESTE DO PROTÓTIPO E COMPARAÇÃO COM O ESPECTROFOTÔMETRO

Para testar o protótipo desenvolvido, outras 252 amostras de gasolinas foram preparadas em laboratório, seguindo os mesmos procedimentos adotados anteriormente na etapa de seleção de variáveis, descritos na seção 3.3.1.1. Nesta nova etapa, foram adquiridas gasolinas tipo C de dezoito postos revendedores, provenientes de seis distribuidores diferentes, das quais quatro são aditivadas. As concentrações iniciais de álcool foram medidas pelo método de referência ABNT NBR 13992 e apenas uma delas apresentou teor de álcool superior à concentração de etanol determinada pela legislação vigente, com 29% v/v. Em seguida, foram realizadas as diluições pela adição de etanol e metanol, as quais resultaram em 83 misturas binárias (sendo 39 gasolina comercial/etanol e 44 gasolina comercial/metanol) e 151 misturas ternárias (gasolina comercial/metanol/etanol). As concentrações de referência foram calculadas a partir dos volumes dos álcoois adicionados. As concentrações finais obtidas estão nas faixas de 23,5 % a 44,4 % v/v, para o etanol, e de 0,0 % a 13,0 % v/v, para o metanol.

Após a preparação das amostras, os dados espectrais foram medidos no protótipo e também no espectrofotômetro de bancada. Para isto, cada amostra foi dividida em duas cubetas para serem analisadas simultaneamente em ambos os instrumentos. Os espectros gerados pelo espectrofotômetro em alta resolução foram submetidos ao processo de redução de variáveis, de forma a se obter apenas os dados referentes aos comprimentos de onda implementados no protótipo. Em seguida, realizou-se o pré-processamento (correção de linha de base)

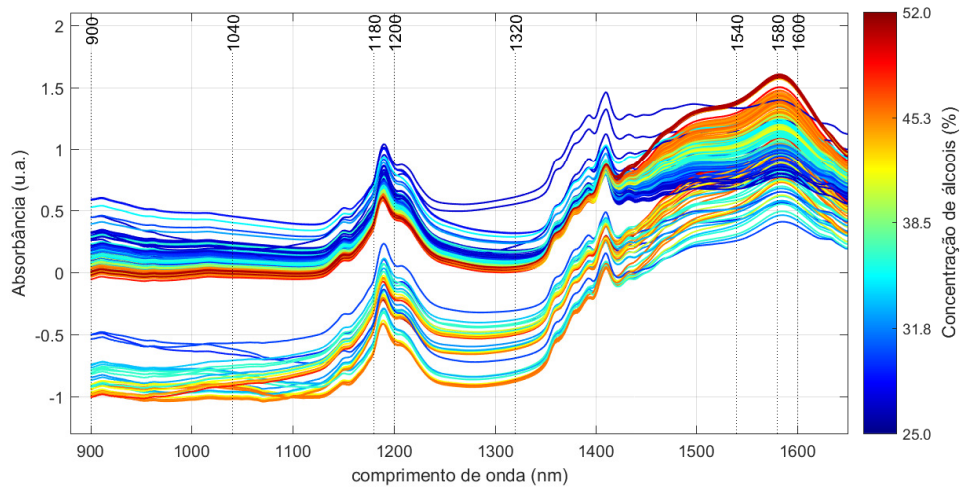
dos dados espectrais de ambos os instrumentos. Também, de forma semelhante ao procedimento adotado na seleção de variáveis, os dados espectrais cada conjunto de amostras foram modelados por MLR e avaliados separadamente, o objetivo de se identificar possíveis *outliers*. Foram identificados, no total, 15 *outliers*. As amostras correspondentes foram então eliminadas, resultando em 237 amostras válidas para a calibração e validação dos modelos matemáticos de medição.

A Figura 4.16(a) mostra os espectros originais das 237 amostras obtidos pelo espectrofotômetro de bancada. Por sua vez, a Figura 4.16(b) apresenta os dados do espectrofotômetro após a redução de variáveis e o pré-processamento espectral. Por fim, na Figura 4.16(c) são apresentados os dados obtidos pelo protótipo. A escala de cores utilizada está relacionada à concentração total dos álcoois. Comparando-se as Figuras 4.16(b) e 4.16(c), é possível observar que os perfis dos espectros obtidos pelo protótipo são diferentes daqueles obtidos pelo espectrofotômetro. Também, as amplitudes das absorbâncias nos dados do protótipo são menores, em comparação com os dados do espectrofotômetro. Acredita-se que isto se deve às diferenças nos sistemas ópticos dos instrumentos, tendo em vista que o sistema óptico implementado no protótipo é bastante simples em comparação com o espectrofotômetro de bancada. Apesar disso, é possível observar claramente nos dados do protótipo as variações nas absorbâncias em função da concentração dos álcoois, as quais são consistentes com aquelas observadas nos dados do espectrofotômetro.

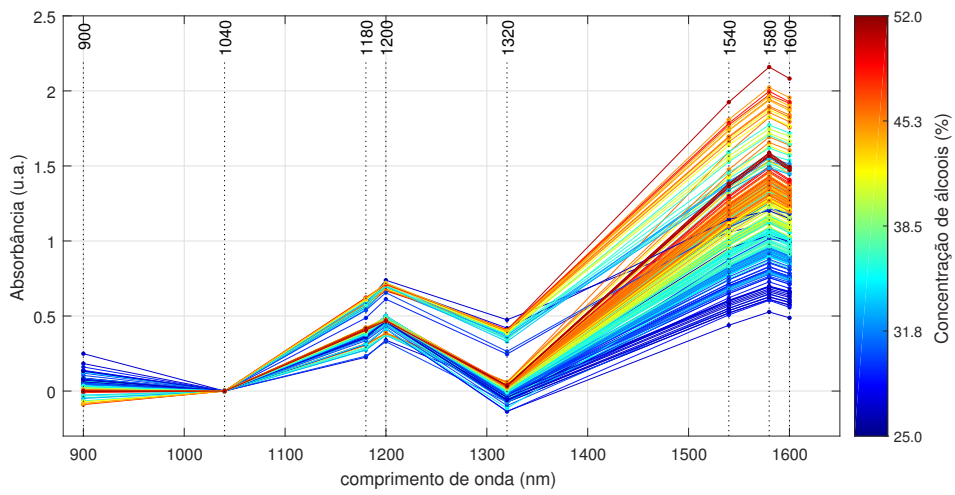
Os dados das Figuras 4.16(b) e 4.16(c) foram utilizados para calibrar e validar os modelos matemáticos de medição com sete variáveis. Para ambos os instrumentos, foram construídos três modelos distintos para as concentrações de etanol, metanol e para a concentração total de ambos os álcoois. Diferentemente do procedimento adotado na seleção de variáveis, que empregou um tipo de validação cruzada, foi realizada nesta etapa uma validação independente. Para isto, o processo de seleção das amostras foi feito como segue.

As amostras foram separadas em dezoito grupos compostos por cada gasolina inicial e suas diluições. Destes, seis grupos foram separados como grupo de teste, e as amostras dos outros doze grupos foram divididas entre um conjunto de calibração (70 %) e outro de validação (30 %) pelo *algoritmo SPXY*. A seleção dos seis grupos para teste foi feita aleatoriamente, respeitando a representatividade de cada distribuidor. Conforme já discutido na seleção de variáveis, ambos os conjuntos de validação e de teste são independentes do grupo de validação; contudo, as matrizes das amostras de teste são distintas daquelas utilizadas na calibração. Dessa forma, os resultados de predição deste grupo fornecem uma estimativa mais realista das performances de predição futura dos modelos. No total, foram empregadas 110 amostras para a calibração, 48 para validação e 79 para teste.

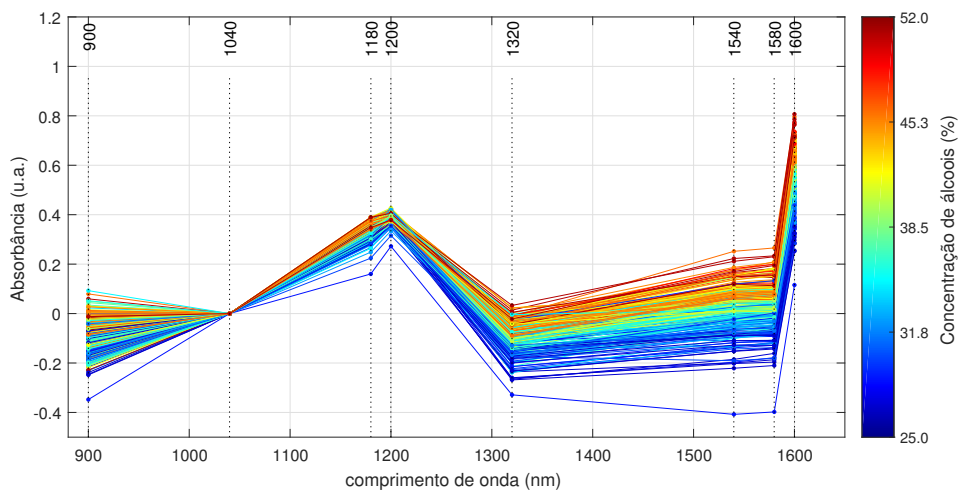
Na etapa de seleção de variáveis, foram definidas duas estatísticas RMSEV e RMSEP para avaliar as performances de predição de dois modelos (para etanol e metanol) conjuntamente, a partir de um dado conjunto de variáveis. Diferentemente, as performances de predição foram avaliadas nesta etapa para cada modelo individualmente. Neste caso, utilizou-se



(a)



(b)



(c)

Figura 4.16: (a) Espectros das 237 amostras, obtidos pelo espectrofotômetro; (b) espectros obtidos pelo espectrofotômetro após pré-processamento e redução para as oito variáveis espectrais do protótipo; (c) Espectros obtidos pelo protótipo.

a estatística raiz quadrada do erro quadrático médio para (RMSE), dada por

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}, \quad (4.9)$$

que foi aplicada aos conjuntos de validação (RMSEV) e teste (RMSEP); e a estatística RMSE aplicada ao conjunto de calibração (RMSEC), dada por

$$\text{RMSEC} = \sqrt{\frac{1}{N-8} \sum_{i=1}^N (y_i - \hat{y}_i)^2}, \quad (4.10)$$

onde y_i e \hat{y}_i denotam os valores especificados e preditos pelo modelo avaliado, respectivamente, e N é o número de amostras do conjunto avaliado.

A Figura 4.17 sintetiza os resultados finais obtidos pelos dois instrumentos. Nas Figuras 4.17(a), 4.17(c) e 4.17(e) (coluna da esquerda), são apresentados os gráficos de concentrações especificadas versus concentrações preditas pelos modelos calibrados a partir dos dados do protótipo, respectivamente para etanol, metanol e total de álcoois. Por sua vez, as Figuras 4.17(b), 4.17(d) e 4.17(f) (coluna da direita) apresentam os resultados obtidos pelos modelos calibrados com os dados do espectrofotômetro.

Pode-se observar na Figura 4.17(a) que o modelo para quantificação do etanol construído a partir dos dados do protótipo produziu pontos bem distribuídos em torno da reta ideal, para os três conjuntos (calibração, validação e teste). Os coeficientes de determinação (R^2) obtidos estão entre 0,76, para as amostras de teste, e 0,81, para as amostras de calibração. Também, a variância dos pontos em torno da reta ideal nos três conjuntos são semelhantes. Isto é, os desvios padrão amostrais dos erros de estimação, indicados pelas estatísticas RMSEC, RMSEV e RMSEP, possuem uma diferença máxima de 0,33 % v/v entre os conjuntos de calibração e de teste. Já a diferença entre RMSEV e RMSEP foi de apenas 0,22 % v/v. Estes resultados indicam que o modelo apresenta boa performance de predição de amostras futuras, uma vez que o modelo foi capaz de prever as concentrações de etanol em amostras de teste e de validação com performances equivalentes. Em comparação com os dados do espectrofotômetro de bancada, mostrados na 4.17(d), observa-se que a variância dos pontos gerados pelo protótipo é maior. A maior diferença obtida foi para as amostras de teste, com uma diferença nos valores de RMSEP de 1,72 % v/v. Este resultado é muito satisfatório, considerando as significativas diferenças construtivas entre o protótipo e o espectrofotômetro comercial.

Por sua vez, analisando-se os resultados obtidos para a estimação do metanol, observa-se que alguns parâmetros de adequação do modelo foram piores em relação ao etanol. Nos dados do protótipo mostrados na Figura 4.17(c), se uma linha de tendências for traçada, existirá uma inclinação significativa entre ela e a reta ideal, o que é indesejável. Verifica-se que os coeficientes de determinação obtidos estão entre 0,49 e 0,62. Também, como a amplitude da faixa de concentrações avaliada para o metanol (0,0 % a 13,0 % v/v) é menor em relação

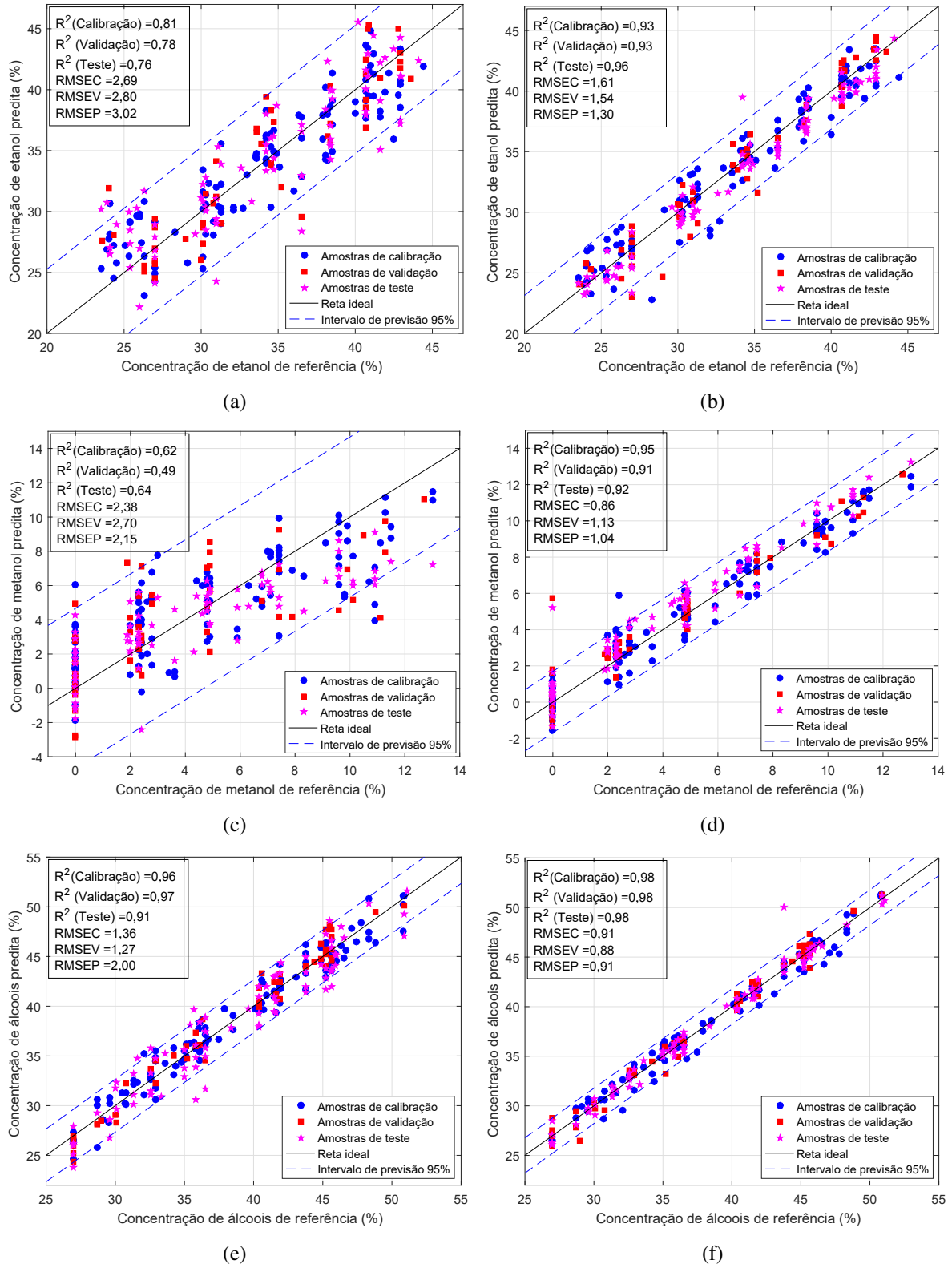


Figura 4.17: Concentrações especificadas versus concentrações preditas pelos modelos MLR com sete variáveis espectrais para concentrações de etanol ((a) e (b)), metanol ((c) e (d)) e total de álcoois ((e) e (f)). Os dados em (a), (c) e (e) são do protótipo; e os dados em (b), (d) e (f) são do espectrofotômetro.

ao etanol (23,5 % a 44,4 % v/v), então a variância dos pontos se torna mais significativa. Isto resulta em valores de RPD iguais a 1,68 para o metanol e 2,05 para o etanol. Apesar disso,

comparando-se os desvios em termos absolutos, os valores das estatísticas RMSEC, RMSEV e RMSEP obtidos pelo protótipo para o metanol foram ainda inferiores aos obtidos para o etanol. Assim como no modelo do etanol, a variância dos erros de predição das amostras de teste foram compatíveis com aquelas obtidas durante a calibração e a validação. Por fim, em comparação com os dados do espectrofotômetro de bancada, mostrados na Figura 4.17(f), a maior diferença foi obtida para o conjunto de validação, em que a estatística RMSEV produzida pelo protótipo foi 1,57 % v/v superior. Assim como no protótipo, o valor de RPD obtido pelo espectrofotômetro para o metanol (RPD = 3,46) também foram inferior ao obtido para o etanol (RPD = 4,78). Todas as demais estatísticas de avaliação obtidas pelo espectrofotômetro foram melhores para o metanol em comparação com o etanol.

Por último, as Figuras 4.17(e) e 4.17(f) mostram os resultados obtidos pelos modelos para quantificação do total de álcoois. Para esta variável de resposta, a performance de predição do protótipo foi mais próxima à do espectrofotômetro. A maior diferença obtida entre os instrumentos foi para a estatística RMSEP, com uma diferença de 1,09 % v/v. Os valores de RPD são 3,27, para o protótipo, e 7,23, para o espectrofotômetro. Observa-se que os pontos estão bem distribuídos em torno da reta ideal, para os três conjuntos de amostras, em ambos os instrumentos. Contudo, para o protótipo, a performance de predição das amostras de teste (RMSEP = 2,00 v/v, $R^2 = 0,91$) foi relativamente inferior em relação aos dois outros conjuntos de amostras. Apesar disso, em termos práticos, a performance desse modelo é muito satisfatória, considerando novamente a simplicidade do protótipo em relação ao espectrofotômetro.

Em todos os gráficos da Figura 4.17, existem algumas amostras que se destacam em cada modelo por estarem mais distantes da reta ideal e, portanto, são potenciais *outliers*. Isto significa que cada modelo individualmente poderia ser melhorado excluindo-se as respectivas amostras das análises. Apesar disso, verificou-se que esta exclusão de *outliers* não gerou uma melhora significativa nas estatísticas apresentadas. Dessa forma, optou-se por apresentar os resultados completos para que fosse possível comparar os três modelos e os dois instrumentos a partir exatamente das mesmas amostras.

Ainda, analisando-se os modelos para etanol e metanol gerados pelo espectrofotômetro, observa-se que os valores da estatística RMSEV foram 1,54 % v/v e 1,13 % v/v, respectivamente. Neste caso, a estatística RMSEV se refere a cada modelo individualmente, conforme definido em (4.9). Contudo, a performance de predição dos dois modelos pode ser avaliada em conjunto, assim como realizado na etapa de seleção de variáveis. Neste caso, a estatística RMSEV conjunta é igual a $\sqrt{1,54^2 + 1,13^2} = 1,91$ % v/v. É importante lembrar que os comprimentos de onda implementados no protótipo foram selecionados no estudo espectral porque geraram um modelo com RMSEV inferior a 1 % v/v. Ou seja, a performance de predição conjunta dos dois modelos apresentados nas Figuras 4.17(b) e 4.17(d) foi significativamente inferior àquela obtida pelo próprio espectrofotômetro na etapa de seleção de variáveis. A mesma análise pode ser estendida para as amostras de teste. Existem duas diferenças fundamentais entre os dois experimentos: a primeira é que, na seleção de variáveis, realizou-se uma validação cruzada, enquanto os resultados deste capítulo se referem a uma validação indepen-

dente; e a segunda diferença se refere à quantidade e à variedade de amostras utilizadas nos dois experimentos. Esta observação evidencia que uma validação cruzada, como a empregada por Correia et al. (2018), não fornece uma estimativa confiável da real capacidade de predição de um modelo. Uma validação independente a partir de uma quantidade e variedade razoáveis de amostras é essencial para que seja possível avaliar a real aplicabilidade prática da solução na predição de amostras futuras.

5 CONCLUSÃO

Neste trabalho, foi apresentado um estudo que avaliou, em termos práticos, a viabilidade de se utilizar componentes espectrais discretos do NIR na determinação analítica de etanol e metanol em gasolinas comerciais. A quantificação simultânea dos dois álcoois em misturas ternárias, por meio de comprimentos de onda discretos, é uma contribuição deste trabalho, já que foram identificados na literatura apenas trabalhos que consideram a quantificação de etanol ou metanol separadamente, em misturas binárias. O principal desafio do trabalho foi identificar os comprimentos de onda adequados para diferenciar duas substâncias que possuem características espectrais muito semelhantes, considerando a real variabilidade de matriz. Para isto, foram realizados experimentos empregando um número de gasolinas suficiente para se obter uma amostragem representativa da variabilidade de matriz real. Foram utilizadas, em todas as etapas da pesquisa, no total, 547 amostras entre binárias e ternárias. Isto confere ao trabalho uma maior qualidade na análise de dados em relação a todos os trabalhos identificados na literatura relacionados ao tema.

O trabalho avaliou experimentalmente sete diferentes métodos de cálculo de absorvâncias existentes na literatura. Os resultados mostraram que o Método E resulta em modelos com melhores performances preditivas, porém não é viável para a aplicação proposta, uma vez que amostras de referência necessárias para a medição de $I_0(\lambda)$ não estão disponíveis na prática. Considerando os resultados dos demais métodos e os aspectos práticos de utilização do instrumento, o Método B foi considerado o mais adequado. Também, discutiu-se conceitos importantes sobre os métodos de pré-processamento de dados espectrais e verificou-se experimentalmente a influência de matriz na gasolina. Os dados experimentais revelaram que existe variação de matriz significativa mesmo entre gasolinas de um mesmo distribuidor, adquiridas em um único posto revendedor e em um curto intervalo de tempo.

No desenvolvido do trabalho, observou-se que a determinação analítica dos dois álcoois em misturas ternárias não seria viável empregando-se apenas LEDs centrados em diferentes comprimentos de onda. As bandas de emissão dos LEDs são relativamente largas. Devido à semelhança espectral dos dois álcoois, foi observada a necessidade de selecionar regiões espectrais de banda estreita que permitam diferenciar os dois compostos. Dados medidos por um espectrofotômetro de bancada foram empregados em um estudo espectral que identificou as dez variáveis mais informativas para a determinação analítica estudada. Então, modelos empregando diferentes variáveis espectrais foram construídos e avaliados por meio de um método de validação cruzada. Mostrou-se que é viável uma solução utilizando LEDs e filtros de interferência ópticos, empregando poucas variáveis espectrais, entre sete e dez.

Posteriormente, projetou-se um protótipo de um instrumento baseado em LEDs e filtros ópticos de interferência para implementar modelos de matemáticos de medição contendo sete variáveis espectrais, e a capacidade preditiva do protótipo foi comparada com a do

espectrofotômetro de bancada. Para esta avaliação, foi proposto um método de validação independente que fornece uma melhor estimativa da real capacidade preditiva futura dos modelos avaliados. Este método avaliou os modelos por meio de dois conjuntos de amostras independentes: um conjunto de validação contendo amostras derivadas do conjunto de calibração; e outro conjunto de teste contendo amostras cujas matrizes não estão representadas no conjunto de calibração.

Os resultados experimentais obtidos pelo protótipo desenvolvido mostraram que modelos com sete variáveis espectrais produziram, de forma geral, performances equivalentes na predição dos conjuntos de validação e de teste, o que demonstra a sua aplicabilidade prática. Estes modelos resultaram em erros preditivos estimados $RMSEP = 3,02 \% v/v$, para o etanol, e $RMSEP = 2,15 \% v/v$, para o metanol, na predição das amostras de teste. Os resultados gerados pelo espectrofotômetro empregando as mesmas regiões espectrais do protótipo foram $RMSEP = 1,30 \% v/v$, para o etanol, e $RMSEP = 1,04 \% v/v$, para o metanol. Considerando as significativas diferenças construtivas entre os dois instrumentos, o resultado é muito satisfatório. Contudo, considera-se que o protótipo pode ser melhorado, e existe um potencial para que os mesmos modelos gerem resultados mais próximos aos do espectrofotômetro. Também, o estudo indica outros comprimentos de onda que podem ser acrescentados para construir modelos com mais variáveis, o que também deve melhorar as performances preditivas. Assim, otimizar a performance do protótipo é uma sugestão para trabalhos futuros. Além disso, protótipos com outras formas construtivas também podem ser implementados e avaliados.

Por fim, o trabalho também aborda aspectos práticos normalmente desconsiderados por outros autores. São discutidos conceitos variados sobre a medição e o processamento dos dados ópticos; sobre o processo de preparação e seleção de amostras de calibração; sobre a apresentação das amostras para o instrumento durante as medições; e também sobre as técnicas quimiométricas utilizadas para calibrar o instrumento. Assim, apesar de tratar de uma aplicação específica, avalia-se que este trabalho é abrangente e contribui ao discutir conceitos aplicáveis também à análise de outros tipos substâncias.

APÊNDICE A
Artigo Publicado na Revista IEEE
Transactions on Instrumentation and
Measurement

Calibration of a Photometer for Quantification of Ethanol in Gasoline: Absorbance Determination

Maurício Moreira, Ana Lúcia de Souza Madureira Felício, and José Alexandre de França^{1b}

Abstract—This paper presents a study about the multivariate calibration of a LED near-infrared photometer applied to the quantification of ethanol in gasoline. Based on the analysis of works of other authors, we implemented seven different methods for determination of absorbances from the raw optical data, which were used for calibration of the prototype. The methods either use different experimental conditions for the background reference measurement: with empty cuvettes, with cuvettes containing a reference sample, or without cuvettes; or are independent of a reference measurement. The results obtained by each method were compared in terms of the root-mean-square errors, coefficients of determination, and residual prediction deviations (RPDs). We discussed characteristics of each method and concepts regarding the modeling of the instrumentation system, spectral preprocessing, and calibration using multiple linear regression, which are applicable both to other photometers and to other types of substances. Experimental results showed that the accuracy of the equipment in quantifying the analyte is determined not only by the accuracy of the light measurement system, but also by the processing of the optical data.

Index Terms—Background reference, baseline correction, light measurement, multiple linear regression (MLR), multivariate calibration.

I. INTRODUCTION

INSTRUMENTS designed for the quantification of substances in liquids and other media using analytical techniques of near-infrared (NIR) spectroscopy has been increasingly widespread. In such systems, one aims to estimate the concentration of a substance by observing the interaction of the electromagnetic radiation with a sample, which is done by measuring the intensity of light either transmitted or reflected by the sample. An example is the equipment developed by Moreira *et al.* [1], a prototype of a LED-based photometer to quantify water added to milk. Besides the low cost, this equipment has simple constructive characteristics and may be easily adapted to quantify other types of substances in liquids.

In this paper, the prototype developed by Moreira *et al.* [1] has been adapted to quantify ethanol in adulterated gasoline samples. For this new application, as predicted by the authors, it has been necessary to modify only the wavelengths of

the three LEDs in the light-emitting system, and the values of the resistances of the instrumentation circuit. Considering this new application, we present a practical study on the multivariate calibration of the equipment, which had not been previously addressed in the quantification of water in milk. In this paper, we focus especially on the method for calculating the absorbances from the optical data, which is an important aspect in the development of an NIR instrument and has a relevant influence on the quality of the calibration. After a literature review, from the descriptions of the experimental procedures found in several papers, we have identified different approaches used in the definition and measurement of the background reference. Based on these observations, we classified seven methods for calculation of the absorbances, which either use different experimental conditions for the background reference measurements (with empty cuvettes, with cuvettes containing a reference sample, or without cuvettes); or are independent of a reference measurement. We compare the seven calculation methods and the practical implications of each of them are discussed. Also, this paper discusses the modeling of the sensor and of the instrumentation circuit used, and evaluates the importance of the spectral preprocessing, which aims to reduce the physical effects and other effects not associated with the chemical information sought in the spectrum.

Studies presented by several authors show prototypes with different constructive characteristics regarding light emission systems, optical structures, monochromator elements, sampling principles (transmittance, reflectance, transreflectance, etc.), types of detectors, and instrumentation circuits [2]–[8]. Although, the constructive aspects of those prototypes are well discussed, the various aspects and procedures involved in the acquisition and processing of the raw optical data for calibration of such instruments are little explored. In this type of technique, the accuracy of the equipment in the analyte quantification is determined not only by the accuracy of the light measurement system, but also involves especially the modeling and processing of optical data as well as the preparation and selection of samples for calibration.

Haxha and Jhoja [3] and Ribeiro *et al.* [9], for example, have presented optical instrumentation systems and demonstrated how the outputs of the detector circuits respond to the variations in the concentration of an analyte of interest. In the same way, Moreira *et al.* [1] demonstrated that the measured light intensities increased with the addition of water to milk, for all the three wavelengths employed. Also, the authors evaluated three univariate calibration curves, which attempted to directly relate the measured voltages to the amount of added water. However, instead of directly associating the voltages provided by the photodetector with the concentration, such data must

Manuscript received August 7, 2017; revised October 11, 2017 and December 26, 2017; accepted February 20, 2018. Date of publication March 19, 2018; date of current version August 9, 2018. This work was supported by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brazil. (Corresponding author: José Alexandre de França).

M. Moreira is with the Coordenadoria de Engenharia Elétrica, Universidade Tecnológica Federal do Paraná, Apucarana 86812-460, Brazil, and also with the Departamento de Engenharia Elétrica, Universidade Estadual de Londrina, Londrina 86057-970, Brazil (e-mail: mmoreira@utfpr.edu.br).

A. L. de Souza Madureira Felício and J. A. de França are with the Departamento de Engenharia Elétrica, Universidade Estadual de Londrina, Londrina 86057-970, Brazil (e-mail: analusou@hotmail.com; franca@la2i.com).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIM.2018.2811285

0018-9456 © 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.

See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

be processed, based on chemometric techniques, and used to develop multivariate calibration models. This is because the techniques based on absorption spectroscopy aim to associate the analyte concentration with the intensity of light absorbed by the analyte. However, the interaction of the electromagnetic radiation with a given sample involves several factors both chemical (light absorption) and physical (including reflection and scattering of light by the particles), so that all these factors together influence on the absolute value of the intensities measured by the system [10]. As discussed by Yadav *et al.* [10], depending on the characteristics of the solution, most of the variation in light intensities measured by the system is due to physical factors, especially light scattering, not due to absorption by the analyte. This explains the results obtained by Moreira *et al.* [1]: in that case, the authors observed an increase in the intensities of light transmitted through milk with the addition of water, at all wavelengths, since it caused a decrease in light scattering. Thus, when attempting to directly relate the measured voltages with the concentration of interest, even in a multivariate way, we will be modeling the physical effects rather than the desired chemical information. This may render the system flawed, for example, in the presence of interfering substances that alter the light scattering characteristics.

In addition to the above, the measurement of the light intensities should consider a proper modeling of the sensor and instrumentation system employed. Thus, a careful interpretation of the output voltages provided by the instrumentation system and their relation to the optical data is required; hence, it is not adequate to directly relate the measured voltages to concentration univariately. More specifically, in this type of technique, calibration models in general must be developed as a function of the absorbances, which are calculated from the optical data at different wavelengths. In turn, different spectral points must be processed, in order to allow the identification of spectral absorption bands, and used multivariately to develop predictive calibration models. However, different methods of calculating the absorbances from the optical information are used in [?], [6], [9], [11]–[14]. Furthermore, the interpretation of the data provided by the equipments and their treatment for the generation of calibration models are little discussed in the literature. The application of an inadequate treatment to the data provided by the prototype can result in an empirical-only calibration model. Such a model may fit well the data of calibration samples but, on the other hand, lead to a poor result in the prediction of future samples, when considering the equipment's practical usage conditions.

In view of all the above, this paper presents a more adequate calibration modeling for the developed prototype, which may also be applied to other systems with discrete wavelengths like the proposed one. We aim to explore especially the interpretation and treatment of the raw optical data, and the practical aspects little discussed in the literature, considering the modeling of both the instrumentation system and the chemometric techniques in which the equipment is based.

II. NIR SPECTRUM MEASUREMENT TECHNIQUE

In general, the various techniques of NIR instrumentation can be classified into two groups: those that observe the light

transmitted through the samples and those that observe the light reflected by them, at different wavelengths. We define as $A^j(\lambda)$ the absorbance at a wavelength λ for the j th sample, which is given by

$$A^j(\lambda) = \log \left(\frac{I_0(\lambda)}{I_S^j(\lambda)} \right) \quad (1)$$

in which $I_0(\lambda)$ is the intensity of light of wavelength λ incident on each sample; $I_S^j(\lambda)$ is the intensity of light transmitted or reflected by the j th sample, at wavelength λ , which reaches the detector; $j = 1, \dots, n$; $\lambda = \lambda_1, \dots, \lambda_m$. The ratio $I_0(\lambda)/I_S^j(\lambda)$ is either the inverse of transmittance, for a technique based on light transmission, or equivalently the inverse of reflectance, for a technique based on light reflection.

Most of the techniques for calibration of spectroscopic instruments aim to establish a linear relationship (usually, a multivariate one) between the concentration of an absorber and the absorbance, based on the Beer–Lambert Law [2]–[4], [10], [15]. This law refers to the observation that the contribution of an absorber to the absorbance of a sample, at a given wavelength, is proportional to the concentration of the absorber. Although there are other types of transformations applicable to the optical data, the transformation given in (1) is the most frequently used one. By calculating the absorbance for each wavelength, one obtains the absorbance spectrum (or spectral points of absorbances, for instruments working with some specific wavelengths like the prototype used). Thus, the light absorption may be associated with the intensities of bands in the absorbance spectrum.

The measurement of absorbance spectral points involves especially two factors. The first one is the definition of a reference experimental condition in which the zero of absorbance is determined. This is necessary because, in practice, $I_0(\lambda)$ corresponds to a light intensity that is also unknown and needs to be measured. The second factor involves the transformation of the output voltages generated by the instrumentation system into optical data, according to the modeling of the sensor and of the conditioning circuit. The definition of the reference has been made in different ways by several authors, as detailed in Section-III. In turn, the relationship between the voltage response and the optical data is discussed in Section IV.

III. DEFINITION OF THE BACKGROUND REFERENCE

As discussed above, the determination of the absorbance spectral points by (1) depends on the definition of a reference condition in which the background reference [$I_0(\lambda)$] is measured. In this paper, seven different methods are considered, which were identified from the analysis of different papers in the literature, as follows.

A. Method A

Lutz *et al.* [6] developed a miniature portable NIR spectrometer for the quantification of ethanol in gasoline, which uses cuvettes to hold the samples. According to the description of their experimental procedure, for each sample, six measurements were performed and a mean spectrum was

considered. In addition, a background reference measurement was recorded every six samples. The authors did not specify in what conditions that measurement was done, but it is likely to have been with an empty cuvette in the light path, which is the mode commonly used in commercial spectrophotometers. Also, Visconti *et al.* [8] developed a prototype for analysis of gases or liquids, which may work in single-beam or dual-beam modes. In the single-beam system, a background reference is measured with the gas chamber empty before analysis of the gas sample. Based on these two cases, we define a method in which a background reference measurement is performed for each sample, and the absorbance is calculated according to

$$A_A^j(\lambda) = \log\left(\frac{I_0^j(\lambda)}{I_S^j(\lambda)}\right) \quad (2)$$

in which $I_0^j(\lambda)$ corresponds to the light intensities measured with an empty cuvette, which are associated with the j th sample.

B. Method B

Gaião *et al.* [11] developed a portable LED-photometer for qualitative analysis of gasoline in order to indicate adulterations by solvents and kerosene. The authors used a quartz cell to hold the samples. In their measurements, 100% of transmittance (or zero of absorbance) was also determined when the cell was empty. However, one can understand from the text that this reference was obtained only once at the beginning of the analysis cycle. Thus, the calculation method performed by Gaião *et al.* [11], unlike Method A, can be represented by

$$A_B^j(\lambda) = \log\left(\frac{I_0'(\lambda)}{I_S^j(\lambda)}\right) \quad (3)$$

wherein $I_0'(\lambda)$ corresponds to a single measurement of the light intensities performed with an empty cuvette, in the beginning of the analysis cycle.

C. Method C

Guchardi and Pasquini [12], in turn, developed an NIR spectrometer that have been tested and evaluated for determination of water in fuel ethanol. Their prototype may work in dual-beam or single-beam modes. According to the authors, in the single-beam system, the reference measurement was taken once when the cuvette was removed from the light path, and the absorbances were calculated by

$$A_C^j(\lambda) = \log\left(\frac{I_0''(\lambda)}{I_S^j(\lambda)}\right) \quad (4)$$

in which $I_0''(\lambda)$ corresponds to a measurement of the light intensities reaching the detector without any obstacle. This method may also be considered analogous to the one used by Isaac and Na [7], based on reflectance, in which $I_0(\lambda)$ is the light reflected by a reference material, such as a ceramic plate.

D. Method D

Chagas [13], in doctoral work, developed different prototypes of both single-beam and dual-beam photometers for fuel analysis. For the dual-beam photometer, the author employed both a measuring detector with an optical filter of wavelength 1480 nm (which corresponds to an absorption band of the substance under test) and a reference detector with a filter at 1300 nm (spectral region with no chemical information). Thus, the absorbance was calculated from the ratio between the voltages generated by both detectors. In practice, it corresponds to a different method in which a wavelength $\lambda_K \in \{\lambda_1, \dots, \lambda_m\}$, where there is no chemical information, is defined as a reference for calculating the absorbance, i.e.,

$$A_D^j(\lambda) = \log\left(\frac{I_S^j(\lambda_K)}{I_S^j(\lambda)}\right). \quad (5)$$

This can be understood as a measure of the magnitude of the attenuated light at each wavelength relative to that at a reference wavelength.

E. Method E

In the tests for quantification of water in fuel alcohol using the single-beam photometer, Chagas [13] described the transmittance calculation as the ratio between the intensities measured for the test sample and for a reference one, which had a specific concentration (94% of alcohol). This is similar to the methods used by Kittipanyangam *et al.* [4], who define $I_0(\lambda)$ as the light intensities transmitted through a solvent; Visconti *et al.* [8], who consider a reference gas when using a dual-beam system; or when the reference is the matrix in which the substance to be determined is inserted. In these cases, a relative absorbance may be calculated, which is given by

$$A_E^j(\lambda) = \log\left(\frac{I_{\text{ref}}(\lambda)}{I_S^j(\lambda)}\right) \quad (6)$$

wherein $I_{\text{ref}}(\lambda)$ corresponds to a measurement of the light intensities reaching the detector after interacting with a solution defined as a reference.

F. Method F

A simpler approach was used by Ribeiro *et al.* [9] and Gentilin *et al.* [14], who developed prototypes based on light reflection. These authors just considered that the intensity of light reflected by the sample is directly proportional to the reflectance. From this consideration, it is possible to assume the transmittance (or reflectance) as the absolute value of the light intensity measured for each sample, which leads to another method for calculating the absorbance

$$A_F^j(\lambda) = \log\left(\frac{1}{I_S^j(\lambda)}\right). \quad (7)$$

Therefore, this method does not require any reference measurements.

G. Method G

Finally, in this paper, we also evaluated a calculation method which is similar to the one defined in (2). It also consists in obtaining a reference measurement with an empty cuvette for each sample. The only difference is that, in this method, the same cuvette used to determine $I_0(\lambda)$ is used in the subsequent measurement of the sample, with the cuvette being held still in the light path when filling it with the sample. Thus, this method for absorbance calculation can be written as

$$A_G^j(\lambda) = \log\left(\frac{I_{\text{still}}^j(\lambda)}{I_S^j(\lambda)}\right) \quad (8)$$

in which $I_{\text{still}}^j(\lambda)$ corresponds to the light intensities measured with an empty cuvette, which is held still for subsequent measurement of the j th sample.

One may notice that methods A, B, C, and G are based on the direct or indirect measurement of the radiation incident on the sample, since $I_0(\lambda)$ is the intensity of radiation transmitted or reflected by air or by a reference material (empty cuvette for transmittance, ceramic plate for reflectance). In turn, the methods D and E are based on the measurement of absorbances relative to a reference wavelength (method D) or to a reference sample (method E), while method F is independent of a reference measurement.

It is evident that the calibration of an equipment from each of these methods will lead to different results, which, in this case, is independent of the accuracy of the instrumentation system. The influence of each of them on the operation of the proposed prototype is discussed in the following sections. Furthermore, it is also evident that the determination of the absorbances is done from the measurement of the light intensities, which, in turn, depends on an appropriate modeling of the sensor and of the instrumentation system. In Section-IV, a modeling for the developed prototype is presented.

IV. MODELING OF THE INSTRUMENTATION SYSTEM

The determination of the absorbances by any of the described methods is done as a function of the light intensities measured by the instrumentation system. An important first consideration should be made with regard to the detector. In the prototype under study, a photodiode of InGaAs (PT511, from *Roithner Lasertechnik*) is used, which is connected to the transimpedance circuit corresponding to the first stage of the circuit of Fig. 1. A description of the operation of this circuit can be found in [1].

In this mode of operation, the photodiode works in a short circuit. Therefore, the photocurrent generated by the sensor is linearly related to the incident light power, and it holds the relation

$$I_r(\lambda) = I_p(\lambda)/r(\lambda) \quad (9)$$

where $I_p(\lambda)$ is the photocurrent generated by the photodiode for an incident light of intensity $I_r(\lambda)$ and wavelength λ , and $r(\lambda)$ is the responsivity of the photodiode at wavelength λ . Note that $I_r(\lambda)$ may correspond either to the light reaching the sensor after interacting with the sample ($I_S^j(\lambda)$) or to the incident light in the reference condition ($I_0(\lambda)$).

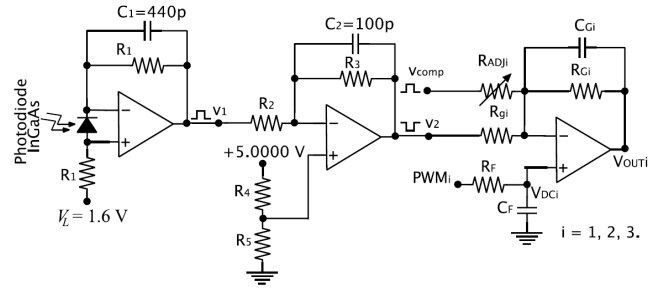


Fig. 1. Detection and signal conditioning circuit [1].

In turn, as described in [1], the circuit of Fig. 1 produces a voltage output ($V(\lambda)$) which is given by

$$V(\lambda) = c_1(\lambda)I_p(\lambda) - c_2(\lambda), \lambda \in \{\lambda_1, \lambda_2, \lambda_3\} \quad (10)$$

where $c_1(\lambda)$ and $c_2(\lambda)$ are constants which may be predetermined from the resistances and reference voltages of the circuit.

From (10), one can observe that the photocurrents may be obtained from the measured output voltages. In turn, by relating (9) to (2), (3), (4), (6), and (8), it can be verified that the light intensities may be replaced by the measured photocurrents in the absorbance calculation, which becomes independent of the sensor's responsivity, in the methods described in Section III, except for methods D and F. On the other hand, in method D, by relating (9) to (5), we obtain

$$A_D^j(\lambda) = \log\left(\frac{r(\lambda)}{r(\lambda_K)}\right) + \log\left(\frac{I_{ps}^j(\lambda_K)}{I_{ps}^j(\lambda)}\right) \quad (11)$$

where $I_{ps}^j(\lambda)$ is the photocurrent generated when the light of wavelength λ reaches the photodiode after interacting with the j th sample, and λ_K is the wavelength used as a reference for this method. In turn, in method F, by relating (9) to (7), we obtain

$$A_F^j(\lambda) = \log(r(\lambda)) + \log\left(\frac{1}{I_{ps}^j(\lambda)}\right). \quad (12)$$

Equations (11) and (12) show that the absorbances, when calculated by methods D and F, become dependent on the sensor's responsivity, which is a function of λ and varies with temperature, for example. Thus, if one ignores this term, by measuring a low photocurrent in a given λ , one might interpret that there is a large absorption associated with that wavelength. In fact, it could be that the sensor's response at that wavelength was lower.

Despite the above, in this paper, the terms dependent on the responsivity in both methods were ignored for comparison purposes between all methods in the performed tests. This is reasonable as the experiments were performed in a relatively short time interval and in a temperature controlled room, so that the responsivity may be considered constant at each λ during the experiment time. Thus, in this paper, the absorbances in all seven methods were calculated by (2) to (8), in which the light intensities were replaced by the photocurrents generated by the photodiode.

V. EVALUATION METHODOLOGY

In order to experimentally evaluate the differences between the methods studied in the calibration of the proposed prototype, we tested the equipment using samples of gasoline diluted with ethanol. Gasoline is a solution which is simple and has homogeneous sample matrix. This reduces the influence of the sample matrix in the process, in relation to a complex natural substance with greater variability such as milk. Thus, it is possible to study the influence of each method itself on the quantification of ethanol. In order to perform the experiments, first, we adjusted the prototype for the quantification of the new substance. Subsequently, samples with different amounts of ethanol were prepared and analyzed by the prototype. After the measurements, the voltages provided by the prototype were converted into absorbances, using each of the methods presented in Section III, to obtain the spectral points. Then, the spectral points were preprocessed and applied to a multivariate linear regression model. Finally, the results obtained after the calibration were compared and discussed. Each of these stages of the experimental procedure is detailed in the following.

A. Adjustments in the Prototype

To carry out the measurements, the prototype of Moreira *et al.* [1], which had been initially developed to quantify water in the milk, were readjusted for the quantification of ethanol in gasoline. Despite being a totally different application, the modifications made in the original system were minimal, and the only significant change required was the selection of new LEDs with wavelengths suitable for the new substance. We made the adjustments as follows.

First, the mechanical structure of the prototype was adapted so that it was possible to use mobile cuvettes to hold the samples, instead of the fixed flowthrough cell which had been used previously. In the tests, we used disposable acrylic cuvettes with optical pathlength of 10 mm.

After, we determined the new wavelengths of the LEDs. For this, after a literature review, we selected three LEDs with emission maxima in the following wavelengths: $\lambda_1 = 970$ nm, $\lambda_2 = 1450$ nm, and $\lambda_3 = 1530$ nm. The last two correspond to spectral bands in which there is absorption by ethanol. On the other hand, the wavelength of 970 nm corresponds to a spectral band where there is no chemical information: the system's response at this wavelength allows it to perform the spectral preprocessing for light scattering correction; also, this information is used as the reference in method D for the absorbance calculation (i.e., $\lambda_K = \lambda_1$).

After some preliminary tests with both commercial gasolines and samples of adulterated gasolines containing 40% of ethanol, the resistances of the circuit of Fig. 1 were adjusted. We performed the adjustments in order to obtain adequate voltage levels at the circuit outputs, so that saturation would not occur for the reference measurements without the cuvettes, in which the light intensities are higher; on the other hand, we aimed to observe a significant variation for different concentrations over the measuring range. From the performed adjustments, we determined the values of $c_1(\lambda)$ and $c_2(\lambda)$

in (10). Further information on the working of this circuit and on its adjustment procedure can be found in [1].

B. Sample Preparation and Measurements

In order to make a proper comparison between the methods described in Section III, it was necessary that all of them were applied to the same samples, analyzed with the same cuvettes and under the same measurement conditions. Thus, we adopted a procedure with the objective of obtaining a single measurement for each sample, which could be used in all evaluated methods. This procedure is described in the following.

First, we prepared samples by adding different volumes of ethanol to commercially available gasolines. The commercial gasolines were collected in four different days, in order to increase sample variability. The calibration model was designed for the range of ethanol concentrations between 25% and 40%, which encompasses possible values of adulteration of gasoline by addition of ethanol in a practical application. Thus, for each day, we performed dilutions in order to span the concentration range.

The measurement procedure was performed as follows. At the beginning of the analyzes of each sample set, the output voltages of the circuit were measured both for an empty cuvette and without any obstacle in the light path. After being transformed to photocurrents by (10), these values correspond to the background references for Methods B and C, respectively. Then, for each sample, the following procedure was adopted.

- 1) One measurement was performed with an empty cuvette and recorded as reference for Method A.
- 2) The same empty cuvette was withdrawn and inserted back into the measuring cell, and a new measurement was performed (which corresponds to the reference for Method G).
- 3) Finally, the sample was carefully inserted into the cuvette without removing it from the light path, and a third measurement was recorded (this is measurement of the sample itself, which is common to all seven methods).

The reference values for Method D were obtained from the voltages for $\lambda_1 = 970$ nm in the measurements with the samples. In turn, for Method E, due to the unavailability of pure (without ethanol) gasoline, we used as reference the values measured for the commercial gasolines without addition of ethanol, that is, the samples with the lowest ethanol concentrations. Last, Method F is independent of a reference measurement, as described previously. After the measurements were carried out, from the voltages measured and transformed into photocurrent by (10), different calculation methods presented in Section III were applied, leading to the absorbance spectral points.

In this paper, we opted to calibrate the equipment and perform an independent validation in order to estimate the future performance of the calibration model by the root-mean-square errors of prediction (RMSEP) parameter. Thus, for calibration, we used samples resulting from the dilutions

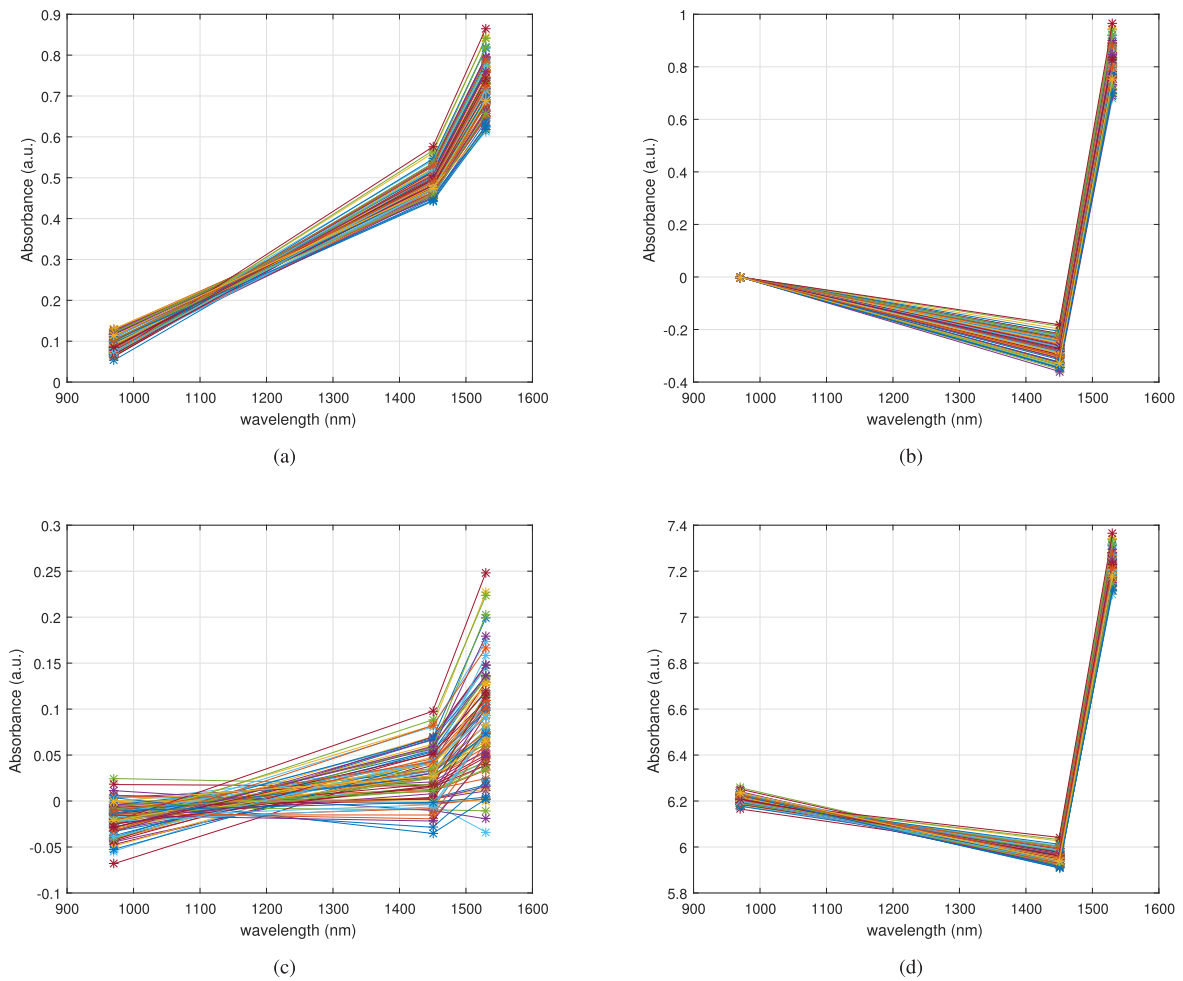


Fig. 2. Spectral points for the 102 calibration and validation samples, without preprocessing, which were obtained from the absorbance calculation by means of (a) method B, (b) method D, (c) method E, and (d) method F.

with concentrations between 24.198% and 40.578% (v/v) in steps of about 0.15%, totaling 104 samples. From this initial set, 17 samples were discarded during the calibration process because they were outliers or corresponded to gross measurement errors. Statistically, the removal of outliers from the model is valid since it does not exceed 20% of the number of samples selected for the model development. After, for the external validation set, we randomly selected 15 independent samples with known concentrations in the range comprised by the model. This number meets the minimum requirements for the development of the chemometric model, in which the number of samples for validation must be at least 10% of the calibration set.

C. Preprocessing of Spectral Data

In order to evaluate the effect of the spectral preprocessing, we used both the spectral points before and after the preprocessing in the regression analysis, and the results obtained were compared. This is because, as discussed by Yadav *et al.* [10], the absolute values of absorbances before the preprocessing vary especially due to the attenuation of the light reaching the detector, caused by the scattering of light by particles. Lutz *et al.* [6], we used a linear baseline

shift correction. Thus, this pretreatment should reduce the effect of the noninformative variation of the baseline, and the preprocessed data might better adhere to the linear calibration model. Other preprocessing methods have not been evaluated because information of only three wavelength are available in the prototype under study.

The adopted baseline correction is applied to each spectral curve by taking a constant curve that passes through the spectral point with the lowest absorbance value, which is defined as the spectral baseline. Then, the correction corresponds to applying a shift in each spectrum in order to match the baselines. Thus, from the absorbance values $A^j(\lambda)$ calculated for the j th sample, after preprocessing, we obtain the corrected absorbances, which are then applied to the linear regression model shown below.

D. Calibration Model

In this paper, we adopted a multivariate calibration by means of multiple linear regression (MLR). In this kind of calibration, the dependent variable (concentration of ethanol) is modeled as a linear combination of the independent variables (absorbances), and the least-square method is used to estimate the regression coefficients.

For all the studied methods, described in Section III, we used as independent variables both the absorbances with and without the data preprocessing, and then different regression coefficients were estimated in each case. Subsequently, the concentrations of ethanol estimated from the seven different calibration curves were determined and used to calculate the root mean square errors of calibration (RMSEC) and RMSEP, as well as the coefficients of determination (R^2) and the RPDs. All data analysis was performed using the software MATLAB 2016. The results obtained are presented in the following section.

VI. RESULTS

Fig. 2 shows the spectral curves obtained for all samples using different methods of absorbance determination under study, without the spectral preprocessing step. We present four curves corresponding to methods B, D, E, and F. Methods A, C, and G led to spectral curves similar to the ones of Fig. 2(a), since in all of them the references are measured with empty cuvettes or with no cuvette; for this reason, they are not shown.

The analysis of Fig. 2 shows that the determination of absorbances by different methods gives rise to spectral curves of different shapes, although the chemical information (i.e., the ethanol concentration in each sample) is the same for all methods. This analysis could be expanded to a system with more points, so it should be noted that the application of spectral preprocessing techniques and the interpretation of spectral bands in comparison to other works in the literature must be performed carefully, depending on the method employed. For example, one may observe that in Fig. 2(a) the spectral points with the lowest absorbance values, which defines the baseline, are at the wavelength 970 nm, while in Fig. 2(b) and (d), the lowest absorbances are found at the wavelength 1450 nm.

Fig. 2 also shows that the absorbances varied at all wavelengths, similar to what Moreira *et al.* [1] had observed for milk. The only exception is the wavelength 970 nm shown in Fig. 2(b), which is used as reference for Method D, and therefore yields zero absorbance values for all samples. This is due to the variations in measured light intensities, whose absolute values do not represent themselves a response to variations in ethanol concentrations. This is because, as already discussed, ethanol does not absorb light around 970 nm, so the measured variation reflects the physical effects that modify the light transmitted by each sample. Instead, one can say that the desired chemical information is in the relationship between the intensities at different wavelengths, i.e., in the shape of the spectrum of each sample. In this sense, one can see in Fig. 2(a) and (c) that method E presents a greater variation in λ_2 and λ_3 in relation to λ_1 , when compared with Method B, for example. This is because method E uses gasoline samples as references, so the effect of the sample matrix is reduced; thus, the ethanol absorption bands at 1450 and 1530 nm are intensified. This characteristic of method E had great influence on the calibration result, as shown below.

Table I presents the results obtained after calibration using the spectral points shown in Fig. 2, without spectral preprocessing. One can see in Table I that Method E reached

TABLE I

COMPARISON OF THE RESULTS FOR THE SEVEN DIFFERENT METHODS OF ABSORBANCE DETERMINATION, CONSIDERING MULTIVARIATE CALIBRATION BY MLR WITHOUT SPECTRAL PREPROCESSING

Method	R^2		RMSEC (%)	RMSEP (%)	RPD
	Calibration	Validation			
A	0.8295	0.8697	1.66	1.61	2.70
B	0.8225	0.8757	1.70	1.56	2.79
C	0.8405	0.8832	1.61	1.51	2.88
D	0.8406	0.8893	1.60	1.49	2.92
E	0.9164	0.9359	1.17	1.12	3.90
F	0.8508	0.8859	1.56	1.49	2.91
G	0.8358	0.8711	1.63	1.61	2.70

TABLE II

COMPARISON OF THE RESULTS FOR THE SEVEN DIFFERENT METHODS OF ABSORBANCE DETERMINATION, CONSIDERING MULTIVARIATE CALIBRATION BY MLR WITH BASELINE CORRECTION

Method	R^2		RMSEC (%)	RMSEP (%)	RPD
	Calibration	Validation			
A	0.8295	0.8698	1.65	1.61	2.70
B	0.8210	0.8779	1.69	1.55	2.80
C	0.8359	0.8856	1.60	1.50	2.90
D	0.8406	0.8893	1.60	1.49	2.92
E	0.9152	0.9406	1.17	1.08	4.04
F	0.8487	0.8897	1.56	1.48	2.95
G	0.8358	0.8735	1.62	1.60	2.72

the best results, with lower values of RMSEC (1.17%) and RMSEP (1.12%); higher coefficients of determination ($R^2 = 0.9164$ for calibration samples and $R^2 = 0.9359$ for the validation ones); and higher RPD (3.90). Meanwhile, the methods with higher RMSEC and RMSEP were Methods B and A, respectively. Also, Method A had the lowest RPD (2.70), along with Method G. In turn, Table II presents the results obtained after application of the baseline correction. It is possible to observe that, again, the lowest errors, the higher coefficients of determination and higher RPD were obtained by method E, while methods A, B, and G reached the worst results. Also, it is possible to verify that, with the exception of the parameter RMSEC in methods E and F and their corresponding coefficients of determination, all other parameters presented equal or better results in relation to the data in Table I. This indicates that the spectral preprocessing improved the adherence of the spectral points to the linear regression model. It can be observed, however, that the difference in the RMSEP and RMSEC values was little significant for this application. This is because gasoline is a homogeneous solution and presents small variability in its composition, so that the physical effects of light scattering are little significant. For other types of substances with more complex matrices and greater sample variability, in which the effect of light scattering is more intense, we believe that the result of spectral preprocessing will be much more evident. However, new experiments must be performed to prove this. Finally, it can be observed in Tables I and II that all methods presented RPD greater than or equal to 2.70, which validates and indicates good predictive performances of all models.

Now, by analyzing the data of Table II, it can be seen that the highest RMSEP and RMSEC values were obtained for

TABLE III

VALIDATION-SET REFERENCE AND PREDICTED CONCENTRATIONS FOR THE SEVEN METHODS OF ABSORBANCE DETERMINATION, CONSIDERING MULTIVARIATE CALIBRATION BY MLR WITH BASELINE CORRECTION. THE MEASUREMENT OF EACH SAMPLE WAS PERFORMED ONLY ONCE AND USED FOR ALL METHODS

Sample	Reference concentration (%)	Method A (%)	Method B (%)	Method C (%)	Method D (%)	Method E (%)	Method F (%)	Method G (%)
1	25.773	26.062	26.028	26.013	26.759	26.431	26.480	26.277
2	26.928	27.155	27.063	27.099	27.766	27.512	27.581	27.058
3	33.093	32.481	32.488	32.661	33.193	33.185	33.234	32.450
4	36.948	35.754	35.592	35.698	35.830	36.438	36.335	35.790
5	25.756	27.545	27.889	27.596	27.408	26.431	27.282	27.483
6	31.615	34.012	34.191	34.046	33.745	33.021	33.840	33.899
7	28.879	31.656	31.833	31.649	31.253	30.554	31.401	31.689
8	36.692	36.368	36.827	36.655	36.756	35.783	36.501	36.708
9	30.592	31.202	30.749	30.901	30.204	31.411	30.697	31.274
10	38.361	38.487	38.113	38.288	38.271	39.121	38.221	37.978
11	36.510	35.833	35.857	36.190	36.142	36.749	36.067	35.841
12	35.402	34.040	33.819	33.982	33.798	34.626	33.835	34.076
13	28.000	26.330	26.428	26.255	26.021	26.431	25.986	26.320
14	32.982	30.729	30.918	30.793	30.674	31.131	30.604	30.769
15	37.958	34.885	35.956	35.801	35.715	36.407	35.709	34.896

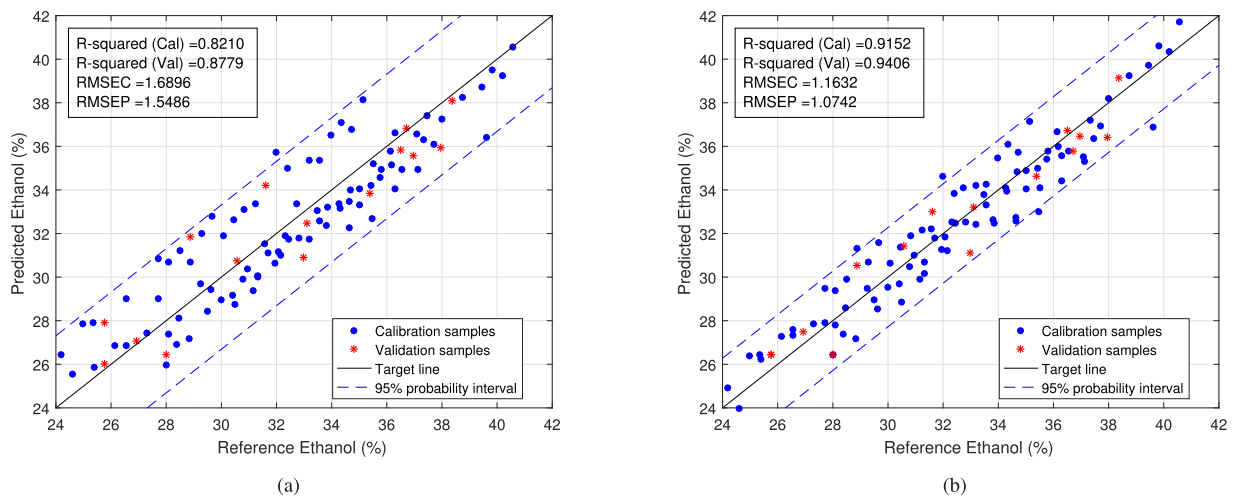


Fig. 3. Relation between specified and predicted concentrations, for the calibration model MLR with baseline correction. (a) Using method B, which takes as reference one measurement with an empty cuvette at the beginning of the analysis cycle. (b) Using method E, which calculates the absorbances relative to a sample or reference standard.

those methods that use reference measurements with empty cuvettes (methods A, B, and G). This shows that the material which constitutes the cuvette is also a significant source of variability in the optical signals. Among these three methods, the better result was obtained when a reference measurement was performed for each sample, keeping the cuvette still in the light path for both the reference measurement and the sample measurement (method G), since it reduces the variability caused by the cuvette. Intermediate results were obtained by method C, which uses a single reference measurement without the cuvette, and by method D, in which the absorbance calculations are relative to the wavelength λ_1 . In turn, the best results were obtained by methods E (in which the absorbances are relative to a reference sample) and F (which is independent of the background reference measurement).

In Table III, we present numerically the specified concentrations and those predicted by the seven methods for the 15 samples of the validation set. In turn, in Fig. 3(a) and (b), we present the relations between the specified concentrations

and those estimated by the prototype from methods B and E, respectively, which presented the higher and the lower RMSEC values. A comparison between the Fig. 3(a) and (b) allows it to graphically evaluate the difference between both methods in estimating the concentrations for both calibration and validation sets. One may observe that the variability of the points around the target line is lower for method E, which indicates a greater accuracy of the prototype when calibrated from this method.

It is important to reaffirm that the data collection of the samples was done only once, for all methods, so that the difference between them is given simply by the definition of different references. Also, it should be noted that the RMSEC and RMSEP values presented correspond to the sample standard deviations of the quantification errors, which are point estimates of the population standard deviations in each method. Nevertheless, in this paper, one should consider that the seven methods led to seven different point estimates for the same sample (that is, for a single set of measurements that

TABLE IV
ADVANTAGES AND DISADVANTAGES OF THE SEVEN METHODS STUDIED, CONSIDERING THEIR CHARACTERISTICS AND QUESTIONS OF PRACTICAL USE. THE METHODS ARE PRESENTED IN ASCENDING ORDER OF RMSEC

	Advantages	Disadvantages
Method E	Requires only one background reference measurement at each new cycle of analysis.	Reference samples or solutions must be available at each new analysis cycle; May be unsuitable for systems that uses fixed flow-through cells.
Method F	Is independent of a background reference measurement.	Should be more dependent on factors such as stability in light emission and opto-mechanical alignment; Is also unsuitable for systems with fixed flow-through cells; Is dependent on sensor responsivity.
Method D	Does not require a reference measurement since the reference is obtained by measuring the sample itself.	Depends on a careful choice of λ_{ref} ; therefore, it requires prior knowledge of the spectral characteristics of the substance to be analyzed; Is also dependent on sensor responsivity.
Method C	Requires only one background reference measurement at each new cycle of analysis; Reference measurement is not influenced by the material of the cuvette.	Once the equipment has been calibrated, it may require a new calibration if one needs to use a different type of cuvette; Unfeasible for systems with fixed flow-through cells.
Method G	Variability in the optical signal caused by the material of the cuvette is minimized, since the reference and the sample are measured with the same cuvette which is held still between the two measurements.	Requires a reference measurement for each new sample; Requires great care in the handling of the equipment, since the sample must be inserted into the cuvette without removing it from the light path.
Method A	Easier handling of samples in comparison to method G, since it allows the movement of the cuvette between reference and sample measurements; Variability caused by the material of the cuvette is also reduced.	Also requires a new reference measurement for each sample.
Method B	Requires only one background reference measurement at each new cycle of analysis.	May also lead to poor results in systems with fixed flow-through cells, since the reference measurement is taken only once at the beginning of the analysis cycle.

were taken only once, not for different sets of observations). Thus, the comparison between the RMSEC and RMSEP values of different methods is valid. In this comparison, we took into account the uncertainties in the estimation of these parameters, with two significant digits, as recommended by JGCM [16]. Thus, the values of RMSEC and RMSEP were rounded up [16] and presented with two decimal digits in order to be consistent with their uncertainties. Finally, for a 99% confidence level, we can estimate the limits of detection (LOD) and limits of quantification (LOQ) as being, respectively, three times and ten times the RMSEC, for each method. Thus, method B presents LOD and LOQ estimated at 5.07% and 16.9%, respectively, whereas method E presents lower limits (LOQ = 3.51%; LOD = 11.7%). However, a more accurate estimate of these limits also requires further experiments.

Besides the numerical results presented, other differences may be inferred from the analysis of the characteristics and practical implications of using each method. Thus, some advantages and disadvantages of each method are summarized in Table IV. In Table IV, the methods are listed in descending order of performance, according to the RMSEC values obtained experimentally.

From the information presented in Table IV and the experimental results presented earlier, we conclude that the most appropriate method of absorbance determination for the quantification of ethanol in gasoline by the proposed prototype was method E. However, depending on the application, the unavailability of reference samples may make this method impractical. Also, despite having reached the second best result, method F has several disadvantages that may cause the need for frequent calibrations, since the equipment response may be affected over time. In turn, method C was demonstrated to be appropriate for this application, provided that the type

of cuvette to be used in the equipment has always the same optical characteristics.

Finally, the analysis presented suggests that, in a system with fixed flowthrough cell such as the one initially proposed by Moreira *et al.* [1], method G may lead to better results than method B, since the measurement of the reference (empty cell) at each new analysis might reduce the effect of the accumulation of residues between subsequent samples. In this sense, methods E and F are probably inadequate for this kind of application, since they use only one reference measurement at the beginning of the analysis cycle, while using method C would be unfeasible.

The tests with gasoline samples could be repeated for other types of solutions. In this case, we believe that the results would be similar for other substances with simple matrices like gasoline, but that the differences in accuracy among the studied methods should be greater for substances with more complex matrices due to greater sample variability and greater influence of the physical factors. However, new experiments are required to confirm this.

In view of all the aspects discussed, it is concluded that the choice of the most appropriate method depends on the characteristics of each application. Evidently, intermediate methods can be implemented in practice, in which a background reference measurement is performed periodically after a given number of samples, which may be good solutions for many applications.

VII. CONCLUSION

In this paper, we presented a study about the multivariate calibration of an NIR digital photometer. The study was carried out using the prototype developed by Moreira *et al.* [1], applied to the quantification of ethanol in gasoline. However,

we discussed important concepts applicable both to other equipment and other types of substances.

For this paper, the prototype originally developed to quantify water in the milk has been easily adapted for the new application, after some simple modifications. We discussed aspects about the modeling of the sensor and of the instrumentation circuit used in the measurement of light. We studied seven different methods of determining the absorbances from the optical data, which use different ways of measuring the background reference. Experimental results showed that the choice of the method has a relevant influence on the determination of the parameters of the calibration curve, which in turn affects the estimation errors. The best results were obtained by method E, in which the background reference is measured from a reference sample. In turn, the highest errors were obtained for those methods in which the background reference is measured from empty cuvettes. Finally, the difference among the methods might be even greater for other substances that cause greater light scattering, such as milk.

The experimental results show that, in this type of technique, the errors in estimating the concentrations are determined by factors other than only the accuracy of the instrumentation system. In this sense, there are other factors that can also affect the accuracy of the equipment, especially the procedures for preparation and selection of the calibration samples, which also deserve to be studied.

REFERENCES

- [1] M. Moreira *et al.*, "A low-cost NIR digital photometer based on ingaas sensors for the detection of milk adulterations with water," *IEEE Sensors J.*, vol. 16, no. 10, pp. 3653–3663, May 2016.
- [2] T. Qiang and J. Chu, "The design of the infrared analyzer for unconventional emission components of ethanol gasoline vehicle," in *Proc. World Autom. Congr.*, Jun. 2012, pp. 1–4.
- [3] S. Haxha and J. Jhoja, "Optical based noninvasive glucose monitoring sensor prototype," *IEEE Photon. J.*, vol. 8, no. 6, pp. 1–11, Dec. 2016.
- [4] S. Kittipanyangam, K. Abe, and K. Eguchi, "Design of a measurement device explaining the relationship between the concentration of solution and the light absorbance for chemical education," in *Proc. 13th Int. Conf. Electr. Eng./Electron., Comput., Telecommun. Inf. Technol. (ECTI-CON)*, Jun./Jul. 2016, pp. 1–6.
- [5] A. V. Fernandes, C. Pinheiro, J. G. Rocha, and G. Minas, "Smart-optical detector array in CMOS for absorbance measurement of physiological fluids," in *Proc. IEEE Int. Symp. Ind. Electron.*, Jun. 2007, pp. 1441–1446.
- [6] O. M. D. Lutz, G. K. Bonn, B. M. Rode, and C. W. Huck, "Reproducible quantification of ethanol in gasoline via a customized mobile near-infrared spectrometer," *Analytica Chimica Acta*, vol. 826, pp. 61–68, May 2014.
- [7] W. Isaac and A. Na, "On-the-go soil nitrogen sensor based on near infrared spectroscopy," in *Proc. Int. Conf. Inf. Technol. (InCITE) Next Generat. IT Summit Theme-Internet Things, Connect Your Worlds*, Oct. 2016, pp. 312–315.
- [8] P. Visconti, A. Lay-Ekuakille, P. Primiceri, G. Ciccicarese, and R. de Fazio, "Hardware design and software development for a white LED-based experimental spectrophotometer managed by a PIC-based control system," *IEEE Sensors J.*, vol. 17, no. 8, pp. 2507–2515, Apr. 2017.
- [9] L. de Souza Ribeiro *et al.*, "Development of a hardware platform for detection of milk adulteration based on near-infrared diffuse reflection," *IEEE Trans. Instrum. Meas.*, vol. 65, no. 7, pp. 1698–1706, Jul. 2016.
- [10] J. Yadav, A. Rani, V. Singh, and B. M. Murari, "Prospects and limitations of non-invasive blood glucose monitoring using near-infrared spectroscopy," *Biomed. Signal Process. Control*, vol. 18, pp. 214–227, Apr. 2015. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1746809415000075>
- [11] E. da Nóbrega Gaião, S. R. B. dos Santos, V. B. Santos, E. C. L. do Nascimento, R. S. Lima, and M. C. U. de Araújo, "An inexpensive, portable and microcontrolled near infrared led-photometer for screening analysis of gasoline," *Talanta*, vol. 75, no. 3, pp. 792–796, 2008. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0039914007008788>
- [12] R. Guchardi and C. Pasquini, "Evaluation of a dual-beam near-infrared spectrometer based on acousto-optic tunable filters," *Appl. Spectrosc.*, vol. 55, no. 4, pp. 454–457, 2001.
- [13] I. P. Chagas, "Desenvolvimento de um Fotômetro Portátil NIR Para Determinação do Teor de Água no Álcool Combustível e do Teor de Etanol na Gasolina," Ph.D. dissertation, Dept. Inst. Chem., Univ. Estadual Campinas, Campinas, Brazil, 2006.
- [14] F. A. Gentilin *et al.*, "Development and testing of a hardware platform for measuring instruments based on near-infrared diffuse reflection," *Measurement*, vol. 86, pp. 14–25, May 2016.
- [15] D. A. Burns and E. W. Ciurczak, *Handbook of Near-Infrared Analysis*, 3rd ed. Boca Raton, FL, USA: CRC Press, 2008.
- [16] JCGM, "Evaluation of measurement data—Guide to the expression of uncertainty in measurement," Joint Committee Guides Metrol., Paris, France, Tech. Rep. JCGM 100:2008, 2008.



Maurício Moreira received the bachelor's and master's degrees in electrical engineering from Universidade Estadual de Londrina, Londrina, Brazil, in 2010 and 2013, respectively, where he is currently pursuing the Doctorate degree in electrical engineering.

He is currently an Assistant Professor with the Federal University of Technology—Paraná, Apucarana, Brazil.



Ana Lúcia de Souza Madureira Felício received the B.S. and M.S. degrees in chemistry from the Universidade Estadual de Londrina, Londrina, Brazil, in 2008 and 2011, respectively, where she is pursuing the Ph.D. degree.



José Alexandre de França received the bachelor's and master's degrees in electrical engineering from the Federal University of Campina Grande, Campina Grande, Brazil, in 1995 and 1997, respectively, and the Doctor's degree in electrical engineering from the Federal University of Santa Catarina, Florianópolis, Brazil, in 2005.

He is currently a Professor and the Leader of the Laboratory of Automation and Intelligent Instrumentation (LA2I), State University of Londrina, Londrina, Brazil. His current research interests include ultrasonic anemometers, soil moisture sensors, near infrared spectroscopy, electronic instrumentation, precision agriculture, calibration of cameras, and computer vision.

APÊNDICE B
Artigo Publicado na Revista Measurement
Science and Technology

Evaluation of the influence of sample variability on the calibration of a NIR photometer for quantification of ethanol in gasoline

M Moreira^{1,2}, A Felício³ and J de França^{2,4}

¹ Coordenadoria de Engenharia Elétrica, Universidade Tecnológica Federal do Paraná, R. Marçílio Dias, 635, Jardim Paraíso, Apucarana, Brazil

² Departamento de Engenharia Elétrica, Universidade Estadual de Londrina, Rodovia Celso Garcia Cid Pr 445 Km 380, Londrina, Brazil

³ Instituto de Tecnologia Desenvolvimento Econômico e Social (ITEDES) Av. Presidente Castelo Branco, 655, 86082-760, Londrina, PR, Brazil

E-mail: franca@la2i.com

Received 6 November 2018, revised 20 February 2019

Accepted for publication 5 April 2019


Published 21 May 2019



Abstract

The calibration of a NIR photometer for analysis of liquid substances aims to find a curve that relates the concentration of some analyte with spectral data collected from the samples. However, the measured optical signals depend not only on the concentration of the analyte of interest, so that the quality of the calibration also depends on the natural variability of samples and on the way the optical data are processed and modeled. This paper presents a study on the influence of these two factors on the calibration of a LED-NIR photometer, which was applied to the quantification of ethanol in gasoline, and compares the results obtained by the prototype with those obtained by a commercial spectrophotometer. We used two distinct procedures for selecting the calibration samples, in order to evaluate the influence of the sample matrix, considering two different methods of determining the absorbances from the optical data. Experimental results showed that the calibration samples must be prepared and selected with uniform distribution in relation to the different sources of variance, in order to make the equipment able to accurately predict future samples. The performance of the prototype was very close to the one of the commercial equipment, with a difference in root mean square errors of less than 0.8% in the concentration (v/v) of ethanol.

Keywords: spectrometers, detector modelling and simulations, liquid detectors, analysis and statistical methods

 Supplementary material for this article is available [online](#)

(Some figures may appear in colour only in the online journal)

1. Introduction

The quantification of substances by means of near-infrared (NIR) spectroscopy is based on the measurement of the light absorbed by a sample, when positioned in the path of a light

beam, in different frequencies of NIR. The performance of an equipment based on this technique is associated both with the opto-electronic characteristics of the instrumentation system, which is responsible for measuring the optical data, and with an appropriate modeling and processing of the optical information. These factors are determinant for a good predictive ability of the calibration model [1].

⁴ Author to whom any correspondence should be addressed.

On the other hand, the quality of the calibration is also determined by the methodology of preparation and selection of the calibration samples. These factors are independent of both the acquisition system and the data processing, and might be determinant for the calibration model to be able to accurately predict the analyte concentration in future samples. In this sense, this paper presents a study on the preparation and selection of samples for the multivariate calibration of a LED-based digital photometer, which was applied to the quantification of ethanol in commercial gasoline. The objective of the study was to quantify the influence of the sample variability on the quality of calibration and their impact on the prediction errors of future samples. Also, the study aimed to compare the performance of the prototype with the one of a commercial spectrophotometer. Although the study deals with the quantification of ethanol in gasoline, we discuss concepts applicable to other types of samples, as well as to other equipment similar to the prototype under study.

In Brazil, as in many other countries, ethanol is mixed with gasoline in amounts fixed by legislation, in order to produce a fuel with lower emission of pollutants. The quantification of ethanol in gasoline in the concentration range between 25% and 40% (v/v) is a requirement for the fuel quality control and fraud identification. Also, this type of measurement is useful for the fuel injection control in engines, for example. In Brazil, standard methods for the measurement of ethanol in gasoline are defined by the Brazilian National Agency of Petroleum, Natural Gas and Biofuels (ANP) in the Technical Standards ABNT NBR 13992 (extraction method) and ASTM D5501 (based on gas chromatography). Also, some studies have already been carried out for the quantification of ethanol in gasolines by NIR spectroscopy, which use bench spectrophotometers [2, 3]. These three methods are limited to laboratory measurements, and the last two require expensive equipment. For this type of applications, it is of great interest the development of analytical instruments that be lightweight, robust, optically stable and of low cost, designated to field usage. In this sense, studies for fuel analysis using simpler analytical instruments, based on NIR spectroscopy, have already been proposed by Lutz *et al* [4] and Correia *et al* [5], which use linear variable optical filters (LVOF) as dispersive elements and detector arrays. Similar instruments can be found in the literature, which are based on diffraction gratings or even on acousto-optic tunable filters (AOTFs), for a wide variety of applications. These constructive types also make it possible to scan a wide range of the electromagnetic spectrum with good spectral resolution. However, these types of optical components make this type of solution still expensive for large-scale use. Finally, another category of equipment are those based on optical filters or LEDs, which enable the design of instruments with discrete wavelengths of narrow spectral bands [6–9]. In this case, they generally correspond to dedicated instruments for specific applications, in which a small number of wavelengths is required for the analytical determination. This type of solution has also been used in the fuel analysis, with proposals of instruments that use two discrete wavelengths [10], and even a single wavelength [11–13]. From the constructive point of view, such systems have desirable

characteristics, such as easier assembly of optical system and good stability over time, which increases the robustness for usage in the field, in addition to the low cost. Despite these advantages, it is evident that the lower the resolution and the spectral range of the available data, the smaller the possibilities of analyzing and processing data to obtain the desired chemical information with satisfactory accuracy. Thus, the impact of sample variability may be greater for equipment of discrete wavelengths. Therefore, in this work, we evaluate the quantification of ethanol in gasoline from this type of solution.

In this type of technique, in order to determine the calibration curve, it is necessary that the calibration samples span the range of variations in the ethanol concentration (analyte to be determined). However, for the calibration curve to be robust, one should consider that there are several other sources of variance related to the chemical composition of gasoline. Automotive gasoline is a type of fuel derived from petroleum that corresponds to a complex mixture of hydrocarbons containing from eight to twelve carbon atoms, the main components of which are branched-chain paraffins, cyclo-paraffins and aromatics; and additives that increase the octane number [2, 3]. Besides, in Brazil, gasolines with additives are also marketed, which have addition of detergents, dispersants and lubricants formulated by each manufacturer. The chemical composition of a particular blend can vary considerably, even for a single manufacturer, due to several factors, such as: origin of the oil used in the production of the gasoline; differences in the oil processing by each refinery; contamination due to transportation and storage in each commercial establishment; and the possible presence of adulterants (such as solvents, kerosene and methanol) [2, 12]. In this sense, the analytical signal normally depends not only on the concentration of the specific analyte for which one wants to develop the model. Variations in the sample matrix may influence on the estimation of the calibration curve and so on the estimation of the analyte concentration after calibration [14]. Measuring the spectra of gasolines samples taken from different suppliers or on different days might result in spectral bands with considerably different intensities, even if they contain the same ethanol concentration [15]. That is because the absorption bands can be intensified or attenuated by either chemical factors (due to variations of different chemical species) or physical ones (especially the scattering of light by particles) [16]. This variability caused by the sample matrix may be even greater in substances with complex sample matrices (composed by many different components) [17, 18], or samples of biological origin such as milk, which has been studied by Moreira *et al* [19] and Ribeiro *et al* [9], and Cassava [20]. In general, the studies already carried out to quantify ethanol in gasoline by NIR neglect or do not adequately consider the influence of sample variability in practice. Consequently, the practical applicability of the instruments already developed may be very limited in a real application. In this sense, this work aims to investigate and quantify this influence in gasoline, in order to guide the later design of a more robust equipment and the preparation of samples for its calibration.

Chagas [11], for example, proposed a single LED based equipment for quantifying ethanol in gasoline and reported

average errors of about 3% (v/v). The use of a single wavelength, while highly desirable, is probably inadequate in practical terms due to sample variability. In this case, even greater prediction errors might be obtained in practice, which may compromise the applicability of the proposed equipment. Harris *et al* [10], in turn, proposed an instrument which uses optical interference filters at two discrete wavelengths in the regions 1500 nm and 1300 nm. The equipment has been proposed to quantify ethanol concentrations from 0 to 85%, which is a very broad range in comparison to the one considered in this paper. Also, the discussion of the analytical method and of the results is superficial, so we believe that the quantification errors would be probably high considering the measurement range in interest and the aspects discussed in this paper. In turn, Lutz *et al* [4] developed a portable NIR spectrometer based on a LVOF and on a detector diode array of 128 pixels, which allows the acquisition of spectral data in the region 1128–2161 nm with resolution of 8 nm. The authors report RMSEP of 0.20%. Despite this excellent result, the authors used only 36 samples, which is a number that may be considered insufficient to cover all important sources of variance in practice. Similarly, Correia *et al* [5] evaluated the ethanol measurement using a portable mini spectrophotometer for a wide range of concentrations (25%–75%) and reported mean quantification errors of 3.81%. However, the authors also used only 16 calibration samples and did not evaluate the influence of matrix. The analysis of the mentioned works shows that it is important to evaluate the practical applicability of analytical instruments that use discrete wavelengths, considering the actual sample variability, in the quantification of ethanol in commercial gasolines.

We used in this study the prototype introduced in [19], a LED-based photometer that is briefly described in section 3. A study has already been carried out by Moreira *et al* [21] about the calibration of this prototype for this application. In that study, the authors addressed the modeling of the instrumentation system; the experimental conditions for measurement of the optical data; and the processing of this data for determination of absorbances by seven different methods. The authors found that the choice of the experimental condition for measuring the background reference has a relevant influence on the determination of the calibration curve, which in turn affects the estimation errors. Also, the authors discussed the influence of physical factors, such as the scattering of light by particles and the influence of the measurement cell on the optical data. Now, in this paper, we have approached the calibration of the equipment regarding the sample variability, in order to guide the preparation and selection of samples and the construction of a calibration model suitable for the practical usage of the equipment. Therefore, the data presented and analyzed in this paper were obtained from the same measurements performed in the experiments presented in [21]. Also, in the present paper, the results obtained by the prototype are analyzed and compared with the results of a bench spectrophotometer.

Considering the above, an adequate calibration depends on a careful evaluation of the population from which samples are taken. Therefore, the preparation and selection of calibration samples should cover all sources of variance that are likely

to be found in the analysis of future samples [15], which also includes the possible presence of adulterants. The knowledge of this factor is essential for the practical usage of the equipment. This is because, once calibrated, the equipment must be accurate in predicting future samples. However, this might not occur if the variability of the sample matrix is not well represented in the calibration samples. An evidence of the spectral variability caused by the sample matrix is the work of Balabin *et al* [22], which use this characteristic to classify gasolines from different refineries. Calibration models generated using samples in which only the concentration of the analyte was varied in a controlled manner, without discussing the variability of the sample matrix, might fit the experimental data accurately for both calibration and validation samples. However, they might be inadequate for a practical usage in which the actual variability of the future samples were greater. In this sense, we aimed to quantify the influence of the variations in the gasoline sample matrices on the prediction of ethanol in future samples. For this, we calibrated the proposed prototype using different procedures of preparation and selection of samples with different matrices, and we compared the prediction errors in each case. Also, we repeated this procedure for two different methods of absorbance determination, in order to evaluate the influence of the sample matrix in each method.

For this study, we considered only gasolines of a same supplier, which have been collected at intervals of two to three days. In this case, the expected variability refers mainly to the variations in the chemical composition of the different lots of gasolines coming from that supplier. Thus, we did not analyze the different possible sources of variability. Instead, we limited the sample space so that the source of variability was known, in order to allow a better interpretation of the results.

Another factor that is determinant for the quality of the calibration is the accuracy of the primary reference method, which is used to quantify the analyte during calibration [15]. Therefore, in this study, we also compared the results obtained by the prototype with those obtained by a commercial spectrophotometer, using the same samples in both equipments. Thus, we evaluated the performance of the prototype by eliminating the possible influence of errors in the primary reference method.

2. Description of the experimental procedure

In order to study the influence of the sample matrix on the calibration of the prototype, we prepared four sets of samples by adding different controlled quantities of absolute ethanol (99.5%) to commercially-available common gasolines. We used gasolines purchased from a same supplier of the region of Londrina, Paraná, Brazil, which were collected at intervals of two to three days. The initial ethanol content in the commercial gasolines were determined by the standard method ABNT NBR 13992. It consists in an extraction test using a solution of water and sodium chloride, wherein the quantification of the alcohol is based on the increase in volume of the aqueous phase. The initial ethanol content measured for the

four gasolines were, respectively, 25%, 24%, 28% e 28%. The concentrations of the samples obtained after the dilutions were calculated according to the amount of added ethanol. Thus, four sets of distinct samples were obtained, differing only by the sample matrix. Samples of each set were prepared in order to span the entire concentration range of interest. Thus, in total, 119 samples with ethanol concentrations between 24.198% and 40.578% (v/v) were obtained. The reference method ABNT NBR 13992 presents uncertainty of $\pm 1\%$ (v/v) due to resolution of the graduated cylinders used in the test, free interpretation of the scale of the graduated cylinders and rounding [2, 23]. In this case, the error in the initial quantification of each gasolines by the method ABNT NBR 13992 affects in the same proportion all samples of a same set. Thus, it is assumed in this study that the total sample variability is composed by both the matrix variability and the quantification errors by the reference method.

After preparation, all samples were analyzed in the prototype in order to obtain the spectral points. In order to study the influence of sample variability on the calibration of the prototype, we used two different procedures to select the samples that would generate the calibration parameters:

- Selection procedure 1: we selected 33 samples for calibration exclusively from the first data set, while 42 test samples were used for validation, in which 29 are from the second set and 13 are from third one;
- Selection procedure 2: the calibration samples were selected randomly from all the four data sets. In this case, we used 87 samples for calibration and 15 test samples for validation. The data from the remaining 17 samples were discarded for being either outliers in the calibration model or gross measurement errors.

Calibration curves were determined considering these two procedures. We compared the concentrations estimated by each calibration model graphically and in terms of the root mean square errors of calibration and prediction. In the selection procedure 1, we chose to use test samples from only two data sets in order to facilitate the distinction between the three sets and the analysis of the results.

Also, in order to relate the influence of sample selection to different forms of processing of the raw optical data, the spectral points used to compose the calibration models were determined in two different ways: we used two of the absorbance determination methods studied by Moreira *et al* [21], which are described in section 3.

As discussed above, besides the influence of the sample matrix, the estimation errors are also related to the accuracy of the primary reference method, which is used to specify the ethanol content in the calibration and validation samples. Anyway, this methodology allows it to evaluate if the influence of the sample selection procedure is more significant in any of the absorbance calculation methods studied. However, in order to eliminate the influence of the primary reference method, we also used a commercial spectrophotometer (FT-IF/NIR Frontier, from PerkinElmer) to analyze the same samples, and the errors obtained by the two equipments were compared. Thus, the performance of the prototype was

evaluated in terms of the difference between the errors, since the influence of the primary reference method, in this case, is the same for both equipment. In this test, we used only the second sample selection procedure.

For the commercial spectrophotometer, the calibration was performed using three spectral regions of width 100nm around the central wavelengths of the LEDs used in the prototype. This width was chosen in order to be similar to the emission bands of the LEDs. So, the calibration of the spectrophotometer was performed only with the spectral content also available in the prototype. The difference, however, is that the commercial equipment measured the spectral points with resolution of 5 nm. Therefore, the calibration of that equipment was performed from about 60 points, unlike the three points available in the prototype. The spectral curves obtained by both equipments are also presented and compared. Lastly, in the spectrophotometer, calibration was performed by partial least squares (PLS) regression with six latent variables. In both equipments, the samples were analyzed in disposable acrylic cuvettes with optical pathlength of 10 mm.

In the next section, a description of the prototype used in the study is made. Also, we present the methods used for absorbance calculation and we describe the calibration model.

3. Prototype calibration model

Figure 1 shows a block diagram of the prototype used in this study. The optical system is simple and does not have elements like lenses or optical filters. The light emission system is composed of three LEDs, which emit radiation at three different wavelengths directly onto the cuvette containing the sample to be analyzed. The light transmitted through the sample is detected by a InGaAs photodiode (PT511 from Roithner Lasertechnik, non-cooled), which has a nominal specification for detection in the wavelength range 850–1650 nm. The samples were analyzed in laboratory, in which the ambient temperature was controlled with maximum variation of 1 °C. Before the measurements, the prototype was kept energized for 30 min for thermal stabilization. In order to quantify ethanol in commercial gasolines, we used three LEDs with emission maxima at wavelengths $\lambda_1 = 970$ nm, $\lambda_2 = 1450$ nm and $\lambda_3 = 1530$ nm. These wavelengths were selected after an analysis of the studies of Harris *et al* [10], Chagas [11], Fernandes *et al* [2], Correia *et al* [5], and other authors who proposed instruments with discrete wavelengths for other applications [8, 9, 12, 17]. Similarly to the methodology used by those authors, the choice was based specifically on the absorption characteristics of the analyte (in this application, the ethanol), which has a broad absorption band between 1400 and 1670 nm related to the first overtone of alcohol OH bond stretching [2]. Within this spectral range, we opted to use the wavelengths 1450 nm and 1530 nm, according to our availability of LEDs. In turn, the wavelength of 970 nm corresponds to a spectral region where there is no chemical information for this analyte. Data from this wavelength are used to perform a spectral pre-processing by baseline correction. It is important to consider that there are also absorptions referring to the first overtone

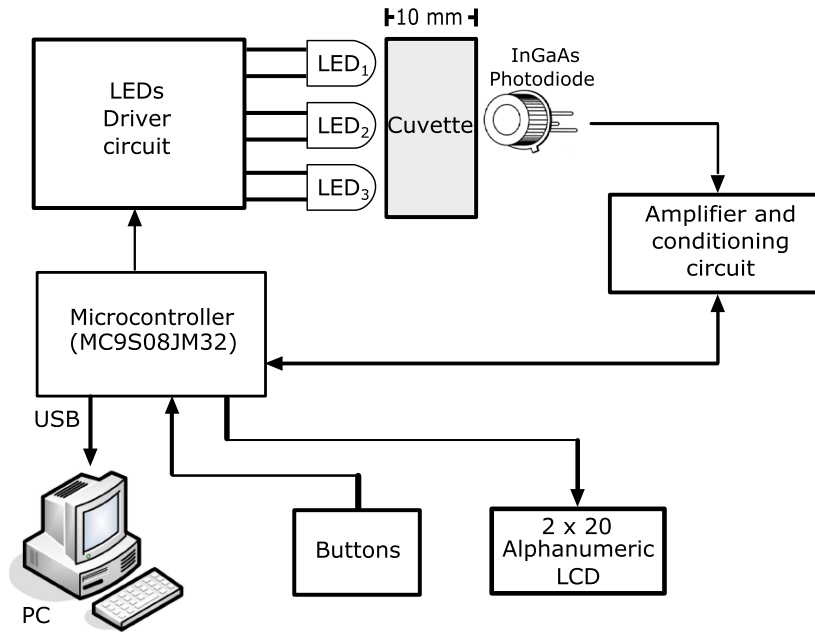


Figure 1. Block diagram of the prototype used in this study. Reprinted with permission from Moreira *et al* [19]. © 2016 IEEE.

of combination bands due to C–H + C–H and C–H + C–C stretching in the region of 1350–1500 nm. In this case, there is overlapping of ethanol and gasoline bands around the wavelength of 1450 nm used in the prototype. However, Fernandes *et al* [2] found that the calibration models constructed with the region 1449–1611 nm showed lower dependence on gasoline composition when compared to other spectral regions.

In order to determine the spectral points, the measured optical data are transformed into absorbances ($A(\lambda)$) by

$$A(\lambda) = \log \left(\frac{I_0(\lambda)}{I_s(\lambda)} \right), \quad (3.1)$$

where $I_s(\lambda)$ is the intensity of light of wavelength λ transmitted by the sample, and $I_0(\lambda)$ is the measured intensity in the reference condition.

Moreira *et al* [21] studied different experimental conditions for the measurement of $I_0(\lambda)$: with empty cuvettes in the light path; with cuvettes containing reference samples; or without cuvettes. In this study, we evaluated two of the methods studied in [21]:

- Method 1: $I_0(\lambda)$ corresponds to a measurement of the light transmitted through an empty cuvette, taken at the beginning of the cycle of analysis [12]. This method was also used in the commercial spectrophotometer;
- Method 2: $I_0(\lambda)$ corresponds to a measurement of the light intensities performed with a sample defined as [6, 11, 24]. In this study, we used as references samples of the commercial gasolines without ethanol addition, that is, with the lowest concentration of the analyte. In this case, we used the specific gasoline associated to each data set, so that the samples from one data set were not influenced by the sample matrices of others.

These methods correspond to the ones referred to as B and E in [21], respectively. We selected them because they presented, respectively, the highest and lowest prediction errors among the methods evaluated in that study.

Gasoline is a homogeneous mixture, since it is a synthetic product processed in laboratory. In spite of this, changes in the baseline of spectra are observed over the whole the NIR region, probably due to the light scattering effect, which requires the application of spectral preprocessing methods. Therefore, as data preprocessing, we used a simple linear baseline correction, which corresponds to the same spectral preprocessing used by Lutz *et al* [4]. It consists in applying a shift in each spectrum in order to match the baselines.

Subsequently, for each of the absorbance calculation methods and sample selection procedures tested, we implemented a multivariate calibration by multiple linear regression (MLR) [15, 18], using the model based on the Beer–Lambert law [15, 24, 25]:

$$Y = \beta_0 + \beta_1 A(\lambda_1) + \beta_2 A(\lambda_2) + \beta_3 A(\lambda_3) + \epsilon, \quad (3.2)$$

in which Y is the concentration of ethanol; β_0 , β_1 , β_2 and β_3 are the regression coefficients of the model; $A(\lambda_1)$, $A(\lambda_2)$ and $A(\lambda_3)$ are the absorbances after baseline correction at wavelengths λ_1 , λ_2 , and λ_3 , respectively; and ϵ is the random error term.

From a set of observations selected for calibration, the estimated regression coefficients ($\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_3$) were calculated, which were determined by the least squares method. Thus, we determined the estimated ethanol concentrations (\hat{Y}), by

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 A(\lambda_1) + \hat{\beta}_2 A(\lambda_2) + \hat{\beta}_3 A(\lambda_3). \quad (3.3)$$

After, we calculated the determination coefficients (R^2) and the root mean squared errors of calibration (RMSEC) and validation (RMSEP), according to (3.4) and (3.5):

$$RMSEC = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - 3}}, \quad (3.4)$$

where: n is the number of samples selected for calibration; y_i is the concentration of ethanol present in the i th calibration sample; \hat{y}_i is the ethanol concentration estimated by the model for the i th calibration sample; and

$$RMSEP = \sqrt{\frac{\sum_{j=1}^m (y_j - \hat{y}_j)^2}{m}}, \quad (3.5)$$

where: m is the number of test samples selected for validation; y_j is the concentration of ethanol present in the j th test sample; \hat{y}_j is the ethanol concentration estimated by the model for the j th test sample.

The independent variables used in the calculations were the absorbances determined by both the previously described methods. It is important to note that, in the spectral preprocessing, we applied a linear shift to all curves in order to match the spectral points in λ_1 . Consequently, $A(\lambda_1)$ is constant for all samples, which leads to $\hat{\beta}_1 = 0$ in all cases. Thus, model (3.3) is reduced to only two regressors and the calibration errors have $n - 3$ degrees of freedom, in accordance with (3.4). All data analysis was performed with the aid of software Matlab 2016.

4. Results

In the following subsections, we present the results of the experiments described in section 2.

4.1. Determination of the spectral points by the prototype and by the spectrophotometer

First, the samples were analyzed in both equipments and the spectral curves shown in figure 2 were obtained, from which the calibration models were built. The figure shows the spectral points of absorbances obtained for all gasoline samples containing different amounts of ethanol. Figures 2(a) and (b) show the spectral points obtained by the prototype by means of Methods 1 and 2 described in section 3, respectively. In turn, figure 2(c) presents the spectra obtained by the commercial spectrophotometer, for the same samples, as well as the emission spectra of the three LEDs of the prototype. In all three graphs, the spectral points underwent the spectral preprocessing based on the wavelength 970 nm. The original curves without preprocessing are provided in the supplementary material (stacks.iop.org/MST/30/075501/mmedia). Also, for comparison, the wavelengths corresponding to the nominal values of the three LEDs of the prototype are indicated.

It can be observed in figure 2 that the characteristic of variation of the absorbances obtained by the prototype in both

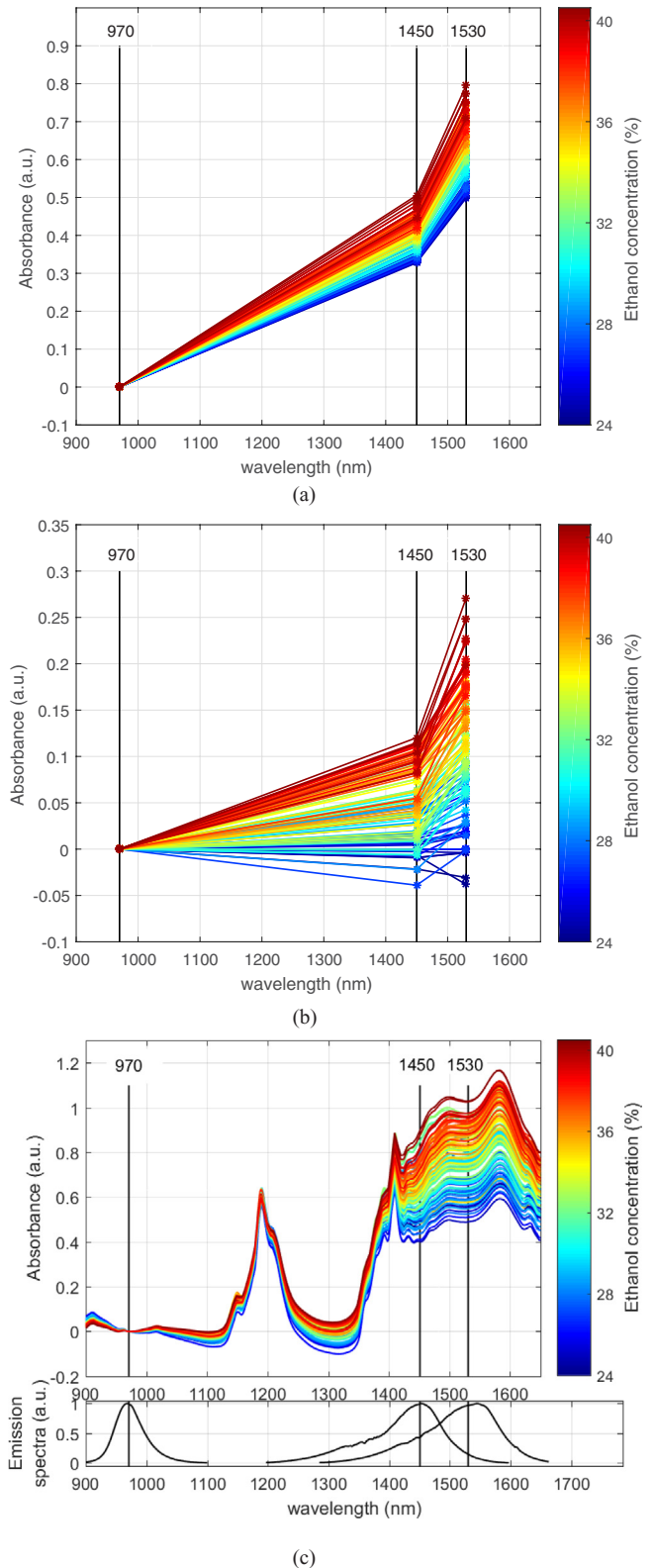


Figure 2. Spectral points of 119 gasoline samples with different ethanol contents, obtained after baseline correction: (a) by the prototype, using Method 1 to calculate the absorbances; (b) by the prototype, using Method 2 to calculate the absorbances; (c) by the commercial spectrophotometer, for the spectral range of 900–1650 nm; and emission spectra of the LEDs.

methods is consistent with that observed in the spectrophotometer, although the amplitudes of variation at 1450 nm and 1530 nm are relatively lower for the prototype. This reveals a lower sensitivity of the prototype in relation to the commercial equipment, which may be associated with the relative simplicity of the optical system used: due to the absence of lenses, filters or other devices in the optical system for collimation of the light, the number of light beams that interact with the sample and reach the detector to compose the observed response is likely to be lower. Also, one should note in figure 2(c) that the LEDs have a relatively broad emission band (spectral half-width of about 100 nm for 1450 and 1530 nm). Thus, unlike the spectrophotometer, the signals arriving at the sensor in the prototype correspond actually to the contributions of wavelengths in a range around the nominal value of each LED. Despite these observations, the graphs show that the proposed prototype is appropriate to analyze substances with relatively wide and well defined absorption bands, as in the case under study, without the need to use lenses, filters or other types of monochromator elements.

4.2. Evaluation of the prototype regarding the selection of calibration samples using method 1

From the spectral points shown in figure 2(a), which were determined by Method 1, we determined the calibration curves for both procedures of sample selection described in section 2. In the first procedure, in which the calibration was done exclusively from the first set of samples, we obtained the calibration curve

$$\hat{Y} = -7.8297 + 96.6083A(\lambda_2) + 4.0831A(\lambda_3). \quad (4.1)$$

The results obtained using (4.1) are shown in figures 3(a)–(c).

Figure 3(a) relates the concentrations of ethanol obtained by the laboratory reference method to those estimated by the prototype according to (4.1). In this figure, the points estimated by the model for the calibration samples (set 1) are plotted, which were used to determine the RMSEC. This corresponds to the sample standard deviation, which is an estimate of the population standard deviation. In addition, the 95% probability interval is also represented, which is defined by $\pm 1,96 \times \text{RMSEC}$, assuming that the error has a Gaussian probability distribution. This interval indicates that, if the calibration curve obtained is a good estimate of the true regression curve, and if the calibration sample represents the population adequately, then about 95 % of the new observations are expected to be within this interval. Finally, figure 3(a) also shows the points estimated by the model for the test samples (sets 2 and 3), which were used to determine RMSEP. In turn, we present in figure 3(b) the graph of the calibration and validation errors. Lastly, in figure 3(c), we plotted histograms of the errors shown in figure 3(b), in order to estimate the probability distribution of the observed errors. Also, we present a normal probability density function with zero mean and standard deviation of RMSEC, which correspond to the

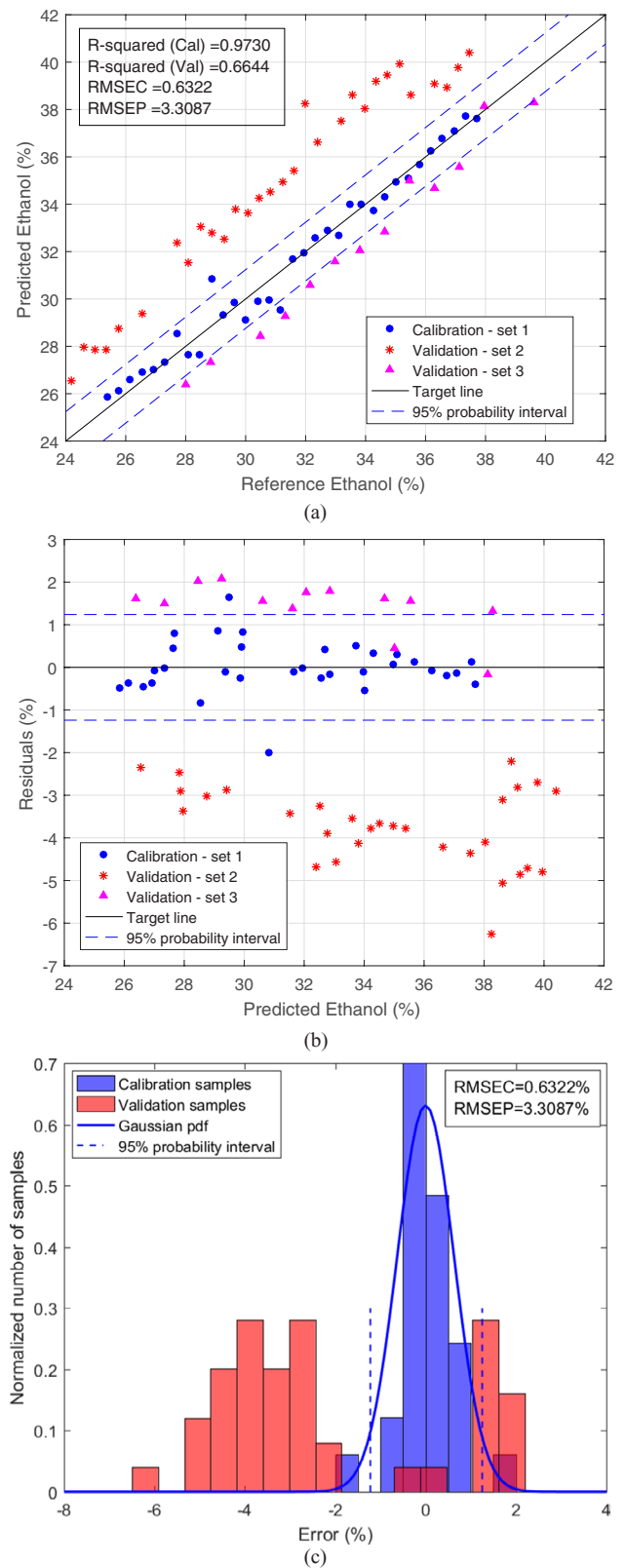


Figure 3. Results after calibration using samples exclusively from the first data set, with spectral points obtained by Method 1: (a) relation between the concentrations specified by the reference method versus those estimated by the calibration model; (b) calibration residues and prediction errors; (c) histograms of the calibration residues and prediction errors.

parameters estimated by the model, along with the 95% probability interval.

The analysis of figure 3(a) indicates that the calibration model fits well the calibration set, leading to a relatively low RMSEC (0.67%) and a $R^2 = 0.9730$. On the other hand, the validation samples are farther from the target line, resulting in $RMSEP = 3.31\%$ and in a R^2 of only 0.6644. Figures 3(a) and (b) show that there is clearly a tendency associated to each of the three data sets. One may observe that the errors for calibration samples are dispersed around the zero, while the errors are all negative for the validation samples of set 2, and predominantly positive for the validation samples of set 3. Also, the prediction errors of set 2 were higher than those of set 3. This indicates that the sample matrix of set 2 probably has spectral characteristics more distinct from that of the calibration samples (set 1). It is superimposed on the matrix variability the possibility that the error in the initial quantification of the gasoline of set 2 was greater in relation to the initial gasoline of set 3. Finally, one can see in figure 3(c) that the calibration errors have a tendency to the Gaussian distribution, which is one of the assumptions of the regression analysis. However, the errors associated with the test samples are far from the estimated 95% probability interval, which indicates that the regression curve (4.1), as well as the probability distribution of errors estimated for the population, are inadequate. From these results, it is also possible to infer that the variability of matrix, associated with the uncertainty of the reference method, had an influence on the variance of the estimation errors that was more significant than the influence of the instrumental errors on the measurement of the optical signals by the instrumentation system.

The analysis of figures 3(a)–(c) indicates that the prediction errors in future samples might be very high if the samples selected for calibration do not represent well the population variability. It is concluded, therefore, that this model is not satisfactory.

Differently, in the second procedure, the calibration and validation samples were selected in a distributed way among the four data sets. In this second test, the calibration curve obtained was

$$\hat{Y} = 1.9174 - 27.5043A(\lambda_2) + 66.0505A(\lambda_3), \quad (4.2)$$

and the results obtained from this calibration curve are shown in figures 4(a)–(c). It is observed in these figures that the spread of points of calibration samples around the target line is higher than that obtained previously, as indicated by the higher RMSEC (1.69%) and lower coefficient of determination ($R^2 = 0.8210$). Consequently, one can also notice that the estimated 95% probability region is wider. On the other hand, now the prediction errors for the validation samples are distributed around zero, and all points are within the range of $\pm 1,96 \times RMSEC$, resulting in a $RMSEP = 1.55\%$. The estimated probability density function of the errors, shown in figure 4(c), is wider and better characterizes the population variability. Therefore, it is concluded that the second procedure of selecting samples for calibration is more appropriate

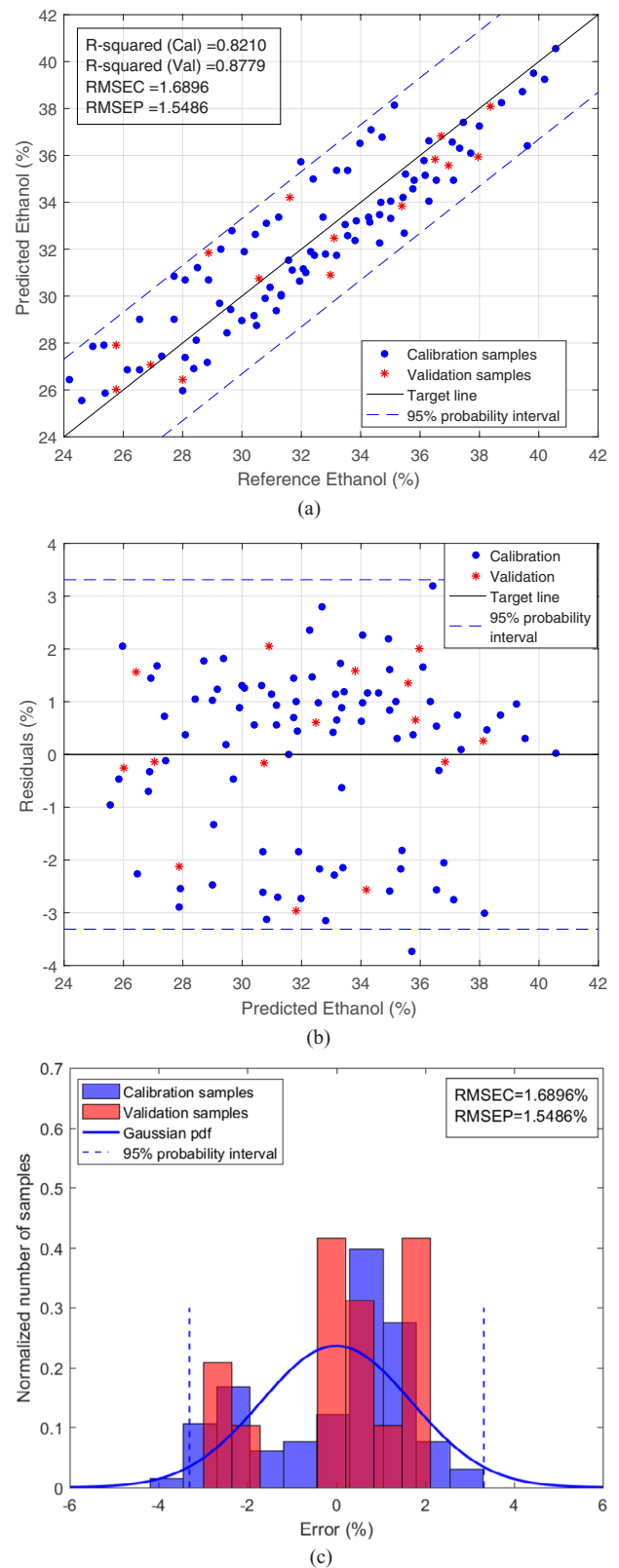


Figure 4. Results after calibration using samples distributed from the four datasets, with spectral points obtained by Method 1: (a) relation between the concentrations specified by the reference method versus those estimated by the calibration model (reprinted with permission from [21]. © 2018 IEEE.); (b) calibration residues and prediction errors; (c) histograms of the calibration residues and prediction errors.

than the previous one. In this case, the population variability was better represented in the calibration, and the prediction errors obtained were consistent with those estimated in the calibration. So, the model is considered to be satisfactory.

In order to relate the effect of sample selection with the absorbance calculation method, both procedures described in this subsection were repeated, for the same samples, this time using the absorbances calculated from Method 2. The results obtained are presented in the following subsection.

4.3. Evaluation of the prototype regarding the selection of calibration samples using method 2

The calibration curves presented in section 4.2 were recalculated, this time using the spectral points shown in figure 2(b). In the first case, using only samples from the first set for calibration, the calibration curve obtained was

$$\hat{Y} = 25.9561 - 24.1222A(\lambda_2) + 87.6033A(\lambda_3), \quad (4.3)$$

and the results obtained from (4.3) are shown in figures 5(a)–(c). On the other hand, for the calibration using samples distributed among the four sets, the calibration curve obtained was

$$\hat{Y} = 26.4313 - 27.8454A(\lambda_2) + 68.5890A(\lambda_3), \quad (4.4)$$

and the results obtained from (4.4) are shown in figures 6(a)–(c).

Comparing figures 5 and 6, it is observed that the selection of calibration samples in a distributed way resulted again in a lower prediction error (RMSEP = 1.08%, $R^2 = 0.9406$) in relation to the calibration with samples from a single set (RMSEP = 1.53%, $R^2 = 0.8538$), as expected. The same previous observations can be made with respect to the distribution observed for the errors.

In turn, by comparing the results presented for both Methods 1 and 2, one may note that the observed difference in the prediction errors between both sample selection procedures was lower for Method 2 (0.45%) in relation to Method 1 (1.76%). This occurred because the variance in the measurements due to the effect of the sample matrix is reduced in Method 2. This is because the background reference is measured in Method 2 using a solution adopted as reference, which in this case includes the matrix in which the ethanol is inserted. Differently, Method 1 uses an empty cuvette. These results show that the accuracy and precision of the prototype in the prediction of future samples may be affected by both the choice of the absorbance calculation method, which has been discussed in [21], and the methodology of selecting the calibration samples, subject addressed in the present study.

Lastly, by comparing the calibration curves obtained in the four cases tested, it can be seen that (4.2) and (4.4) have estimated regression partial coefficients ($\hat{\beta}_2$ and $\hat{\beta}_3$) relatively close. This indicates that both absorbance calculation methods resulted in calibration curves with similar sensitivities to the variations of $A(\lambda_2)$ and $A(\lambda_3)$, when the calibration samples were selected from the four sample sets. In contrast, when we used samples only from the first day for calibration, the same methods gave rise to calibration curves ((4.1) and (4.3)) with totally different $\hat{\beta}_2$ and $\hat{\beta}_3$ parameters. This is an indication that this way of selecting the samples generates an imprecision in

the regression coefficient estimates. It should be noted in this comparison, however, that the number of samples used for calibration in both procedures were different.

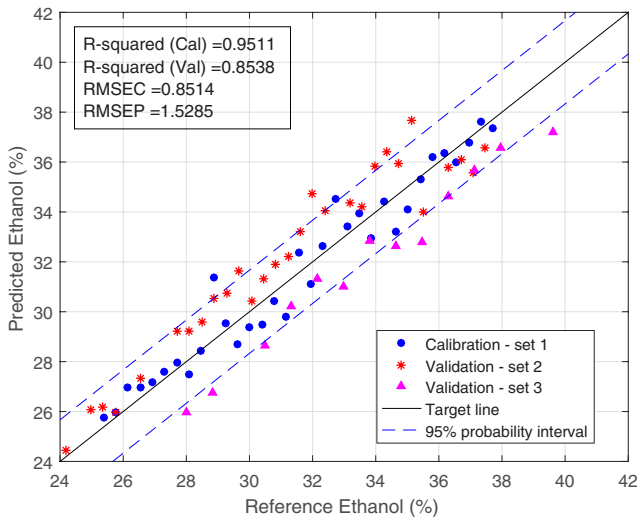
In this type of application, it is impracticable that the equipment be calibrated only by means of samples collected in the field, since it is necessary to span the entire range of analyte concentrations. Instead, samples are generally prepared in laboratory by the controlled addition of ethanol. In this sense, the results discussed in sections 4.2 and 4.3 indicate that the preparation and selection of calibration samples must be done with uniform distribution with respect to both the range of the analyte concentrations and the other sources of variance. It means that any sample has the same probability of being selected. Otherwise, it will be more probable that only a relatively small proportion of the population be selected. In this case, many calibration samples will be spectrally similar, which may insert a tendency in the calibration curve. This is what happens in the situation shown in figure 3. The analysis of the data presented in figures 3–6 shows that the variability of matrix in gasoline, associated to the uncertainty of the primary reference method, has a significant influence on the determination of the calibration parameters and, consequently, on the prediction of future samples by the photometer. Therefore, it is concluded that these factors must be carefully evaluated in order to make it possible to build a robust and useful calibration curve for practical usage in the field. We believe that the same result could be observed from other photometers already described in the literature, since these factors are independent of the system construction.

Finally, it is important to note that the gasolines used in the four days were purchased from the same supplier. Thus, it is expected that the sample variability and its effects might be even more significant when considering distinct suppliers, or in other applications with different kinds of samples. The results show that these factors must be carefully evaluated in a practical application, and that different data treatments can lead to significantly different results, for the same electrical signals measured by the instrumentation circuit.

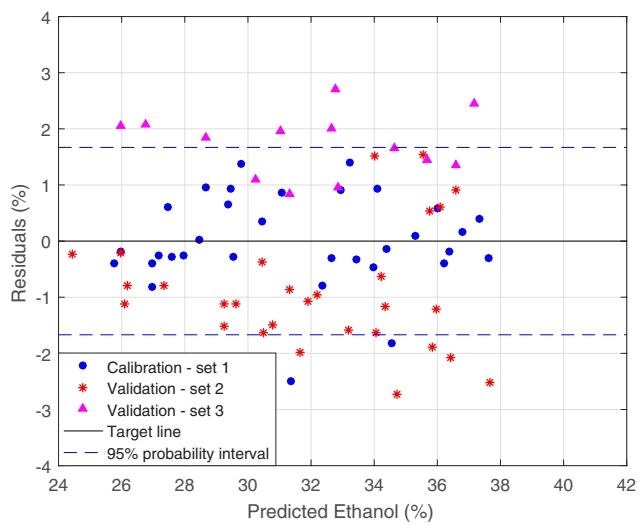
4.4. Comparison with the commercial spectrophotometer

As in any regression analysis, it is assumed that there is no error in the dependent variable, i.e. in the determination of the concentrations by the primary reference method. However, it is well known that this is also a limiting factor for the accuracy of the equipment [15]. Therefore, the results obtained by the prototype were compared with the ones of a commercial spectrophotometer, in terms of the estimation errors and coefficients of determination obtained, using the same samples. In this comparison, only the second sample selection procedure was used, which proved to be the most appropriate one. We employed the same number of samples for calibration and validation used in the prototype.

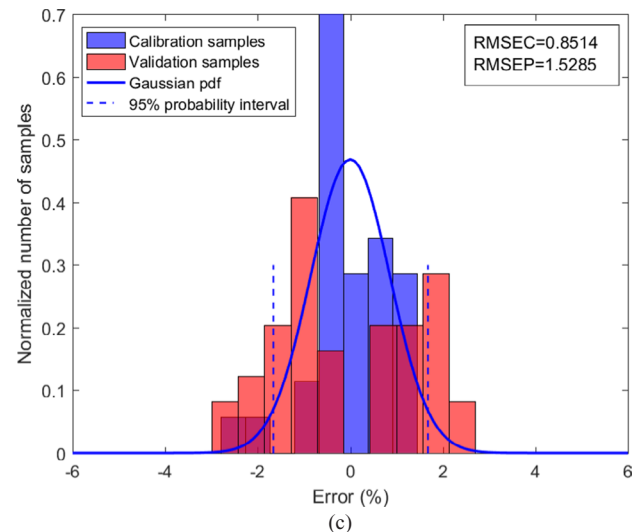
Figure 7 shows the results obtained after calibration of the commercial spectrophotometer. It is observed in figure 7(a) that the scattering of the points around the target line is lower when compared to the prototype, as indicated by the lower



(a)

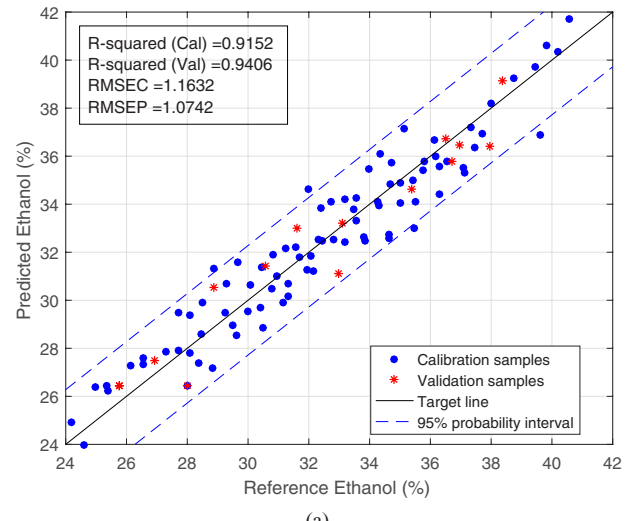


(b)

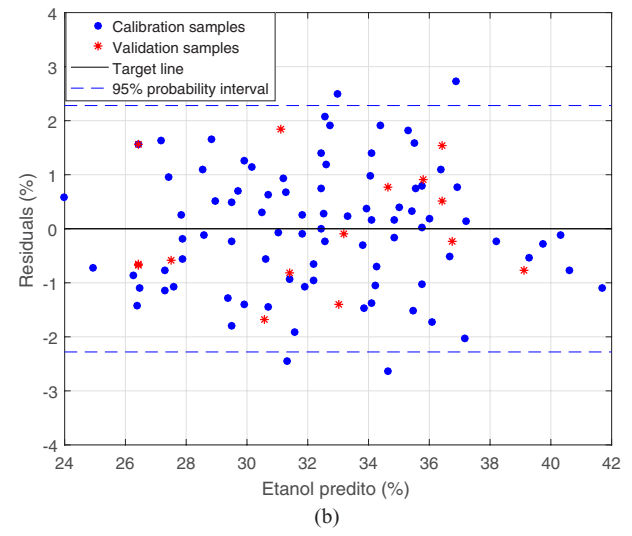


(c)

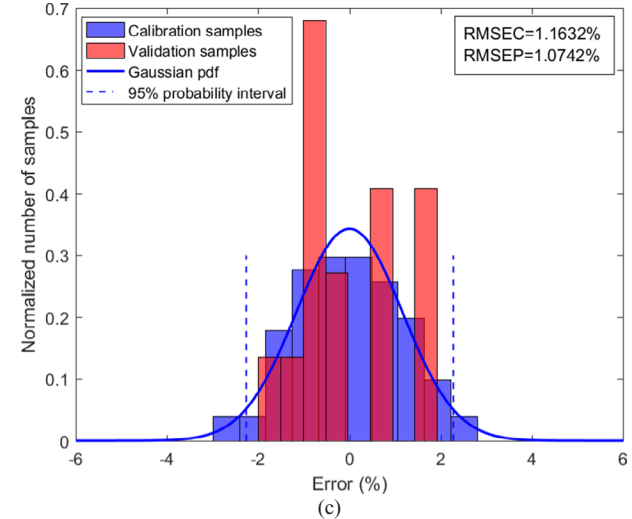
Figure 5. Results after calibration using samples exclusively from the first data set, with spectral points obtained by Method 2: (a) relation between the concentrations specified by the reference method versus those estimated by the calibration model; (b) calibration residues and prediction errors; (c) histograms of the calibration residues and prediction errors.



(a)



(b)



(c)

Figure 6. Results after calibration using samples distributed from the four datasets, with spectral points obtained by Method 2: (a) relation between the concentrations specified by the reference method versus those estimated by the calibration model (reprinted with permission from [21]. © 2018 IEEE.); (b) calibration residues and prediction errors; (c) histograms of the calibration residues and prediction errors.

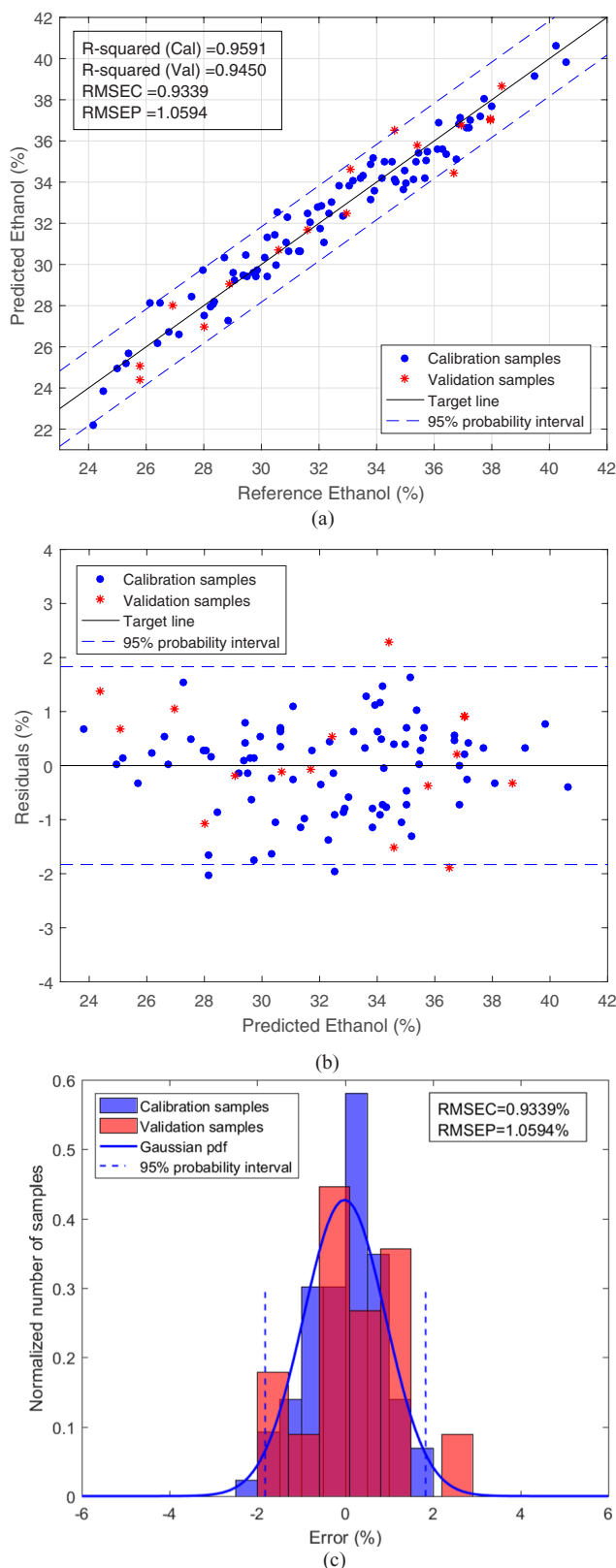


Figure 7. Results obtained by the commercial spectrophotometer after calibration using samples distributed from the four datasets: (a) relation between the concentrations specified by the reference method versus those estimated by the calibration model; (b) calibration residues and prediction errors; (c) histograms of the calibration residues and prediction errors.

RMSEC (0.94). The coefficients of determination obtained for the calibration and validation samples were, respectively, 0.9591 and 0.9450. In turn, both calibration and validation errors are well distributed around zero, as shown in figure 7(b), and are consistent with the estimated Gaussian distribution, as shown in figure 7(c).

Comparing the results of the commercial equipment (figure 7(a)) with those of the prototype for Method 2 (figure 6(a)), one can verify that the difference in the RMSEC was only 0.23% of ethanol, and the difference in the RMSEP was about 0.02% of ethanol. Thus, it is estimated that, on average, the difference between the errors of measurements performed by the prototype and by the spectrophotometer corresponds to less than 0.3% in the concentration (v/v) of ethanol. It should be considered, however, that the background reference is measured in the spectrophotometer for an empty cuvette, unlike the prototype in Method 2. In this sense, one can more adequately compare the results of the commercial equipment with those shown in figure 4(a), which uses the data calculated by Method 1. In this case, it can be seen that the difference was 0.76% for the calibration samples and of 0.49% for the validation ones. Thus, it is estimated that the mean error in ethanol quantification by the prototype, in magnitude, is higher than the commercial spectrophotometer in less than 0.8% (v/v) ethanol only.

5. Conclusion

In this paper, we presented a study about the influence of the sample variability on the multivariate calibration of a NIR digital photometer, applied to the quantification of ethanol in gasoline. We discussed concepts applicable to other similar equipment, as well as to other types of samples. Tests were performed in order to estimate the errors in the prediction of future samples when different sampling procedures were used to calibrate the equipment. Also, the parameters of the calibration curves were obtained from the absorbances determined by two different methods. The results showed that the variance of the estimation error due to sample variability is more significant than the one caused by errors in the measurement of the optical signals. Also, we showed that the samples used in the calibration of the prototype should best represent the expected sample variability, under the conditions in which the equipment is designed to operate. Otherwise, prediction errors in future samples may be very high. Finally, the comparison with the spectrophotometer showed that the performance of the prototype was very close to the one of the commercial equipment, with a difference in root mean square errors of less than 0.8% in the concentration (v/v) of ethanol. Considering the simplicity of the prototype when compared to the spectrophotometer, the results are very satisfactory.

The study presents limitations regarding the total variability of the gasoline studied and the standard method used in the determination of the reference concentrations. We consider that it is still necessary to evaluate the other sources of variability and determine an optimal set of wavelengths for this analytical determination. This is because

the selection of wavelengths based exclusively on the analyte's absorption characteristics, usually adopted by many authors, is also a factor that can make the system vulnerable to matrix variations, since absorption bands of ethanol and of other components of high variability may overlap. Thus, it is proposed to investigate the wavelength selection in order to optimize the quantification of ethanol in gasoline by the future equipment. Therefore, it is proposed in the future to extend the study using a larger number of samples in order to cover the different sources of variability, and to use a reference method by gas chromatography, which has lower uncertainty in comparison to the ABNT NBR 13992. Finally, it should be considered that the hardware of the equipment used in this study should be improved to allow practical usage in the field, especially with regard to thermal issues, since the tests were performed only in the laboratory. Despite these limitations, the results of this study showed that a prototype with discrete wavelengths may be adequate for quantification of ethanol in gasolines, provided that a suitable methodology is used for preparation and selection of calibration samples.

Acknowledgments

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES)—Finance Code 001.

ORCID iDs

M Moreira  <https://orcid.org/0000-0002-4211-4458>

A Felício  <https://orcid.org/0000-0003-4997-8897>

J de França  <https://orcid.org/0000-0002-1402-5642>

References

- [1] McArthur L and Greensill C 2007 Impact of resolution on nir pls calibration of kaolin content with weipa bauxite *Meas. Sci. Technol.* **18** 1343
- [2] Fernandes H L, Raimundo I M Jr, Pasquini C and Rohwedder J J 2008 Simultaneous determination of methanol and ethanol in gasoline using nir spectroscopy: effect of gasoline composition *Talanta* **75** 804–10
- [3] Ouyang A and Liu J 2013 Classification and determination of alcohol in gasoline using nir spectroscopy and the successive projections algorithm for variable selection *Meas. Sci. Technol.* **24** 025502
- [4] Lutz O M D, Bonn G K, Rode B M and Huck C W 2014 Reproducible quantification of ethanol in gasoline via a customized mobile near-infrared spectrometer *Anal. Chim. Acta* **826** 61–8
- [5] Correia R M et al 2018 Portable near infrared spectroscopy applied to fuel quality control *Talanta* **176** 26–33
- [6] Visconti P, Lay-Ekuakille A, Primiceri P, Ciccarese G and de Fazio R 2017 Hardware design and software development for a white led-based experimental spectrophotometer managed by a pic-based control system *IEEE Sensors J.* **17** 2507–15
- [7] Gentilin F A et al 2016 Development and testing of a hardware platform for measuring instruments based on near-infrared diffuse reflection *Measurement* **86** 14–25
- [8] Fonseca A and Raimundo I M 2007 A simple method for water discrimination based on a light emitting diode (led) photometer *Anal. Chim. Acta* **596** 66–72
- [9] Ribeiro L S, Gentilin F A, de França J A, Felício A L D S M and de França M B M 2016 Development of a hardware platform for detection of milk adulteration based on near-infrared diffuse reflection *IEEE Trans. Instrum. Meas.* **65** 1698–706
- [10] Harris S J, Swarin S, Sultan M, Lambert D K and Jack M 1993 Sensor for Measuring Alcohol Content of Alcohol/Gasoline Fuel Mixtures *US Patent Specification* 5,239,860
- [11] Chagas I P 2006 Desenvolvimento de um fotômetro portátil NIR para determinação do teor de Água no Álcool combustível e do teor de etanol na Gasolina *PhD Thesis* State University of Campinas, Institute of Chemistry, Campinas, SP, Brazil
- [12] Gaião E N, Santos S R B, Santos V B, Nascimento E C L, Lima R S and Araújo M C U 2008 An inexpensive, portable and microcontrolled near infrared led-photometer for screening analysis of gasoline *Talanta* **75** 792–6 (Special section: remote sensing)
- [13] Nascimento E C L 2008 Um fotômetro microcontrolado LED-NIR, portátil e de baixo custo para análise screening de Gasolinas tipo C *PhD Thesis* Federal University of Paraíba João Pessoa
- [14] Ni N and Chan C C 2009 Identification and measurement of gas mixture by using the support vector regression technique *Meas. Sci. Technol.* **20** 115601
- [15] Burns D A and Ciurczak E W 2008 *Handbook of Near-Infrared Analysis* 3rd edn (Boca Raton, FL: CRC Press)
- [16] Trevisan S, Bavera M and Giardini M E 2007 Portable instrument for *in vivo* infrared oxymetry using spread-spectrum modulation *J. Instrum.* **2** P04005
- [17] Haxha S and Jhoja J 2016 Optical based noninvasive glucose monitoring sensor prototype *IEEE Photonics J.* **8** 1–11
- [18] Sun D-W (ed) 2009 *Infrared Spectroscopy for Food Quality Analysis and Control* 1st edn (New York: Academic)
- [19] Moreira M et al 2016 A low-cost NIR digital photometer based on ingaas sensors for the detection of milk adulterations with water *IEEE Sensors J.* **16** 3653–63
- [20] Harnsoongnoen S and Siritariwat A 2015 Electrical and absorption properties of fresh cassava tubers and cassava starch *J. Instrum.* **10** P09004
- [21] Moreira M, de Souza Madureira A L and de França J A 2018 Calibration of a photometer for quantification of ethanol in gasoline: absorbance determination *IEEE Trans. Instrum. Meas.* **67** 2119–28
- [22] Balabin R M, Safieva R Z and Lomakina E I 2010 Gasoline classification using near infrared (NIR) spectroscopy data: comparison of multivariate techniques *Anal. Chim. Acta* **671** 27–35
- [23] Associação Brasileira de Normas Técnicas Rio de Janeiro, Brazil 2015 ABNT-NBR 13992: Motor gasoline—determination of anhydrous ethanol fuel content
- [24] Kittipanyangam S, Abe K and Eguchi K 2016 Design of a measurement device explaining the relationship between the concentration of solution and the light absorbance for chemical education *13th Int. Conf. on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology* pp 1–6
- [25] Qiang T and Chu J 2012 The design of the infrared analyzer for unconventional emission components of ethanol gasoline vehicle *World Automation Congress* pp 1–4

APÊNDICE C
Diagramas Esquemáticos das Placas
Eletrônicas Desenvolvidas

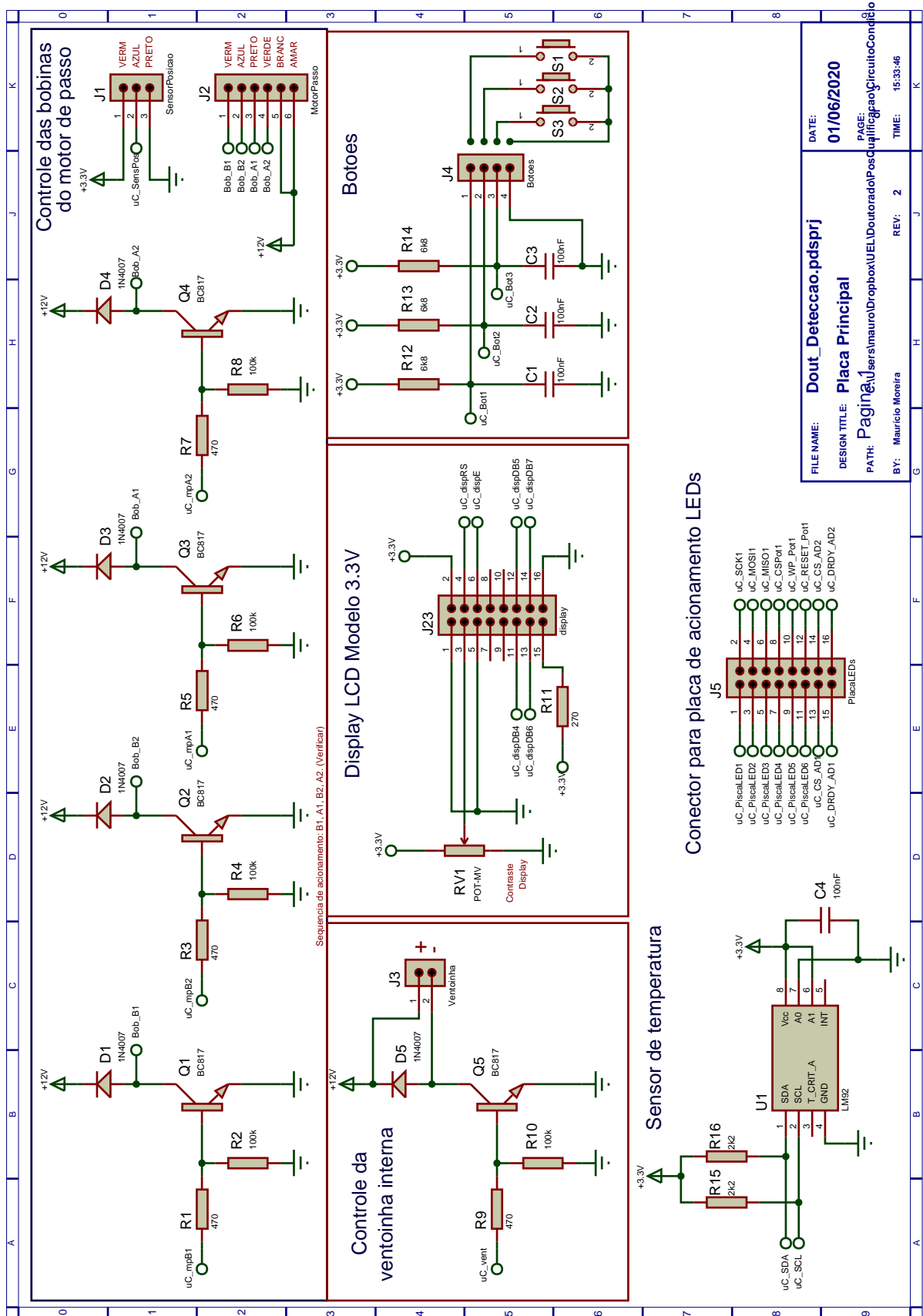


Figura C1: Diagrama esquemático da placa do circuito de detecção e condicionamento de sinais - folha 1.

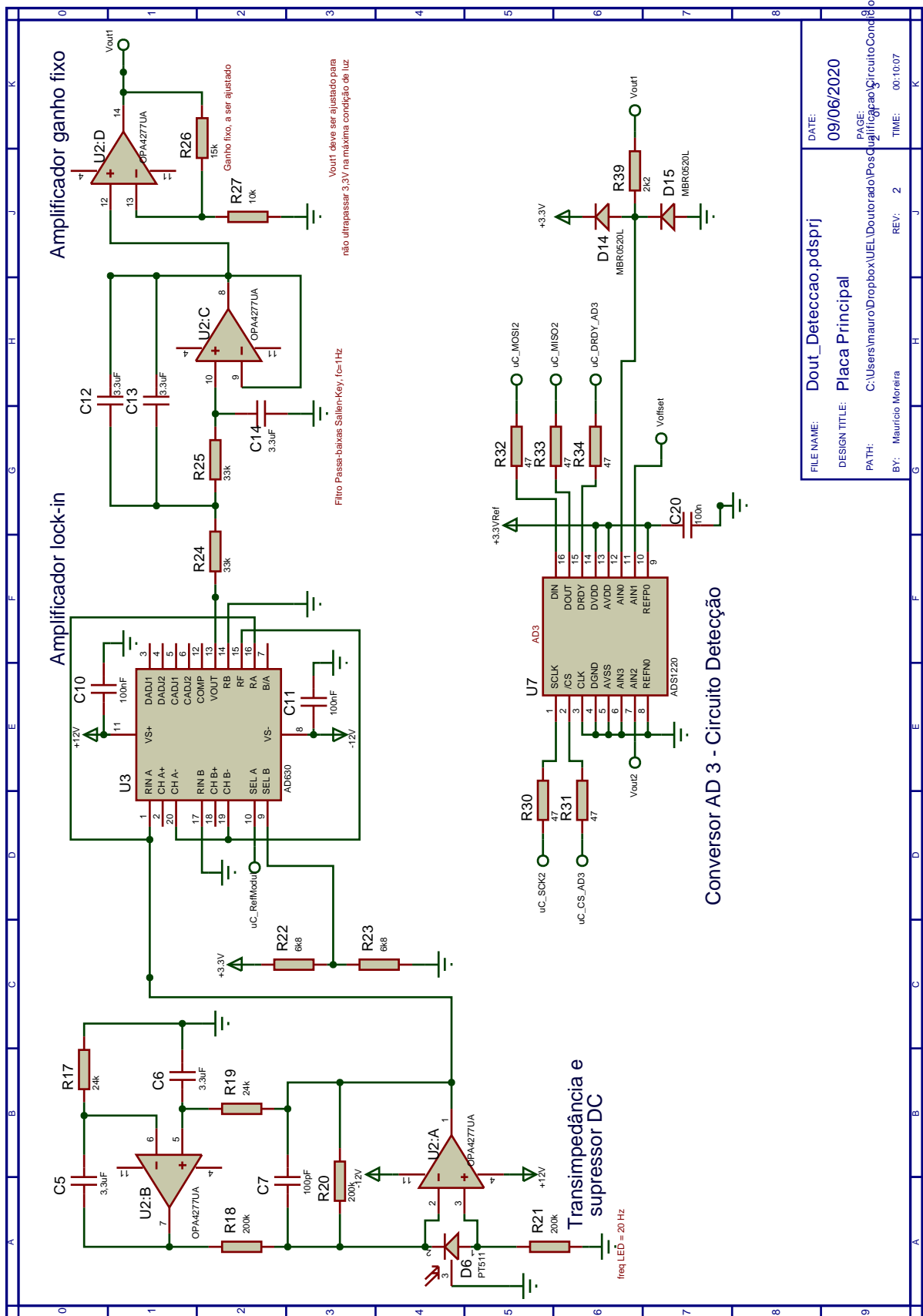
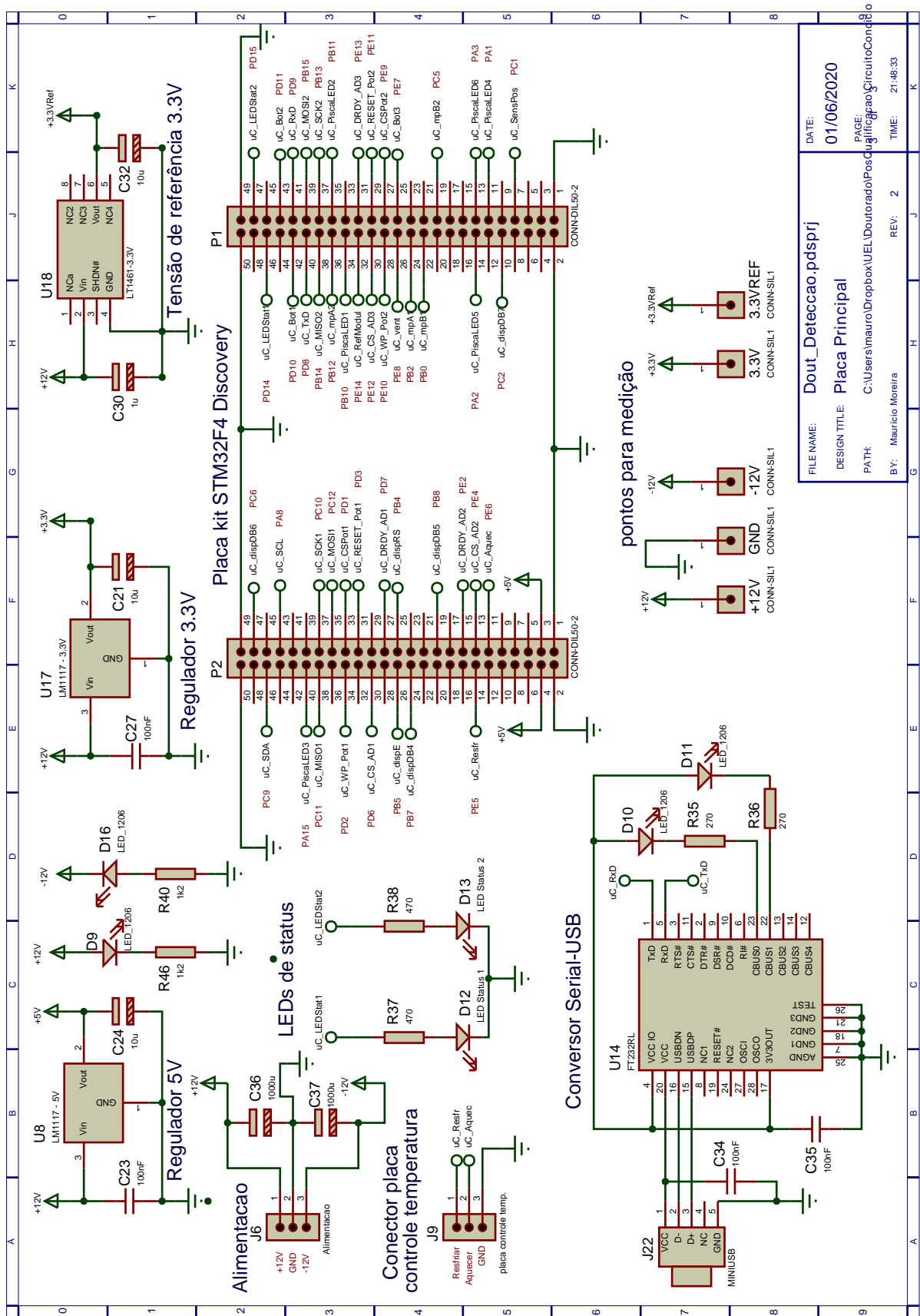


Figura C2: Diagrama esquemático da placa do circuito de detecção e condicionamento de sinais - folha 2.



FILE NAME:	Dout_Deteccao.pdsprj
DESIGN TITLE:	Placa Principal
PATH:	C:\Users\mauro\Dropbox\UEEL\Doctorado\PosQualificacao\CircuitoCond
BY:	Mauricio Moreira
REV:	2
DATE:	01/06/2020
PAGE:	
TIME:	21:48:33

Figura C3: Diagrama esquemático da placa do circuito de detecção e condicionamento de sinais - folha 3.

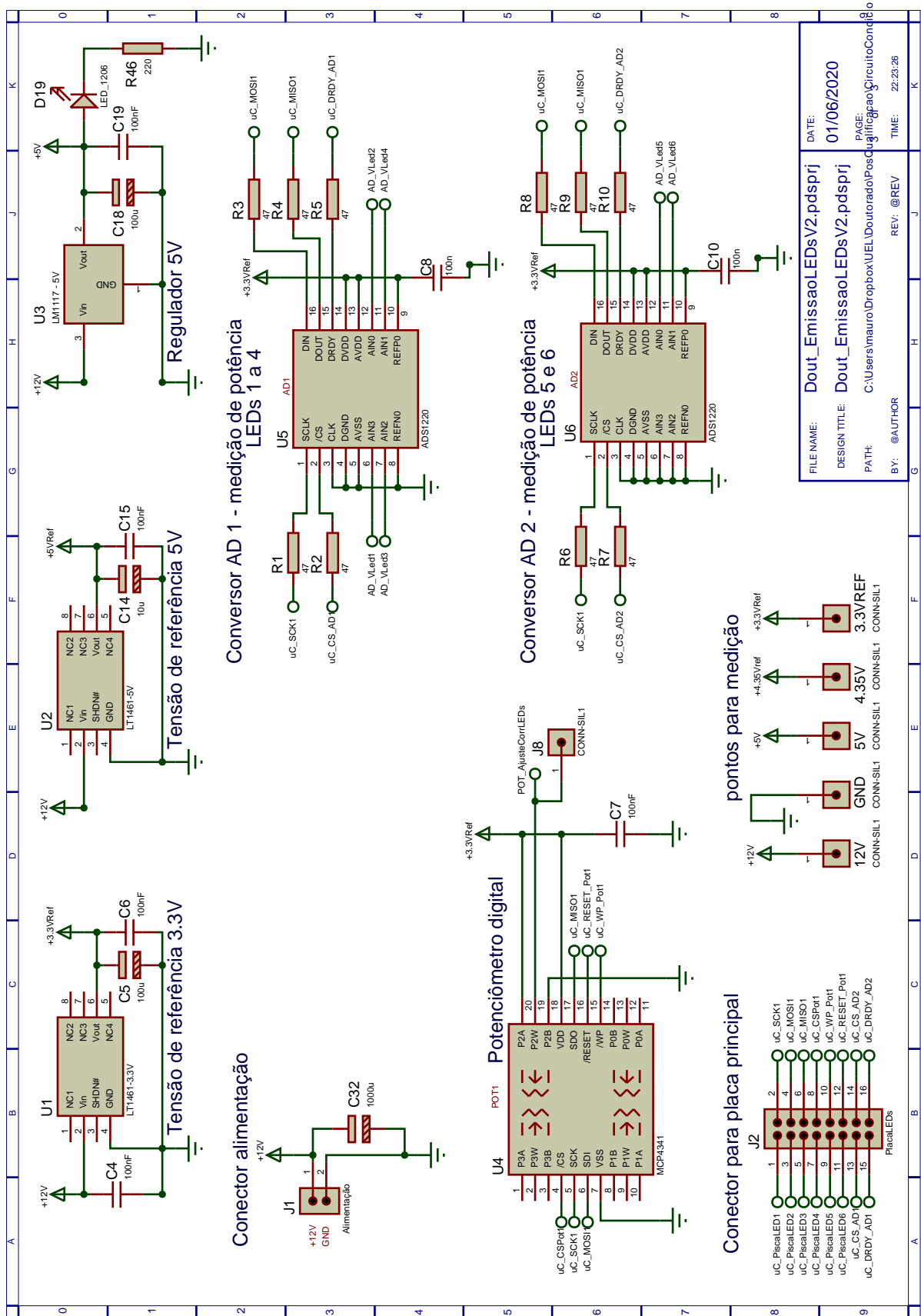


Figura C4: Diagrama esquemático da placa de acionamento dos LEDs infravermelhos - folha 1.



Figura C5: Diagrama esquemático da placa de acionamento dos LEDs infravermelhos - folha 2.

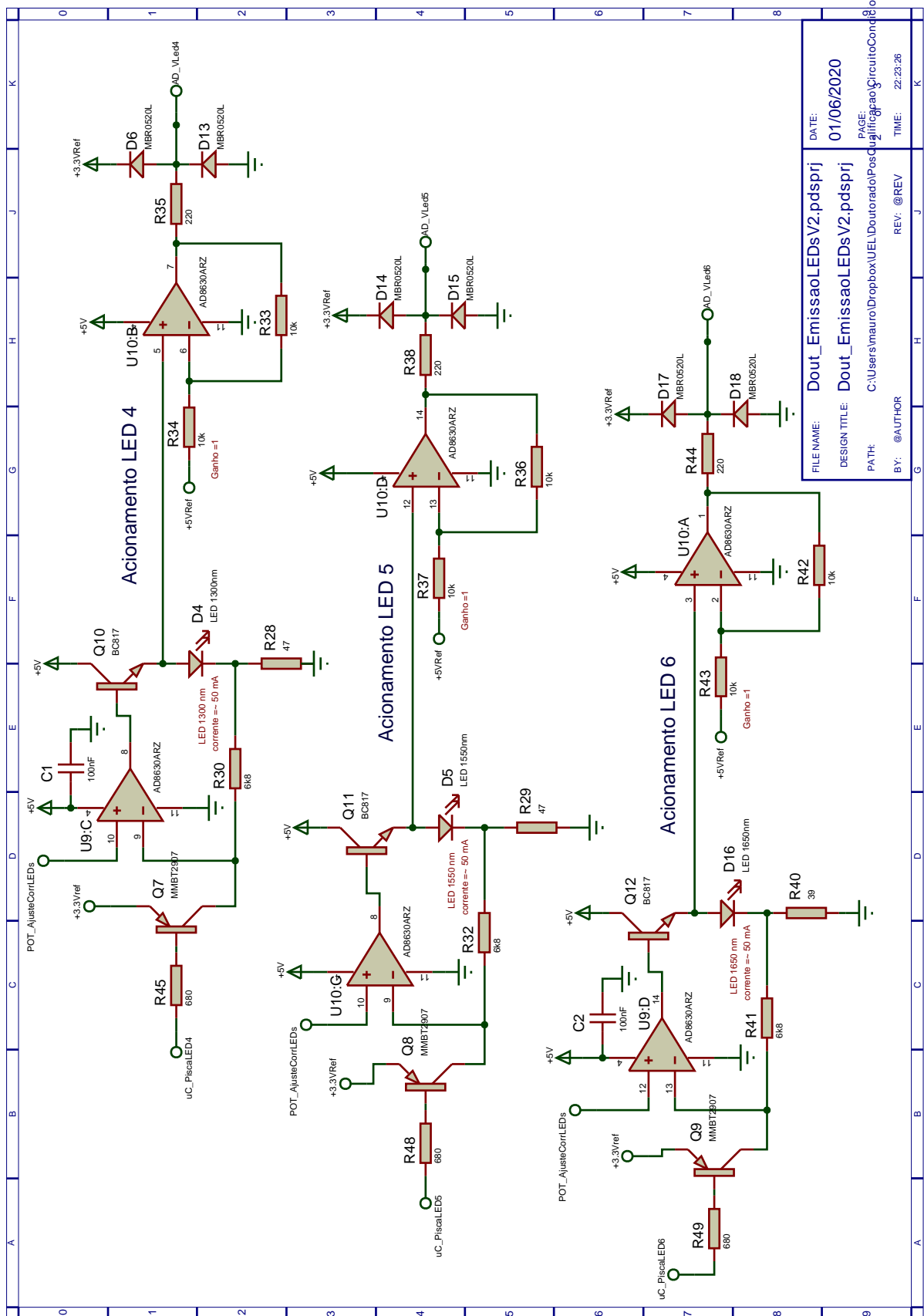
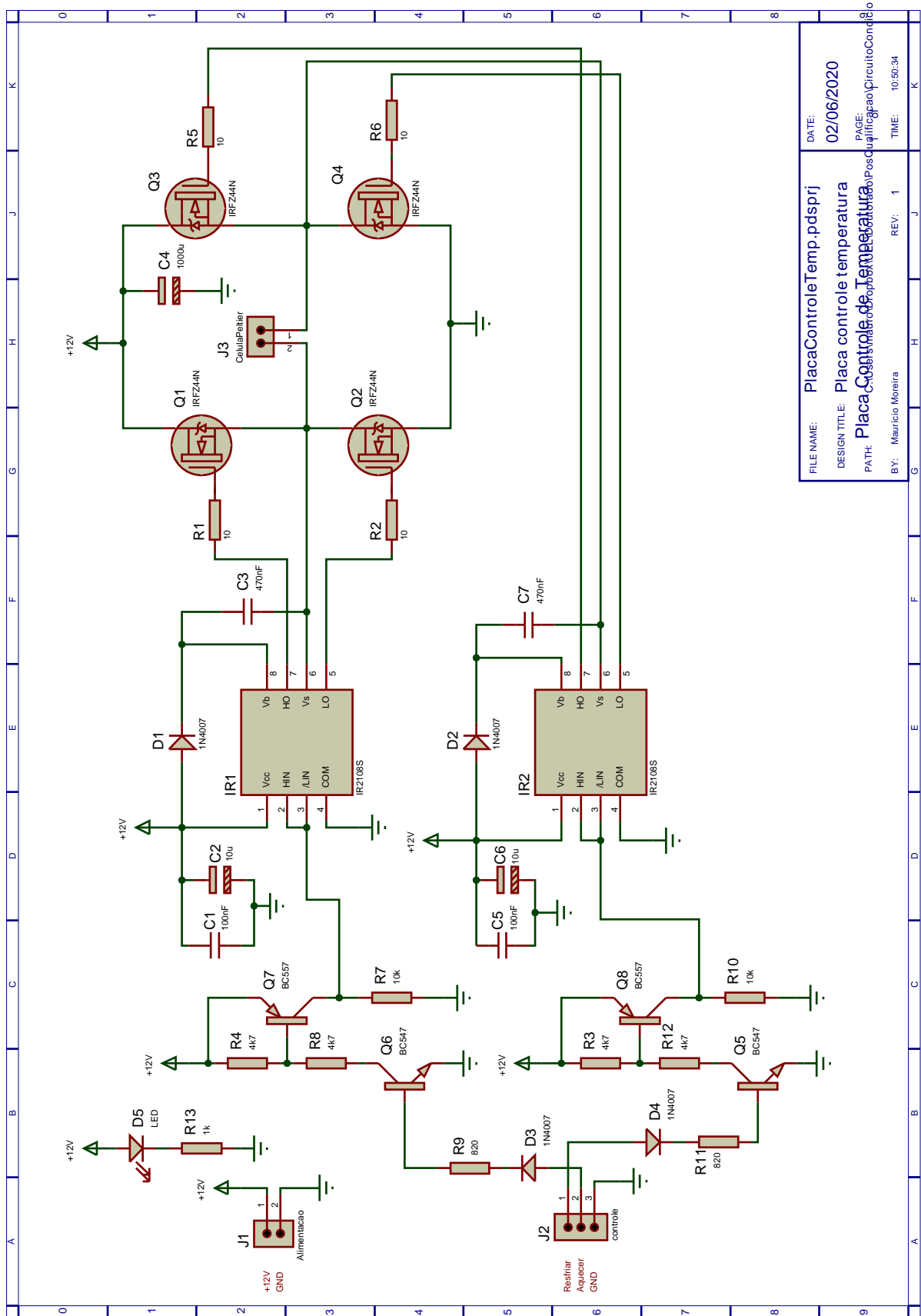


Figura C6: Diagrama esquemático da placa de acionamento dos LEDs infravermelhos - folha 3.



FILE NAME:	PlacaControleTemp.pdsprj	DATE:	02/06/2020
DESIGN TITLE:	Placa controle temperatura	PAGE:	1
PATH:	C:\top\proj\pos\qualificao\CircuitoControl	REV:	1
BY:	Mauricio Moreira	TIME:	10:50:34

Figura C7: Diagrama esquemático da placa de acionamento dos LEDs infravermelhos - folha 3.

ANEXO A

**Folhas de Dados dos LEDs Infravermelhos
e do Fotodetector**



LED910-01

- Infrared LED
- 910 nm, 13 mW
- Chip: AlGaAs, 350 x 350 μm
- 5 mm Clear Molding, Epoxy Resin
- Viewing Angle: 18°



Description

LED910-01 contains one AlGaAs LED chip die mounted on a lead frame hermetically sealed with a clear epoxy lens.

On forward bias, it emits a power radiation of typical **13 mW** at a peak wavelength at **910 nm**.

Maximum Ratings ($T_{\text{CASE}}=25^{\circ}\text{C}$)

Parameter	Symbol	Values		Unit
		Min.	Max.	
Power Dissipation	P_D		160	mW
Forward Current	I_F		100	mA
Pulse Forward Current *1	I_{FP}		500	mA
Reverse Voltage	V_F		5	V
Thermal Resistance	R_{THJA}		250	KW
Junction Temperature	T_J		120	$^{\circ}\text{C}$
Operating Temperature	T_{CASE}	- 40	+ 100	$^{\circ}\text{C}$
Storage Temperature	T_{STG}	- 40	+ 100	$^{\circ}\text{C}$
Lead Solder Temperature *2	T_{SLD}		+ 265	$^{\circ}\text{C}$

*1 duty=1%, pulse width = 10 μs

*2 must be completed within 3 seconds

Electro-Optical Characteristics ($T_{\text{CASE}}=25^{\circ}\text{C}$)

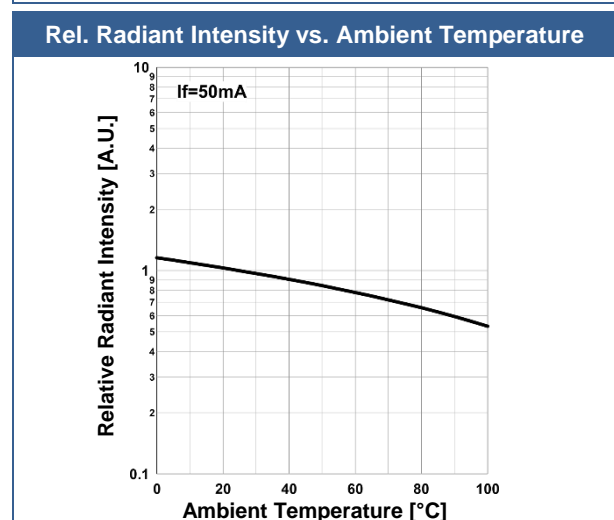
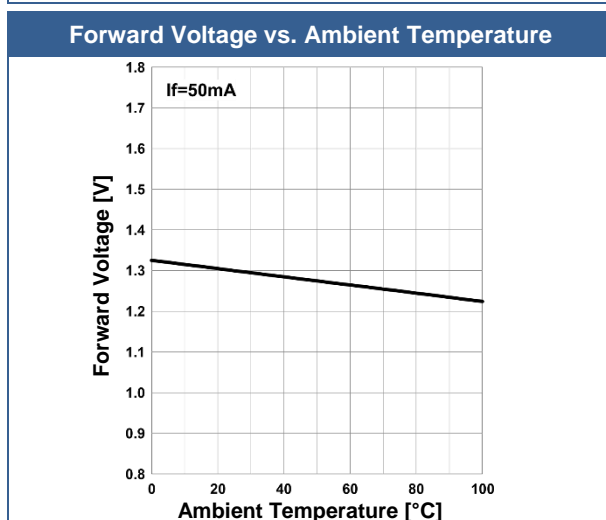
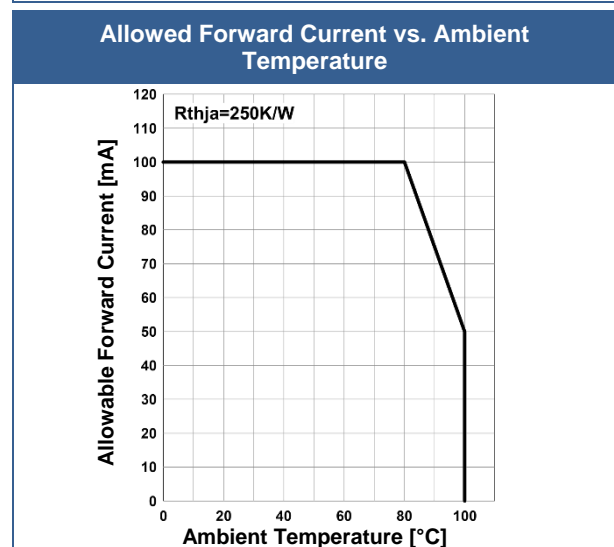
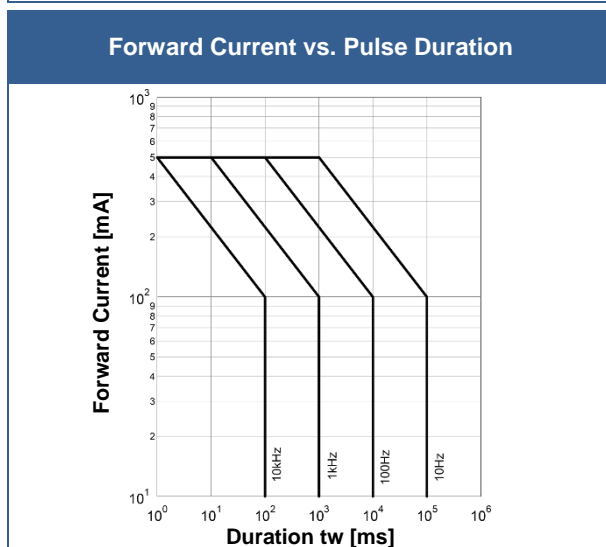
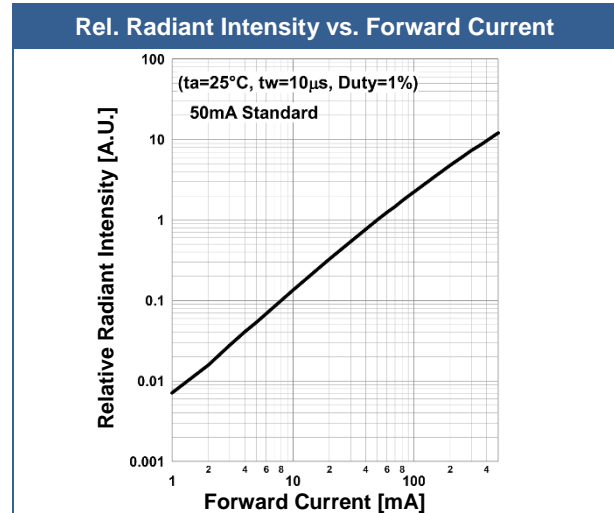
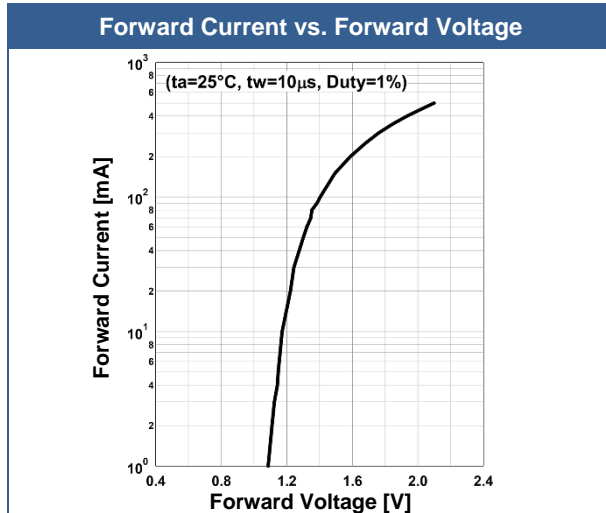
Parameter	Symbol	Conditions	Values			Unit
			Min.	Typ.	Max.	
Peak Wavelength	λ_P	$I_F=50\text{mA}$	895		925	nm
Half Width	$\Delta\lambda$	$I_F=50\text{mA}$		46		nm
Forward Voltage	V_F	$I_F=50\text{mA}$		1.3	1.6	V
	V_{FP}	$I_{FP}=500\text{mA}$		2.1		
Radiated Power *1	P_O	$I_F=50\text{mA}$		13		mW
		$I_{FP}=500\text{mA}$		150		
Radiant Intensity *2	I_E	$I_F=50\text{mA}$		75		mW/sr
		$I_{FP}=500\text{mA}$		900		
Viewing Angle	φ	$I_F=50\text{mA}$		18		deg.
Rise Time	t_R	$I_F=50\text{mA}$		30		ns
Fall Time	t_F	$I_F=50\text{mA}$		40		ns

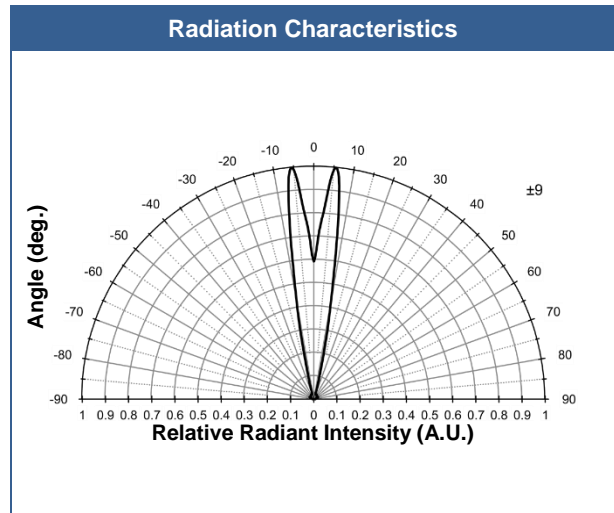
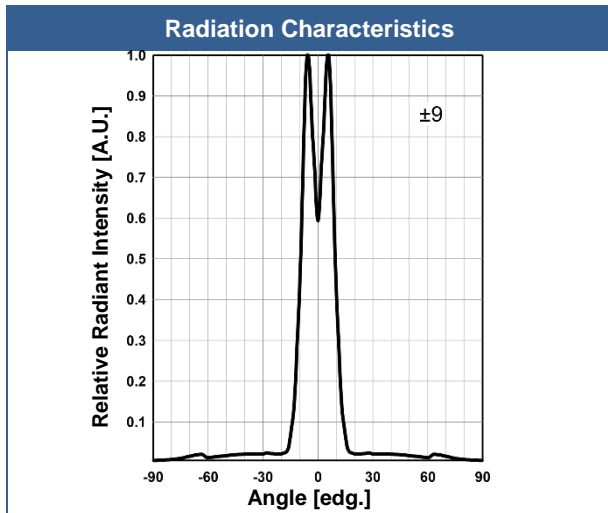
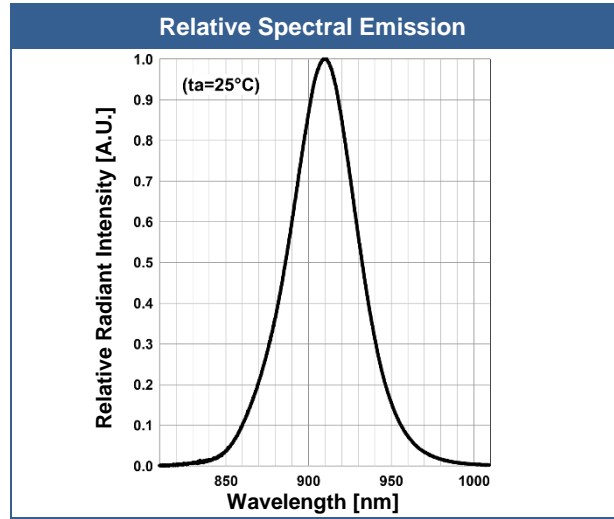
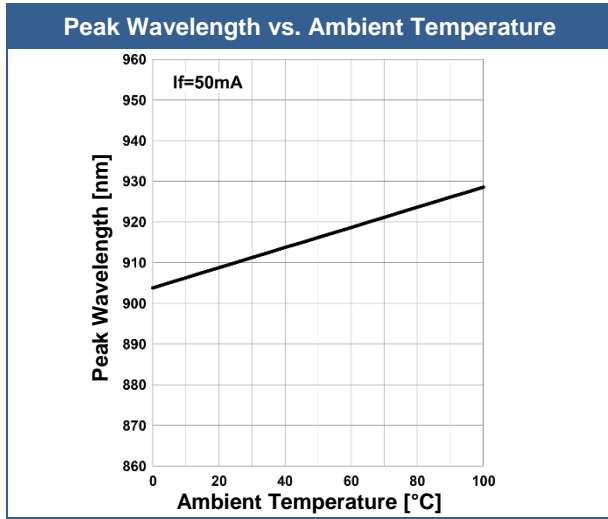
*1 measured by S3584-08

*2 measured by CIE127-2007 Condition B

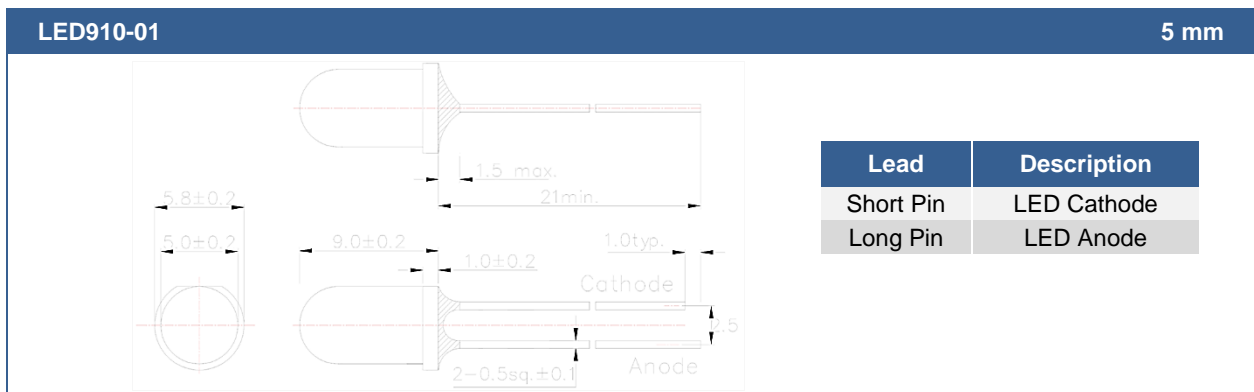


Typical Performance Curves





Outline Dimensions



All Dimensions in mm

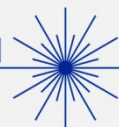


ROITHNER LASERTECHNIK GmbH

WIEDNER HAUPTSTRASSE 76
TEL. +43 1 586 52 43 -O. FAX. -44

1040 VIENNA
OFFICE@ROITHNER-LASER.COM

AUSTRIA



DIL5104

- **INFRARED** Light Emitting Diode
- **1040 nm, 9 mW**
- **AlGaAs**
- **5 mm epoxy package**



Description

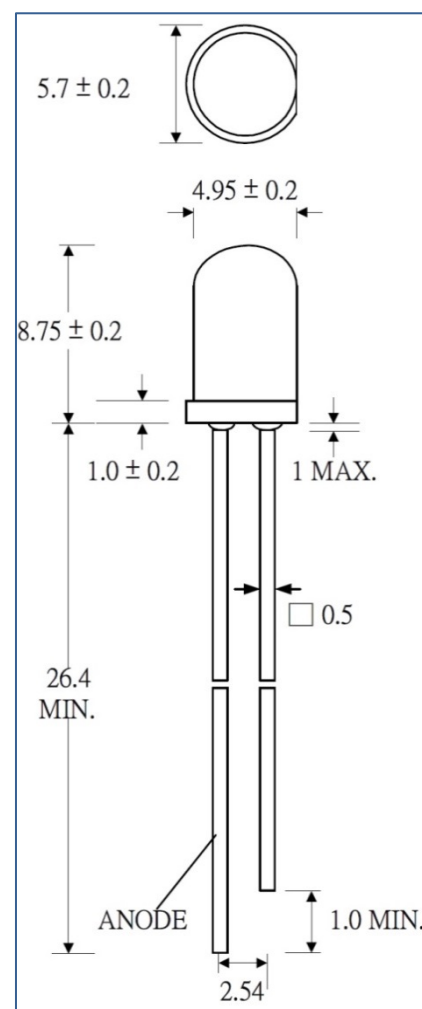
DIL5104 is an infrared LED, typically emitting at 1040 nm with an optical output power of 9 mW. It comes in a hermetically sealed clear 5 mm epoxy resin.

Maximum Rating (T_{CASE} = 25°C)

Parameter	Symbol	Values		Unit
		Min.	Max.	
Power Dissipation, DC	P _D		60	mW
Forward Current	I _{FM}		40	mA
Operating Temperature	T _{OPR}	- 20	+ 80	°C
Storage Temperature	T _{STG}	- 40	+ 85	°C
Soldering Temperature (max 3s)	T _{SOL}		+ 260	°C

Electro-Optical Characteristics (T_{CASE} = 25°C, I_F = 20 mA)

Parameter	Symbol	Values			Unit
		Min.	Typ.	Max.	
Peak Wavelength	λ _P		1040		nm
Spectral Width (FWHM)	Δλ		80		nm
Forward Voltage @ 20 mA	V _F		1.3	1.5	V
Reverse Voltage (I _R = 10 μA)	V _R			5	V
Radiant Power @ 20 mA	ϑ _e		9		mW
Viewing Half Angle	ϑ _{1/2}		15		deg.



All dimensions in mm



ROITHNER LASERTECHNIK GmbH

WIEDNER HAUPTSTRASSE 76

1040 VIENNA

AUSTRIA

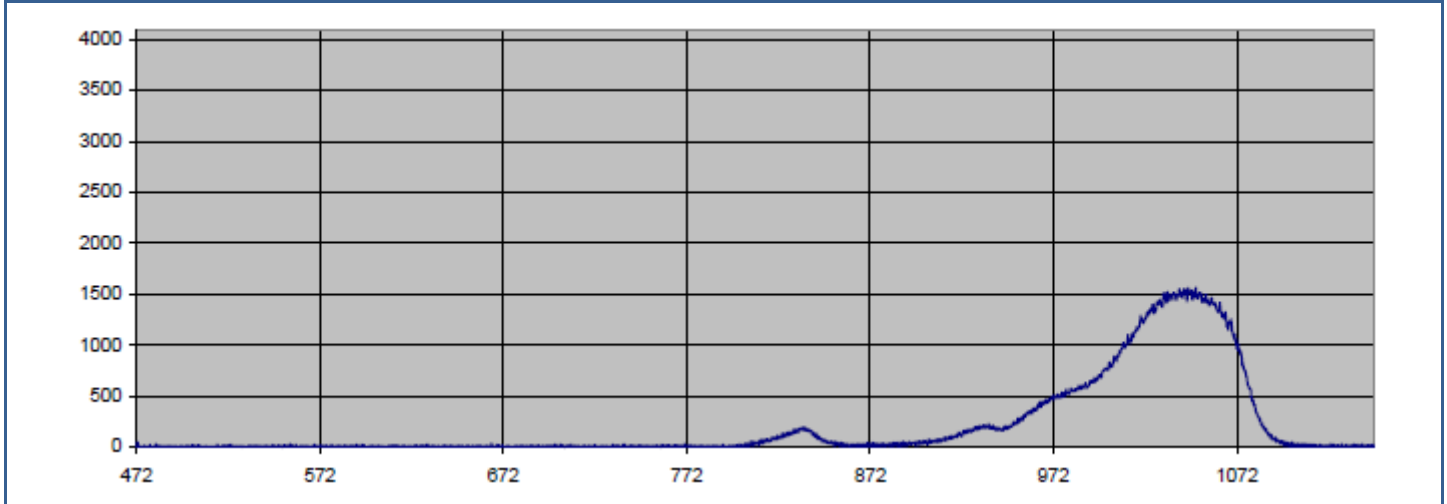
TEL. +43 1 586 52 43 -0. FAX. -44

OFFICE@ROITHNER-LASER.COM



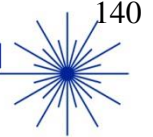
Performance Characteristics

Spectral Emission



© All Rights Reserved

The above specifications are for reference purpose only and subjected to change without prior notice

**PRELIMINARY**

LED1200S-03

Rev A1

- Infrared LED
- 1200 nm, 9.2 mW
- Chip: InGaAsP, 300 x 300 μm , 1 pc.
- 5 mm Clear Molding, Epoxy Resin
- Viewing Angle: 28°



Description



LED1200S-03 contains one InGaAsP LED chip die with a typical peak wavelength of **1200 nm** and radiation power of **9.2 mW**. It comes in $\varnothing 5$ mm clear molding package with soldered lead frame (lead free) and lens molded with epoxy resin.

Maximum Ratings ($T_{\text{CASE}}=25^{\circ}\text{C}$)

Parameter	Symbol	Values		Unit
		Min.	Max.	
Power Dissipation	P_D		130	mW
Forward Current	I_F		100	mA
Pulse Forward Current *1	I_{FP}		1000	mA
Reverse Voltage	V_F		3	V
Thermal Resistance	R_{THJA}		250	K/W
Junction Temperature	T_J		120	$^{\circ}\text{C}$
Operating Temperature	T_{CASE}	- 40	+ 100	$^{\circ}\text{C}$
Storage Temperature	T_{STG}	- 40	+ 100	$^{\circ}\text{C}$
Lead Solder Temperature *2	T_{SLD}		+ 265	$^{\circ}\text{C}$

*1 duty=1%, pulse width = 10 μs

*2 must be completed within 5 seconds

Electro-Optical Characteristics ($T_{\text{CASE}}=25^{\circ}\text{C}$)

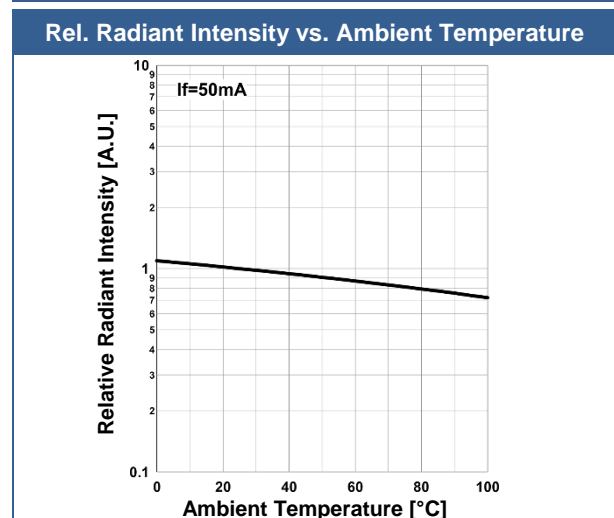
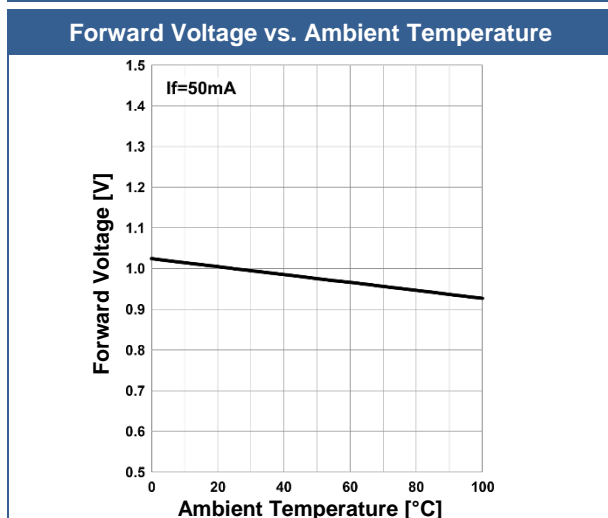
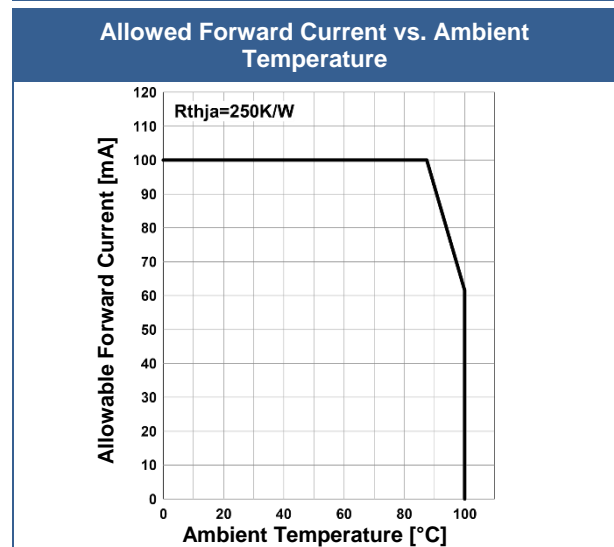
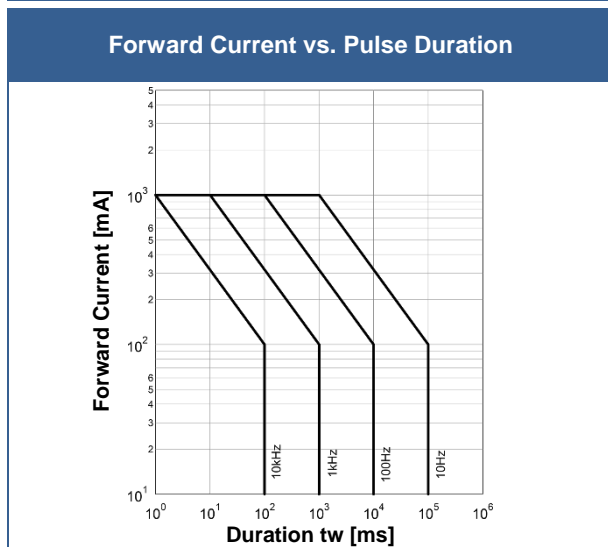
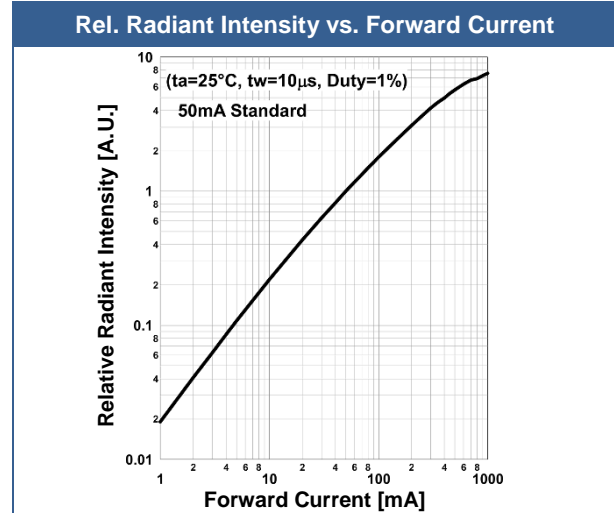
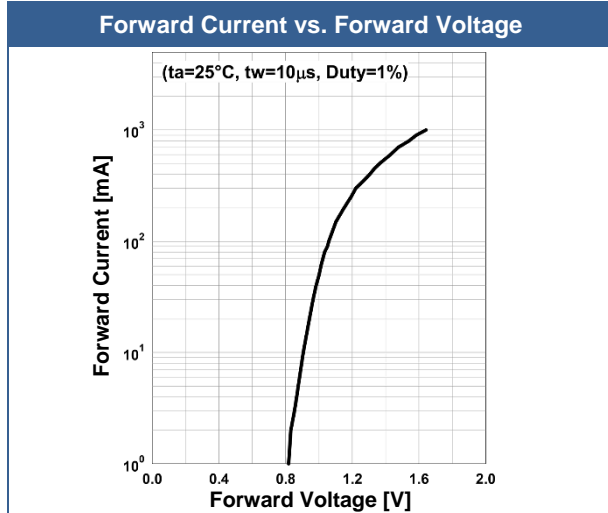
Parameter	Symbol	Conditions	Values			Unit
			Min.	Typ.	Max.	
Peak Wavelength	λ_P	$I_F=50\text{mA}$	1150		1250	nm
Half Width	$\Delta\lambda$	$I_F=50\text{mA}$		90		nm
Forward Voltage	V_F	$I_F=50\text{mA}$		1.0	1.3	V
	V_{FP}	$I_{FP}=1\text{A}$		1.6		
Radiated Power *1	P_O	$I_F=50\text{mA}$		9.2		mW
		$I_{FP}=1\text{A}$		69		
Radiant Intensity *2	I_E	$I_F=50\text{mA}$		67		mW/sr
		$I_{FP}=1\text{A}$		500		
Viewing Angle	$2\theta_{1/2}$	$I_F=50\text{mA}$		28		deg.
Rise Time	t_r	$I_F=50\text{mA}$		30		ns
Fall Time	t_f	$I_F=50\text{mA}$		70		ns

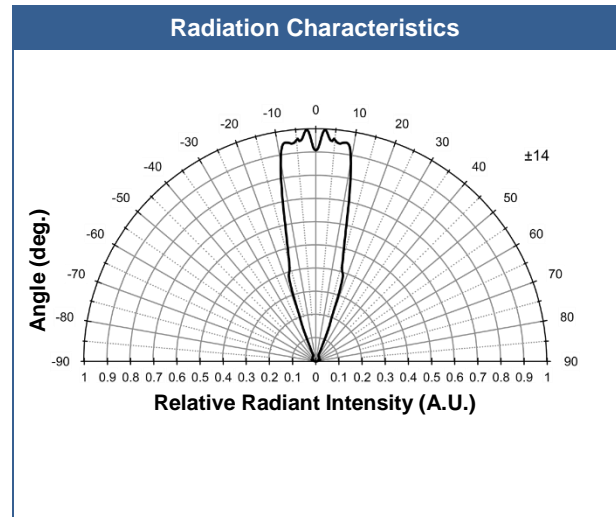
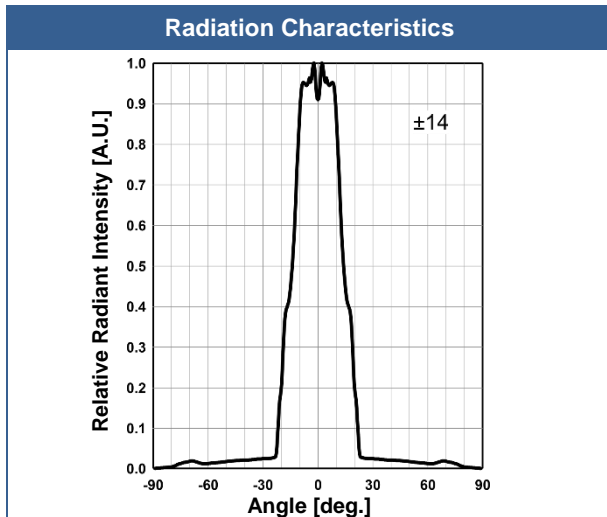
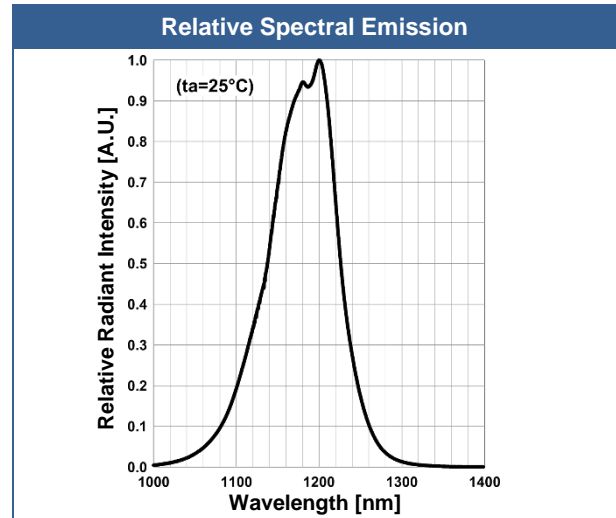
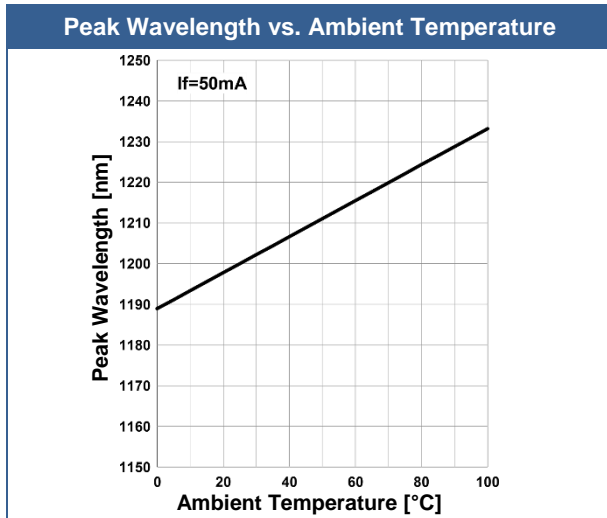
*1 measured by G8370-85

*2 measured by Ando Optical Multi Meter AQ2140 & AQ2742

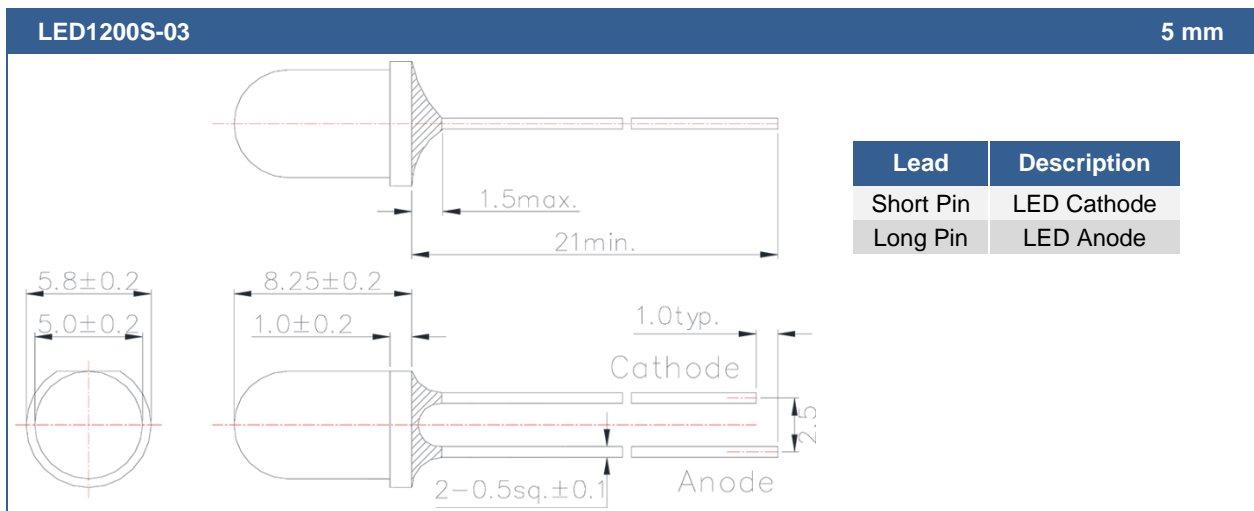


Typical Performance Curves





Outline Dimensions



All Dimensions in mm

LED1300-35xx High Power InGaAsP NIR LED

LED1300-35xx is an InGaAsP LED mounted on a metal stem and covered with epoxy resin or glass lens can.

On forward bias it emits a spectral band of radiation, which peaks at 1300 nm.

Absolute Maximum Ratings

Item	Symbol	Maximum Rated Value	Unit	Ambient Temperature
Power Dissipation	P _D	120	mW	T _a =25°C
Forward Current	I _F	100	mA	T _a =25°C
Pulse Forward Current	I _{FP}	1000	mA	T _a =25°C
Reverse Voltage	V _R	5	V	T _a =25°C
Operating Temperature	T _{OPR}	-20 ~ +90	°C	
Storage Temperature	T _{STG}	-30 ~ +100	°C	
Soldering Temperature	T _{SOL}	260	°C	

‡ Pulse Forward Current condition: Duty = 1% and Pulse Width = 1 μs.

‡ Soldering condition: Soldering condition must be completed within 3 seconds at 260°C

Electro-Optical Characteristics [T_a=25°C]

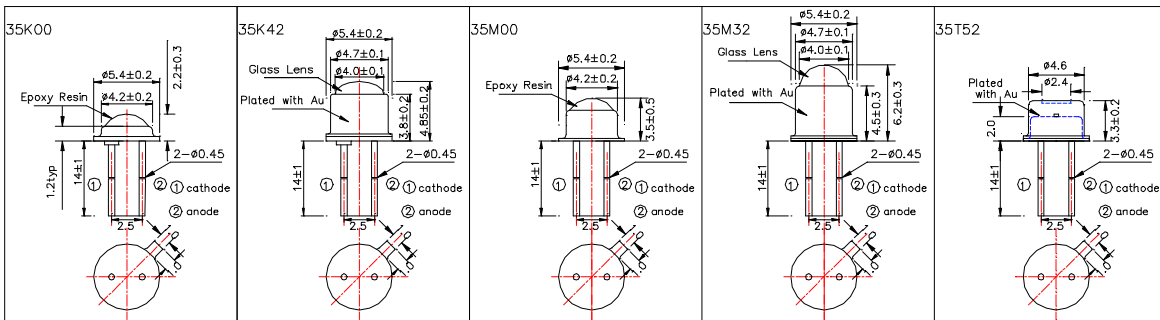
Item	Symbol	Condition	Minimum	Typical	Maximum	Unit
Forward Voltage	V _F	I _F =50mA		1.10	1.40	V
Reverse Current	I _R	V _R =5V			10	uA
Peak Wavelength	λ _P	I _F =50mA	1250	1300	1350	nm
Half Width	Δλ	I _F =50mA		100		nm
Rise Time	t _r	I _F =50mA		10		ns
Fall Time	t _f	I _F =50mA		10		ns

Radiated Power [T_a=25°C]

Type No.	Radiated Power at I _F =50mA unit:mW			Viewing Half Angle: Q _{1/2}
	Minimum	Typical	Maximum	
L1300-35K00		1.8		±50°
L1300-35K42		2.0		±6°
L1300-35M00		3.0		±50°
L1300-35M32		3.0		±15°
L1300-35T52		1.6		±55°

‡ Radiated Power is measured by Ando Optical Multi Meter AQ2140 & AQ2742

Outer dimension (Unit: mm)



**PRELIMINARY**

LED1550S-03

- Infrared LED
- 1550 nm, 3.8 mW
- Chip: InGaAsP, 300 x 300 μm , 1 pc.
- 5 mm Clear Molding, Epoxy Resin
- Viewing Angle: 26°



Description



LED1550S-03 contains one InGaAsP LED chip die with a typical peak wavelength of **1550 nm** and radiation power of **3.8 mW**. It comes in $\varnothing 5$ mm clear molding package with soldered lead frame (lead free) and lens molded with epoxy resin.

Maximum Ratings ($T_{\text{CASE}}=25^{\circ}\text{C}$)

Parameter	Symbol	Values		Unit
		Min.	Max.	
Power Dissipation	P_D		130	mW
Forward Current	I_F		100	mA
Pulse Forward Current *1	I_{FP}		1000	mA
Reverse Voltage	V_F		5	V
Thermal Resistance	R_{THJA}		250	K/W
Junction Temperature	T_J		120	$^{\circ}\text{C}$
Operating Temperature	T_{CASE}	- 40	+ 100	$^{\circ}\text{C}$
Storage Temperature	T_{STG}	- 40	+ 100	$^{\circ}\text{C}$
Lead Solder Temperature *2	T_{SLD}		+ 265	$^{\circ}\text{C}$

*1 duty=1%, pulse width = 10 μs

*2 must be completed within 5 seconds

Electro-Optical Characteristics ($T_{\text{CASE}}=25^{\circ}\text{C}$)

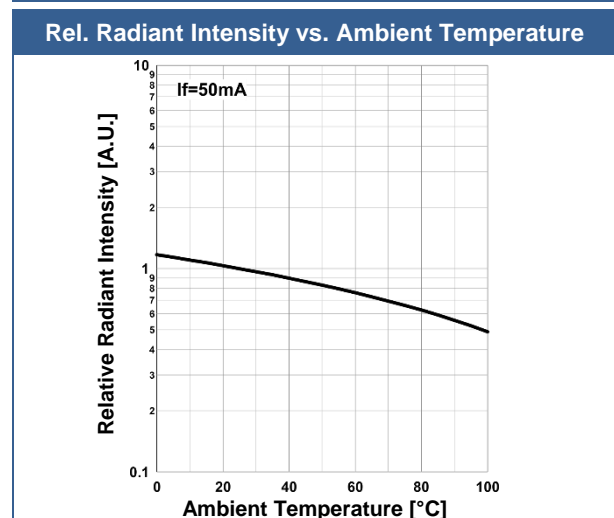
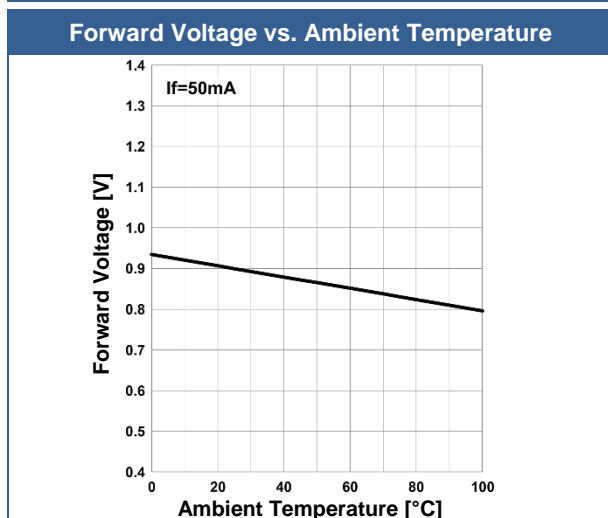
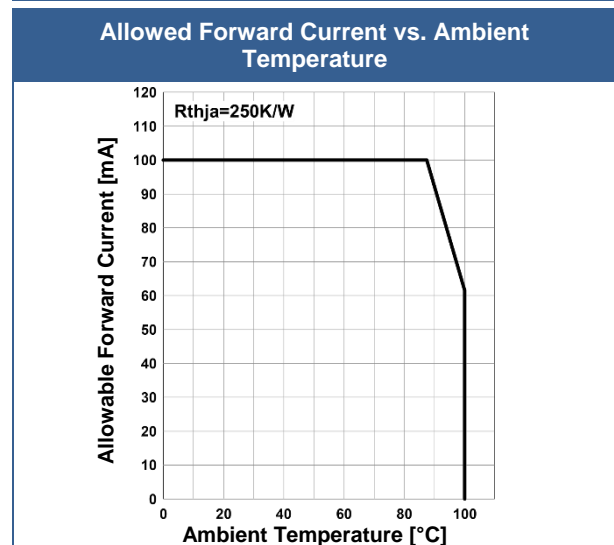
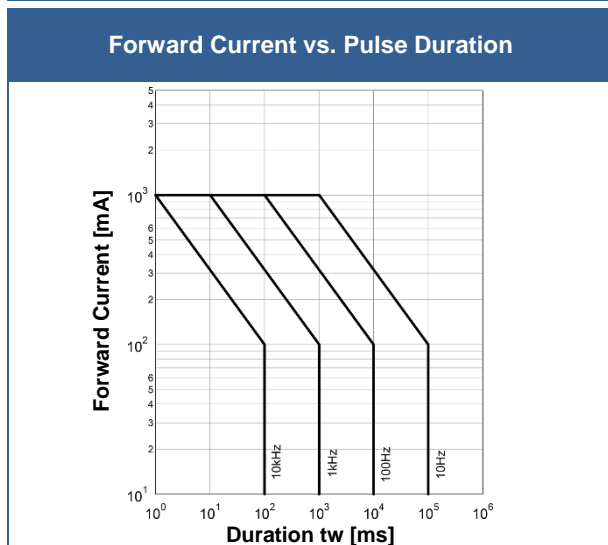
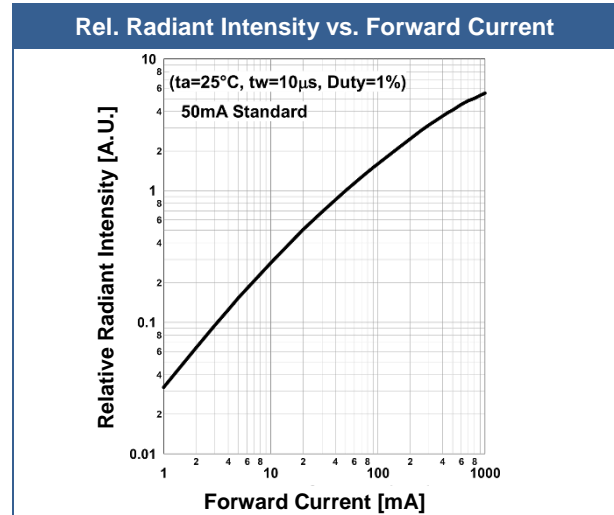
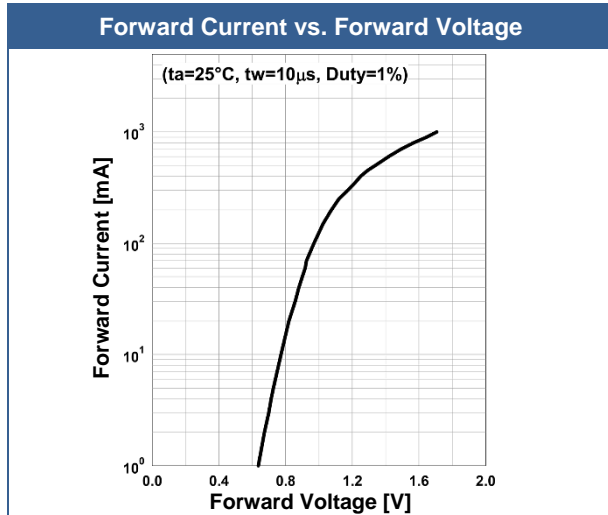
Parameter	Symbol	Conditions	Values			Unit
			Min.	Typ.	Max.	
Peak Wavelength	λ_P	$I_F=50\text{mA}$	1500		1600	nm
Half Width	$\Delta\lambda$	$I_F=50\text{mA}$		120		nm
Forward Voltage	V_F	$I_F=50\text{mA}$		0.9	1.3	V
	V_{FP}	$I_{FP}=1\text{A}$		1.7		
Radiated Power *1	P_O	$I_F=50\text{mA}$		3.8		mW
		$I_{FP}=1\text{A}$		21		
Radiant Intensity *2	I_E	$I_F=50\text{mA}$		49		mW/sr
		$I_{FP}=1\text{A}$		270		
Viewing Angle	$2\theta_{1/2}$	$I_F=50\text{mA}$		26		deg.
Rise Time	t_r	$I_F=50\text{mA}$		30		ns
Fall Time	t_f	$I_F=50\text{mA}$		70		ns

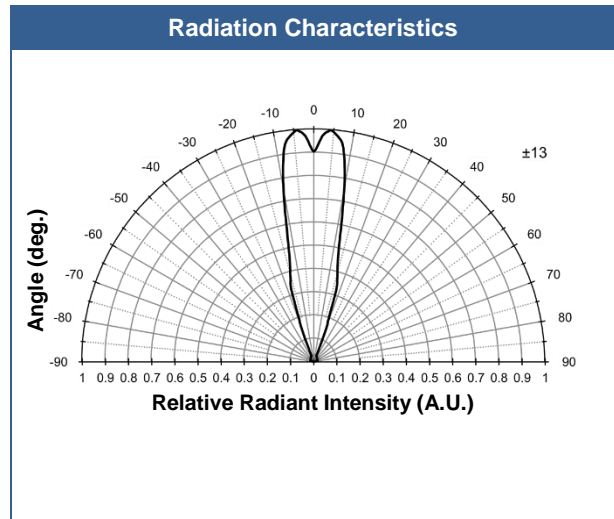
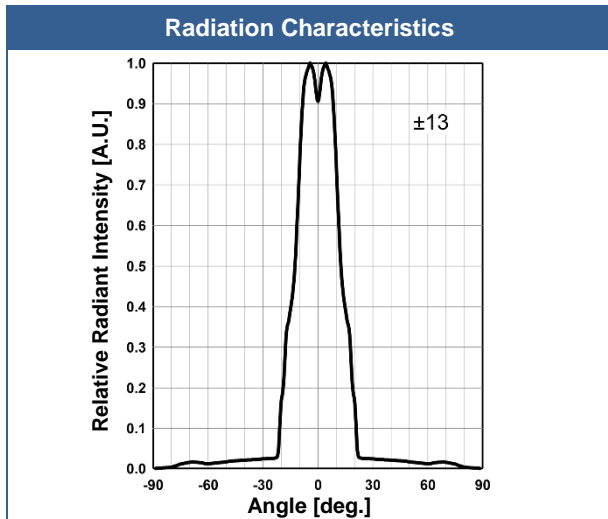
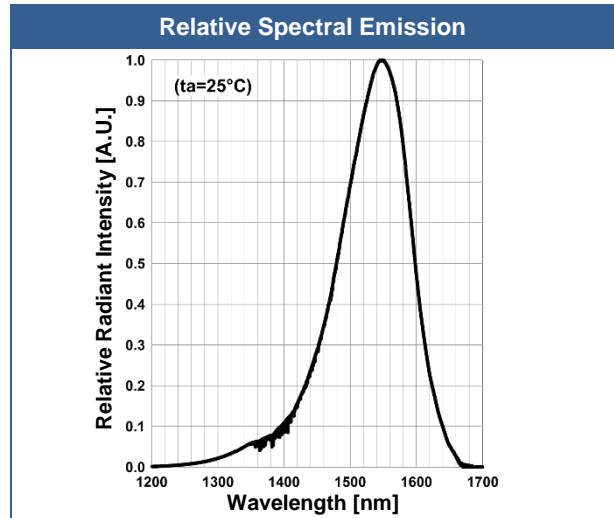
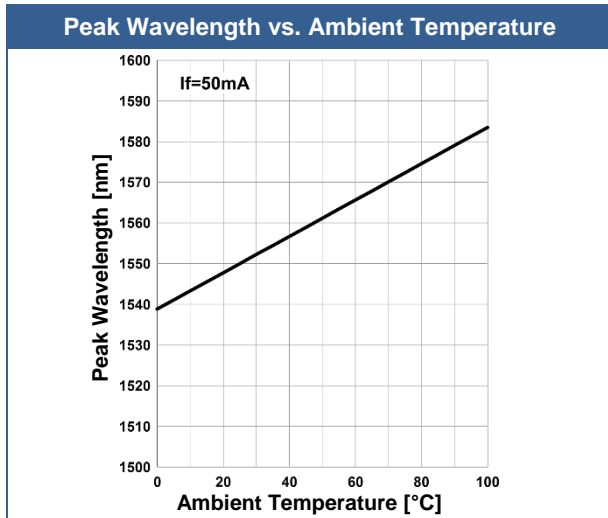
*1 measured by G8370-85

*2 measured by Ando Optical Multi Meter AQ2140 & AQ2742

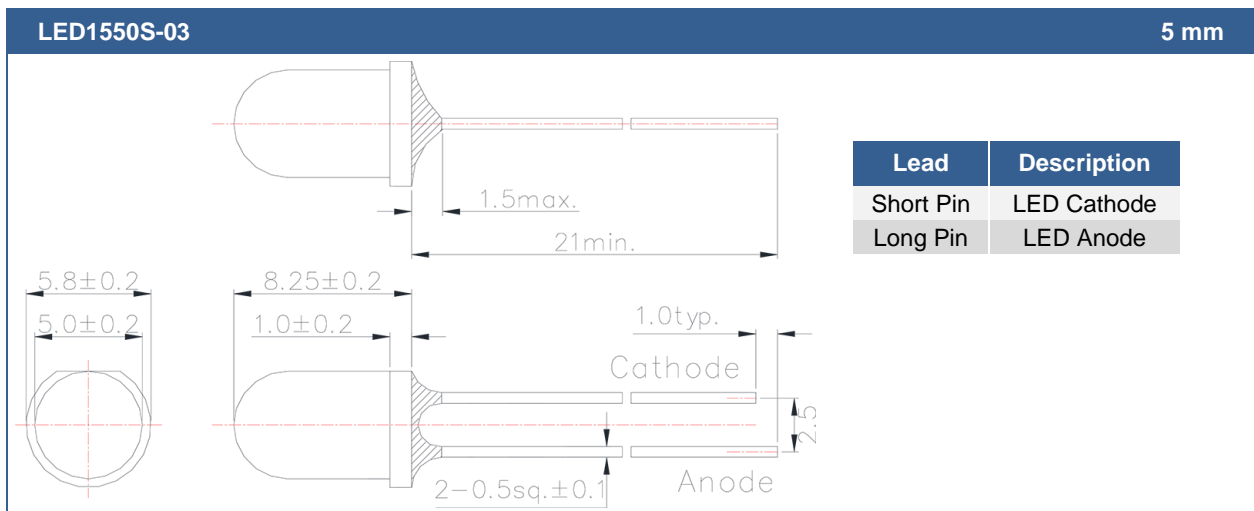


Typical Performance Curves





Outline Dimensions



All Dimensions in mm

**PRELIMINARY**

LED1650S-03

Rev A1

- Infrared LED
- 1650 nm, 2.3 mW
- Chip: InGaAsP, 300 x 300 μm , 1 pc.
- 5 mm Clear Molding, Epoxy Resin
- Viewing Angle: 24°



Description



LED1650S-03 contains one InGaAsP LED chip die with a typical peak wavelength of **1650 nm** and radiation power of **2.3 mW**. It comes in $\varnothing 5$ mm clear molding package with soldered lead frame (lead free) and lens molded with epoxy resin.

Maximum Ratings ($T_{\text{CASE}}=25^{\circ}\text{C}$)

Parameter	Symbol	Values		Unit
		Min.	Max.	
Power Dissipation	P_D		130	mW
Forward Current	I_F		100	mA
Pulse Forward Current *1	I_{FP}		1000	mA
Reverse Voltage	V_F		3	V
Thermal Resistance	R_{THJA}		250	K/W
Junction Temperature	T_J		120	$^{\circ}\text{C}$
Operating Temperature	T_{CASE}	- 40	+ 85	$^{\circ}\text{C}$
Storage Temperature	T_{STG}	- 40	+ 100	$^{\circ}\text{C}$
Lead Solder Temperature *2	T_{SLD}		+ 265	$^{\circ}\text{C}$

*1 duty=1%, pulse width = 10 μs

*2 must be completed within 5 seconds

Electro-Optical Characteristics ($T_{\text{CASE}}=25^{\circ}\text{C}$)

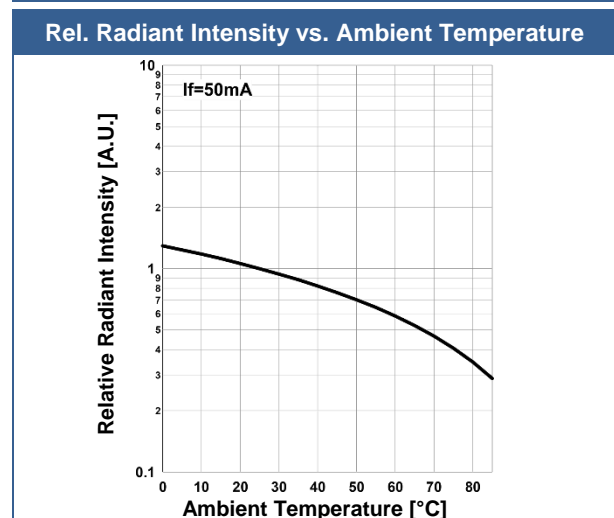
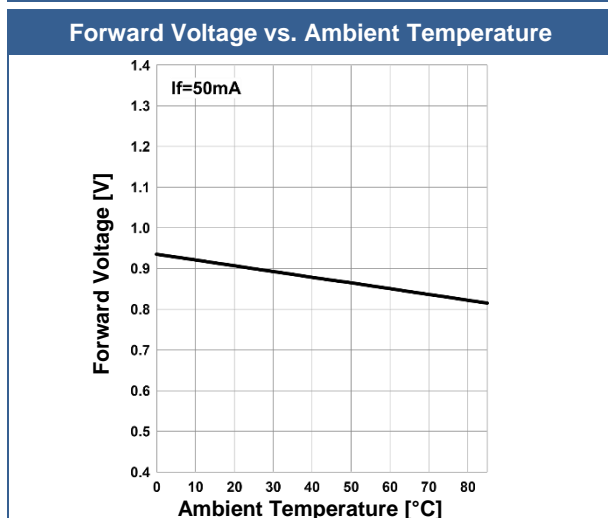
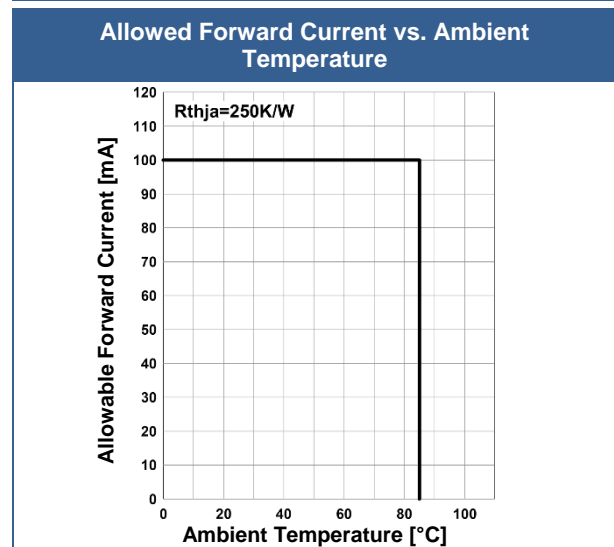
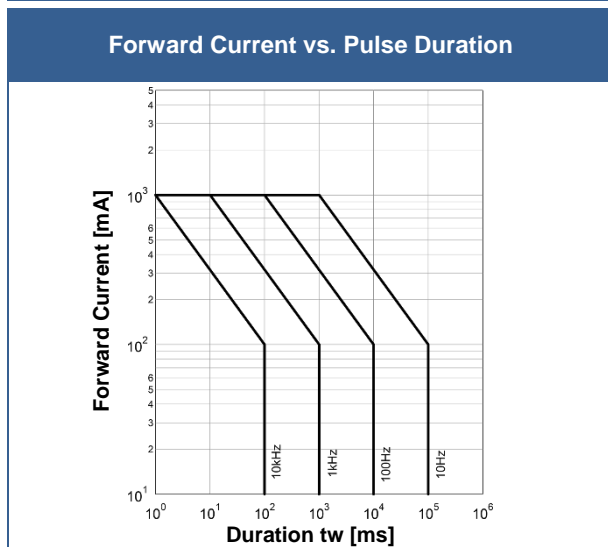
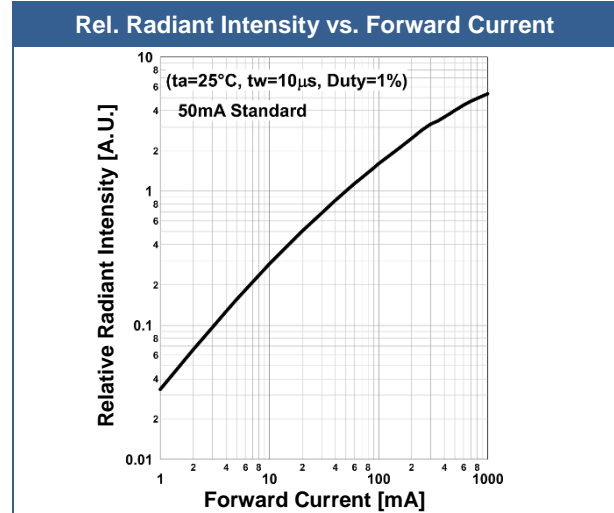
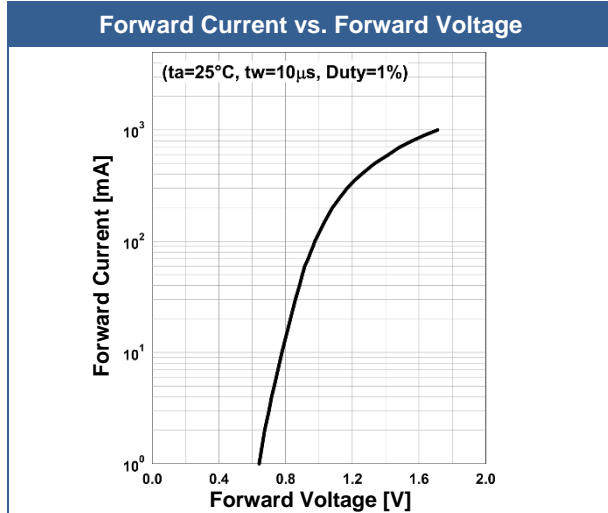
Parameter	Symbol	Conditions	Values			Unit
			Min.	Typ.	Max.	
Peak Wavelength	λ_P	$I_F=50\text{mA}$	1600		1700	nm
Half Width	$\Delta\lambda$	$I_F=50\text{mA}$		120		nm
Forward Voltage	V_F	$I_F=50\text{mA}$		0.9	1.3	V
	V_{FP}	$I_{FP}=1\text{A}$		1.7		
Radiated Power *1	P_O	$I_F=50\text{mA}$		2.3		mW
		$I_{FP}=1\text{A}$		12		
Radiant Intensity *2	I_E	$I_F=50\text{mA}$		37		mW/sr
		$I_{FP}=1\text{A}$		190		
Viewing Angle	$2\theta_{1/2}$	$I_F=50\text{mA}$		24		deg.
Rise Time	t_r	$I_F=50\text{mA}$		30		ns
Fall Time	t_f	$I_F=50\text{mA}$		70		ns

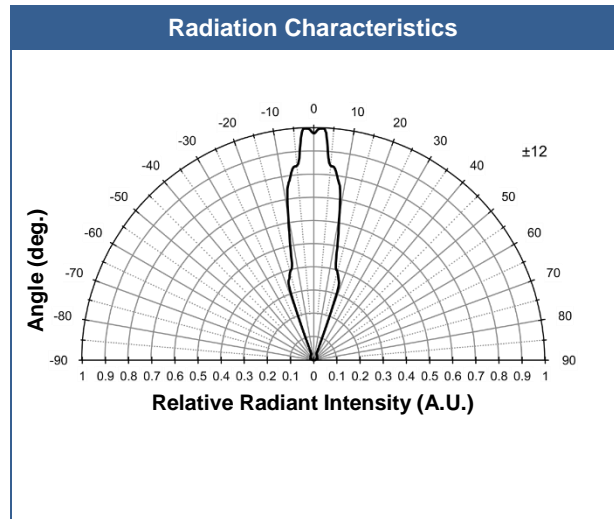
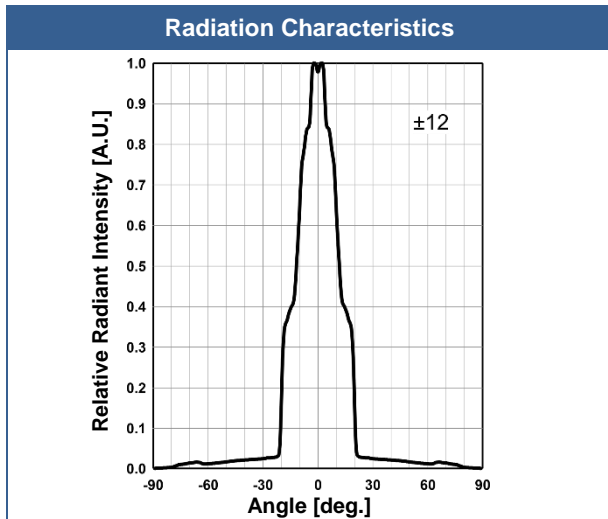
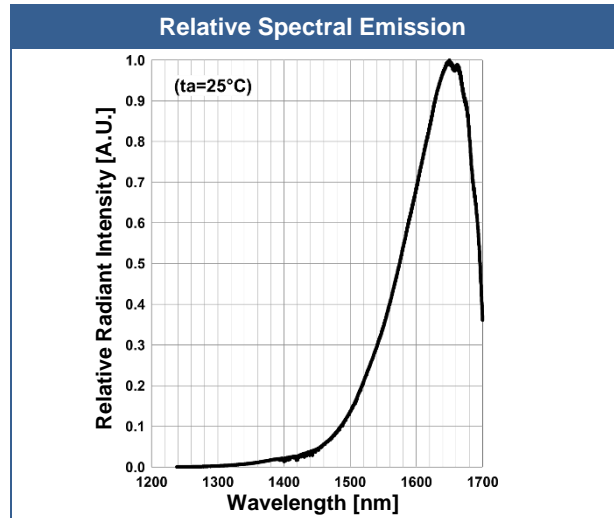
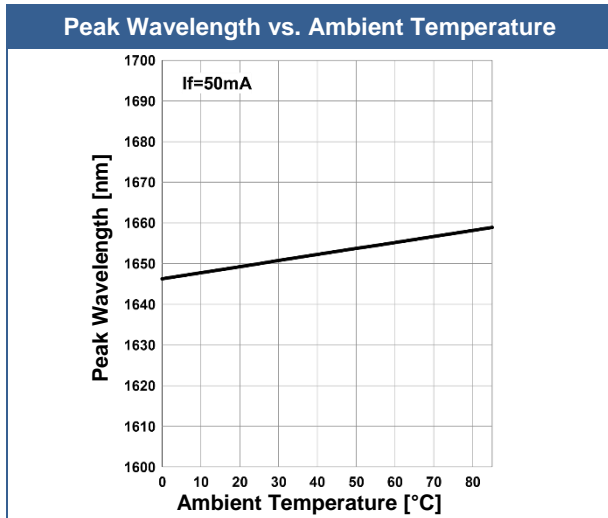
*1 measured by G8370-85

*2 measured by Ando Optical Multi Meter AQ2140 & AQ2742

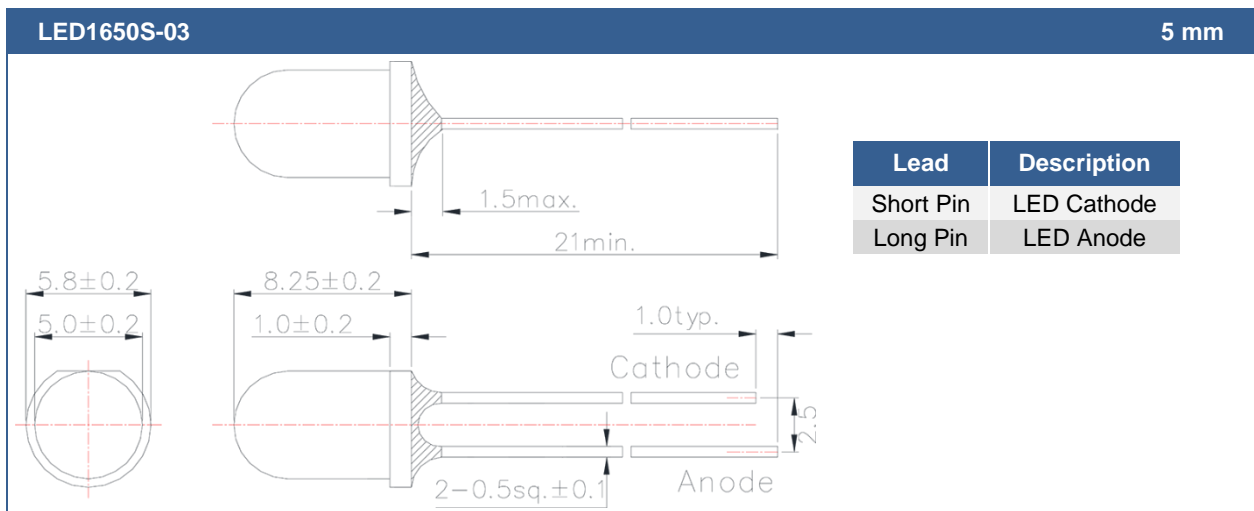


Typical Performance Curves





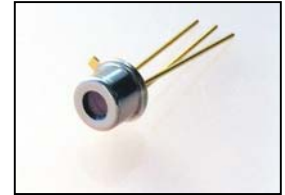
Outline Dimensions



All Dimensions in mm



PT511-2



TECHNICAL DATA

TO-Can PIN Photodiode

PT511-2 is an InGaAs pin structure based photodiode on InP by MOCVD method and planar diffusing technology. The sensitive area is $\varnothing 300\mu\text{m}$ respectively. TO-46 package with flat glass lens.



Absolute Maximum Ratings

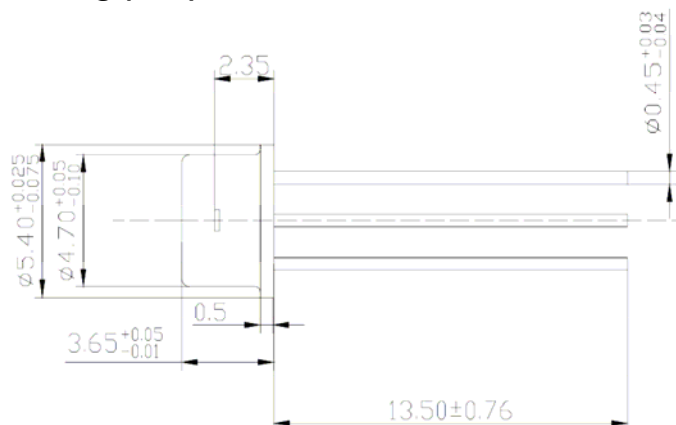
Item	Symbol	Value	Unit
Reverse Voltage	U_R	30	V
Reverse current	I_R	1000	μA
Forward current	I_F	5	mA
Power Dissipation	P_D	100	mW
Operating Temperature	T_{opr}	-40 ... +85	$^{\circ}\text{C}$
Storage Temperature	T_{stg}	-40 ... +125	$^{\circ}\text{C}$
Soldering Temperature *	T_{sol}	260	$^{\circ}\text{C}$

* must be completed within 10 seconds

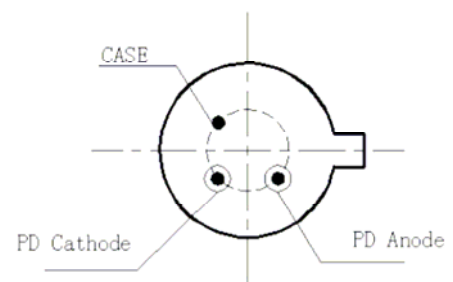
Specifications

Item	Value	Unit	
Wavelength Range	850 .. 1650	nm	
Sensitive Area	$\varnothing 300$	μm	
Unsaturated Power	10	dBm	
Responsivity	$\lambda=1300 \text{ nm}$	0.85	A/W
Capacitance (-5 V)	2	pF	
Dark Current (-5 V)	≤ 1	nA	
Rise/Fall time	3	ns	
Saturation Power	> 10	dB	

Drawing (mm)



Bottom View



ANEXO B

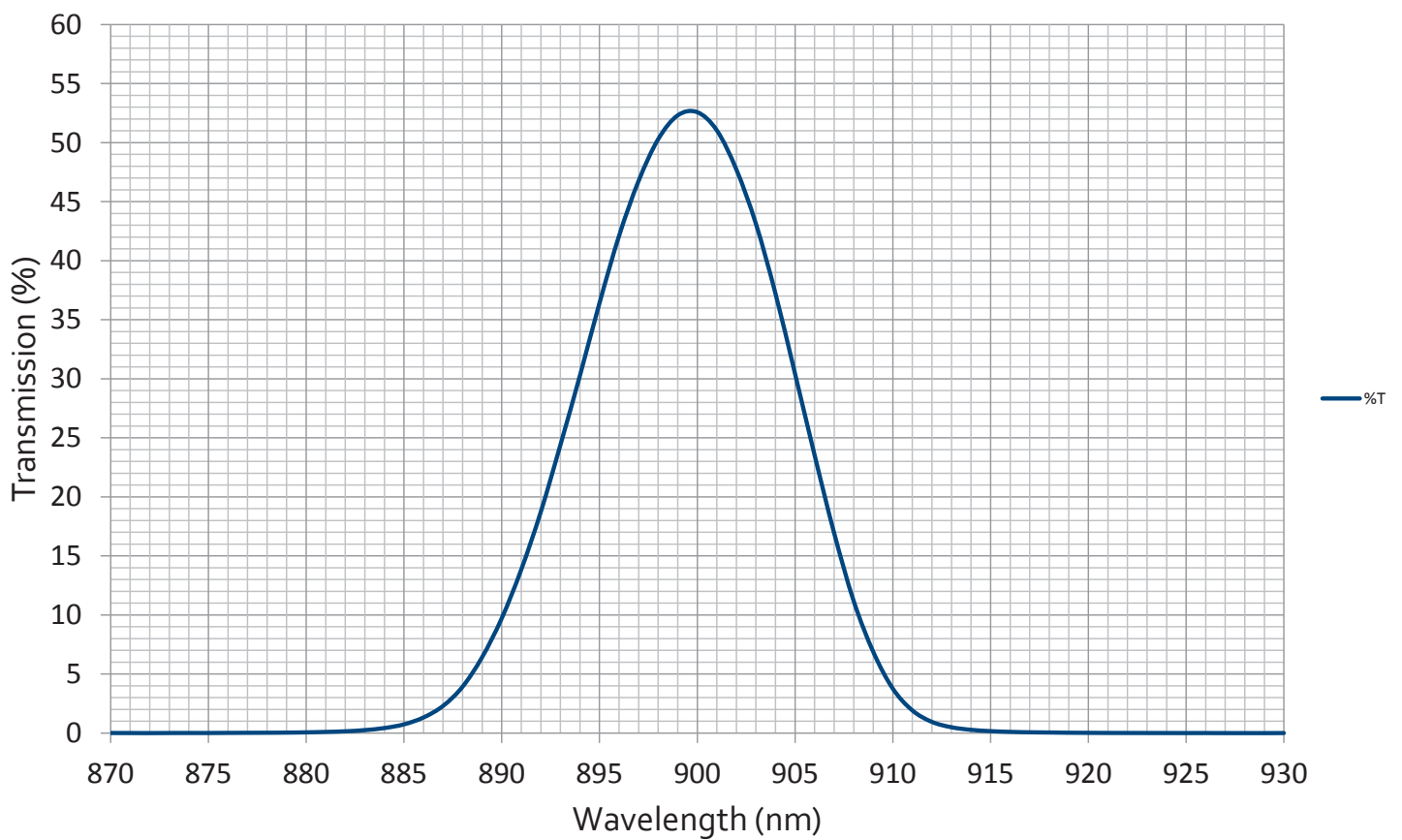
**Espetros de Transmitância dos Filtros de
Interferência Ópticos**

OPTICAL FILTERS

Title: Interference band-pass filter

Material/Specification: 900nm - Standard IR range (10nm FWHM)

Range/Description: 900FIB



+44 (0)1622 859444
info@knightoptical.co.uk
www.knightoptical.com

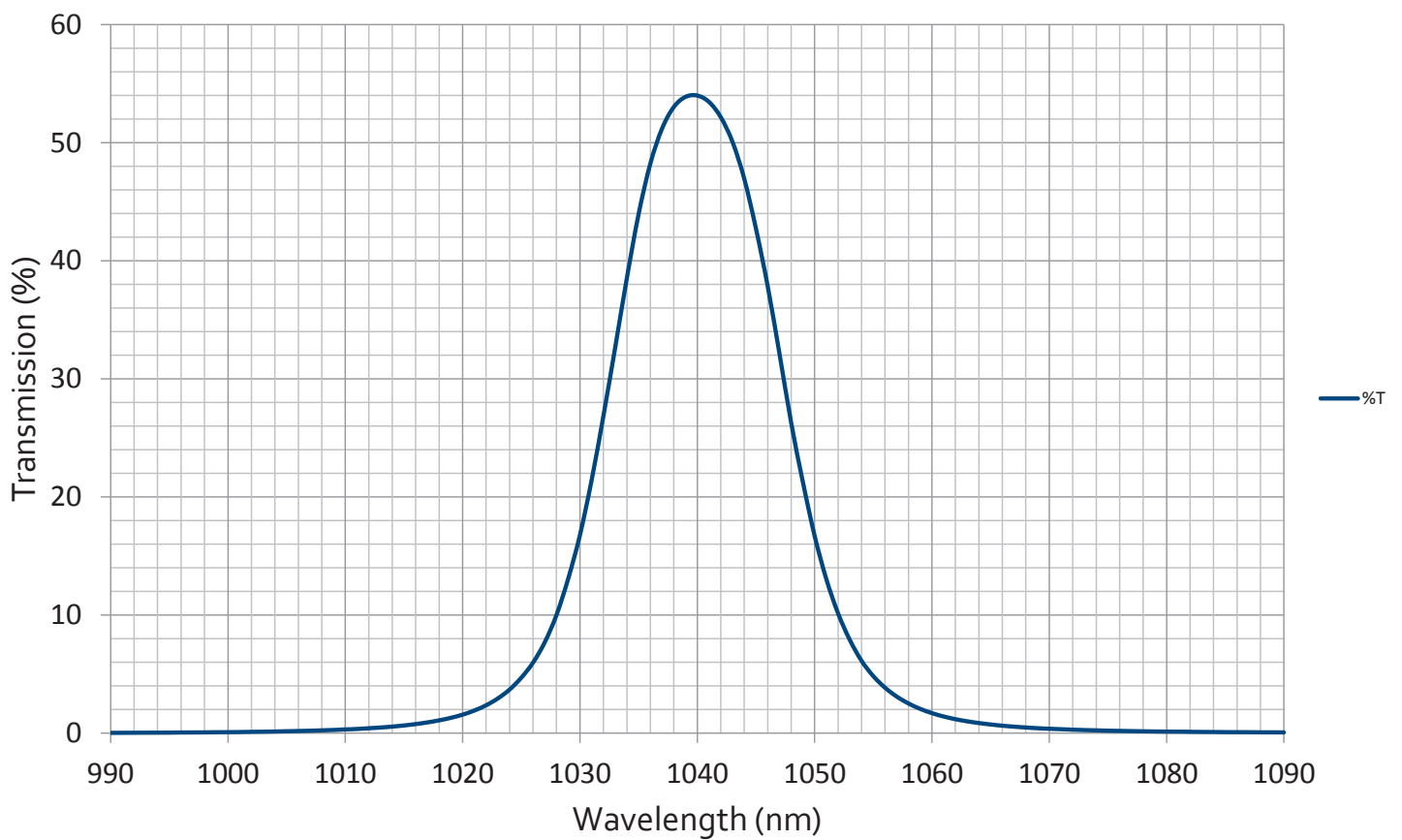


OPTICAL FILTERS

Title: Interference band-pass filter

Material/Specification: 1040nm - Standard IR range (15nm FWHM)

Range/Description: 1040FIB



+44 (0)1622 859444
info@knightoptical.co.uk
www.knightoptical.com

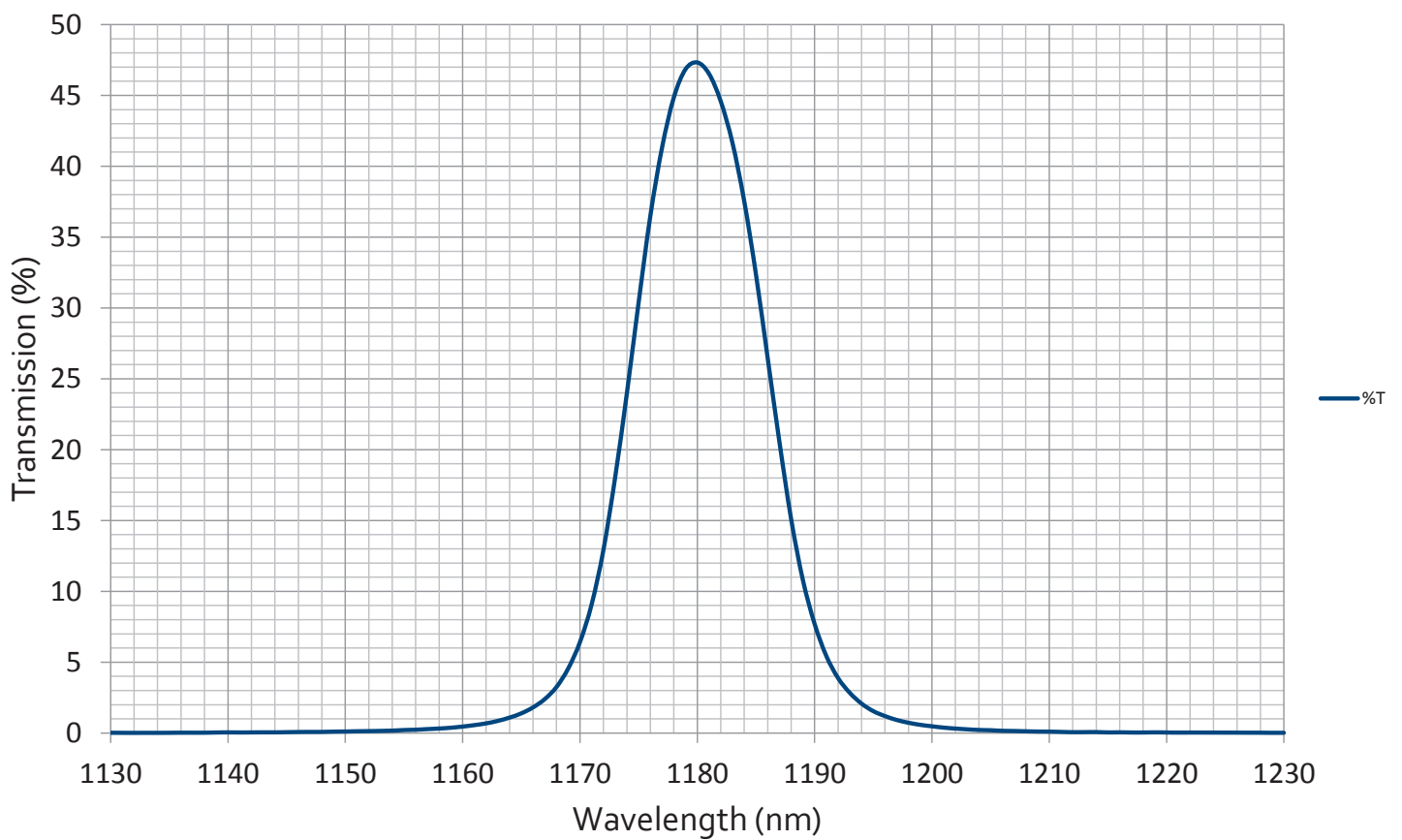


OPTICAL FILTERS

Title: Interference band-pass filter

Material/Specification: 1180nm - Standard IR range (11nm FWHM)

Range/Description: 1180FIB



+44 (0)1622 859444
info@knightoptical.co.uk
www.knightoptical.com

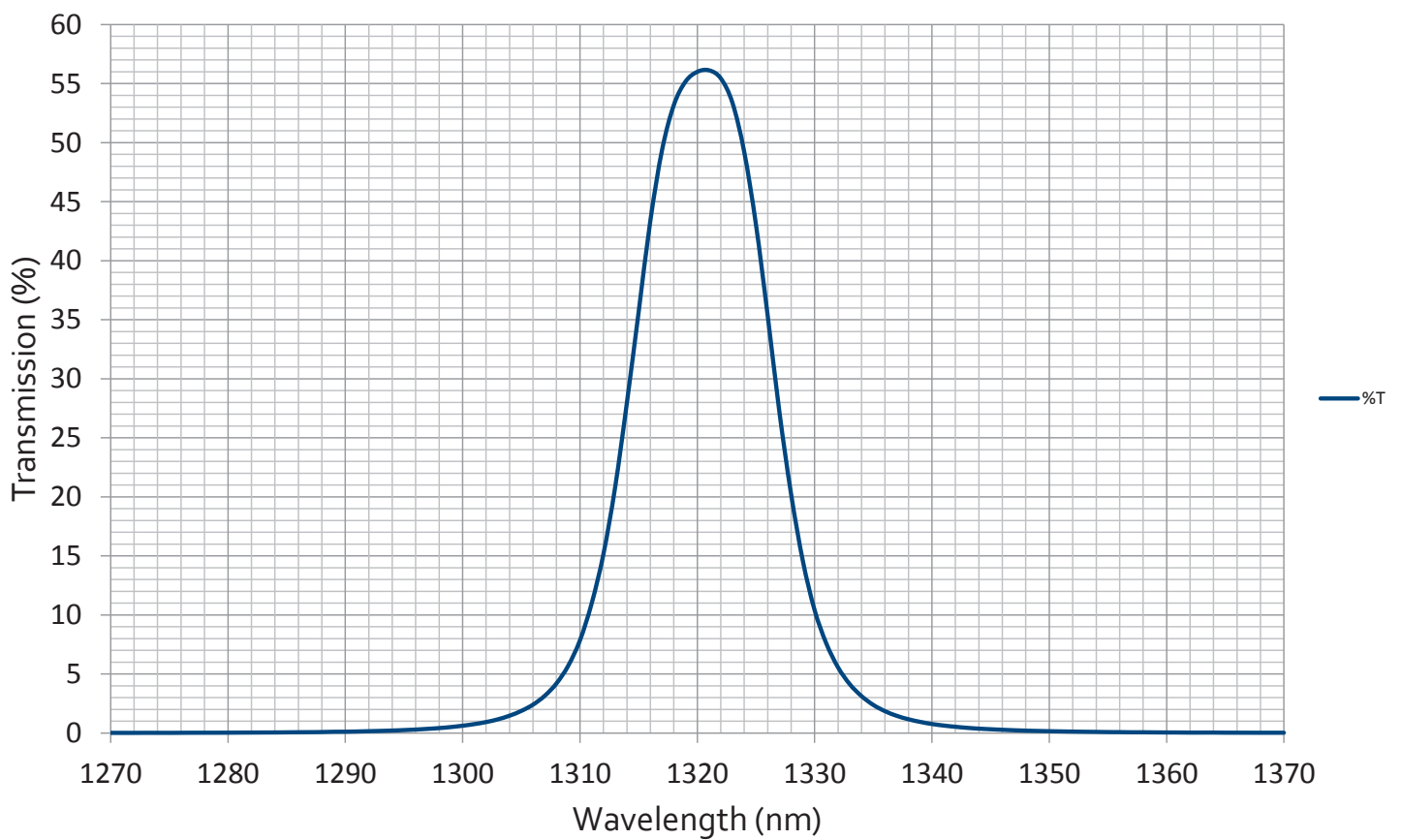


OPTICAL FILTERS

Title: Interference band-pass filter

Material/Specification: 1320nm - Standard IR range (12.5nm FWHM)

Range/Description: 1320FIB



+44 (0)1622 859444
info@knightoptical.co.uk
www.knightoptical.com

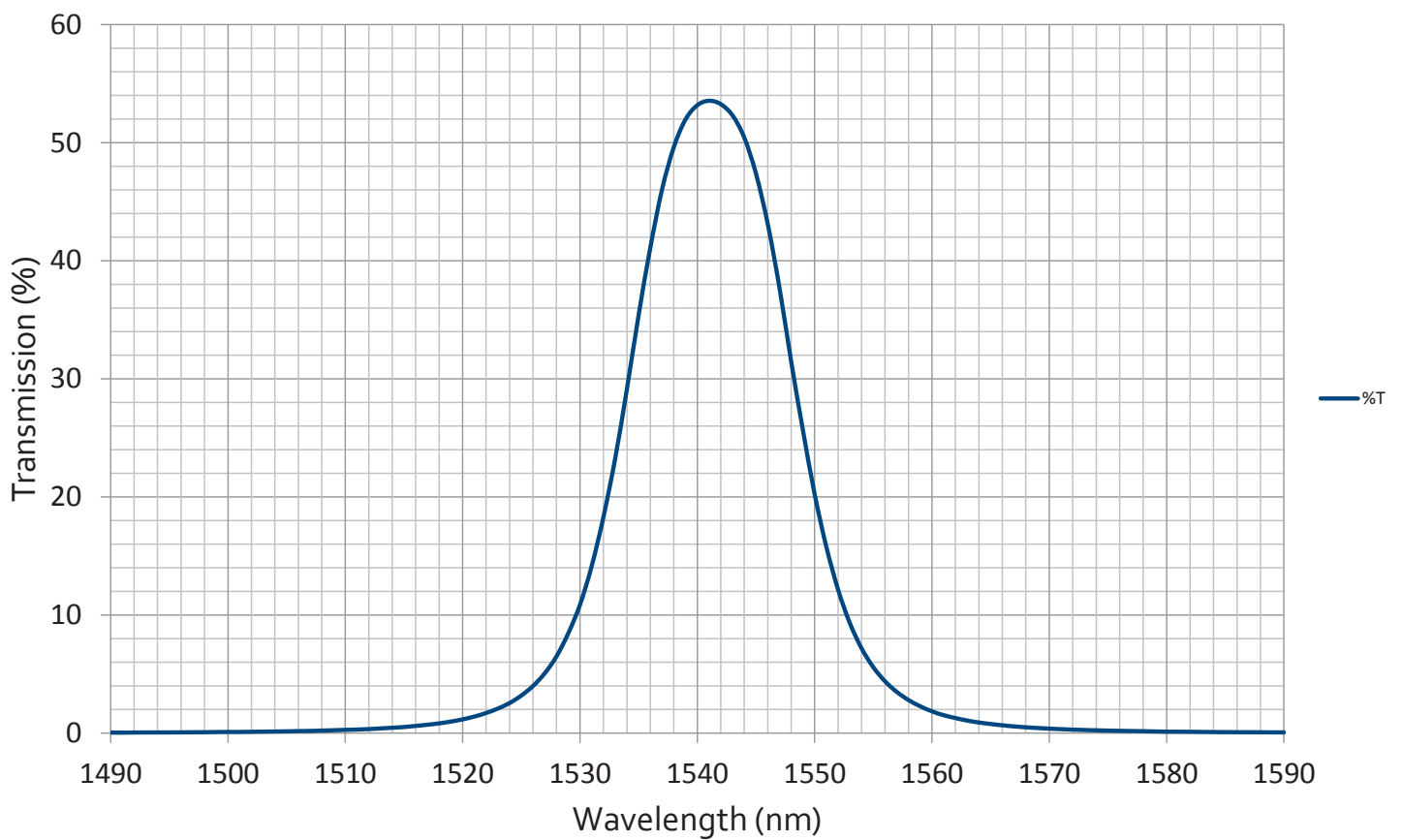


OPTICAL FILTERS

Title: Interference band-pass filter

Material/Specification: 1540nm - Standard IR range (15nm FWHM)

Range/Description: 1540FIB



+44 (0)1622 859444
info@knightoptical.co.uk
www.knightoptical.com

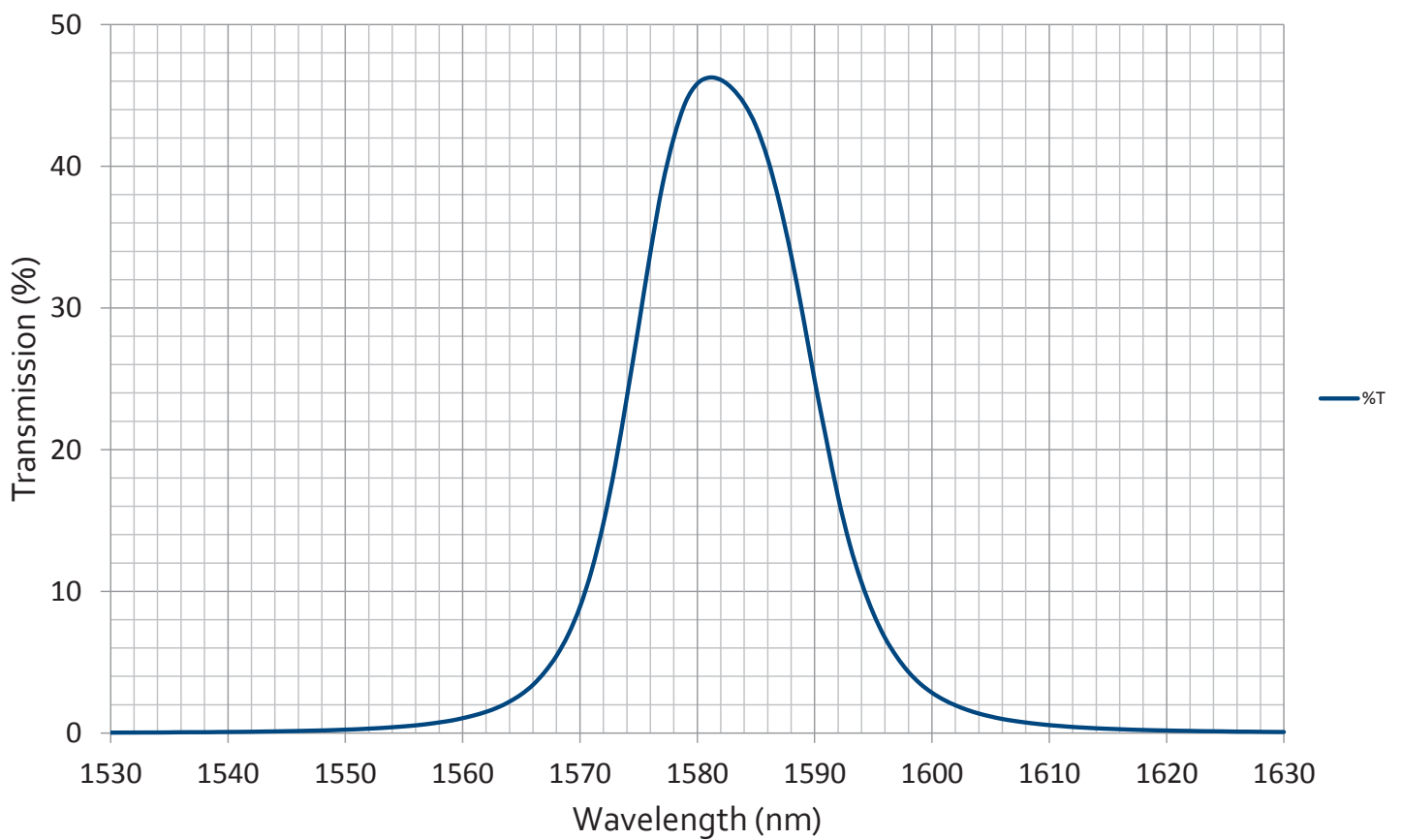


OPTICAL FILTERS

Title: Interference band-pass filter

Material/Specification: 1580nm - Standard IR range (16nm FWHM)

Range/Description: 1580FIB



+44 (0)1622 859444
info@knightoptical.co.uk
www.knightoptical.com

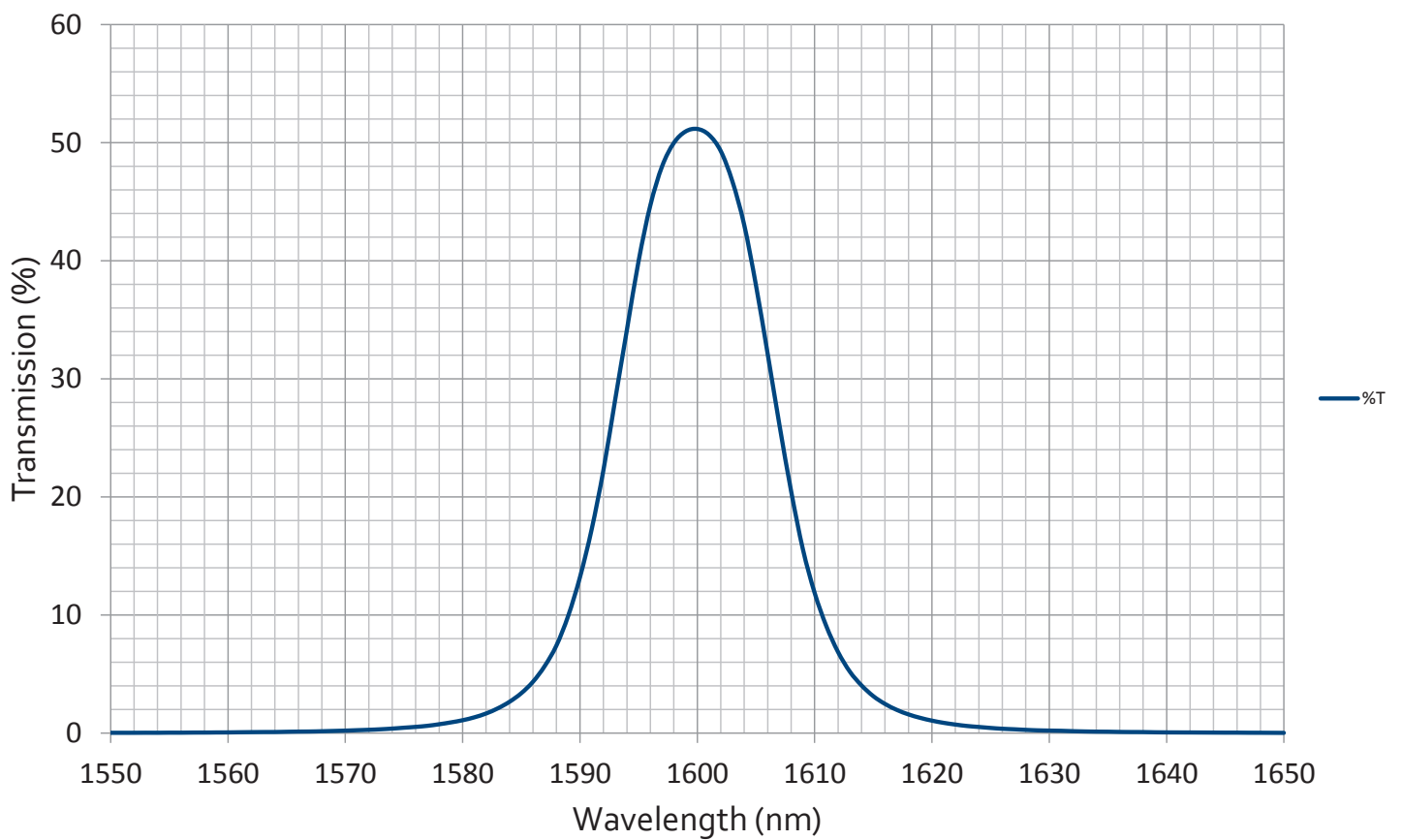


OPTICAL FILTERS

Title: Interference band-pass filter

Material/Specification: 1600nm - Standard IR range (13nm FWHM)

Range/Description: 1600FIB



+44 (0)1622 859444
info@knightoptical.co.uk
www.knightoptical.com



ANEXO C

Folhas de Dados dos Principais
Componentes Eletrônicos Utilizados



Balanced Modulator/Demodulator

Data Sheet

AD630

FEATURES

Recovers signal from 100 dB noise
 2 MHz channel bandwidth
 45 V/ μ s slew rate
 Low crosstalk: -120 dB at 1 kHz, -100 dB at 10 kHz
 Pin programmable, closed-loop gains of ± 1 and ± 2
 0.05% closed-loop gain accuracy and match
 100 μ V channel offset voltage (AD630)
 350 kHz full power bandwidth
 Chips available

APPLICATIONS

Balanced modulation and demodulation
 Synchronous detection
 Phase detection
 Quadrature detection
 Phase sensitive detection
 Lock in amplification
 Square wave multiplication

GENERAL DESCRIPTION

The AD630 is a high precision balanced modulator/demodulator that combines a flexible commutating architecture with the accuracy and temperature stability afforded by laser wafer trimmed thin film resistors. A network of on-board applications resistors provides precision closed-loop gains of ± 1 and ± 2 with 0.05% accuracy (AD630B). These resistors may also be used to accurately configure multiplexer gains of 1, 2, 3, or 4. External feedback enables high gain or complex switched feedback topologies.

The AD630 can be thought of as a precision op amp with two independent differential input stages and a precision comparator that is used to select the active front end. The rapid response time of this comparator coupled with the high slew rate and fast settling of the linear amplifiers minimize switching distortion.

The AD630 is used in precision signal processing and instrumentation applications that require wide dynamic range. When used as a synchronous demodulator in a lock-in amplifier configuration, the AD630 can recover a small signal from 100 dB of interfering noise (see the Lock-In Amplifier Applications section). Although optimized for operation up to 1 kHz, the circuit is useful at frequencies up to several hundred kilohertz.

FUNCTIONAL BLOCK DIAGRAM

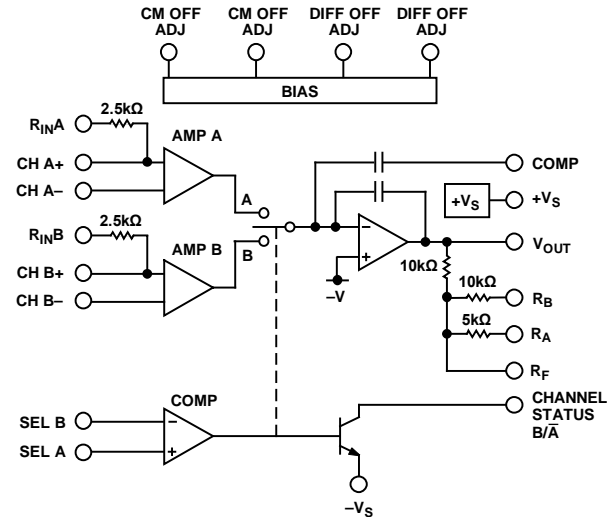


Figure 1.

Other features of the AD630 include pin programmable frequency compensation; optional input bias current compensation resistors, common-mode and differential-offset voltage adjustment, and a channel status output that indicates which of the two differential inputs is active.

PRODUCT HIGHLIGHTS

1. The application flexibility of the AD630 makes it the best choice for applications that require precisely fixed gain, switched gain, multiplexing, integrating-switching functions, and high speed precision amplification.
2. The 100 dB dynamic range of the AD630 exceeds that of any hybrid or IC balanced modulator/demodulator and is comparable to that of costly signal processing instruments.
3. The op amp format of the AD630 ensures easy implementation of high gain or complex switched feedback functions. The application resistors facilitate the implementation of most common applications with no additional parts.
4. The AD630 can be used as a 2-channel multiplexer with gains of 1, 2, 3, or 4. The channel separation of 100 dB at 10 kHz approaches the limit achievable with an empty IC package.
5. Laser trimming of the comparator and amplifying channel offsets eliminate the need for external nulling in most cases.

Rev. G

Document Feedback

Information furnished by Analog Devices is believed to be accurate and reliable. However, no responsibility is assumed by Analog Devices for its use, nor for any infringements of patents or other rights of third parties that may result from its use. Specifications subject to change without notice. No license is granted by implication or otherwise under any patent or patent rights of Analog Devices. Trademarks and registered trademarks are the property of their respective owners.

One Technology Way, P.O. Box 9106, Norwood, MA 02062-9106, U.S.A.
 Tel: 781.329.4700 ©2015–2016 Analog Devices, Inc. All rights reserved.
 Technical Support www.analog.com

ADS1220 4-Channel, 2-kSPS, Low-Power, 24-Bit ADC with Integrated PGA and Reference

1 Features

- Low Current Consumption:
As Low as 120 μA (typ) in Duty-Cycle Mode
- Wide Supply Range: 2.3 V to 5.5 V
- Programmable Gain: 1 V/V to 128 V/V
- Programmable Data Rates: Up to 2 kSPS
- Up to 20-Bits Effective Resolution
- Simultaneous 50-Hz and 60-Hz Rejection at 20 SPS with Single-Cycle Settling Digital Filter
- Two Differential or Four Single-Ended Inputs
- Dual Matched Programmable Current Sources: 10 μA to 1.5 mA
- Internal 2.048-V Reference: 5 ppm/ $^{\circ}\text{C}$ (typ) Drift
- Internal 2% Accurate Oscillator
- Internal Temperature Sensor: 0.5 $^{\circ}\text{C}$ (typ) Accuracy
- SPI-Compatible Interface (Mode 1)
- Package: 3.5-mm \times 3.5-mm \times 0.9-mm VQFN

2 Applications

- Temperature Sensor Measurements:
 - Thermistors
 - Thermocouples
 - Resistance Temperature Detectors (RTDs):
2-, 3-, or 4-Wire Types
- Resistive Bridge Sensor Measurements:
 - Pressure Sensors
 - Strain Gauges
 - Weigh Scales
- Portable Instrumentation
- Factory Automation and Process Control

3 Description

The ADS1220 is a precision, 24-bit, analog-to-digital converter (ADC) that offers many integrated features to reduce system cost and component count in applications measuring small sensor signals. The device features two differential or four single-ended inputs through a flexible input multiplexer (MUX), a low-noise, programmable gain amplifier (PGA), two programmable excitation current sources, a voltage reference, an oscillator, a low-side switch, and a precision temperature sensor.

The device can perform conversions at data rates up to 2000 samples-per-second (SPS) with single-cycle settling. At 20 SPS, the digital filter offers simultaneous 50-Hz and 60-Hz rejection for noisy industrial applications. The internal PGA offers gains up to 128 V/V. This PGA makes the ADS1220 ideally-suited for applications measuring small sensor signals, such as resistance temperature detectors (RTDs), thermocouples, thermistors, and resistive bridge sensors. The device supports measurements of pseudo- or fully-differential signals when using the PGA. Alternatively, the device can be configured to bypass the internal PGA while still providing high input impedance and gains up to 4 V/V, allowing for single-ended measurements.

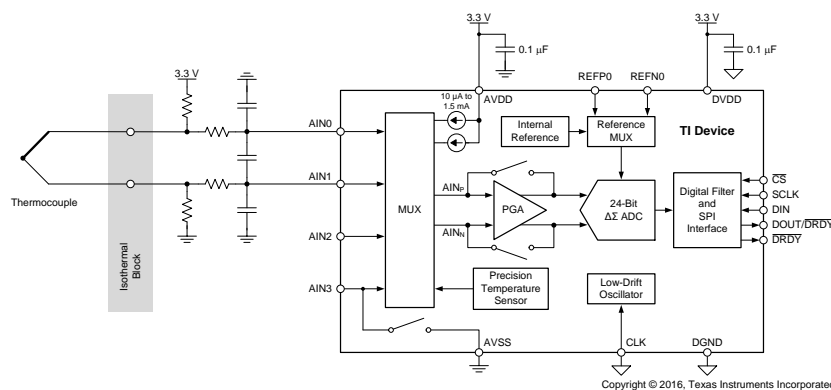
Power consumption is as low as 120 μA when operating in duty-cycle mode with the PGA disabled. The ADS1220 is offered in a leadless VQFN-16 or a TSSOP-16 package and is specified over a temperature range of -40°C to $+125^{\circ}\text{C}$.

Device Information⁽¹⁾

PART NUMBER	PACKAGE	BODY SIZE (NOM)
ADS1220	VQFN (16)	3.50 mm \times 3.50 mm
	TSSOP (16)	5.00 mm \times 4.40 mm

(1) For all available packages, see the orderable addendum at the end of the data sheet.

K-Type Thermocouple Measurement



Copyright © 2016, Texas Instruments Incorporated





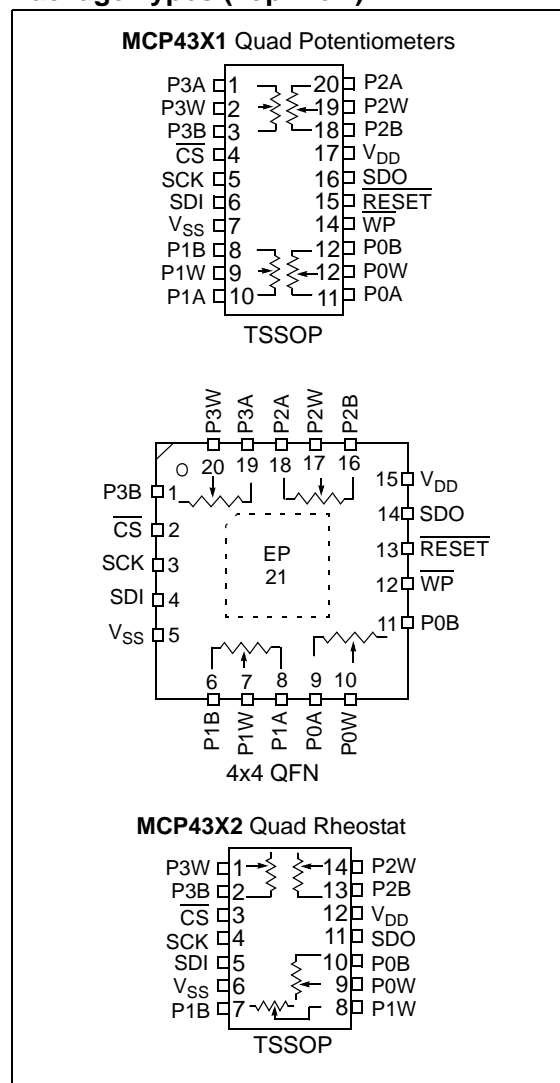
MCP434X/436X

7/8-Bit Quad SPI Digital POT with Non-Volatile Memory

Features

- Quad Resistor Network
- Potentiometer or Rheostat configuration options
- Resistor Network Resolution
 - 7-bit: 128 Resistors (129 Taps)
 - 8-bit: 256 Resistors (257 Taps)
- R_{AB} Resistances options of:
 - 5 k Ω
 - 10 k Ω
 - 50 k Ω
 - 100 k Ω
- Zero Scale to Full Scale Wiper operation
- Low Wiper Resistance: 75 Ω (typical)
- Low Tempco:
 - Absolute (Rheostat): 50 ppm typical (0°C to 70°C)
 - Ratiometric (Potentiometer): 15 ppm typical
- Non-volatile Memory
 - Automatic Recall of Saved Wiper Setting
 - WiperLock™ Technology
- SPI serial interface (10 MHz, modes 0,0 & 1,1)
 - High-Speed Read/Writes to wiper registers
 - Read/Write to Data EEPROM registers
 - Serially enabled EEPROM write protect
- Resistor Network Terminal Disconnect Feature via Terminal Control (TCON) Register
- Reset input pin
- Write Protect Feature:
 - Hardware Write Protect (\overline{WP}) Control pin
 - Software Write Protect (WP) Configuration bit
- Brown-out reset protection (1.5V typical)
- Serial Interface Inactive current (2.5 μ A typical)
- High-Voltage Tolerant Digital Inputs: Up to 12.5V
- Supports Split Rail Applications
- Internal weak pull-up on all digital inputs
- Wide Operating Voltage:
 - 2.7V to 5.5V - Device Characteristics Specified
 - 1.8V to 5.5V - Device Operation
- Wide Bandwidth (-3 dB) Operation:
 - 2 MHz (typical) for 5.0 k Ω device
- Extended temperature range (-40°C to +125°C)

Package Types (Top View)





Zero-Drift, Single-Supply, Rail-to-Rail Input/Output Operational Amplifier

Data Sheet

AD8628/AD8629/AD8630

FEATURES

- Lowest auto-zero amplifier noise**
- Low offset voltage: 1 μV**
- Input offset drift: 0.002 $\mu\text{V}/^\circ\text{C}$**
- Rail-to-rail input and output swing**
- 5 V single-supply operation**
- High gain, CMRR, and PSRR: 130 dB**
- Very low input bias current: 100 pA maximum**
- Low supply current: 1.0 mA**
- Overload recovery time: 50 μs**
- No external components required**
- Qualified for automotive applications**

APPLICATIONS

- Automotive sensors**
- Pressure and position sensors**
- Strain gage amplifiers**
- Medical instrumentation**
- Thermocouple amplifiers**
- Precision current sensing**
- Photodiode amplifiers**

GENERAL DESCRIPTION

This amplifier has ultralow offset, drift, and bias current. The [AD8628/AD8629/AD8630](#) are wide bandwidth auto-zero amplifiers featuring rail-to-rail input and output swing and low noise. Operation is fully specified from 2.7 V to 5 V single supply (± 1.35 V to ± 2.5 V dual supply).

The [AD8628/AD8629/AD8630](#) provide benefits previously found only in expensive auto-zeroing or chopper-stabilized amplifiers. Using Analog Devices, Inc., topology, these zero-drift amplifiers combine low cost with high accuracy and low noise. No external capacitor is required. In addition, the [AD8628/AD8629/AD8630](#) greatly reduce the digital switching noise found in most chopper-stabilized amplifiers.

With an offset voltage of only 1 μV , drift of less than 0.005 $\mu\text{V}/^\circ\text{C}$, and noise of only 0.5 μV p-p (0 Hz to 10 Hz), the [AD8628/AD8629/AD8630](#) are suited for applications where error sources cannot be tolerated. Position and pressure sensors, medical equipment, and strain gage amplifiers benefit greatly from nearly zero drift over their operating temperature range. Many systems can take advantage of the rail-to-rail input and output swings provided by the [AD8628/AD8629/AD8630](#) to reduce input biasing complexity and maximize SNR.

Rev. K

[Document Feedback](#)

Information furnished by Analog Devices is believed to be accurate and reliable. However, no responsibility is assumed by Analog Devices for its use, nor for any infringements of patents or other rights of third parties that may result from its use. Specifications subject to change without notice. No license is granted by implication or otherwise under any patent or patent rights of Analog Devices. Trademarks and registered trademarks are the property of their respective owners.

PIN CONFIGURATIONS

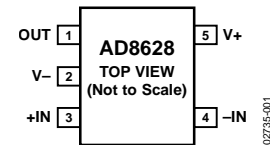


Figure 1. 5-Lead TSOT (UJ-5) and 5-Lead SOT-23 (RJ-5)

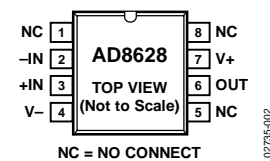


Figure 2. 8-Lead SOIC_N (R-8)

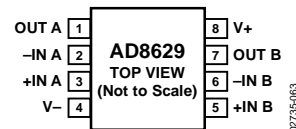


Figure 3. 8-Lead SOIC_N (R-8) and 8-Lead MSOP (RM-8)

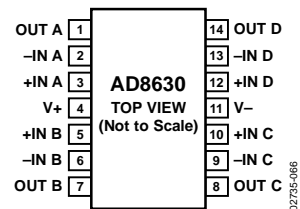


Figure 4. 14-Lead SOIC_N (R-14) and 14-Lead TSSOP (RU-14)

The [AD8628/AD8629/AD8630](#) are specified for the extended industrial temperature range (-40°C to $+125^\circ\text{C}$). The [AD8628](#) is available in tiny 5-lead TSOT, 5-lead SOT-23, and 8-lead narrow SOIC plastic packages. The [AD8629](#) is available in the standard 8-lead narrow SOIC and MSOP plastic packages. The [AD8630](#) quad amplifier is available in 14-lead narrow SOIC and 14-lead TSSOP plastic packages. See the Ordering Guide for automotive grades.

OPAx277 High Precision Operational Amplifiers

1 Features

- Ultralow Offset Voltage: 10 μV
- Ultralow Drift: $\pm 0.1 \mu\text{V}/^\circ\text{C}$
- High Open-Loop Gain: 134 dB
- High Common-Mode Rejection: 140 dB
- High Power Supply Rejection: 130 dB
- Low Bias Current: 1-nA maximum
- Wide Supply Range: $\pm 2 \text{ V}$ to $\pm 18 \text{ V}$
- Low Quiescent Current: 800 μA /amplifier
- Single, Dual, and Quad Versions
- Replaces OP-07, OP-77, and OP-177

2 Applications

- Transducer Amplifiers
- Bridge Amplifiers
- Temperature Measurements
- Strain Gage Amplifiers
- Precision Integrators
- Battery-Powered Instruments
- Test Equipment

3 Description

The OPAx277 series precision operational amplifiers replace the industry standard OP-177. They offer improved noise, wider output voltage swing, and are twice as fast with half the quiescent current. Features include ultralow offset voltage and drift, low bias current, high common-mode rejection, and high power supply rejection. Single, dual, and quad versions have identical specifications, for maximum design flexibility.

OPAx277 series operational amplifiers operate from $\pm 2\text{-V}$ to $\pm 18\text{-V}$ supplies with excellent performance. Unlike most operational amplifiers which are specified at only one supply voltage, the OPAx277 series is specified for real-world applications; a single limit applies over the $\pm 5\text{-V}$ to $\pm 15\text{-V}$ supply range. High performance is maintained as the amplifiers swing to their specified limits. Because the initial offset voltage ($\pm 20 \mu\text{V}$ maximum) is so low, user adjustment is usually not required. However, the single version (OPA277) provides external trim pins for special applications.

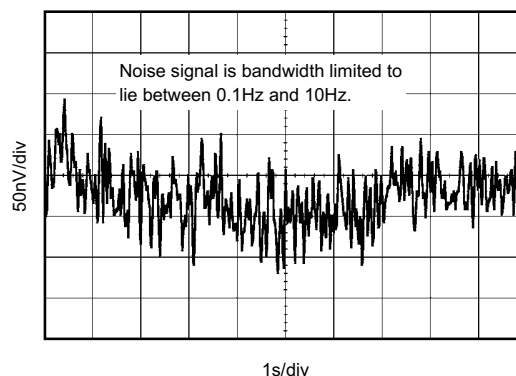
OPA277 operational amplifiers are easy to use and free from phase inversion and the overload problems found in some other operational amplifiers. They are stable in unity gain and provide excellent dynamic behavior over a wide range of load conditions. Dual and quad versions feature completely independent circuitry for lowest crosstalk and freedom from interaction, even when overdriven or overloaded.

Device Information⁽¹⁾

PART NUMBER	PACKAGE	BODY SIZE (NOM)
OPA277 OPA2277	VSON (8)	4.00 mm x 4.00 mm
	SOIC (8)	3.91 mm x 4.90 mm
	PDIP (8)	6.35 mm x 9.81 mm
OPA4277	SOIC (14)	3.91 mm x 8.65 mm
	PDIP (14)	6.35 mm x 19.30 mm

(1) For all available packages, see the orderable addendum at the end of the data sheet.

0.1 Hz to 10 Hz Noise



±0.33°C Accurate, 12-Bit + Sign Temperature Sensor and Thermal Window Comparator With Two-Wire Interface

Check for Samples: [LM92](#)

FEATURES

- Window Comparison Simplifies Design of ACPI-Compatible Temperature Monitoring and Control
- Serial Bus Interface
- Separate Open-Drain Outputs for Interrupt and Critical Temperature Shutdown
- Shutdown Mode to Minimize Power Consumption
- Up to Four LM92s can be Connected to a Single Bus
- 12-Bit + Sign Output
- Operation up to 150°C

APPLICATIONS

- HVAC
- Medical Electronics
- Electronic Test Equipment
- System Thermal Management
- Personal Computers
- Office Electronics
- Automotive

KEY SPECIFICATIONS

- Supply Voltage 2.7V to 5.5V
- Supply Current operating 350 μ A (typ) 625 μ A (max) shutdown 5 μ A (typ)
- Temperature Accuracy
 - 30°C, ±0.33°C (max)
 - 10°C to 50°C, ±0.50°C (max)
 - -10°C to 85°C, ±1.0°C (max)
 - 125°C, ±1.25°C (max)
 - -25°C to 150°C, ±1.5°C (max)
- Linearity ±0.5°C (max)
- Resolution 0.0625°C

DESCRIPTION

The LM92 is a digital temperature sensor and thermal window comparator with an I²C™ Serial Bus interface and an accuracy of ±0.33°C. The window-comparator architecture of the LM92 eases the design of temperature control systems. The open-drain Interrupt (INT) output becomes active whenever temperature goes outside a programmable window, while a separate Critical Temperature Alarm (T_CRIT_A) output becomes active when the temperature exceeds a programmable critical limit. The INT output can operate in either a comparator or event mode, while the T_CRIT_A output operates in comparator mode only.

The host can program both the upper and lower limits of the window as well as the critical temperature limit. Programmable Hysteresis as well as a fault queue are available to minimize false tripping. Two pins (A0, A1) are available for address selection. The sensor powers up with default thresholds of 2°C T_{HYST}, 10°C T_{LOW}, 64°C T_{HIGH}, and 80°C T_CRIT.

The LM92's 2.7V to 5.5V supply voltage range, Serial Bus interface, 12-bit + sign output, and full-scale range of over 128°C make it ideal for a wide range of applications. These include thermal management and protection applications in personal computers, electronic test equipment, office electronics, automotive, medical and HVAC applications.



Please be aware that an important notice concerning availability, standard warranty, and use in critical applications of Texas Instruments semiconductor products and disclaimers thereto appears at the end of this data sheet.

I²C is a trademark of dcl_owner.

All other trademarks are the property of their respective owners.

Micropower Precision Low Dropout Series Voltage Reference Family

FEATURES

- **Trimmed to High Accuracy: 0.04% Max**
- **Low Drift: 3ppm/°C Max**
- **Low Supply Current: 50µA Max**
- High Output Current: 50mA Min
- Low Dropout Voltage: 300mV Max
- Excellent Thermal Regulation
- Power Shutdown
- Thermal Limiting
- All Parts Guaranteed Functional from -40°C to 125°C
- Voltage Options: 2.5V, 3V, 3.3V, 4.096V and 5V
- AEC-Q100 Qualified for Automotive Applications

APPLICATIONS

- A/D and D/A Converters
- Precision Regulators
- Handheld Instruments
- Power Supplies

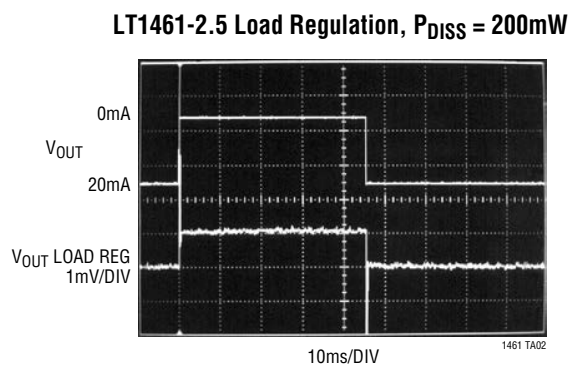
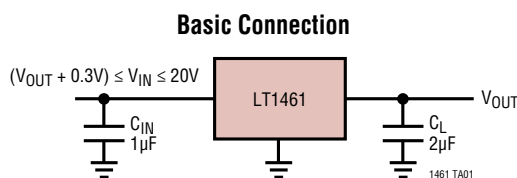
DESCRIPTION

The **LT[®]1461** is a family of low dropout micropower band-gap references that combine very high accuracy and low drift with low supply current and high output drive. These series references use advanced curvature compensation techniques to obtain low temperature coefficient and trimmed precision thin-film resistors to achieve high output accuracy. The LT1461 family draws only 35µA of supply current, making them ideal for low power and portable applications, however their high 50mA output drive makes them suitable for higher power requirements, such as precision regulators.

In low power applications, a dropout voltage of less than 300mV ensures maximum battery life while maintaining full reference performance. Line regulation is nearly immeasurable, while the exceedingly good load and thermal regulation will not add significantly to system error budgets. The shutdown feature can be used to switch full load currents and can be used for system power down. Thermal shutdown protects the part from overload conditions. The LT1461 is available in 2.5V, 3V, 3.3V 4.096V and 5V options.

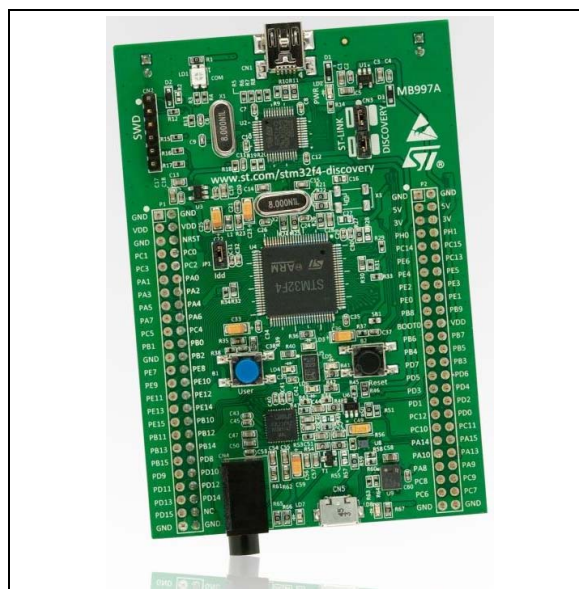
All registered trademarks and trademarks are the property of their respective owners.

TYPICAL APPLICATION



Features

- STM32F407VGT6 microcontroller featuring 32-bit ARM® Cortex®-M4 with FPU core, 1-Mbyte Flash memory, 192-Kbyte RAM in an LQFP100 package
- On-board ST-LINK/V2 on STM32F4DISCOVERY (old reference) or ST-LINK/V2-A on STM32F407G-DISC1 (new order code)
- USB ST-LINK with re-enumeration capability and three different interfaces:
 - Debug port
 - Virtual Com port (with new order code only)
 - Mass storage (with new order code only)
- Board power supply: through USB bus or from an external 5 V supply voltage
- External application power supply: 3 V and 5 V
- LIS302DL or LIS3DSH ST MEMS 3-axis accelerometer
- MP45DT02 ST-MEMS audio sensor omni-directional digital microphone
- CS43L22 audio DAC with integrated class D speaker driver
- Eight LEDs:
 - LD1 (red/green) for USB communication
 - LD2 (red) for 3.3 V power on
 - Four user LEDs, LD3 (orange), LD4 (green), LD5 (red) and LD6 (blue)
 - 2 USB OTG LEDs LD7 (green) VBUS and LD8 (red) over-current
- Two push-buttons (user and reset)
- USB OTG FS with micro-AB connector
- Extension header for all LQFP100 I/Os for quick connection to prototyping board and easy probing
- Comprehensive free software including a variety of examples, part of STM32CubeF4 package or STSW-STM32068 to use legacy standard libraries.



1. Picture is not contractual.

Description

The STM32F4DISCOVERY kit leverages the capabilities of the STM32F407 high performance microcontrollers, to allow users to easily develop applications featuring audio.

It includes an ST-LINK embedded debug tool, one ST-MEMS digital accelerometer, a digital microphone, one audio DAC with integrated class D speaker driver, LEDs, push-buttons and an USB OTG micro-AB connector.

To expand the functionality of the STM32F4DISCOVERY kit with ethernet connectivity, LCD display and more, visit the www.st.com/stm32f4dis-expansion webpage.

With the latest board enhancement, the new order code STM32F407G-DISC1 has replaced the old reference STM32F4DISCOVERY.

REFERÊNCIAS

- ABNT (1997). *ABNT NBR 13992 - Motor gasoline: determination of fuel anhydrous ethylic alcohol content*. Associação Brasileira de Normas Técnicas, Rio de Janeiro, Brazil.
- Balabin, R. M., Safieva, R. Z., e Lomakina, E. I. (2010). Gasoline classification using near infrared (NIR) spectroscopy data: comparison of multivariate techniques. *Anal Chim Acta*, 671:27 – 35.
- Burns, D. A. e Ciurczak, E. W. (2008). *Handbook of Near-infrared Analysis*. CRC Press.
- Chagas, I. P. (2006). *Desenvolvimento de um Fotômetro Portátil NIR para Determinação do Teor de Água no Álcool Combustível e do Teor de Etanol na Gasolina*. Tese de Doutorado, State University of Campinas, Institute of Chemistry, Campinas, SP, Brazil.
- Chia-Hung Wang, Tzu-Lun Hung, e Tai-Ping Sun (2010). Pwm photo-power control for light-emitting diode lighting system without photo-diode detection. Em *2010 5th IEEE Conference on Industrial Electronics and Applications*, pp. 1313–1316.
- Correia, R. M., Domingos, E., Cáo, V. M., Araujo, B. R., Sena, S., Pinheiro, L. U., Fontes, A. M., Aquino, L. F. M., Ferreira, E. C., Filgueiras, P. R., e Romão, W. (2018). Portable near infrared spectroscopy applied to fuel quality control. *Talanta*, 176:26 – 33.
- Fernandes, H. L., Jr, I. M. R., Pasquini, C., e Rohwedder, J. J. (2008). Simultaneous determination of methanol and ethanol in gasoline using nir spectroscopy: Effect of gasoline composition. *Talanta*, 75(3):804 – 810.
- Fonseca, A. e Raimundo, I. M. (2007). A simple method for water discrimination based on an light emitting diode (led) photometer. *Analytica Chimica Acta*, 596:66–72.
- Gaião, E. N., Medeiros, E. P., Lyra, W. S., Moreira, P. N. T., Vasconcelos, P. C., Silva, E. C., e Araújo, M. C. U. (2005). Um fotômetro multi-led microcontrolado, portátil e de baixo custo. *Quim. Nova*, 28(6):1102–2205.
- Gaião, E. N., Santos, S. R. B., Santos, V. B., Nascimento, E. C. L., Lima, R. S., e Araújo, M. C. U. (2008). An inexpensive, portable and microcontrolled near infrared led-photometer for screening analysis of gasoline. *Talanta*, 75(3):792 – 796. Special Section: Remote Sensing.
- Galvão, R. K. H., Araújo, M. C. U., Silva, E. C., José, G. E., Soares, S. F. C., e Paiva, H. M. (2007). Cross-validation for the selection of spectral variables using the successive projections algorithm. *Journal of the Brazilian Chemical Society*, 18:1580 – 1584.

- Gentilin, F. A. (2012). Sistema de aquisição eletrônico baseado na reflectância do infravermelho próximo. Dissertação de Mestrado, Universidade Estadual de Londrina.
- Gentilin, F. A., Kobo, B. T., de Souza Ribeiro, L., de França, J. A., de S.M. Felício, A. L., de M. França, M. B., e de O. Toginho Filho, D. (2016). Development and testing of a hardware platform for measuring instruments based on near-infrared diffuse reflection. *Measurement*, 86(Supplement C):14 – 25.
- Guchardi, R. e Pasquini, C. (2001). Evaluation of a dual-beam near-infrared spectrometer based on acousto-optic tunable filters. *Applied Spectroscopy*, 55(4):454 – 457.
- Harris, S. J., Swarin, S., Sultan, M., Lambert, D. K., e Jack, M. (1993). Sensor for measuring alcohol content of alcohol/gasoline fuel mixtures.
- Haxha, S. e Jhoja, J. (2016). Optical based noninvasive glucose monitoring sensor prototype. *IEEE Photonics Journal*, 8(6):1–11.
- Hegedüs, J., Hantos, G., e Poppe, A. (2017). Light output stabilisation of led based streetlighting luminaires by adaptive current control. *Microelectronics Reliability*, 79:448 – 456.
- Isaac, W. e Na, A. (2016). On-the-go soil nitrogen sensor based on near infrared spectroscopy. Em *2016 International Conference on Information Technology (InCITe) - The Next Generation IT Summit on the Theme - Internet of Things: Connect your Worlds*, pp. 312–315.
- JGCM (2008). Jcgm 100: Evaluation of measurement data - guide to the expression of uncertainty in measurement.
- Johnson, M. (2003). *Photodetection and Measurement: Maximizing Performance in Optical Systems*. McGraw-Hill.
- Kittipanyangam, S., Abe, K., e Eguchi, K. (2016). Design of a measurement device explaining the relationship between the concentration of solution and the light absorbance for chemical education. Em *2016 13th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, pp. 1–6.
- Lima, K. M. G. (2012). A portable photometer based on led for the determination of aromatic hydrocarbons in water. *Microchemical journal*, 103:62–67.
- Lin, Y.-W., Deng, B.-C., Wang, L.-L., Xu, Q.-S., Liu, L., e Liang, Y.-Z. (2016). Fisher optimal subspace shrinkage for block variable selection with applications to nir spectroscopic analysis. *Chemometrics and Intelligent Laboratory Systems*, 159:196 – 204.
- Lutz, O. M. D., Bonn, G. K., Rode, B. M., e Huck, C. W. (2014). Reproducible quantification of ethanol in gasoline via a customized mobile near-infrared spectrometer. *Analytica Chimica Acta*, 826:61–68.

- Marois, M., L. Jacques, S., e D. Paulsen, K. (2018). Optimal wavelength selection for optical spectroscopy of hemoglobin and water within a simulated light-scattering tissue. *Journal of Biomedical Optics*, 23:1.
- Moreira, M. (2013). Fotômetro digital nir para quantificação de Água adicionada ao leite. Dissertação de Mestrado, Universidade Estadual de Londrina, Londrina.
- Moreira, M., de França, J. A., de Oliveira Toginho Filho, D., Beloti, V., Yamada, A. K., de M. França, M. B., e de Souza Ribeiro, L. (2016). A low-cost NIR digital photometer based on ingaas sensors for the detection of milk adulterations with water. *IEEE Sensors Journal*, 16(10):3653–3663.
- Nascimento, E. C. L. (2008). *Um Fotômetro Microcontrolado LED-NIR, Portátil e de Baixo Custo para Análise Screening de Gasolinas Tipo C*. Tese de Doutorado, Universidade Federal da Paraíba, João Pessoa.
- Neto, A. C., Oliveira, E. C., Lacerda, V., Castro, E. V., Romão, W., Silva, R. C., Pereira, R. G., Sten, T., Filgueiras, P. R., e Poppi, R. J. (2014). Quality control of ethanol fuel: Assessment of adulteration with methanol using ^1H nmr. *Fuel*, 135:387 – 392.
- Ouyang, A. e Liu, J. (2013). Classification and determination of alcohol in gasoline using nir spectroscopy and the successive projections algorithm for variable selection. *Measurement Science and Technology*, 24(2):025502.
- Paiva, E. M., Rohwedder, J. J. R., Pasquini, C., e Pereira, C. F. (2019). Method for building a portable near infrared photometer based on leds and interference filters chosen by a spectral variable selection algorithm. *Microchemical Journal*, 146:842 – 849.
- Pasquini, C. (2003). Near infrared spectroscopy: Fundamentals, practical aspects and analytical applications. *J. Braz. Chem. Soc.*, 14(2):198–219.
- Pinheiro, P. P. (2018). Projeto e desenvolvimento de um sistema opto-eletrônico para detecção de substâncias através da reflexão difusa na faixa do infravermelho próximo. Dissertação de Mestrado, Universidade Estadual de Londrina, Londrina.
- Pinheiro, P. P., Dos Santos, J. C. F., e De Moraes França, M. B. (2019). Development, testing, and validation of a prototype for qualification of substances based on near-infrared spectroscopy. *IEEE Access*, 7:25650–25659.
- Qiang, T. e Chu, J. (2012). The design of the infrared analyzer for unconventional emission components of ethanol gasoline vehicle. Em *World Automation Congress 2012*, pp. 1–4.
- Ribeiro, L. S., Gentilin, F. A., d. França, J. A., d. S. M. Felício, A. L., e d. M. França, M. B. (2016). Development of a hardware platform for detection of milk adulteration based on near-infrared diffuse reflection. *IEEE Transactions on Instrumentation and Measurement*, 65(7):1698–1706.

- Rui, Z., Yuanyuan, C., Zhibin, W., e Kewu, L. (2017). A novel ensemble l1 regularization based variable selection framework with an application in near infrared spectroscopy. *Chemometrics and Intelligent Laboratory Systems*, 163:7 – 15.
- Sun, D.-W. (2009). *Infrared Spectroscopy for Food Quality Analysis and Control*. Academic Press.
- Visconti, P., Lay-Ekuakille, A., Primiceri, P., Ciccarese, G., e de Fazio, R. (2017). Hardware design and software development for a white led-based experimental spectrophotometer managed by a pic-based control system. *IEEE Sensors Journal*, 17(8):2507–2515.
- Yadav, J., Rani, A., Singh, V., e Murari, B. M. (2015). Prospects and limitations of non-invasive blood glucose monitoring using near-infrared spectroscopy. *Biomedical Signal Processing and Control*, 18:214 – 227.
- Zucchini, R. R., Gioseffi, C. S., Nascimento, C. R., Torres, E. S., e Franco, J. B. S. (2008). Estudo da precisão e exatidão do método brasileiro para determinação de etanol e gasolina. Rio Oil & Gas Expo and Conference 2008.