



**UEPB**

**UNIVERSIDADE ESTADUAL DA PARAÍBA  
CAMPUS I – CAMPINA GRANDE  
CENTRO DE CIÊNCIA E TECNOLOGIA  
DEPARTAMENTO DE QUÍMICA  
CURSO DE LICENCIATURA EM QUÍMICA**

**MARIA FERNANDA ARAUJO SILVA**

**ESPECTROSCOPIA NIR E REGRESSÃO POR MÁQUINA DE VETORES DE  
SUPORTE PARA A DETERMINAÇÃO DO TEOR ALCOÓLICO  
EM BEBIDA MISTA A BASE DE CERVEJA**

**CAMPINA GRANDE - PB  
2020**

MARIA FERNANDA ARAUJO SILVA

**ESPECTROSCOPIA NIR E REGRESSÃO POR MÁQUINA DE VETORES DE  
SUPORTE PARA A DETERMINAÇÃO DO TEOR ALCOÓLICO  
EM BEBIDA MISTA A BASE DE CERVEJA**

Trabalho de Conclusão de Curso em  
Licenciatura em Química da Universidade  
Estadual da Paraíba, como requisito parcial à  
obtenção do título de Licenciada em Química.

**Orientadora:** Prof<sup>ª</sup>. Dr<sup>ª</sup>. Simone da Silva Simões.

**CAMPINA GRANDE - PB  
2020**

É expressamente proibido a comercialização deste documento, tanto na forma impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que na reprodução figure a identificação do autor, título, instituição e ano do trabalho.


S586e Silva, Maria Fernanda Araújo.  
Espectroscopia NIR e regressão por máquina de vetores de suporte para a determinação do teor alcoólico em bebida mista a base de cerveja [manuscrito] / Maria Fernanda Araujo Silva. - 2020.  
42 p.  
Digitado.  
Trabalho de Conclusão de Curso (Graduação em Química) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2020.  
"Orientação : Profa. Dra. Simone da Silva Simões, Coordenação do Curso de Licenciatura em Química - CCT."  
1. Cerveja. 2. Quimiometria. 3. Regressão. 4. Máquinas de Vetores de Suporte. I. Título  
21. ed. CDD 641.23

ESPECTROSCOPIA NIR E REGRESSÃO POR MÁQUINA DE VETORES DE  
SUPORTE PARA A DETERMINAÇÃO DO TEOR ALCOÓLICO  
EM BEBIDA MISTA A BASE DE CERVEJA.

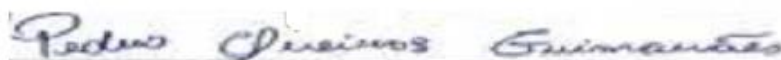
Trabalho de Conclusão de Curso em  
Licenciatura em Química da  
Universidade Estadual da Paraíba, como  
requisito parcial à obtenção do título de  
Licenciada em Química.

Aprovada em: 27/11/2020.

**BANCA EXAMINADORA**



Prof. Dr. Simone da Silva Simões (Orientadora)  
Universidade Estadual da Paraíba (UEPB)



Msc. Pedro Queiroz Guimarães  
Universidade Federal de Campina Grande (UFCG)



Prof. Dr. Verônica Evangelista de Lima  
Universidade Estadual da Paraíba (UEPB)

Aos meus pais, que sempre me apoiaram para  
que pudesse realizar todos os meus sonhos,  
DEDICO.

## AGRADECIMENTOS

A Deus, pelo o dom da vida e por ter me dado sabedoria nos momentos difíceis.

Aos meus pais Hélio e Josefa, pelo o esforço, apoio e incentivo para que pudesse realizar esse sonho.

A minha orientadora, Simone Simões, pela excelente orientação, dedicação e paciência e pelo o apoio para que este trabalho fosse realizado.

Ao meu namorado e amigo Álvaro, por todo incentivo e apoio e por ter estado ao meu lado durante todos os momentos da graduação.

As minhas irmãs Fabiana e Flávia pela força e pela compreensão nos momentos de ausência.

Aos amigos do curso, Messias, Gabriel, Jaquely, Izabela, Elane e Welio, pelos momentos de estudos, pela as brincadeiras e pela a amizade.

Aos professores do Curso de Licenciatura em Química, pelo os ensinamentos durante o período da graduação.

Ao Núcleo de Química Analítica Avançada no Estado de Pernambuco – NUQAAPE por disponibilizar o MicroNir portátil.

A toda minha família, que torceram pelo meu sucesso.

A todos que de forma direta ou indireta fizeram parte dessa conquista.

*“O sucesso nasce do querer, da determinação e persistência em se chegar a um objetivo.”*

*(José de Alencar)*

## RESUMO

Uma das características de qualidade de bebidas alcoólicas é a sua porcentagem de álcool (%ABV). Esta característica pode estar relacionada a qualidade e preço dessas bebidas. Com a crescente difusão de cervejarias e bebidas a base de cerveja no Brasil, faz-se necessário o desenvolvimento de métodos práticos e rápidos para determinação desta característica. Este trabalho propõe a utilização de dados analíticos fornecidos por um espectrômetro NIR portátil e métodos de calibração multivariada para a determinação de %ABV em uma bebida alcoólica mista a base de cerveja. Para isso adquiriu-se 4 lotes distintos desta bebida, com teor alcoólico rotulado de 7,9%, e produziu-se 25 amostras com %ABV variando de 1,58 a 17,11%, a partir de diluições e fortificações da bebida base. Os espectros NIR de todas as 29 amostras foram registrados e modelos de regressão PLS e SVM foram construídos. Diversos pré-processamentos foram testados para remoção de características espectrais indesejadas. A performance dos modelos desenvolvidos foi comparada e o SVM foi o que apresentou melhor capacidade preditiva apresentando um RMSEP de 0,8373. Deste modo, acredita-se que o uso da espectroscopia NIR associada a regressão SVM representa um método rápido e prático para a determinação de %ABV na bebida mista a base de cerveja analisada.

**Palavras-Chave:** Cerveja. Máquinas de vetores de suporte. Quimiometria. Regressão.



## ABSTRACT

One of the quality characteristics of alcoholic beverages is their percentage of alcohol (%ABV). This characteristic may be related to the quality and price of these drinks. With the increasing diffusion of breweries and beer-based drinks in Brazil, it is necessary to develop practical and quick methods to determine this characteristic. This work proposes the use of analytical data provided by a portable NIR spectrometer and multivariate calibration methods for the determination of %ABV in a mixed alcoholic beverage based on beer. For this purpose, 4 distinct lots of this drink were purchased, with a labeled alcohol content of 7,9%, and 25 samples were produced with %ABV ranging from 1,58 to 17,11%, based on dilutions and fortifications of the drink base. The NIR spectra of all 29 samples were recorded and PLS and SVM regression models were constructed. Several pre-processes have been tested to remove unwanted spectral characteristics. The performance of the developed models was compared and the SVM was the one that had the best predictive capacity with a RMSEP of 0,8373. Thus, it is believed that the use of NIR spectroscopy associated with SVM regression represents a quick and practical method for the determination of %ABV in the analyzed beer-based mixed drink.

**Keywords:** Beer. Support vector machines. Chemometrics. Regression.

## LISTA DE ILUSTRAÇÕES

Figura 1 –	Organização de dados na calibração multivariada .....	21
Figura 2 –	Espectros brutos para todas as amostras analisadas e faixa de trabalho destacada em verde .....	29
Figura 3a –	Gráfico dos escores .....	30
Figura 3b –	Influência para as amostras brutas .....	30
Figura 4 –	Gráfico dos valores medidos vs valores previstos pelo modelo PLS/MS- JK na etapa de (a) calibração (em azul) e validação cruzada (em vermelho) e (b) etapa de predição .....	32
Figura 5 –	Gráfico de valores medidos vs valores preditos para o modelo SVM construído com os dados pré - processados com MSC .....	35

## LISTA DE TABELAS

Tabela 1 – Parâmetros da água de boa qualidade para produção de cerveja .....	15
Tabela 2 – Composição média do grão de cevada e do malte de cevada .....	16
Tabela 3 – Composição do lúpulo usado nas cervejarias .....	17
Tabela 4 – Figuras de mérito para o modelo PLS .....	31
Tabela 5 – Desvios padrões para os modelos PLS/MSD - full e PLS/MSD – JK .....	33
Tabela 6 – Comparação do valores medidos e preditos pelos modelos SVM construídos com diversos pré-processamentos e seus respectivos desvios padrões .....	34
Tabela 7 – Parâmetros de comparação entre os modelos PLS e SVM construídos .....	35

## LISTA DE ABREVIATURAS E SIGLAS

ABV	Álcool por volume (do inglês, <i>Alcohol by Volume</i> )
IR	Infravermelho (do inglês, <i>Infrared</i> )
JK	Algoritmo Jack-Knife
MAPA	Ministério da Agricultura, Pecuária e Abastecimento
MIR	Infravermelho Médio (do inglês, <i>Mid-Infrared</i> )
MLR	Regressão Linear Múltipla (do inglês, <i>Multiple Linear Regression</i> )
MSC	Correção do Sinal Multiplicativo (do inglês, <i>Multiplicative Scatter Correction</i> )
NIPALS	Mínimos Quadrados Parciais Não Lineares (do inglês, <i>Non Linear Iterative Partial Least Squares</i> )
NIR	Infravermelho Próximo (do inglês, <i>Near-Infrared</i> )
OSC	Correção de Sinal Ortogonal (do inglês, <i>Orthogonal Signal Correction</i> )
PCA	Análise de Componentes Principais (do inglês, <i>Principal Component Analysis</i> )
PCR	Regressão por Componentes Principais (do inglês, <i>Principal Component Regression</i> )
PLS	Regressão por Mínimos Quadrados Parciais (do inglês, <i>Partial Least Squares Regression</i> )
RMSE	Raiz do Erro Médio Quadrático (do inglês, <i>Root Mean Square Error</i> )
RMSEC	Raiz do Erro Médio Quadrático de Calibração (do inglês, <i>Root Mean Square Error of Calibration</i> )
RMSECV	Raiz do Erro Médio Quadrático de Validação Cruzada (do inglês, <i>Root Mean Square Error of Cross Validation</i> )
RMSEP	Raiz do Erro Médio Quadrático de Predição (do inglês, <i>Root Mean Squared Error of Prediction</i> )
SNV	Varição Normal Padrão (do inglês, <i>Standard Normal Variate</i> )
SPXY	<i>Sample set Partitioning based on joint x-y distances</i>
SVC	Classificação de Vetores de Suporte (do inglês, <i>Support Vector Classification</i> )
SVM	Máquinas de Vetores de Suporte (do inglês, <i>Support Vector Machine</i> )
SVR	Regressão de Vetores de Suporte (do inglês, <i>Support Vector Regression</i> )
VL	Variáveis Latentes

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> .....	12
<b>2</b>	<b>OBJETIVOS</b> .....	13
<b>2.1</b>	<b>Objetivo Geral</b> .....	13
<b>2.2</b>	<b>Objetivos Específicos</b> .....	13
<b>3</b>	<b>REFERENCIAL TEÓRICO</b> .....	14
<b>3.1</b>	<b>Cerveja</b> .....	14
<b>3.2</b>	<b>Matéria Prima</b> .....	15
<b>3.3</b>	<b>Espectroscopia no Infravermelho Próximo (NIR)</b> .....	19
<b>3.4</b>	<b>Quimiometria</b> .....	20
<b>3.4.1</b>	<b><i>Calibração Multivariada</i></b> .....	20
<b>3.4.1.1</b>	<b><i>Regressão por Mínimos Quadrados Parciais (PLS)</i></b> .....	22
<b>3.4.1.2</b>	<b><i>Máquinas de Vetores de Suporte (SVM)</i></b> .....	23
<b>3.4.1.2.1</b>	<b><i>Classificação de Vetores de Suporte (SVC)</i></b> .....	23
<b>3.4.1.2.2</b>	<b><i>Regressão de Vetores de Suporte (SVR)</i></b> .....	24
<b>4</b>	<b>METODOLOGIA</b> .....	27
<b>4.1</b>	<b>Localização do Experimento</b> .....	27
<b>4.2</b>	<b>Dados Analíticos</b> .....	27
<b>4.3</b>	<b>Espectroscopia NIR</b> .....	27
<b>4.4</b>	<b>Construção e Validação dos Modelos</b> .....	28
<b>4.4.1</b>	<b><i>Etapas para a Construção dos Modelos</i></b> .....	28
<b>5</b>	<b>RESULTADOS E DISCUSSÕES</b> .....	29
<b>6</b>	<b>CONCLUSÃO</b> .....	36
<b>7</b>	<b>PERSPECTIVAS</b> .....	37
	<b>REFERÊNCIAS</b> .....	38

## 1 INTRODUÇÃO

A cerveja é considerada uma bebida carbonada, de teor alcoólico entre 3 e 8% (v/v), produzida com malte de cevada, lúpulo, fermento e água de boa qualidade (SILVA, 2005). De acordo com o Art. 38 do Decreto nº 6.871, de 4 de junho de 2009, as cervejas podem ser classificadas levando em consideração 5 critérios: quanto ao extrato primitivo, quanto à cor, quanto ao teor alcoólico, quanto à proporção de malte de cevada e quanto à fermentação (BRASIL, 2009). É uma das bebidas mais produzidas em todo mundo, sendo consumida desde a antiguidade por diversos povos de diferentes culturas (ROSA;AFONSO, 2015).

No ano de 2019, o Brasil possuía cerca de 1.209 cervejarias registradas. Segundo o Ministério da Agricultura, Pecuária e Abastecimento (MAPA), houve um crescimento de 36% em relação a 2018. Sendo terceiro maior fabricante mundial, com 13,3 bilhões de litros produzidos por ano (SINDICERV, 2020). As cervejas no Brasil em relação a produção e o consumo são caracterizadas como da família de cerveja *lager*, por se adaptar melhor ao clima brasileiro (SILVA, 2005).

O mercado cervejeiro brasileiro proporciona aos consumidores experimentar produtos diferenciados, com o objetivo de oferecer uma grande variedade de produtos de boa qualidade (BRUNELLI, 2012). Para obtenção de cerveja de qualidade, é necessário a realização de análises dos parâmetros físico-químicos, regulamentados pela legislação nacional, que estabelece os padrões de identidade e qualidade da bebida (SERRÃO, 2014). Todo processo de produção da cerveja inicia-se pela maltagem, no qual os grãos de cevada germinados selecionados passam por uma secagem e moagem. O malte, possui enzimas desenvolvidas no processo de germinação, que atua como catalisador na transformação do amido em açúcares fermentescíveis. Por isso, este processo é indispensável, pois sem a presença de amilases, as leveduras não serão capazes de fermentar o amido (ROSA;AFONSO, 2015).

O teor alcoólico é um dos parâmetros que conferem valor agregado as cervejas e neste sentido, a espectroscopia no infravermelho próximo (NIR), aparece como uma ótima alternativa, por se tratar de um método analítico rápido, não destrutivo, reprodutível e de custo baixo, visto que se pode eliminar ou reduzir a quantidade de solvente/reagente utilizado e diminuir o tempo de análise (BLANCO *et al.*, 2006). Deste modo, o objetivo deste trabalho é desenvolver um método utilizando a espectroscopia NIR e a regressão por máquina de vetores de suporte (SVM) para a quantificação do teor de álcool em bebida alcoólica mista a base de cerveja, de forma rápida e precisa. Paralelamente pretende-se comparar a performance dos modelos obtidos com a de modelos de regressão por mínimos quadrados parciais (PLS).

## **2 OBJETIVOS**

### **2.1 Objetivo Geral**

Desenvolver um método utilizando a espectroscopia NIR e a regressão por máquina de vetores de suporte (SVM) para a quantificação do teor de álcool em bebida alcoólica mista a base de cerveja.

### **2.2 Objetivos Específicos**

- Construir modelos PLS e SVM.
- Determinar os melhores pré-processamentos espectrais para os modelos PLS e SVM.
- Avaliar a capacidade preditiva dos modelos utilizando amostras de predição.
- Comparar o modelo PLS com o modelo SVM.

### 3 REFERENCIAL TEÓRICO

#### 3.1 Cerveja

Atualmente a cerveja é a bebida alcoólica mais antiga, comparando-se com os outros tipos de bebidas fermentadas e destiladas que conhecemos (BEAUMONT, 2017). A origem das bebidas fermentadas se dá no território que hoje corresponde ao Iraque, antiga Mesopotâmia, no qual as escavações arqueológicas encontraram potes com resquícios de bebida fermentada a partir de cereais, estimadas de 5.000 a.C. Além do valor histórico, esses achados têm grande importância para determinar as origens da cerveja (MORADO, 2017).

Devido a revolução industrial, a produção de cerveja sofreu muitas mudanças, fazendo com que se estabelecessem grandes fábricas na Alemanha, Inglaterra e no Império Austro-Húngaro. As inovações tecnológicas possibilitaram um avanço na história da cerveja, como por exemplo, a refrigeração artificial desenvolvida por Carl Linde, a criação da máquina a vapor idealizada por James Watt e os estudos sobre fermentação de microrganismos desenvolvidos por Louis Pasteur (MORADO, 2009).

A primeira fábrica de cerveja no Brasil possivelmente foi fundada em 1842 por um imigrante alemão licenciado pelo Império brasileiro. Em 1858 haviam seis fábricas de cerveja no país, que produziam a “cerveja de barbante”, assim chamada pois liberava uma enorme quantidade de gás carbônico no interior da garrafa, gerando assim uma grande pressão e a rolha era amarrada com barbante para que não saltasse da garrafa (LIMBERGER, 2013).

No ano de 2019, o Brasil chegou a 1.209 cervejarias registradas, segundo o Ministério da Agricultura, Pecuária e Abastecimento (MAPA), houve um crescimento de 36% em relação a 2018. Sendo o terceiro maior fabricante mundial, com 13,3 bilhões de litros produzidos por ano (SINDICERV, 2020).

De acordo com a Lei nº 8.918, de 4 de setembro de 1994 regulamentada pelo Decreto nº 9.902, de 8 de julho de 2019, entende-se por cerveja a bebida produzida da fermentação, a partir da levedura cervejeira, do mosto de cevada malteada ou de extrato de malte, submetido a um processo de cocção adicionado de lúpulo ou extrato de lúpulo, hipótese em que uma parte da cevada malteada ou do extrato de malte poderá ser substituída parcialmente por adjunto cervejeiro (BRASIL, 2019).

O teor alcoólico da cerveja pode variar entre 3 e 67,5%. No Brasil não se encontra cervejas em que o teor alcoólico se apresente acima de 14%. A unidade de medida para o teor alcoólico é representada por % ABV, que significa álcool por volume (MAGRI, 2016).



Nas últimas décadas, no processo de fabricação de cerveja foi introduzida novas técnicas bioquímicas, microbiológicas e genéticas, com o intuito de produzir cervejas capazes de agradar o paladar e o olfato dos seus consumidores. Diferenciando-se por meio dos ingredientes e métodos de elaboração, percebendo cada vez mais a necessidade de produzir cervejas com novos e sofisticados aromas por exemplo de limão, morango, abacaxi, kiwi, maçã, chocolate, entre outros (CARVALHO, 2009).

### 3.2 Matéria prima

Segundo a Lei da Pureza da Cerveja Alemã, as cervejas são compostas por malte de cevada, lúpulo, água e levedo (MORADO, 2009).

A água, por ser apresentada em grande quantidade no processo cervejeiro, corresponde a uma faixa de 92 a 95% da massa da cerveja. Por isso, as indústrias tendem a se localizar em regiões onde a composição da água seja de boa qualidade (VENTURINI FILHO, 2016).

Os requisitos básicos para uma boa água cervejeira precisam cumprir os padrões de potabilidade, apresentar alcalinidade no máximo de 50 mg/L (preferencialmente inferior a 25 mg/L) e possuir concentração de cálcio em torno de 50 mg/L (VENTURINI FILHO, 2016).

Geralmente, a água de boa qualidade utilizada para produção de cerveja deve apresentar os parâmetros descritos na Tabela 1.

**Tabela 1** – Parâmetros da água de boa qualidade para produção de cerveja.

<i>Parâmetro</i>	<i>Unidade</i>	<i>Especificação</i>
Sabor	-	Insípida
Odor	-	Inodora
Ph	Ph	6,5-8,0
Turbidez	NTU	< 0,4
Matéria Orgânica	mg O <sub>2</sub> /L	0 – 0,8
Dureza Total	mg CaCO <sub>3</sub> /L	18 – 79
Cloretos	mg Cl/L	1 – 20
Sólidos Totais Dissolvidos	mg/L	50 – 150
Sulfatos	mg SO <sub>4</sub> /L	1 – 30
Nitratos	mg NO <sub>3</sub> /L	Ausente
Magnésio	mg Mg <sup>+2</sup> /L	1 – 6
Cálcio	mg Ca <sup>+2</sup> /L	5 – 22
CO <sub>2</sub> Livre	mg CO <sub>2</sub> /L	0,5 – 5

**Fonte:** Adaptado de LAZZARI, 2009.

A cevada apresenta dois tipos principais, uma de duas fileiras e a outra de seis fileiras. A de duas fileiras (2 fileiras de grãos na mesma espiga), é a mais utilizada pois abrange grãos maiores e mais uniformes e se subdivide em dois grupos principais, a cevada de haste ereta (*Hordeum distichum erectum*) e a cevada de haste curva (*Hordeum distichum nutans*) (REINOLD,1997). A cevada apresenta algumas características como por exemplo, o alto teor de amido, o alto teor de enzimas, proteínas para equilibrar os efeitos de espuma, corpo e estabilidade coloidal e possui baixo teor de lipídios (OLIVEIRA, 2011).

O malte utilizado em cervejarias é obtido através da cevada, uma gramínea do gênero *Hordeum*, que logo após a colheita as sementes são guardadas em silos com temperatura e umidade adequadas (VENTURINI FILHO, 2016). Uma função importante do malte é fornecer açúcares e nutrientes como alimento ao fermento, consequentemente produzindo álcool, gás carbônico e outros subprodutos de fermentação da cerveja (PASSARELLI, 2009).

Para obtenção do malte o processo é dividido em três etapas, a maceração, no qual o amido do cereal é convertido em proteínas, fermentescíveis e aminoácidos necessários no processo de fermentação realizado pela a leveduras. Em seguida, vem a etapa de germinação, quando o cereal é espalhado em superfície de piso, cimento ou asfalto, criando um colchão de 30 cm a uma temperatura de 15 °C, sem exposição da luz, durante oito dias (EVANGELISTA, 2012). Por último, a etapa da secagem que serve para determinar a coloração (quanto mais torrado o malte, mais escura a cerveja), o aroma e diversas características que são importantes no produto final. Na fabricação de algumas cervejas utiliza-se mais de um tipo de malte (MORADO, 2009). A composição média do grão de cevada e do malte, está descrito na Tabela 2.

**Tabela 2** – Composição média do grão de cevada e do malte de cevada.

<b>Características</b>	<b>Grão de Cevada</b>	<b>Malte de Cevada</b>
Massa do Grão (mg)	32 a 36	29 a 33
Umidade (%)	10 a 14	4 a 6
Amido (%)	55 a 60	50 a 55
Açúcares (%)	0,5 a 1,0	8 a 10
Nitrogênio Total (%)	1,8 a 2,3	1,8 a 2,3
Nitrogênio Solúvel (% de N total)	10 a 12	35 a 50
Poder Diastásico, °% L <sup>1</sup>	50 a 60	100 a 250
Enzima $\alpha$ -amilase, 20° unidades <sup>2</sup>	Traços	30 a 60
Atividade Proteolítica	Traços	15 a 30

<sup>1</sup>Lintner (índice de atividade das amilases), <sup>2</sup>em unidades de dextrinas produzidas

Fonte: SILVA, 2005.

O lúpulo é uma planta trepadeira da família das canabáceas (*Humulus lupulus*), geralmente cultivada em climas frios do hemisfério norte (SANTOS; DINHAM; ADAMES, 2006). É considerada como planta dióica, ou seja, possui exemplares fêmeas e machos, mas na produção de cerveja utiliza-se a flor feminina que apresenta a lupulina que caracteriza o amargo (DELIBERALLI, 2015). Existem dois tipos importantes de lúpulo, o amargo que beneficia a estabilidade da espuma e aumenta sua estabilidade biológica, evitando o desenvolvimento de alguns microrganismos e proporciona o sabor da cerveja (KUCK, 2008). E o outro tipo, é o aromático que é utilizado como lúpulo de acabamento ou condicionador que é inserido ao mosto geralmente nos últimos minutos de fervura para conferir aroma a cerveja (OLIVEIRA, 2011). A composição média do lúpulo utilizado na indústria cervejeira, está ilustrado na Tabela 3.

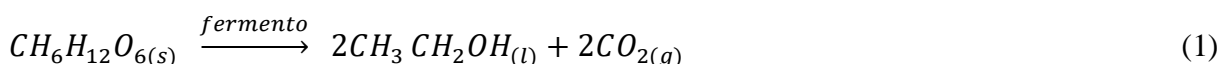
**Tabela 3** - Composição do lúpulo usado nas cervejarias.

<i>Composição química</i>	<i>Concentração (%)</i>
Água	10,0
Resinas totais	15,0
Óleos essenciais	0,5
Taninos	4,0
Monossacarídeos	2,0
Pectina	2,0
Aminoácidos	0,1
Proteína bruta	15,0
Lípido e ceras	3,0
Cinzas	8,0
Celulose, lignina, etc	40,4
Total	100,0

**Fonte:** Adaptado de Evangelista, 2012

As leveduras são microrganismos eucariontes, unicelulares e pertencentes ao Reino *Fungi* (SICARDI; LEGRAS, 2011). Realiza um processo chamado de fermentação, convertendo açúcares simples, como a glicose e maltose, em dióxido de carbono e álcool etílico. (DELIBERALLI, 2015).

Na produção de cerveja as leveduras são do gênero *Saccharomyces*, sendo mais utilizada a *Saccharomyces cerevisiae* como levedura para produzir cervejas da família das Ales e a *Saccharomyces uvarum* ou a *Saccharomyces carlsbergensis* para a família das Lagers (PALMER, 2006). Geralmente, o processo de fermentação ocorre pela adição de fermento a mistura de açúcares e água, devido o fermento conter enzimas que promovem uma série de reações que convertem o açúcar em etanol e dióxido de carbono, como está representado na Equação 1 (SANTOS JUNIOR, 2012):



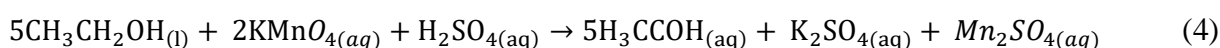
O etanol é um líquido incolor, de odor aromático, de sabor ardente e por ser muito higroscópico retira a umidade das mucosas. É também usado em bebidas como cervejas, vinhos, licores, destilados e derivados (ZARPELON, 2013). De acordo com suas propriedades físicas-químicas, apresenta ponto de ebulição de 78,3 °C (1 atm), ponto de fusão de -117 °C (1 atm) e densidade de 0,789 g.mL<sup>-1</sup> (SANTOS JUNIOR, 2012).

No processo de fermentação, não é capaz a produção de bebidas com teor álcool maior que 12-15%, pois em concentrações altas as enzimas do fermento são inutilizadas. Assim sendo, para produzir bebidas com um teor alcoólico com concentrações maiores, a solução passa por um processo de destilação (SANTOS JUNIOR, 2012).

O álcool etílico pode ser obtido pela hidratação do eteno catalisada por ácido, de acordo com a Equação 2.



Geralmente os álcoois podem ser oxidados por muitos agentes oxidantes, no qual se pode destacar o dicromato de potássio (K<sub>2</sub>Cr<sub>2</sub>O<sub>7</sub>) e o permanganato de potássio (KMnO<sub>4</sub>), ambos na presença de ácido sulfúrico (H<sub>2</sub>SO<sub>4</sub>), formam o aldeído acético sendo observado pela mudança na coloração e pelo o odor característico, de acordo com as Equações 3 e 4, respectivamente (SANTOS JUNIOR, 2012).



O teor alcoólico em cervejas ou bebidas à base de cerveja pode ser medido por picnometria, que é a determinação de massa e volume de substâncias sólidas e líquidas; pelo o método da refratometria, no qual o índice de refração de uma solução varia com a concentração do soluto (REIS, 2006) ou ainda, pelo método de densimetria que utiliza a densidade, relacionando as massas de um corpo com o seu volume (BRASIL, 2010).

Esses métodos geralmente são demorados e, deste modo, a busca por métodos mais rápidos e precisos se faz necessário. Neste sentido, o uso da espectroscopia no infravermelho próximo pode ser uma boa alternativa já que é uma técnica rápida, segura, não destrutiva e de baixo custo (PASQUINI, 2003).

### **3.3 Espectroscopia no infravermelho próximo (NIR)**

A espectroscopia estuda a interação da matéria com a radiação eletromagnética e um dos seus principais objetivos a determinação dos níveis de energia de moléculas ou átomos. Quase todos os compostos orgânicos ou inorgânicos, que apresenta ligações covalentes, absorvem na região do infravermelho frequências de radiação eletromagnética (CARVALHO, 2017). Os sinais obtidos relacionados aos sobretons e combinações de bandas das vibrações C-H, O-H, N-H e S-H das moléculas presentes na amostra, devido a interação de matéria com a radiação, geram espectros com informações sobrepostas (NOBREGA, 2018).

A região nomeada de infravermelho (IR) apresenta 3 regiões, infravermelho próximo com uma faixa de números de ondas entre  $14290\text{ cm}^{-1}$  a  $200\text{ cm}^{-1}$ , o infravermelho médio com uma faixa de números de ondas entre  $4000\text{ cm}^{-1}$  a  $400\text{ cm}^{-1}$  e a faixa com números de ondas entre  $200\text{ cm}^{-1}$  a  $10\text{ cm}^{-1}$  é denominada por infravermelho distante (CARVALHO, 2017).

A espectroscopia NIR é uma técnica espectroscópica vibracional, em que a energia de fótons pode variar de 750 a 2500 nm em comprimento de onda, promovendo uma interação da amostra com as ondas eletromagnéticas do NIR. Antes era utilizada como uma técnica com um único tipo de mediação obtida por uma amostra sólida a partir da radiação difusa refletida. Atualmente existem várias formas de medição, como por exemplo, a transmitância, transfectância, reflectância difusa, interactância e transmitância através de espalhamento médio (PASQUINI, 2003)

As bandas de absorção NIR normalmente são largas, sobrepostas e de 10 a 100 vezes mais fracas que as bandas de absorção no infravermelho médio. Com a baixa absorvidade permite alta profundidade de penetração, análises diretas de fortes absorventes e nivela amostras com espalhamento, por exemplo, os líquidos turvos ou sólidos nos modos de

refletância ou transmitância sem necessidade de pré-tratamento (SIMÕES, 2008). Deste modo, para utilizar os dados NIR como sinais analíticos é, geralmente, necessário o uso de técnicas quimiométricas.

### **3.4 Quimiometria**

A partir dos anos 60, os dados foram obtidos em grande quantidade e de maneira rápida e com menor esforço por parte dos analistas. Precisando analisar todos esses dados e extrair maior quantidade de informações relevantes, vários métodos estatísticos, matemáticos e computacionais estão sendo desenvolvidos no qual é possível o processamento e interpretação de dados, que antes dificilmente eram analisadas. De modo que hoje em dia pode-se afirmar que um equipamento para determinações analíticas por esta técnica utiliza hardware como instrumento e software com a quimiometria (SIMÕES, 2008).

A quimiometria pode ser dividida em três grandes áreas: planejamento e otimização de experimentos, calibração multivariada e reconhecimento de padrões (GELADI, 2003).

#### ***3.4.1 Calibração multivariada***

Calibração Multivariada é o procedimento que desenvolve modelos para quantificar uma propriedade de interesse da amostra (FERREIRA, 2015). Em uma amostra, para determinar sua concentração, o espectrofotômetro registra as absorvâncias que depende da concentração do analito e com o auxílio de um modelo matemático e estatístico é possível prever a propriedade do analito (FERREIRA et al, 1999).

Na calibração univariada, o sinal analítico é relacionado em apenas um comprimento de onda com a propriedade de interesse. Já na análise multivariada os modelos são construídos utilizando uma faixa de comprimentos de onda, atribuindo mais informação e mais robustez aos modelos (PEREIRA, 2011).

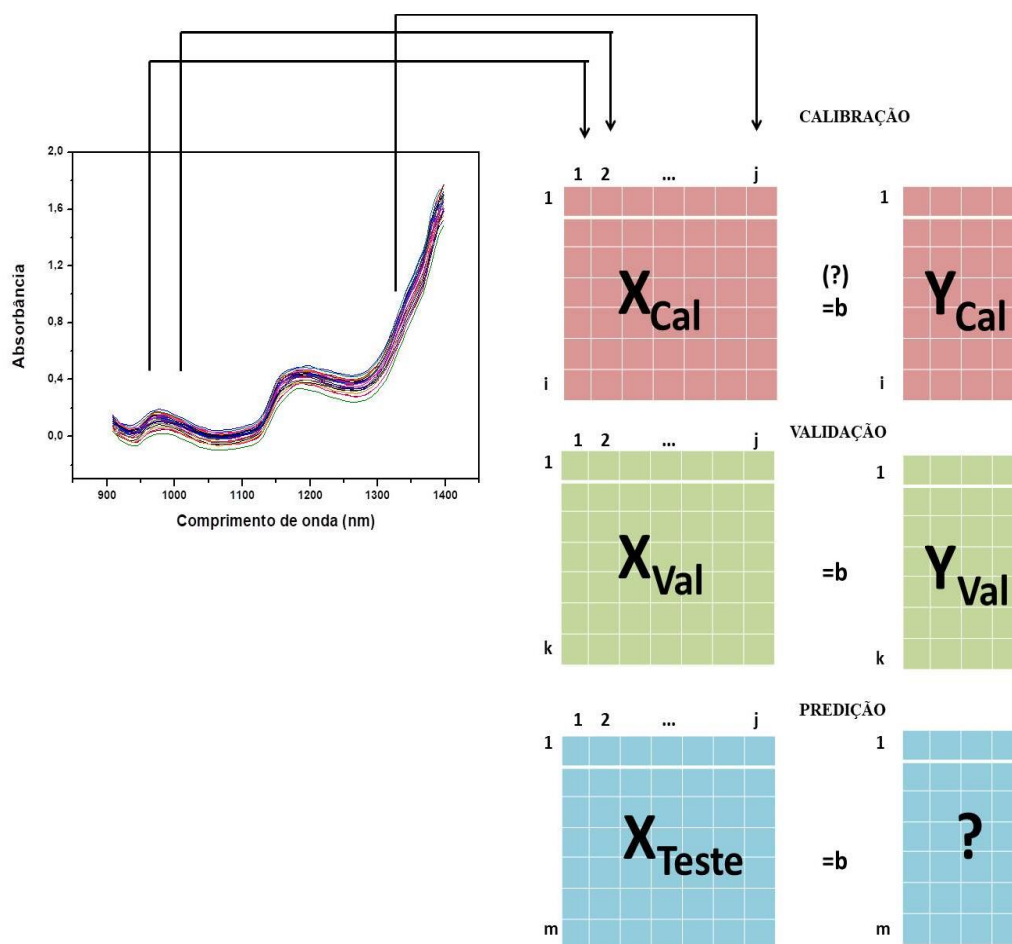
O processo de calibração multivariada é desenvolvido em três fases: a calibração, a validação e a predição. Na fase de calibração são construídos os modelos matemáticos, na fase de validação estes modelos são avaliados usando algumas ferramentas estatísticas para o diagnóstico e na fase de predição os modelos matemáticos validados são aplicados para prever as concentrações ou propriedades de amostras desconhecidas (SIMÕES, 2008).

Existem muitas técnicas de calibração multivariada, usadas para métodos de regressão, por exemplo, regressão por componentes principais (PCR - *Principal Component Regression*),

regressão linear múltipla (MLR - *Multiple Linear Regression*), regressão por mínimos quadrados parciais (PLS - *Partial Least Squares Regression*) e regressão de vetores de suporte (SVR - *Support Vector Regression*) (SIMÕES, 2008).

Os dados para a calibração multivariada são organizados de acordo com a Figura 1, no qual a matriz X, corresponde aos valores instrumentais do sinal analítico (Ex.: absorvância dos espectros a cada comprimento de onda, variáveis independentes) e o vetor Y, corresponde a propriedade de interesse (Ex.: as concentrações das amostras) estabelecida pelo o método de referência (variáveis dependentes) (GRANGEIRO JÚNIOR, 2011).

**Figura 1** - Organização de dados na calibração multivariada.



Fonte: Silva, 2018

Espectros de “ $i$ ” amostras do conjunto de calibração e de “ $k$ ” amostras de validação são obtidos em “ $j$ ” valores de comprimento de onda diferentes, formando uma matriz  $X_{cal}$  ( $i \times j$ ) para a etapa de calibração e outra matriz  $X_{val}$  ( $k \times j$ ) para a etapa de validação. As duas matrizes  $Y_{cal}$  ( $i \times q$ ) e  $Y_{val}$  ( $k \times q$ ) também são formadas com “ $q$ ” colunas, relacionadas aos diferentes

componentes presentes nas amostras e os valores de concentração são conhecidos por métodos de análise de referência (SIMÕES, 2008).

#### 3.4.1.1 Regressão por mínimos quadrados parciais (PLS)

A regressão por mínimos quadrados parciais foi desenvolvida por Herman Hold no ano de 1975 com uma modelagem de conjunto de dados complexos (ZIMMER; ANZANELO, 2014). Nela as matrizes  $X_{cal}$  e  $Y_{cal}$  (Figura 1) são decompostas em suas matrizes de pesos e escores (SIMÕES, 2008; GUIMARÃES, 2017) respectivamente, como mostrado nas Equações 5 e 6.

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} \quad (5)$$

$$\mathbf{Y} = \mathbf{UQ}^T + \mathbf{F} \quad (6)$$

onde:

$\mathbf{T}$  e  $\mathbf{U}$  são as matrizes dos escores;

$\mathbf{P}$  e  $\mathbf{Q}$  são as matrizes dos pesos das matrizes  $\mathbf{X}$  e  $\mathbf{Y}$ , respectivamente;

$\mathbf{E}$  representa a matriz de resíduos espectrais;

$\mathbf{F}$  representa a matriz dos resíduos de concentração.

Na decomposição em escores e pesos pode ser empregados muitos tipos de algoritmos, mas no resultado final eles são similares. Como uma representação desses algoritmos tem-se o NIPALS (do inglês, *Non-linear Iterative Partial Least Squares*) (SIMÕES, 2008). O modelo PLS final consiste em relacionar linearmente os *scores* da matriz  $\mathbf{X}$  com os *scores* da matriz  $\mathbf{Y}$ , de acordo com as Equações 7 e 8, respectivamente.

$$\mathbf{U} = \mathbf{BT} + \mathbf{G} \quad (7)$$

$$\mathbf{Y} = \mathbf{BTQ}^T + \mathbf{H} \quad (8)$$

onde:

$\mathbf{B}$  é a matriz dos coeficientes de regressão;

$\mathbf{G}$  é a matriz de resíduos dos escores;

$\mathbf{H}$  a matriz de resíduos de concentração.



### 3.4.1.2 Máquinas de Vetores de Suporte (SVM)

As máquinas de vetor de suporte (SVM, do inglês *Support Vector Machine*) são técnicas quimiométricas utilizadas em calibração multivariada e em problemas de reconhecimento de padrões e classificação de dados (LENARDON, 2016). Desenvolvidas na Rússia, na década de 60 por Vapnik, Lerner e Chervonenkis. As vantagens do uso das máquinas de vetores de suporte são sua flexibilidade, pois pode ser desenvolvida tanto pra classificação quanto pra regressão; capacidade de tratar os conjuntos de dados lineares e não lineares; apresenta ótima habilidade de generalização para pequeno conjunto de amostras. O SVM inicialmente, foi desenvolvido para classificação binária, mas, também é utilizado para problemas de classificação em múltiplas classes e resolve problemas de regressão, mantendo as principais propriedades que caracterizam o algoritmo de máxima margem, por exemplo, a dualidade, esparsidade, utilização da função Kernel e convexidade (ALVES, 2012). As máquinas de vetores de suporte são divididos em classificação de vetores de suporte (SVC, do inglês *Support Vector Classification*) e regressão de vetores de suporte (SVR, do inglês *Support Vector Regression*) (GUNN, 1998).

#### 3.4.1.2.1 Classificação de Vetores de Suporte (SVC)

O objetivo da classificação de vetores de suporte é realizar com que a máquina tenha a capacidade de classificar objetos de classes diferentes, classificando um objeto a classe que ele pertença. A aplicação é apresentada de forma linear, quando os dados lineares são separado por uma reta (GUNN, 1998). O classificador é conhecido por hiperplano, mostrado na Equação 9 (CHAMASEMANI; SINGH, 2011).

$$f(x) = w \cdot x + b = 0 \quad (9)$$

onde:

**w**: vetor de peso;

**x**: pontos sobre o hiperplano;

**b**: distância entre o hiperplano e o ponto de origem.

O hiperplano divide as classes em duas regiões, sendo elas:  $f(x) > 0$  ou  $f(x) < 0$  (LORENA; CARVALHO, 2007). Para obtenção dessa classificação é desenvolvida a função de sinal, ilustrado na Equação 10.

$$h(x) = \text{sgn}(f(x)) = \begin{cases} -1 & \text{se } w \cdot x + b < 0 \\ +1 & \text{se } w \cdot x + b > 0 \end{cases} \quad (10)$$

O -1 pertence a objetos de uma classe e o +1 pertence a objetos de outra classe.

Na forma não linear, corresponde para dados não lineares, ou seja, quando os dados não são separados por uma reta (GUNN, 1998), o mapeamento desses dados para espaços de dimensões maiores é desenvolvido pela a Equação 11 (HEARST,1998):

$$\Phi(\text{Obj}) = \Phi(x, y) = (x^2, \sqrt{2}xy, y^2) \quad (11)$$

onde:

**Obj:** objeto que se deseja mapear para um plano de dimensão maior;

**x,y:** coordenadas do Obj no espaço original.

Ao encontrar um classificador, são aplicadas as funções Kernel (CHAMASEMANI; SINGH, 2011). O objetivo destas funções é projetar dados do espaço de entrada para um espaço que tem como características alta dimensão, classificando estes dados de forma linear. Esse tipo de função é desenvolvida pela a Equação 12 (ARAÚJO, 2015):

$$K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j) = (x_i \cdot x_j)^2 \quad (12)$$

onde:

**K:** função Kernel;

**x<sub>i</sub>, x<sub>j</sub>:** pontos do espaço original.

A vantagem da função Kernel é que não precisa identificar o mapeamento  $\Phi$ , que é gerado. Deste modo, a utilização destas funções está na facilidade do seu cálculo e na sua habilidade de representar espaços abstratos (LORENA; CARVALHO, 2007).

#### 3.4.1.2.2 Regressão de Vetores de Suporte (SVR)

O SVR propõe determinar um hiperplano ótimo em que as amostras de treinamento estejam tão próximas quanto possível, o objetivo é encontrar uma função  $f(x)$ , com uma margem de erro determinado pelo intervalo  $[y_i - \epsilon, y_i + \epsilon]$  onde são permitidos desvios desde que não

ultrapassem a margem especificada (ABREU, 2016). A linha de regressão é encontrada utilizando a mesma equação apresentada para classificação descrita na Equação 13 (GUNN, 1998).

$$f(x) = w \cdot x + b \quad (13)$$

onde:

**w**: vetor de peso;

**x**: pontos sobre o hiperplano;

**b**: distância entre o hiperplano e o ponto de origem.

**w.x**: produto cartesiano de w.x.

A função de regressão está representada na Equação 14 (SMOLA; SCHOLKOPF, 1998).

$$f(x) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) (x_i \cdot x) + b \quad (14)$$

A variável **w** (vetor de pesos) está ilustrada na Equação 15 (SMOLA; SCHOLKOPF, 1998).

$$w = \sum_{i=1}^n (\alpha_i - \alpha_i^*) x_i \quad (15)$$

A variável **b** (bias) é definida de acordo com a Equação 16 (GUNN, 1998).

$$b = -\frac{1}{2} (w \cdot (x_i + x_s)) \quad (16)$$

A regressão dos vetores de suporte não linear pode ser utilizada para prever valores de dados não lineares, ou seja, dados que não são separados por uma reta, então é necessário fazer o mapeamento ( $\Phi$ ) dos dados não lineares, utilizando a mesma maneira como é desenvolvido na classificação de vetores de suporte não linear (BASAK; PAL; PATRANABIS, 2007).

A SVR precisa somente do produto escalar dos pontos do conjunto de dados (LORENA; CARVALHO, 2007). Para a cálculo do produto escalar é utilizado as funções Kernel (GUNN,1998).

O cálculo de  $w$  é definido de acordo com a Equação 17 (BASAK;PAL;PATRANABIS, 2007).

$$w = \sum_{i=1}^n (\alpha_i - \alpha_i^*) \Phi(x_i) \quad (17)$$

A equação  $b$  é definida de acordo com a Equação 18 (GUNN, 1998).

$$b = -\frac{1}{2} (\alpha_i - \alpha_i^*) \cdot K(x_i, x_s) \quad (18)$$

A função de regressão de vetores de suporte para dados não lineares, está ilustrado na Equação 19 (SMOLA; SCHOLKOPF, 1998).

$$f(x) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(x_i, x) + b \quad (19)$$

## **4 METODOLOGIA**

### **4.1 Localização do experimento**

O trabalho foi desenvolvido no Laboratório de Síntese Inorgânica e Quimiometria – LabSIQ, localizado na Universidade Estadual da Paraíba – UEPB.

### **4.2 Dados analíticos**

A matriz de dados utilizada neste trabalho é composta por espectros de bebidas mistas a base de cerveja, com %ABV variando de 1,58 – 17,11%. Os dados espectrais são fruto do trabalho de SILVA, 2018. Para a preparação da matriz de dados, foram adquiridos 4 lotes distintos de bebidas alcoólicas mistas a base de cerveja com sabor cítrico em estabelecimentos comerciais na cidade de Campina Grande-PB. O valor de %ABV rotulado foi de 7,9%. Para ampliar a faixa de concentrações, um dos lotes comerciais da bebida mista foi utilizada para preparar amostras com o teor alcoólico em %ABV superior e inferior ao rotulado. Para obter as amostras com a concentração inferior ao valor rotulado de 7,9% ABV, preparou-se diluições da cerveja base, adicionando-se diferentes quantidades de água deionizada com o auxílio de uma micropipeta. As amostras com a concentração superior ao valor rotulado de 7,9 % ABV foram preparadas a partir de uma solução estoque de cerveja com etanol PA 99,8% (8mL: 2mL). A matriz espectral, X, foi composta por 29 amostras, sendo 14 diluições, 11 fortificações e 4 amostras puras.

### **4.3. Espectroscopia NIR**

As análises dos espectros de absorvância, foram realizados através de um espectrômetro portátil MicroNir™ 1700 Spectrometer fabricado pela JDSU (Estados Unidos) com uma faixa espectral de 900 nm a 1650 nm e resolução de 12,5 nm. Também foi utilizada uma cubeta de quartzo de 1,003 cm.

#### 4.4 Construção e validação dos modelos

O software The Unscrambler® X 10.4 da CAMO Process AS, foi utilizado para construção e validação dos modelos quimiométricos e para a realização dos pré-processamentos dos espectros.

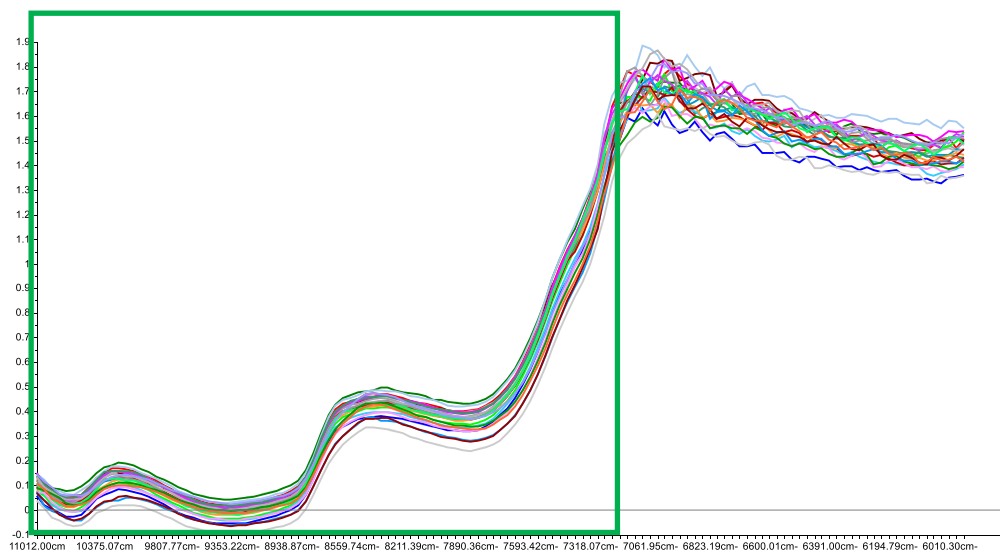
##### 4.4.1 Etapas para a construção dos modelos

Depois do registro dos espectros NIR, selecionou-se a faixa de trabalho. Neste trabalho utilizou-se uma estratégia diferente da adotada por SILVA (2018), que utilizou como amostras de calibração apenas amostras sintéticas e como amostras de predição apenas amostras reais. Deste modo, utilizou-se o algoritmo SPXy (GALVÃO *et al.*, 2005) para dividir os conjuntos de calibração (20 amostras) e teste (9 amostras), que continham ambas as amostras reais e sintéticas. Diferentes métodos de pré-processamentos foram testados para a construção dos modelos PLS e SVM com as amostras de calibração. E em seguida a capacidade preditiva destes modelos foram testadas utilizando as amostras de predição. A performance dos modelos PLS e SVM foram descritas e comparadas utilizando figuras de mérito como RMSE das etapas de calibração, validação e teste, número de variáveis latentes utilizadas para a construção do modelo,  $r^2$ , bias e desvio padrão.

## 5 RESULTADOS E DISCUSSÕES

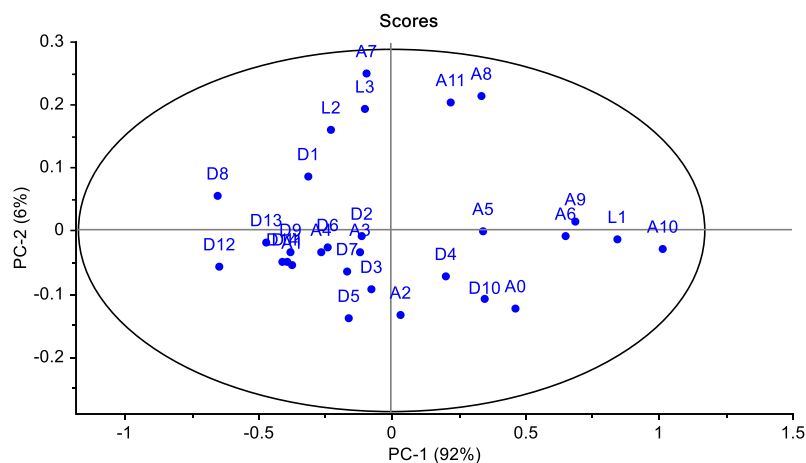
Inicialmente foi realizada uma análise prévia dos dados espectrais. A figura 2 apresenta os espectros brutos para todas as amostras analisadas.

**Figura 2** - Espectros brutos para todas as amostras analisadas e faixa de trabalho destacada em verde.

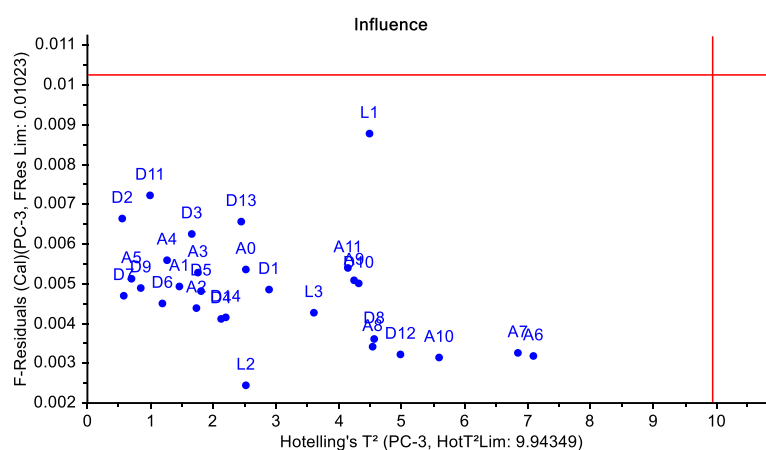


**Fonte:** Elaborada pelo autor, 2020

Nela pode-se observar uma região com alto ruído espectral que pode prejudicar a capacidade preditiva do modelo. Esta região foi excluída e a região espectral utilizada para a construção dos modelos foi de  $11.012 \text{ cm}^{-1}$  a  $7.187 \text{ cm}^{-1}$ . Além disso, pode-se observar a presença de um deslocamento da linha de base, ruído espectral e espalhamento do sinal. Deste modo, alguns pré-processamentos foram testados para remover estas fontes de erro para o modelo. Os pré-processamentos avaliados foram a derivada com polinômio Savitzky-Golay (variando o polinômio e o tamanho da janela de 5 a 21), variação normal padrão (SNV), correção do sinal multiplicativo (MSC), correção de sinal ortogonal (OSC) e correção de linha de base. Antes da construção do modelo de quantificação foi feita uma análise em componentes principais (PCA) com os dados brutos para observar o comportamento das amostras. A Figura 3 mostra (a) o gráfico dos escores e (b) de influência para as amostras brutas. De acordo com estes gráficos pode-se afirmar que não há a presença de amostras anômalas, pois todas as amostras se encontram dentro da elipse de confiança (3a) e nenhuma tem valores elevados de  $T^2$  de Hotelling ou  $F_{\text{residual}}$  (3b).

**Figura 3a** - Gráfico dos escores.

Fonte: Elaborada pelo autor, 2020

**Figura 3b** - Influência para as amostras brutas.

Fonte: Elaborada pelo autor, 2020

Deste modo, prosseguimos para a construção dos modelos PLS e SVM.

Foram construídos modelos PLS com todas as variáveis espectrais (PLS/MSC-full) e com as variáveis mais significativas selecionadas pelo algoritmo Jack-Knife (PLS/MSC-JK). Dos pré-processamentos testados, os que apresentaram melhor desempenho foram o MSC e SNV. Os dois modelos apresentaram valores de figuras de mérito semelhantes, deste modo, serão apresentados apenas os resultados para o MSC, de acordo com a Tabela 4.



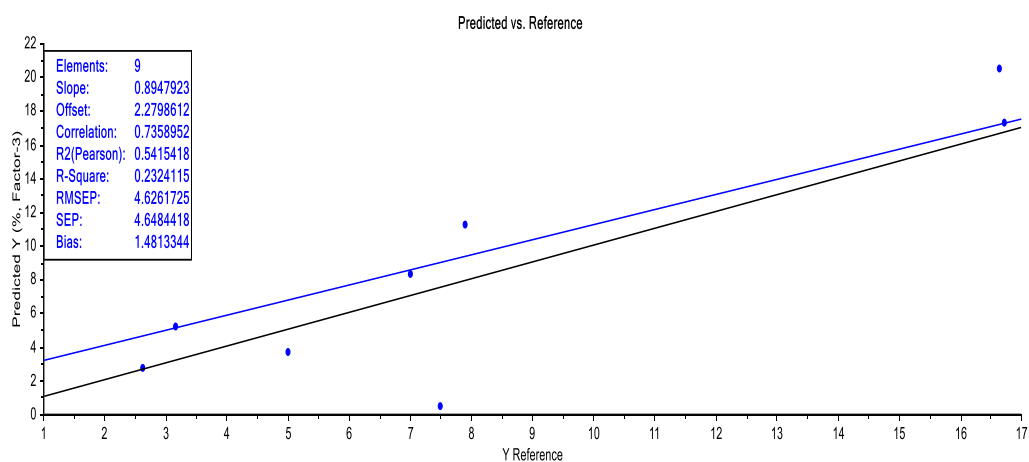
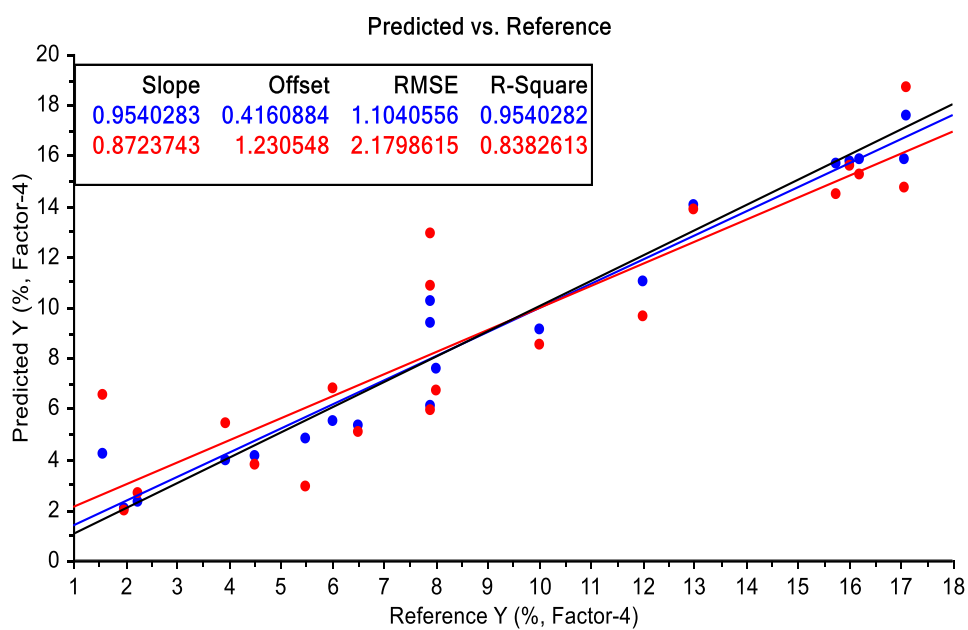
**Tabela 4** - Figuras de mérito para o modelo PLS

	PLS/MSC-full	PLS/MSC-JK
VL	4	3
RMSEC	1,10	2,04
RMSECV	2,18	2,63
RMSEP	3,9	4,62
Correlação	0,8385	0,7355
$r^2$	0,7031	0,2324
Bias	0,8825	1,48
Slope	1,0977	0,8547

**Fonte:** Elaborada pelo autor, 2020

De acordo com os valores de figuras de mérito observados, podemos notar que há uma discordância entre os valores de RMSEC, RMSECV e RMSEP para o modelo PLS/MSC-full, o que pode indicar um sobreajuste do modelo. Já o modelo PLS/MSC-JK mostra uma melhor concordância nos valores de RMSEC e RMSECV, no entanto este valor de erro praticamente dobra na etapa de predição (RMSEP), indicando a baixa capacidade preditiva do modelo. A Figura 4 mostra o gráfico dos valores medidos vs valores previstos pelo modelo PLS/MSC-JK (a) na etapa de calibração (em azul) e validação (em vermelho) e (b) na etapa de predição. Pode-se observar a grande dispersão das amostras em torno da bissetriz (em preto) que mostra a baixa performance do modelo. Um comportamento similar é observado para o modelo PLS/MSC-full.

**Figura 4-** Gráfico dos valores medidos vs valores previstos pelo modelo PLS/MS-CK na etapa de (a) calibração (em azul) e validação cruzada (em vermelho) e (b) etapa de predição.



Fonte: Elaborada pelo autor, 2020

A Tabela 5, mostra os valores medidos e preditos para os modelos PLS/MSK-full e PLS/MSK-JK, bem como seus respectivos desvios padrões.

**Tabela 5** – Desvios padrões para os modelos PLS/MSK-full e PLS/MSK-JK.

Valores medidos (%ABV)	Valores preditos (%ABV)		Desvio Padrão	
	PLS/MSK- full	PLS/MSK- JK	PLS/MSK- full	PLS/MSK- JK
7.0	6.6	8.3	0.26	0.91
5.0	3.5	3.7	1.07	0.94
7.5	1.8	0.5	4.01	4.95
1.7	9.6	12.2	5.54	7.39
3.2	3.8	5.2	0.45	1.45
2.6	0.4	2.7	1.55	0.08
16.7	17.9	17.2	0.84	0.37
16.6	21.7	20.5	3.57	2.74
7.9	10.9	11.3	2.11	2.38
	<b>Desvio padrão médio</b>		<b>2.16</b>	<b>2.36</b>

Fonte: Elaborada pelo autor 2020

Os valores observados atestam a baixa capacidade preditiva dos modelos PLS construídos, que pode estar relacionada ao baixo número de amostras que compõe o conjunto de dados. Uma solução para este problema é o uso da SVR.

Assim, fim de desenvolver um modelo com melhor capacidade preditiva testou-se o desempenho da regressão SVM. Os modelos SVM foram construídos com diversos pré-processamentos, a Tabela 6 mostra os valores medidos e preditos para os modelos construídos.

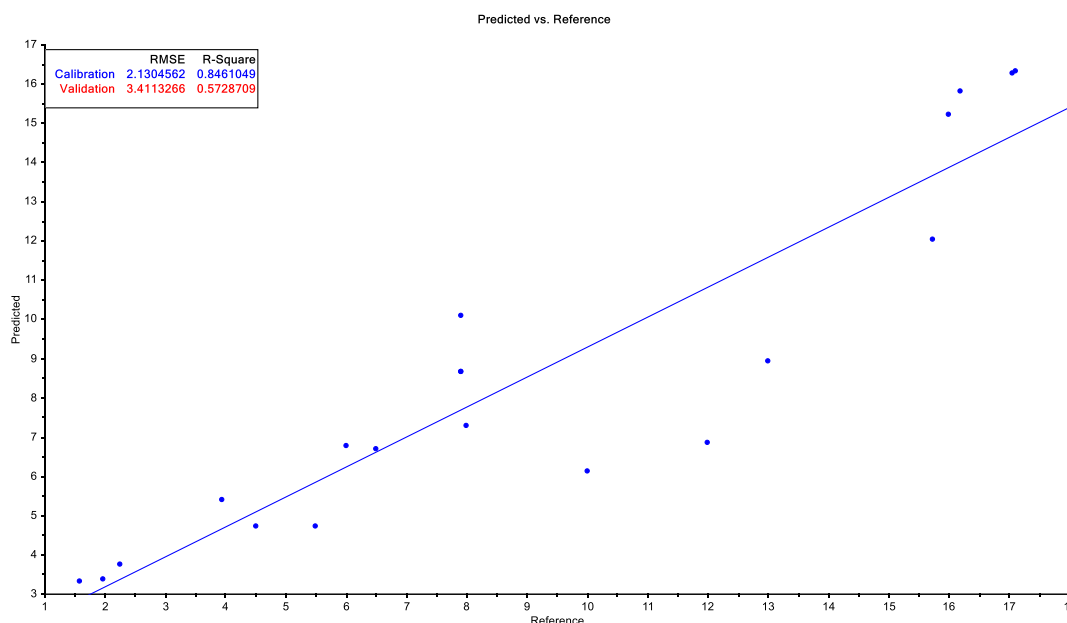
**Tabela 6** - Comparação dos valores medidos e preditos pelos modelos SVM construídos com diversos pré-processamentos e seus respectivos desvios padrões.

Valor Medido %ABV	Valores Preditos (%ABV)				Desvio Padrão			
	Correção de linha				Correção de linha			
	SNV	de base	MSC	Derivada	SNV	de base	MSC	Derivada
7	7.1	9.8	6.2	9.9	0.10	1.96	0.53	2.03
5	5.0	6.4	4.2	9.9	0.02	0.99	0.53	3.44
7.5	7.0	8.1	6.7	9.9	0.36	0.40	0.53	1.67
1.7	5.2	10.3	2.7	9.9	2.46	6.06	0.68	5.74
3.1	6.4	9.7	4.8	9.9	2.33	4.60	1.16	4.74
2.6	5.3	7.0	3.4	9.9	1.88	3.07	0.53	5.12
16.7	15.1	15.6	16.0	9.9	1.14	0.79	0.53	4.85
16.6	17.1	16.5	15.9	9.9	0.34	0.06	0.53	4.80
7.9	9.7	11.4	8.3	9.9	1.29	2.48	0.31	1.39
<b>Desvio Padrão Médio</b>					1.10	2.27	0.59	3.75

Fonte: Elaborada pelo autor, 2020

De acordo com os valores observados, pode-se notar que o pré-processamento que apresentou a melhor capacidade preditiva foi o MSC já o pré-processamento de derivada mostrou baixa capacidade preditiva visto que não consegue fazer distinção entre as diferentes concentrações. De fato, os valores preditos pelo modelo SVM-MS são concordantes com os valores medidos, apresentando um baixo desvio padrão médio. A Figura 5, mostra o gráfico de valores medidos vs valores preditos para a etapa de calibração para o modelo SVM construído com os dados pré-processados com MSC.

**Figura 5** - Gráfico de valores medidos vs valores preditos para o modelo SVM construído com os dados pré-processados com MSC.



**Fonte:** Elaborada pelo autor, 2020

Foi obtido um RMSEC de 2,13 ( $r^2 = 0,8461$ ) e um RMSEV de 3,41 ( $r^2 = 0,5728$ ). Pode-se observar uma grande dispersão dos valores e um afastamento da reta de regressão para amostras na faixa de 10 a 16 %ABV. No entanto, para o valor das amostras de interesse (7,0% ABV) o modelo mostrou um bom ajuste.

A Tabela 7 traz alguns parâmetros de comparação entre os modelos PLS e SVM construídos.

**Tabela 7** - Parâmetros de comparação entre os modelos PLS e SVM construídos.

	PLS/MSC-full	PLS/MSC-JK	SVM/MSC
RMSEC	1,10	2,04	2,13
RMSECV	2,18	2,63	3,41
RMSEP	3,9	4,62	0,84
Desvio padrão médio	2,16	2,36	0,59

**Fonte:** Elaborada pelo autor, 2020

A partir deles é possível perceber que apesar de ter uma performance comparável aos modelos PLS, os modelos SVM apresenta uma melhor capacidade preditiva.

## 6 CONCLUSÃO

A metodologia desenvolvida utilizando um espectrômetro NIR e as técnicas de calibração multivariada se mostrou viável para a quantificação de álcool em uma bebida mista a base de cerveja. Pode-se observar que a escolha do pré-processamento espectral interfere na capacidade preditiva do modelo. Para o banco de dados utilizado o SVM mostrou uma melhor performance na etapa de predição quando se utilizou o pré-processamento MSC. O método desenvolvido é prático, rápido e não utiliza reagentes químicos. Além disso, pelas características apresentadas, acredita-se que seja uma boa alternativa para a medida rápida da %ABV em cervejarias ou para ações de fiscalização. Além de ser de fácil implementação para medidas em linha.

## 7 PERSPECTIVAS

- Ampliar o número de amostras na faixa de concentração de 2 até 9% para a construção de modelos destinados a análise de cervejas com teor alcóolico nesta faixa de %ABV;
- Ampliar o número de amostras na faixa de concentração de 9 até 20% para a construção de modelos destinados a análise de cervejas com teor alcóolico nesta faixa de %ABV;
- Construção de modelos classificação de máquinas de vetores de suporte (SVM) para distinção entre cervejas originais e adulteradas

## REFERÊNCIAS

- ABREU, André Luiz Emidio de. **Bootstrap e modelos de support vector machine – SVM**. Orientador: Anselmo Chaves Neto. 2016. Tese (Doutorado em Métodos Numéricos) – Universidade Federal do Paraná, Curitiba, 2016.
- ALVES, Julio Cesar Laurentino. **Máquinas de vetores de suporte aplicada a dados de espectroscopia NIR de combustíveis e lubrificantes para o desenvolvimento de modelos de regressão e de classificação**. Orientador: Ronei Jesus Poppi. 2012. Tese (Doutorado em Ciências) – Universidade Estadual de Campinas, Campinas, 2012.
- ARAÚJO, Kevin Martins. **Utilização do algoritmo de máquinas de vetores de suporte (SVM) para predição de dados climáticos**. Orientador: Edeilson Milhomem da Silva. 2015. Monografia (Graduação em Ciência da Computação) – Centro Universitário Luterano de Palmas, Palmas, 2015.
- BASAK, Debasish; PAL, Srimanta; PATRANABIS, Dipak. Support vector regression. **Neural Information Processing**, v.11, n.10, p. 203-224, oct. 2007.
- BEAUMONT, Stephen. **Cerveja e comida: Princípios de harmonização, receitas e guia de degustação**. 1. ed. São Paulo: Publifolha, 2017.
- BLANCO, M. *et al.*, Application of multivariate curve resolution to chemical process control of an esterification reaction monitored by near-infrared spectroscopy. **Applied Spectroscopy**, v.60, p. 641-647, jun. 2006.
- BRASIL. **Decreto nº 6.871, de 04 de junho de 2009**: regulamenta a Lei nº 8.918, de 14 de julho de 1994, que dispõe sobre a padronização, a classificação, o registro, a inspeção, a produção e a fiscalização de bebidas. Diário Oficial da União, Brasília, 04 de junho de 2009.
- BRASIL. **Decreto nº 9.902, de 08 de julho de 2019**: regulamenta a Lei nº 8.918, de 14 de julho de 1994, que dispõe sobre a padronização, a classificação, o registro, a inspeção, a produção e a fiscalização de bebidas. Diário Oficial da União, Brasília, 08 de julho de 2019.
- BRASIL. Agência Nacional de Vigilância Sanitária. **Farmacopéia Brasileira**. Disponível em: < [https://www.gov.br/agricultura/pt-br/assuntos/inspecao/produtos-vegetal/legislacao-1/biblioteca-de-normas-vinhos-e-bebidas/farmacopeia\\_volume-2\\_2010\\_monografias.pdf](https://www.gov.br/agricultura/pt-br/assuntos/inspecao/produtos-vegetal/legislacao-1/biblioteca-de-normas-vinhos-e-bebidas/farmacopeia_volume-2_2010_monografias.pdf) >. Acesso em: 12/08/2020.
- BRUNELLI, Luciana Trevisan. **Produção de cerveja com mel: Características físico-químicas, energética e sensorial**. Orientador: Waldemar Gastoni Venturini Filho. 2012. Dissertação (Mestrado em Agronomia) – Universidade Estadual Paulista “Julio de Mesquita Filho”, Botucatu, 2012.
- CARVALHO, Emanuel da Silva. **Modelos quimiométricos para a identificação e quantificação de cafeína em baixas concentrações em meio aquoso**. Orientadora: Simone da Silva Simões. 2017. Dissertação (Mestrado em Ciência e Tecnologia Ambiental) – Universidade Estadual da Paraíba, Campina Grande, 2017.



CARVALHO, Giovani Brandão Mafra. **Obtenção de cerveja usando banana como adjunto e aromatizante**. Orientador: Orientador: João Batista de Almeida e Silva. 2009. Tese (Doutorado em Biotecnologia Industrial) – Escola de Engenharia de Lorena, Lorena, 2009.

CHAMASEMANI, Fereshteh Falah; SINGH, Yashwant Prasad. Multi-class Support Vector Machine (SVM) classifiers-An Application in Hypothyroid detection and Classification. *In: Sixth International Conference on Bio-Inspired Computing: Theories and Applications*, 2011.

DELIBERALLI, Camilo Camargo. **Cervejas artesanais no Brasil: Análise da comunicação integrada de marketing da cervejaria bodebrown**. Orientador: Aryovaldo de Castro Azevedo Júnior. 2015. Trabalho de Graduação (Curso de Comunicação Social – Publicidade e Propaganda) – Universidade Federal do Paraná, Curitiba, 2015.

EVANGELISTA, Rafaela Ribeiro. **Análise do processo de fabricação industrial de cerveja**. Orientador: Renato Tadeu Guerreiro. 2012. Trabalho de Graduação (Curso de Tecnologia em biocombustíveis) – Faculdade de Tecnologia de Araçatuba, Araçatuba, 2012.

FERREIRA, Márcia Miguel Castro. **Quimiometria: Conceitos, métodos e aplicações**. 1. ed. São Paulo: Editora da Unicamp, 2015.

FERREIRA, Márcia Miguel Castro *et al.* Quimiometria I: Calibração multivariada, um tutorial. **Química Nova**, São Paulo, v. 22, n. 5, p.724-731, jan. 1999.

GALVÃO, Roberto Kawakami Harrop *et al.* A method for calibration and validation subset partitioning. **Talanta**, v.67, n.4, p. 736-740, oct.2005.

GELADI, Paul. Chemometrics in spectroscopy. Part 1. Classical chemometrics. **Spectrochimica Acta Part B**, v.58, p. 767-782, 2003.

GRANGEIRO JÚNIOR, Severino. **Aplicação de métodos quimiométricos em análises de formas farmacêuticas sólidas Lamivudina (3TC) + Zidovudina (AZT) baseados na espectrometria de absorção na região do ultravioleta e infravermelho próximo**. Orientadora: Miracy Muniz de Albuquerque. 2011. Tese (Doutorado em Ciências Farmacêuticas) – Universidade Federal de Pernambuco, Recife, 2011.

GUIMARÃES, Pedro Queiroz. **Metodologia baseada em NIRS e quimiometria para a determinação de parâmetros de qualidade da quitosana para fins biomédicos**. Orientadora: Simone da Silva Simões. 2017. Dissertação (Mestrado em Química) – Universidade Estadual da Paraíba, Campina Grande, 2017.

GUNN, Steve. **Support vector machines for classification and regression**. Technical report, Image Speech & Intelligent Systems Group, University of Southampton, 1998.

HEARST, Marti A. *et al.* Trends and controversies - support vector machines. **IEEE Intelligent Systems**. v. 13, n.4, p. 18–28, 1998.

KUCK, Luiza Siede. **Cerveja: Sabor e Aroma**. Orientadora: Carla Mendonça. 2008. Monografia (Graduação em Química dos Alimentos) – Universidade Federal de Pelotas, Pelotas, 2008.

LAZARRI, L. M. et. Al. **Produção de cerveja**. Santa Catarina, UFSC, p. 18, 2009

LENARDON, Layane. **Ferramentas quimiométricas aplicadas no desenvolvimento de metodologia analítica para avaliação de adulteração em leite bovino por adição de soro do queijo por ATR-FTIR**. Orientadora: Ana Maria Bergold. 2016. Dissertação (Mestre em Ciências Farmacêuticas) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 2016.

LIMBERGER, Sílvia Cristina. O setor cervejeiro no Brasil: Gênese e evolução. **CardeNAU**, Rio Grande do Sul, v.6, n.1, 2013.

LORENA, Ana Carolina; CARVALHO, André Carlos Ponce de Leon Ferreira de. Uma Introdução às Support Vector Machines. **Revista de Informática Teórica e Aplicada (RITA)**. Rio Grande do Sul, v.14, n.2, p.43-67, 2007.

MAGRI, Daniel. **As 10 mais do mundo**: Cervejas com alto teor alcoólico. Disponível em: <<http://chefcomcerveja.com.br/cervejas-com-o-maior-teor-alcoolico/>> Acesso em: 27/08/2020

MAPA, Ministério da Agricultura, Pecuária e Abastecimento. **Anuário da Cerveja**. Disponível em: <<https://www.gov.br/agricultura/pt-br/assuntos/noticias/total-de-cervejarias-registradas-no-mapa-cresceu-36-em-2019-e-chegou-a-1.209>> Acesso em: 12/08/2020.

MORADO, Ronaldo. **Larousse da Cerveja**. 1. ed. São Paulo: Lafonte, 2009.

MORADO, Ronaldo. **Larousse da Cerveja**: A história e as curiosidades de uma das bebidas mais populares do mundo. 1. ed. São Paulo: Editora Alaúde, 2017.

NÓBREGA, Rossana Oliveira. **Aplicação da espectroscopia NIR para a determinação de polifenóis totais em pólen de abelha produzido na província de Buenos Aires – Argentina**. Orientador: José Germano Véras Neto. 2018. Monografia (Graduação Química Industrial) - Universidade Estadual da Paraíba, Campina Grande, 2018.

OLIVEIRA, Nayara Aline Muniz de. **Levedura utilizadas no processo de fabricação da cerveja**. Orientador: Carlos Augusto Rosa. 2011. Monografia (Especialista em Microbiologia Ambiental e Industrial) – Universidade Federal de Minas Gerais, Belo Horizonte, 2011.

PALMER, John. **How to brew**. 3. ed. Nova Iorque: Breweres Publications, 2006.

PASQUINI, Celio. Near Infrared Spectroscopy: Fundamentals, Practical Aspects and Analytical Applications. **Journal of Brazilian Chemistry Society**, v.14, n.2, p.198-219, 2003.

PASSARELLI, Edu. **Malte**: O elemento base da cerveja. 2009. Disponível em: <[http://www.beerlife.com.br/portal/default.asp?id\\_texto=22](http://www.beerlife.com.br/portal/default.asp?id_texto=22)>. Acesso em: 12/08/2020.

PEREIRA, C. M. **A Espectroscopia NIR no Controle de Qualidade de Solventes Industriais**. 2011. Dissertação (Mestrado em Engenharia Química) - Universidade Técnica de Lisboa, Lisboa, 2011

REINOLD, Matthias Rembert. **Manual Prático de Cervejaria**. 1. ed. São Paulo: Aden, 1997.

REIS, Juliana Pereira Zanon. **Dosagem de etanol utilizando álcool desidrogenase de levedura de panificação**. Orientadora: Edwil Ap. de Lucca Gattás. 2006. Dissertação (Mestrado em Ciências dos Alimentos) – Universidade Estadual Paulista “Julio de Mesquita Filho”, Araraquara, 2006.

ROSA, Natasha Aguiar; AFONSO, Júlio Carlos. A Química da Cerveja. **Quím. nova esc.**, São Paulo, v.37, n.2, p.98-105, mai. 2015.

SANTOS, José Ivan; DINHAM, Robert; ADAMES, Cesar. **O essencial em cervejas e destilados**. 2. Ed. São Paulo: Senac, 2006.

SANTOS JÚNIOR, Felisberto Gonçalves. **Determinação de uma nova versão do injetor comutador automático**: Determinação fotométrica de etanol em bebidas destiladas. Orientador: Boaventura Freire dos Reis. 2012. Dissertação (Mestre em Ciências) – Universidade de São Paulo, São Carlos, 2012.

SERRÃO, Keyla Freitas. **Determinação do teor alcoólico de cachaças por imagens digitais**. Orientador: Valdomiro Lacerda Martins. 2014. Dissertação (Mestrado em Ciências e Tecnologia de Recursos Amazônicos) – Universidade Federal do Amazonas, Itacoatiara, 2014.

SICARD, Delphine; LEGRAS, Jean-Luc. **Bread, beer and wine: yeast domestication in the Saccharomyces sensu stricto complex**. 3. ed. Comptes rendus biologiques, 2011.

SILVA, Adrielle Gomes da. **Metodologia alternativa para a determinação do teor alcoólico do produto misto a bases de cerveja utilizando espectroscopia NIR e PLS**. Orientadora: Simone da Silva Simões. 2018. Monografia (Graduação em Química Industrial) – Universidade Estadual da Paraíba, Campina Grande, 2018.

SILVA, Daniel Pereira. **Produção e avaliação sensorial de cerveja obtida a partir de mostos com elevadas concentrações de açúcares**. Orientador: João Batista de Almeida e Silva. 2005. Tese (Doutorado em Biotecnologia Industrial) – Faculdade de Engenharia Química de Lorena, Lorena, 2005.

SILVA, João Batista de Almeida e. Cerveja. *In*: VENTURINI FILHO, Waldemar Gastoni. **Tecnologia de bebidas**. São Paulo: Edgard Blücher, 2005. p. 44-78.

SIMÕES. Simone da Silva. **Desenvolvimento, de métodos validados para a determinação de captopril usando espectrometria NIR e calibração multivariada**. Orientador: Mário César Ugulino de Araujo. 2008. Tese (Doutorado em Química) - Universidade Federal da Paraíba, João Pessoa, 2008.

SINDICERV, Sindicato Nacional da Indústria da Cerveja. **O setor em números**. Disponível em: <<https://www.sindicerv.com.br/o-setor-em-numeros/>> Acesso em: 12/08/2020

SMOLA, Alex J.; SCHOLKOPF, Bernhard. **A tutorial on support vector regression**. Technical Report NC2-TR-1998-030, NeuroCOLT2, 1998.

VENTURINI FILHO, Waldemar Gastoni. **Bebidas alcoólicas: Ciência e tecnologia**. 2. ed. São Paulo: Blücher, 2016.

ZARPELON, Florenal. **As especificações do Álcool Focadas Para o Mercado Mundial**. Disponível em: < [https://www.agencia.cnptia.embrapa.br/Repositorio/Especificacoes\\_do-Alcool\\_Focado\\_para\\_Mercado\\_Mundial\\_000fxgfcrtu02wyiv80soht9hal6t8qx.pdf](https://www.agencia.cnptia.embrapa.br/Repositorio/Especificacoes_do-Alcool_Focado_para_Mercado_Mundial_000fxgfcrtu02wyiv80soht9hal6t8qx.pdf)>. Acesso em: 25/08/2020.

ZIMMER, Juliano; ANZANELO, Michel José. Um novo método para seleção de variáveis preditivas com base em índices de importância. **Production**, v. 24, n. 1, p. 84 – 93, jan./ mar. 2014.