

**Universidade Estadual Paulista
“Júlio de Mesquita Filho”**

Faculdade de Ciências Farmacêuticas

**Métodos espectroscópicos para classificação e
análise de qualidade de tomates**

Karla Rodrigues Borba

Tese apresentada ao Programa de Pós-graduação em Alimentos e Nutrição para obtenção do título de Doutor em Alimentos e Nutrição.

Área de Concentração: Ciência de Alimentos.

Orientador: Prof. Dr. Marcos David Ferreira
Coorientador: Prof. Dr. Luiz Alberto Colnago

Araraquara
2020

Métodos espectroscópicos para classificação e análise de qualidade de tomates

Karla Rodrigues Borba

Tese apresentada ao Programa de Pós-graduação em Alimentos e Nutrição para obtenção do título de Doutor em Alimentos e Nutrição.

Área de Concentração: Ciência de Alimentos.

Orientador: Prof. Dr. Marcos David Ferreira
Coorientador: Prof. Dr. Luiz Alberto Colnago

Araraquara
2020

B726m

Borba, Karla Rodrigues.

Métodos espectroscópicos para classificação e análise de qualidade de tomates / Karla Rodrigues Borba. – Araraquara, [s/n.]. 2020. 121 f. : il.

Tese (Doutorado). Faculdade de Ciências Farmacêuticas. Programa de Pós Graduação em Alimentos e Nutrição. Área de Concentração em Ciência de Alimentos.

Orientador: Marcos David Ferreira.

Coorientador: Luiz Alberto Colnago.

1. Espectroscopia. 2. Infravermelho. 3. Ressonância magnética nuclear. 4. Espectrômetro portátil. 5. Tomate. 6. Pasta de tomate. 7. Qualidade. 8. Quimiometria. 9. Machine learning. I. Ferreira, Marcos David, orient. II. Colnago, Luiz Alberto, coorient. III. Título.

Diretoria do Serviço Técnico de Biblioteca e Documentação - Faculdade de Ciências Farmacêuticas
UNESP – Campus de Araraquara

CAPES: 33004030055P6

Esta ficha não pode ser modificada

**Dedico ao meu filho Dante, ao meu amor Ricardo Verhaeg e à minha
família.**

Agradecimentos

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES, pelas bolsas concedidas.

Ao professor Dr. Marcos David Ferreira pela orientação ao longo de todo o doutorado.

Ao professor Dr. Luiz Alberto Colnago pela coorientação e pelos ensinamentos em espectroscopia.

Ao professor Dr. Luis E. Rodriguez-Saona, por me dar a oportunidade de passar parte do doutorado no seu laboratório na The Ohio State University.

Aos analistas, bolsistas e estagiários da Embrapa Instrumentação pelo apoio durante a realização dos experimentos. Em especial: Silviane, Sabrina, Nayara, Marcela e Fernanda.

À toda a equipe do departamento de alimentos na Universidade Estadual de Ohio. Em especial: Maria Izabel, Poliana e Peren Aykas.

À Faculdade de Ciências Farmacêuticas - UNESP.

À Embrapa Instrumentação pela infraestrutura.

Ao Departamento de Ciência e tecnologia de Alimentos, da Universidade Estadual de Ohio (The Ohio State University).

Obrigada!

Resumo

Objetivo: Durante a produção, processamento e comercialização de tomates e seus derivados é importante a avaliação de alguns atributos que irão influenciar diretamente na composição final e também na aceitação do consumidor. Teor de sólidos solúveis (SS), pH (potencial hidrogeniônico), ácidos orgânicos, açúcares, firmeza, consistência e viscosidade são alguns exemplos de parâmetros de qualidade geralmente avaliados e controlados durante toda a cadeia de produção e processamento de tomate. Os métodos utilizados, nos dias atuais, para esse controle são invasivos, demandam tempo, muitas amostras, reagentes e geram resíduos. Técnicas espectroscópicas, como a espectroscopia no infravermelho e de ressonância magnética nuclear, têm sido propostas como métodos alternativos. As principais vantagens da utilização de espectroscopia são: rapidez; possibilidade de determinação de vários atributos simultaneamente; e preparação mínima da amostra, sendo que, na maior parte do tempo não há necessidade de algum tipo de preparação prévia. Este projeto teve como objetivo: desenvolver um modelo de calibração, a partir da espectroscopia de infravermelho de médio e próximo, e ressonância magnética com o auxílio das ferramentas quimiométricas e *machine learning* para determinar e classificar a qualidade de tomates – frutos e pasta. **Métodos:** Os 3 primeiros capítulos consistem em modelos de predição e classificação desenvolvidos a partir de espectros de infravermelho e sinais de RMN-dt coletados de tomates intactos, enquanto o capítulo 4 foi realizado a partir da análise infravermelha de pasta de tomate. Capítulo 1: foi avaliada a utilização de um NIR portátil na determinação da qualidade de diferentes variedades de tomates no campo. Capítulo 2: espectros de infravermelho próximo e médio e ferramentas quimiométricas foram combinadas com o desenvolvimento de modelos de predição e classificação da qualidade de tomates destinados à indústria. Capítulo 3: Sinais de decaimento de RMN de tomates intactos juntamente com ferramentas quimiométricas de classificação e algoritmos de ‘machine learning’ foram utilizados para a criação de modelos de classificação de tomates, destinados ao processamento, de acordo com índice de maturação, SS e firmeza. Capítulo 4: Pastas de tomate, processadas por empresas na Califórnia - EUA, foram avaliadas com a utilização de um espectrômetro de infravermelho próximo portátil e os espectros foram correlacionados, por meio de PLS, com dados de qualidade (NTSS, pH, acidez titulável, consistência de Bostwick e Viscosidade) realizados por cada empresa. **Resultados:** Em todos os experimentos foram obtidos bons modelos de predição e/ou classificação da qualidade de tomate e pasta de tomate. No capítulo 1, modelos apresentaram boa predição, com coeficientes de correlação (r) acima de 0,81 (exceto para o modelo de determinação de glicose). A determinação de SS e firmeza (capítulo 2) resultaram em modelos com altos valores de r (0,79 e 0,93, respectivamente) e erros médios padrões baixos na determinação de pH. Os modelos de classificação do capítulo 3 apresentaram ótimos valores de acurácia (em %), próximos de 100%. Por fim, no capítulo 4, o uso de um NIR portátil combinado com PLS na análise de NTSS, consistência de

Bostwick e viscosidade resultou em modelos com alta precisão ($r \geq 0,93$).

Conclusão: A aplicação de espectroscopia na determinação e classificação de parâmetros de qualidade mostrou-se aplicável e uma ótima alternativa para produtores e empresas de processamento de tomate na análise de qualidade. Obtendo resultados rápidos, simultâneos, precisos e sem a necessidade de preparação das amostras.

Palavras-chave: Espectroscopia, infravermelho, ressonância magnética nuclear, espectrômetro portátil, tomate, pasta de tomate, qualidade, quimiometria, 'Machine learning'.

Abstract

Objective: During the production, processing, and marketing of tomatoes and their processed products, it is important to control some attributes that will directly influence the final composition and consumer acceptance. Content of soluble solids (SS), pH (hydrogen potential), organic acids, sugars, texture, consistency, and viscosity are some examples of quality parameters generally evaluated and controlled during the entire tomato production and processing stage. The methods used, nowadays, for this control are invasive, require time, many samples, reagents, and generate waste. Spectroscopic techniques, such as infrared and nuclear magnetic resonance spectroscopy, have been proposed as alternative methods. The main advantages of using spectroscopy are being a fast method; possibility of determining several attributes simultaneously; and minimal sample preparation, most of the time there is no need for any prior preparation. This project aimed to develop a calibration model, using medium and near-infrared spectroscopy, and magnetic resonance with the help of chemometric tools and machine learning to determine and classify the quality of intact tomatoes and paste. **Methods:** The first 3 chapters consist of prediction and classification models developed from infrared spectra and ¹H NMR signals collected from intact tomatoes, while chapter 4 was carried out from the infrared analysis of tomato paste. Chapter 1: the use of a portable NIR in determining the quality of different varieties of tomatoes in the field was evaluated. Chapter 2: near and medium infrared spectra and chemometric tools were combined and models for the prediction and quality classification of tomatoes for the industry were developed. Chapter 3: NMR decay signals from intact tomatoes along with chemometric grading tools and machine learning algorithms were used to create tomato grading models for processing according to ripeness index, SS, and firmness. Chapter 4: Tomato pastes, processed by companies in California - USA, were evaluated using a portable near-infrared spectrometer and the spectra were correlated, through PLS, with quality data (NTSS, pH, titratable acidity, consistency, and viscosity) performed by each company. **Results:** All experiments resulted in good models for predicting and / or classifying the quality of tomato and tomato paste. In chapter 1, models showed good prediction, with correlation coefficients (r) above 0.81 (except for the glucose determination model). The determination of SS and firmness (chapter 2) resulted in models with high values of r (0.79 and 0.93, respectively) and mean standard errors Low in pH determination. The classification models, based on NMR signals, in chapter 3 showed excellent accuracy values (in%), close 100%. Finally, in chapter 4, the use of a portable NIR combined with PLS in the analysis of NTSS, Bostwick consistency, and viscosity resulted in models with high precision (r ≥ 0.93). **Conclusion:** The application of spectroscopy in the determination and classification of quality parameters proved to be applicable and a great alternative for producers and companies of tomato processing in the quality analysis. Obtaining fast, simultaneous, accurate results and without the need for sample preparation.

Key-words: Spectroscopy, infrared, nuclear magnetic resonance, portable spectrometer, tomato, tomato paste, quality, chemometric, 'Machine learning'.

Sumário

	Página
Resumo	iv
Abstract	vi
Introdução	9
Capítulo 1. Portable Near infrared spectroscopy as a tool for fresh tomato quality control analysis in the field.	24
Introdução	26
Material e Métodos	28
Resultados e Discussão	31
Conclusões	38
Referências	38
Capítulo 2. Espectroscopia NIR e MIR na análise e classificação da qualidade de tomates destinados ao processamento.	44
Introdução	46
Material e Métodos	47
Resultados e Discussão	50
Conclusões	60
Referências	61
Capítulo 3. Selection of industrial tomatoes using TD-NMR data and computational classification methods.	66
Introdução	68
Material e Métodos	71
Resultados e Discussão	73
Conclusões	85
Referências	85
Capítulo 4. Rapid quality control of tomato paste based on a miniaturized NIR spectroscopy system.	91

Introdução	93
Material e Métodos	95
Resultados e Discussão	96
Conclusões	102
Referências	103
Considerações Finais	106
Referências	107

Introdução

O tomate é um fruto originário das Américas, Central e do Sul, que pertence à família das solanáceas (1), mundialmente cultivado, sendo a segunda hortaliça mais consumido no mundo, logo após a batata (2). É um fruto climatérico, ou seja, após a colheita continua passando por transformações fisiológicas de maturação. No final do estágio de maturação ocorre um aumento na respiração e depois um decréscimo (3,4). O fruto é macio, caracterizado por uma polpa suave, pele fina e muitas sementes (5). Cerca de 95% da composição do fruto de tomate é água, nos 5% restantes encontram-se os compostos inorgânicos e sólidos orgânicos solúveis e insolúveis (celulose por exemplo) (4,6). O tomate apresenta grande valor nutricional, com uma diversidade de compostos e vitaminas, contendo na sua composição vitamina C, provitamina A (β -caroteno) e, também, traços de potássio, fósforo e ferro, além dos açúcares (7).

A produção mundial de tomates em 2019 foi de aproximadamente 197 milhões de toneladas, em uma área de produção aproximada de 5 milhões de hectares (8). A Ásia é continente com maior produção do fruto, responsável por 52% do total. O continente americano ocupa a segunda colocação, com 18,7%, desta produção, onde Brasil é o nono maior produtor, representando quase 3,5% da produção mundial, com 4,3 milhões de toneladas em 2019 (8). O estado de São Paulo é o segundo maior estado produtor, com 22,29% da produção ficando atrás de Goiás que representa 26,27 % da produção brasileira (8).

Cultivares distintas são utilizadas na produção para o mercado fresco e de processamento. O tomate rasteiro, ou tomate industrial, é destinado para ser processado, sendo matéria-prima dos mais diversos produtos atomatados (extratos, molhos, sucos e outros derivados). Por sua vez, o tomate estaqueado é cultivado com o objetivo de ser destinado ao mercado fresco. Do total produzido, nos dois modelos de cultivo, cerca de 30% é utilizado para processamento e o 70% é destinado para a comercialização *in natura* (9).

A qualidade de um produto vegetal é importante tanto para tomates que serão consumidos *in natura* quanto para aqueles que serão processados, pois irá interferir diretamente no resultado final do produto assim como nos custos de produção. Portanto, seu controle garante que o produto final satisfaça as exigências e requisitos do mercado. Alguns parâmetros que devem ser controlados para uma boa qualidade são agrupados em da seguinte forma: aparência, textura, “flavor” (sabor e aroma), valor nutritivo e rendimento (3,4).

Para uma maior aceitação do consumidor, frutos destinados ao mercado fresco devem apresentar alguns padrões de qualidade, principalmente em atributos como, formato do fruto, teores de açúcar (sólidos solúveis e açúcares redutores), acidez total, teores de ácidos orgânicos (principalmente ácido ascórbico e cítrico), dentre outros. Ao longo dos anos programas de melhoramento em tomate focaram em ações para, principalmente, atingir altos níveis de rendimento e objetivando algumas características no formato do fruto (10). Nas últimas três décadas, devido a mudanças nas demandas dos consumidores por produtos com maiores teores nutricionais e qualidade sensorial, estes programas precisaram se adaptar e

focar numa melhora da qualidade do fruto além de aumentar o rendimento da produção (10,11). Tal mudança trouxe uma maior diversidade de cultivares, onde cada um deles apresenta características diferentes em relação a açúcares, ácidos e carotenoides.

Na indústria de processamento de tomate alguns atributos de qualidade, como a coloração, firmeza, consistência e o teor de sólidos solúveis são de maior importância para garantir um produto final nos padrões desejados (12). Os métodos de classificação da matéria prima utilizados pela indústria são, na grande maioria, manuais e não muito eficazes devido aos limites de seleção de frutos na esteira de processamento. A adição de uma etapa automatizada de classificação e exclusão de frutos impróprios resultaria num maior rendimento e qualidade final.

Durante o processamento a cor do tomate tem grande importância, pois irá influenciar diretamente na coloração final do produto atomatado (13). O consumidor é exigente quanto à cor do extrato de tomate sendo que a coloração vermelho intensa é preferível (9,13). Portanto deve-se evitar uma grande proporção de tomates verdes, para que o produto final se mantenha vermelho intenso e satisfaça o consumidor. Além disso, tomates com coloração vermelho intenso apresentam altos teores de carotenoides, mais especificamente licopeno, aumentando os teores de compostos bioativos do fruto (13) e conseqüentemente do tomate processado. Esse composto tem característica antioxidante e alguns estudos mostram que seu consumo pode diminuir os riscos de doenças crônicas como câncer e doenças cardiovasculares (14,15). O teor de licopeno, em geral, no fruto de tomate *in*

natura é de 30 mg.kg⁻¹, no suco de tomate 159 mg.kg⁻¹ e no ketchup, em média, 100 mg.kg⁻¹ do produto (7). Mesmo apresentando grande importância na qualidade, o teor de carotenoides não tem sido usualmente analisado ou controlado durante a cadeia de processamento, isso porque seus métodos de análise demandam muito tempo e manipulação das amostras, além de serem destrutivos.

Juntamente com a coloração, a consistência final do molho ou pasta de tomate é fator crucial para o consumidor (16). A consistência final depende de fatores como grau de amadurecimento do fruto e diferentes cultivares (17). Ao longo do processo de amadurecimento ocorre a perda progressiva da consistência (ou firmeza) em decorrência de um processo complexo que envolve diferentes mecanismos como: perda do turgor celular, redução do tamanho e distribuição dos polímeros da parede celular, ação de enzimas hidrolíticas e outros mecanismos não enzimáticos (3,4).

Uma das principais características avaliadas e controladas durante o processamento de tomates é o teor de sólidos solúveis (SS), o qual é medido por refratometria, em °Brix, e é utilizado como medida estimada do teor de açúcares nos frutos, podendo indicar o seu grau de amadurecimento (12,18). Os SS estão, diretamente, relacionados ao rendimento industrial, quanto maior o seu valor, maior será o rendimento e menor o gasto de energia no processo de concentração da polpa (19). Isso porque, quanto maior a quantidade de SS do fruto menor será a quantidade de água, gerando um aumento do rendimento da polpa concentrada (20). Segundo estudos da Embrapa Hortaliças (2006), para cada aumento de 1° Brix na matéria-prima

existe um incremento de 20% no rendimento industrial. As indústrias brasileiras recebem tomates com SS médio de 4,5° Brix (21).

Além do SS, a acidez do produto é outro parâmetro do atributo 'flavor' que deve ser avaliado e controlado durante a etapa de comercialização e processamento. A acidez é atribuída, principalmente, aos ácidos orgânicos que se encontram dissolvidos nos vacúolos das células. A acidez da polpa influencia no sabor do fruto e interfere também no período de aquecimento durante a esterilização da polpa. Um pH inferior a 4,5 é necessário para impedir a proliferação de micro-organismos no produto final, sendo que valores de pH acima de 4,5 requerem períodos mais longos de esterilização, levando a um maior consumo de energia e maior custo do processamento (7). Portanto, realizar uma classificação inicial selecionando tomates com maiores teores de SS e adequado pH pode aumentar o rendimento industrial.

Além do controle da matéria prima é essencial que, durante todo o processamento, seja avaliada e garantida a qualidade do produto industrializado. Laboratórios de qualidade dentro das indústrias fazem análises, nas suas linhas de produção, 24h por dia. As análises de consistência, viscosidade, teor de sólidos solúveis, acidez e coloração são realizadas por meio de métodos convencionais, como titulação, refratometria, colorimetria, dentre outras. Esse tipo de metodologia destrutiva requer despesas mais elevadas com amostras e reagentes, além de demandar muito tempo e mão de obra especializada para análise.

Portanto, todos os atributos descritos são importantes de ser avaliados desde o cultivo até o processamento dos tomates. O desenvolvimento e

utilização de novos métodos de análises, mais rápidos e não destrutivos, podem otimizar todo o processo de análises laboratoriais e de campo. Além disso, com técnicas rápidas e de menor manipulação das amostras alguns parâmetros importantes, e que em geral não são controlados como aqueles que dizem a respeito ao valor nutricional do produto, poderão ser analisados e controlados agregando valor ao produto final (12).

Técnicas rápidas, baratas e não destrutivas, como a utilização de espectroscopia de infravermelho e ressonância magnética nuclear, são uma ótima alternativa durante todas as etapas do processamento e também para a avaliação da qualidade de tomates destinados ao mercado fresco. Nestes métodos a manipulação/preparo da amostra é a mais reduzida possível (na maioria das vezes não é necessário sequer algum preparo), e na maioria dos casos, possibilita a avaliação simultânea de todos os atributos de qualidade (23–28).

A espectroscopia é o estudo da interação da radiação eletromagnética com a matéria, essa radiação pode ser dividida em regiões de energia (29). No caso da radiação de infravermelho, que corresponde à parte do espectro eletromagnético situada entre a região do visível e das micro-ondas, ela está subdividida em infravermelho próximo - NIR ($4000 - 12500 \text{ cm}^{-1}$), infravermelho médio - MIR ($400 - 4000 \text{ cm}^{-1}$) e distante ($100 - 400 \text{ cm}^{-1}$) (29). O comprimento de onda de uma absorção depende das massas relativas dos átomos, das constantes de força das ligações e da geometria dos átomos (31). Por sua vez, a região do espectro eletromagnético que corresponde a ressonância magnética nuclear é a região de rádio frequência (30) e a interação ocorre

quando as spins nucleares dos elementos contidos nas amostras estão na presença de um campo magnético.

As técnicas de infravermelho e ressonância magnética nuclear (RMN), aliadas com ferramentas matemáticas e computacionais, podem ser uma alternativa para desenvolvimento de métodos rápidos e precisos para análise de qualidade. Mostrando grande potencial para monitoramento da qualidade de produtos vegetais em linhas de processamento.

Os espectros de infravermelho próximos - NIR são complexos, pois apresentam informações de sobretons e de bandas de combinação. Isso torna os espectros na faixa NIR muito mais complexos do que os espectros de MIR o que dificulta a atribuição de bandas específicas de absorção (32,33) e pode tornar difícil a interpretação de espectros de misturas complexas como frutas e hortaliças. Levando a necessidade de recorrer a métodos estatísticos multivariados para gerar modelos utilizando muitos comprimentos de onda (34).

A técnica de infravermelho próximo tem aplicações na análise quantitativa e qualitativa de materiais (35), fármacos (36), alimentos (37–40) e também no controle de processos (41). Muitos estudos mostraram a viabilidade e eficiência na utilização do NIR na predição da qualidade interna de frutas (38,42–46) e hortaliças (47). Modelos de calibração, a partir da análise dos espectros obtidos por infravermelho próximo, foram desenvolvidos para estimar firmeza e sólidos solúveis, de forma não destrutiva, em pêssegos (48,49) e citros (45,50–53).

Clément e colaboradores (2008) analisando parâmetros de qualidade em tomates, de diferentes cultivares, de forma não destrutiva por meio de NIR obtiveram um modelo de predição de licopeno e coloração com alta acurácia ($r^2 = 0,96$ e $r^2 = 0,98$, respectivamente). Porém, para a predição da qualidade interna (pH, SST, acidez titulável e condutividade elétrica) os modelos foram menos precisos, devido a baixa variabilidade desses parâmetros entre as amostras.

Em estudo da análise simultânea de SS, açúcares e acidez em produtos concentrados de tomate (20) foram obtidos modelos com habilidade de previsão, por meio da ferramenta quimiométrica PLS (Regressão por mínimos quadrados parciais, do inglês *Partial Least Squares*). He e colaboradores (2005) comprovaram o potencial de aplicação da espectroscopia de reflectância no visível e infravermelho próximo como método não destrutivo na medição de características do tomate 'Heatwave' (SS, pH, e firmeza), eles desenvolveram modelos com alta correlação e baixo erro.

Assim como NIR, a aplicação da técnica de MIR já vem sendo amplamente utilizada, sendo a principal região da espectroscopia vibracional é a do infravermelho médio (MIR) (30). Espectros de MIR fornecem informações de frequência de vibrações moleculares fundamentais (56), que envolvem o estado fundamental e o primeiro estado vibracional excitado (30,57). Essa técnica permite a identificação de moléculas orgânicas e a caracterização de grupos funcionais, tais como aqueles presentes em ácidos orgânicos e polissacarídeos (32). Até alguns anos atrás, a sua utilização

estava restrita a análise qualitativa de compostos orgânicos, relacionada principalmente à identificação de ligações moleculares. Com o desenvolvimento de espectrofotômetros com transformada de Fourier (FT) houve um aumento no número e tipo de aplicações da espectroscopia na região do MIR (57,58). A partir de então, passou também a ser usado para análise quantitativa de amostras complexas (24).

Os espectros obtidos na região do MIR, diferentemente dos obtidos por NIR, apresentam bandas bem definidas, facilitando a interpretação de seus sinais (12). Esta faixa espectral é bastante útil para a identificação de compostos, pequenas diferenças na estrutura e na constituição das moléculas, as quais resultam em mudanças significativas do perfil e na distribuição das bandas de absorção, tornando possível a identificação de um composto orgânico comparando seu espectro com banco de dados pré-existentes (59).

Apesar do espectro de infravermelho médio (MIR) ser o mais utilizado na química orgânica, pois fornece informações diversas que auxiliam na análise da composição das moléculas químicas, existem ainda poucos estudos que relatam o uso de MIR para determinar composição de frutas e hortaliças intactas (60,61).

Ao comparar a utilização da técnica de MIR e NIR na análise de ácidos orgânicos (ácido cítrico e ácido málico) e açúcares (sacarose, frutose e glicose) em amostras liofilizadas de maracujá, De Oliveira e colaboradores (2014) encontraram que modelos de MIR apresentaram melhores resultados do que aqueles construídos com NIR.

Outra técnica baseada em espectroscopia, a ressonância magnética nuclear (RMN) se fundamenta no registro da interação da radiação eletromagnética na região de rádio frequência, 2 a 1200 MHz, com núcleos atômicos (30). Os núcleos absorvem radiação eletromagnética quando expostos a um campo magnético. A técnica de RMN permite a observação de núcleos com momento magnético nuclear ou spin $\neq 0$ (62,63). A espectroscopia de RMN pode ser subdividida em aplicações em alta e baixa resolução. Espectrômetros de baixo campo (ou baixa resolução), também chamada de RMN no domínio do tempo (RMN-DT) vêm sendo utilizados para análise de qualidade de produtos agroindustriais desde a década de 1970 (64,65). A rapidez na análise sem necessidade de destruição da amostra, além da facilidade na execução, são as principais vantagens dos espectrômetros de baixo campo sobre os métodos convencionais (27).

A utilização de RMN para análise de produtos alimentícios vem crescendo ao longo dos últimos anos (66). A técnica vem sendo aplicada na análise de parâmetros de qualidade em frutas como SS em ameixas (28) e atributos sensoriais e qualidade físico-química em laranjas (42,67,68), mangas (69), bananas (62) e também na determinação de adulteração de alimentos, como leites (64) e azeites (65). Em um estudo de classificação de frutos a partir de atributo sensorial, Flores e colaboradores (2016) classificaram laranjas em dois grupos de acordo com o SS de seus sucos. Essa classificação foi obtida por meio dos sinais de decaimento de RMN das laranjas intactas e posteriormente foi realizada análise sensorial que corroborou com a classificação instrumental.

Devido a complexidade dos dados gerados tanto a partir das técnicas de infravermelho quanto de RMN, elas são dependentes de ferramentas estatísticas e matemáticas elaboradas e complexas para processar os sinais analíticos (espectros ou sinais de decaimento por exemplo) (23), retirando o máximo de informações importantes possíveis. Nos últimos anos, o número de estudos com técnicas instrumentais mais sofisticadas para análises de alimentos têm aumentando, isso tem sido possível graças ao surgimento de métodos mais complexos e robustos para tratamento dos dados, como é o caso de ferramentas quimiométricas e algoritmos de aprendizado de máquinas.

A quimiometria, ou análise multivariada, tem como objetivo facilitar a interpretação de dados multivariados, retirando o máximo de informação possível (23). Dentre as diversas áreas da análise multivariada podemos citar quatro principais: planejamento e otimização de experimentos, pré-tratamentos matemáticos, métodos de classificação e métodos de regressão ou calibração multivariada. De uma maneira geral, calibração multivariada é composta por três etapas: calibração, predição e validação (70). Neste trabalho os métodos de calibração multivariada foram utilizados, pois possibilitam a realização de determinações quantitativas de compostos de interesse presentes na amostra.

A ferramenta base dos métodos de calibração é a Análise de Componentes Principais (PCA, do inglês Principal Component Analysis), ela consiste em uma maneira de identificar a relação entre características extraídas de dados complexos quando os vetores de características têm

muitas dimensões (71). A PCA é uma análise exploratória dos dados não supervisionada, que transforma um conjunto de variáveis originais em outro conjunto de variáveis da mesma dimensão, denominadas de componentes principais (72).

Outra ferramenta utilizada na quimiometria é a regressão em mínimos quadrados parciais (PLS, do inglês *Partial Least Squares*). PLS é um método matemático baseado em uma relação linear entre variáveis instrumentais (X) e as variáveis de interesse (Y) (73). Esse é um método eficiente que consegue lidar com ruídos experimentais, colinearidade e não linearidade (74), além de ser considerado um método robusto, ou seja, os parâmetros do modelo desenvolvido não se alteram de maneira significativa quando novas amostras são acrescentadas ou retiradas do conjunto de calibração. Na construção dos modelos PLS, as variáveis X e Y são decompostas simultaneamente em uma soma de h variáveis latentes. Onde X é uma matriz $n \times p$ onde n é o número de amostras e p o número de variáveis de medida (absorbância em determinados comprimentos de ondas, por exemplo) e Y é um vetor da propriedade de interesse a ser medida pelo modelo (74,75).

A análise discriminante pelos mínimos quadrados parciais (PLS-DA) é um método de reconhecimento de padrão supervisionado baseado no algoritmo PLS, técnica que busca a relação entre a resposta instrumental e propriedades de interesse (76). A relação é estabelecida por meio de uma matriz contendo as repostas instrumentais, X, com um vetor ou uma matriz contendo as propriedades de interesse, Y. No caso do PLS-DA à matriz X

permanece com os dados instrumentais (espectros) e a matriz Y recebe a designação da classe a qual a amostra pertence (77).

Assim como PLS-DA, a ferramenta SIMCA, Modelagem suave independente por analogia de classe (do inglês, *Soft Independent Modeling by Class Analogy*), é outro método quimiométrico supervisionado utilizado para o desenvolvimento de modelos de classificação. Foi a primeira técnica de modelagem de classe introduzida em quimiometria, SIMCA desenvolve modelos de classificação com base em uma PCA realizada utilizando apenas as amostras da categoria estudada (71). Ele assume que os valores medidos para um grupo de amostras parecidas tenderão para uma distribuição uniforme e modelável (78).

Além da quimiometria, tem sido implementado técnicas de *Machine learning*, ou em sua tradução aprendizado de máquina, no desenvolvimento de modelos para controle de qualidade em alimentos e demais setores. Aprendizado de máquina é umas das grandes aplicações da área de Inteligência Artificial (IA) (79). Existem diversos algoritmos de IA, em geral eles são utilizados para gerar classificadores para um conjunto de exemplos (80). Alguns autores caracterizam *Machine learnig* como programas que automaticamente melhoram com a experiência do algoritmo (81), outros definem como o processo de fazer o computador modificar ou adaptar suas ações para que obtenham maior acurácia (82), sendo que acurácia é uma possível medida de desempenho que representa a medida do quão bem as ações acolhidas refletem as corretas (79).

O *support vector machine* (SVM), ou máquinas de vetores suporte, é um dos principais algoritmos de *Machine Learning* para desenvolvimento de modelos de classificação (82) e foi introduzido na área pelo pesquisador Vapnik em 1992. Algumas características do algoritmo SVM são: boa capacidade de generalização, a qual é medida por sua eficiência na classificação de dados que não pertençam ao conjunto utilizado em seu treinamento (portanto, boa eficiência no conjunto de teste) (83); robustez em grandes dimensões, SVMs são ferramentas robustas quando utilizam dados com muitas dimensões, por exemplo imagens e espectros de infravermelho com centenas de variáveis; convexidade da função objetivo; e teoria bem definida dentro da matemática e estatística (80).

De maneira simplificada, o SVM separa um conjunto de objetos com diferentes classes utilizando um conceito de planos de decisão. O objetivo do algoritmo é encontrar um hiperplano ótimo de classificação (80). Para isso, o conjunto de dados original é compactado em conjuntos de vetores de suporte (84). SVM é um importante algoritmo de abordagens orientadas a dados para monitoramento de processos industriais, e vem sendo utilizada em diversas áreas como no processamento de informação biológica (85,86), reconhecimento e recuperação de imagens (87), diagnóstico de falhas industriais (88,89) e classificação em diversas áreas do agronegócio e indústria alimentícia (67,90–92).

Assim como SVM, o algoritmo *k- Nearest Neighbor* (kNN), ou *k* – vizinhos mais próximos, também é um algoritmo de aprendizado supervisionado dentro da área de *Machine learning* utilizado para

classificação (93). Ele é um dos mais simples de todos os algoritmos de aprendizado de máquina. O conceito geral consiste em classificar um novo exemplo calculando a distância entre os k exemplos mais próximos a ele (67).

Em resumo, um objeto é classificado por maioria dos votos de seus vizinhos, com o objeto sendo atribuído à classe mais comum entre os k vizinhos mais próximos (k é um número inteiro positivo, tipicamente pequeno). Se $k = 1$, o objeto é simplesmente atribuído à classe do seu vizinho mais próximo (94). Esse algoritmo vêm sendo utilizado em diversos trabalhos no setor da agricultura (95) e mais especificamente na indústria de alimentos (67,90,92,96,97).

O objetivo deste trabalho foi avaliar a utilização de técnicas espectroscópicas, de infravermelho próximo e médio e de ressonância magnética nuclear no domínio do tempo, combinadas com ferramentas quimiométricas (PLS, PLS-DA e SIMCA) e algoritmos de aprendizado de máquinas (SVM e kNN), para monitoramento da qualidade de tomates para consumo *in natura* e aqueles destinados como matéria prima para processamento, como também em produtos derivados desse processamento industrial.

Capítulo 1.

**Portable Near infrared spectroscopy as a tool for fresh tomato
quality control analysis in the field**

Artigo submetido em 17 de julho de 2020 para a revista Food Control.

Portable Near infrared spectroscopy as a tool for fresh tomato quality control analysis in the field

Karla R. Borba^a, Didem P. Aykas^{b, c}, Maria I. Milani^d, Luiz A. Colnago^e, Marcos D. Ferreira^e and Luis E. Rodriguez-Saona^b.

^aDepartment of Food and Nutrition, School of Pharmaceutical Sciences, São Paulo State University-UNESP, Araraquara – Jaú, Km 1, 14801-902 - Araraquara-SP, Brazil.

^bDepartment of Food Science and Technology, The Ohio State University, 100 Parker Food Science and Technology Building, 2015 Fyffe Road, Columbus, Ohio – USA.

^cDepartment of Food Engineering, Faculty of Engineering, Adnan Menderes University, Aydin, Turkey.

^dDepartment of Analytical Chemistry, Institute of Chemistry, UNESP – São Paulo State University, 55 Prof. Francisco Degni Street, Araraquara, SP, Brazil.

^e Embrapa Instrumentation, XV de Novembro, 1452, 13561-206 - São Carlos-SP, Brazil.

*Corresponding author < marcos.david@embrapa.br >

ABSTRACT

Portable spectrometer devices are promising tools that can be an alternative for various purposes of analyzing food quality, such as monitoring in a few seconds the internal quality during fruit ripening on the field. A portable/handheld (palm sized) Near-infrared spectrometer (Neospectra, Si-ware) with a spectral range of 3800-7700 cm^{-1} , equipped with a micro-electro-mechanical system (MEMs), was used to develop prediction models to evaluate tomato quality attributes nondestructively. Soluble solid content (SSC), fructose, glucose, titratable acidity (TA), ascorbic and citric acid contents of different types of fresh tomatoes were analyzed with standard methods and those values were correlated to spectra data by Partial Least Squares Regression (PLSR). Fresh tomato samples were obtained in the years of 2018 and 2019 crops in a commercial production, and 4 fruit types were evaluated: roma, round, grape, and cherry tomatoes. This large tomato types variation and fruits from distinct years resulted in a wide range in quality parameters enabling robust PLSR models. The results showed accurate prediction and a good correlation (R_{pred}) for SSC=0.87, fructose=0.91, glucose=0.73, ascorbic acid=0.81, and citric acid=0.86. Our results support that handheld NIR spectrometer have high potential to determine simultaneously, several quality attributes of different types of tomatoes in a practical and fast way.

Keywords: Infrared spectroscopy; handheld spectrometer; fresh tomato quality; Partial Least Squares (PLS); field production.

1. Introduction

The great popularity of the tomato is mainly related to the fact that it can be eaten in multiple fresh or processed forms, being the second most important vegetable crop produce worldwide. In 2019, the tomato global production was approximately 197 million tons (FAO, 2020), with 75% for the fresh market and 25% for processing. USA is the world's third-largest tomato producer. California and Florida are the biggest

producers representing two-thirds of the USA total production area (Costa & Heuvelink, 2018).

For a long time, tomato breeding programs has focused on crop yield, fruit shape, and shelf-life (Bai & Lindhout, 2007). However, over the past 30 years, tomato breeders have focus on the development of high-quality tomato to meet customer demands for fresh fruits with good visual appearance, higher organoleptic and nutritional characteristics (Bai & Lindhout, 2007; Ibáñez, Valcárcel, Cebolla-Cornejo, & Roselló, 2019). Therefore, to guarantee internal quality of the fruit, its titratable acidity, soluble solids, sugars, organic acids, and lycopene contents have to be taken into consideration. The conventional techniques used to determine these quality parameters are laborious and time-consuming, determine a single parameter and some of them requires the use of toxic reagents and trained staff to performed the measurement (Rodriguez-Saona & Aykas, 2019).

Recently, non-invasive methods based on physical principles, couple with multivariate analysis have shown the potential to measure several parameters in a single measurement. Most of these methods are based vibrational/rotational (infrared), electronic (UV-visible) and nuclear magnetic resonance spectroscopies (Reuhs & Simsek, 2017). Among these methods near infrared spectroscopy (NIR) from 10,000 to 4000 cm^{-1} is one of the most popular. NIR absorption are assigned to the overtone or combination bands of the fundamental vibrational and rotational transitions (Nicolai et al., 2007). The NIR spectrum shows broad bands associated with combinations of vibration modes (O–H, N–H, and C–H) and overtones of molecular vibrations can be used to determine the parameters of concentrations of fruit quality (Wang, Sun, Pu, & Cheng, 2017).

Handheld NIR sensors (NIRS) have been used to provide non-destructive real-time and in the field information to optimize harvesting times for greater quality during on-tree ripening (Rodriguez-saona, Aykas, Borba, & Urbina, 2020). Rapid screening and characterization of major quality parameters of tomatoes have been increasing due to spectrometer portability and miniaturization of optical sensor technology. Novel low cost-effective, non-invasive, robust, and faster spectrometers have been developed, becoming the leading instrument for fruit and vegetables quality screening in the field,

(Crocombe, 2018; Erfan et al., 2016). One of those advancements was the development of micro-electro-mechanical systems (MEMS) to miniaturize optical, mechanical and hardware components of large and stationary spectrometers (Rodriguez-saona et al., 2020; Sorak et al., 2012). MEMS is the integration of mechanical elements, sensors, actuators, and electronics technologies that enables rotation of miniature gratings, measure spectra at millisecond time resolution and co-add several scans to one spectrum, generating good signal-to-noise ratio (Kenda et al., 2009).

The application of portable NIR systems to determine the quality attributes of fruits and vegetables has widely increased in the last few years (Pérez-marín et al. 2019; M. Sánchez et al. 2018; Marques et al. 2016; Beghi et al. 2013; Sánchez et al. 2011). Portable NIR spectrometers have been applied to measure firmness, SSC, and/or pH in strawberries (Sánchez et al. 2012), plums (Pérez-Marín, Paz, Guerrero, Garrido-Varo, & Sánchez, 2010), oranges (Liu, Gao, Hao, Sun, & Ouyang, 2012; Torres, Pérez-Marín, De la Haba, & Sánchez, 2017), tomatoes (Castrignanò et al., 2019), and bell peppers (Sánchez et al. 2019). Also, handheld spectrometers were successfully applied to evaluate moisture and fat content of fresh olives (Cirilli et al., 2016; Fernández-Espinosa, 2016) and to classify the maturity stage in avocados (Blakey, 2016).

The aim of this study was to develop robust and accurate prediction multivariate models using a handheld NIR spectrometer data to determine SSC, sugars, titratable acidity (TA) and organic acids of different types of fresh tomatoes (Roma, round, cherry, and grape) and evaluate its feasibility in the field application.

2. Material and methods

2.1. Samples

A total of 319 fresh market tomato samples including different types (Roma, round, grape and cherry) and ripening stages (green, breakers, turning, pink, light red, and red), according to United States Standards for Grades of Fresh Tomatoes (USDA, 2020), were analyzed in this study. Samples were harvested from Florida and California farms, during 2018 and 2019 growing seasons. In 2019 part of external validation (prediction step) was done in Florida directly in the field.

2.2. Spectra measurement

Near-infrared (NIR) spectrum for each intact tomato sample were collected at room temperature (20 ± 5 °C) using a portable NIR spectrometer (NeoSpectra – Module, Sci-ware Systems, Cairo, Egypt) (Fig. 1) that is based on a Micro Electro Mechanical System spectrometer. Spectra were collected in the $3800\text{--}7700$ cm^{-1} range, 16 cm^{-1} resolution, and 20 seconds scan. During the spectra acquisition, samples were placed in a rotating stage and scanned in the equatorial region. The field analyses were performed at the Lipman Family farm in Florida, USA. Those spectra were added to the prediction dataset.

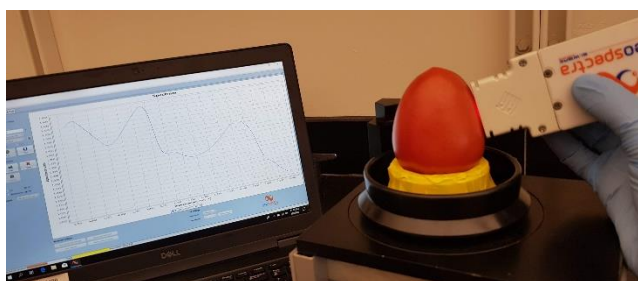


Figure 1: Experimental setup for acquiring the NIR spectrum of equatorial region of a rotating tomato using a handheld NIR spectrometer (NeoSpectra-Module)

2.3. Chemical analysis

Destructive measurements were performed immediately after the spectral acquisition to determine the quality attributes used as reference values. Roma and round tomatoes were blended by using a laboratory blender (6646 Oster 12 speed blender, Sunbeam Products, Inc., Boca Raton, FL), and grape and cherry types were used a handheld homogenizer (Benchmark Scientific Inc. NJ, USA), each tomato was individually blended. Tomato juice samples were extracted to estimate the total soluble solids content (SSC) and titratable acidity (TA). The SSC was determined using a digital refractometer ATAGO (ATAGO, Bellevue, WA) at 20 °C with $\pm 0.1\%$, 2 mL of juice were added into a centrifuge tube and centrifuged during 15 min at $13\,000$ rpm. A supernatant aliquot was placed to refractometer detector and the results were expressed as °BRIX. TA analysis was performed with an automatic titrator (Easy pH Titrator - Mettler Toledo, Columbus, OH) with 0.5 mol L^{-1} NaOH until pH 8.1 , using 2 grams

of juice and 40 mL of ultrapure water. The TA was determined based on the citric acid percentage, using the equation (1).

EQUATION

$$TA [g_{citricacid} / 100g_{ofsample}] = \left(\frac{A \times B \times C}{D} \right) \times 100 / E$$

where A is the volume of NaOH consumed in the titration (in L), B is the normality of NaOH (0.5 N), C is the molecular weight of citric acid (192.1 g mol⁻¹), D is the weight of the sample (2 g) and E is the valence of citric acid (E = 3).

Glucose and fructose concentrations were quantified using reverse-phase HPLC. Tomato juice aliquots (1.5 mL) were centrifuged at 13,000 rpm for 15 min at 25 °C; supernatant (1 mL) was filtered through 0.45 µm pore into the HPLC vials. Filtrated supernatant was injected (25µL) in a Shimadzu UFLC (Shimadzu, Columbia, MD) equipped with dual LC-6AD, SIL-20AHT auto-sampler, a CTO-20A column oven and RID-10A refractive index detector. A stainless steel Aminex®HPX-87C carbohydrate column with 7.8 mm ID300 mm dimensions and a micro-guard Carbo-C cartridge with 4.6 mm30 mm dimensions (Bio-Rad laboratories, Hercules, CA) was used to carry out an isocratic separation at 80 °C with HPLC grade water as the solvent at 0.6 mL/min flow rate during 30 min runs. The concentration of each sugar was determined using a calibration curve that was prepared with standards of sucrose, glucose and fructose (Fisher Scientific, Fair Lawn, NJ) in a range of concentrations between 1.6 and 25 mg/mL. Once glucose and fructose concentrations were obtained, the reducing sugars were calculated by adding these two values.

Organic acids (ascorbic and citric acid) were determined by reverse-phase high performance liquid chromatography (HPLC). Tomato juice (0.5 g) was extracted with 4.5% metaphosphoric acid and 100 mmol/L tris [2- carboxyethyl] phosphine (TCEP) was added to improve the stability of ascorbic acid (Wechtersbach & Cigić, 2007). The separation was achieved by using a C18 based Prevail organic acids column (5 µ, 150 × 4.6 mm) (Hichrom, Berkshire, UK) using pH 2.2 ultrapure water running at a flow rate of 0.8 mL/min. The calibration curve was developed using different concentrations of reagent grade ascorbic acid and citric (Fisher Scientific, Hampton,

NH) was used to determine acid concentrations. The elution of analytes was monitored at 210 nm (citric acid) and 245 nm (ascorbic acid) using a photodiode array detector.

2.4 Multivariate analysis

Partial least squares regression (PLSR) was used to develop tomato quality prediction models using the Pirouette software (Infometrix, Bothell, WA, USA). The NIR spectra were normalized and mean centralization. The second derivative and orthogonal signal correction were applied to correct baseline shifts, decrease noise regions and evidence specific absorption bands. The PLSR models were built by using the reference values as dependent variables and the NIR spectral data as independent variables. To calibrate, validate (cross-validation), and predict (external validation) the PLSR model's data was randomly separated in two sets, 80% of samples were used in the calibrate and validate (cross-validation) step and the remaining 20% to predict and validate the performance of models externally. Cross-validation by using leave-one-out algorithm was applied to do an internal validation in each calibration model. In the leave-one-out cross validation approach, each spectrum is dropped from the data set, and the remaining spectra are used for prediction of left-out data point until every spectrum is left-out once. The prediction set consisted in a blend between tomatoes analyzed at the laboratory and samples which spectra were collected in the field.

Models performance were expressed and evaluated by the statistical terms as the correlation coefficient of cross validation (r_v), the correlation coefficient of prediction (r_{pred}), the standard error of cross validation (SECV), and the standard error of prediction (SEP) that shows the magnitude of error expected when the concentration of the independent/unknown samples are predicted, and number of factors.

3. Results and discussion

3.1. Reference values

Reference compositional average, standard deviation, and ranges obtained for all quality parameters are summarized in Table 1, based on the tomato types. The values were similar to those reported previously in the literature (Akpola, Barineau, Jackson, Aykas, & Rodriguez-saona, 2020). The wide range in the compositional levels was due to the great diversity of tomato types, cultivar genotypes, maturity index, harvest from different seasons, and growing under different environmental and soil conditions.

The composition data showed that grape and cherry tomatoes had higher SSC, sugar, and organic acids levels, similar to reported by Akpolat et. al (2020). These tomatoes are small and have less water content, making them more concentrated, mainly, in relation to sugars (Bertin & Génard, 2018).

Table 1. Values of the quality parameters determined in tomato samples, from 2018 and 2019 growing seasons, using reference methods.

Parameters	Tomato type	N	Average	SD	Range	CV (%)
Soluble solids (° BRIX)	all	319	5.72	1.68	2.92 - 11.22	29.47
	roma	106	4.45	0.82	2.92 - 9.93	11.52
	round	47	4.59	0.53	3.49 - 5.95	11.54
	grape	93	6.81	1.24	4.29 - 9.36	18.25
	cherry	73	6.91	1.8	3.96 - 11.22	26.13
Glucose (g. L ⁻¹)	all	319	17.12	7.77	4.70 - 40.37	45.40
	roma	106	11.97	3.39	4.70 - 26.02	28.36
	round	47	13.97	3.59	7.70 - 27.08	27.25
	grape	93	21.19	7.35	8.08 - 36.93	34.68
	cherry	73	21.93	8.94	8.85 - 40.37	40.79
Fructose (g. L ⁻¹)	all	319	13.86	5.4	3.48 - 29.08	39.00
	roma	106	10.21	2.79	3.48 - 26.41	27.32
	round	47	10.24	2.19	6.69 - 18.98	21.47
	grape	93	16.54	4.99	7.23 - 27.96	30.17
	cherry	73	18.11	5.29	9.61 - 29.08	29.24
TA (g citric acid. 100g ⁻¹)	all	204	0.41	0.2	0.12 - 1.40	49.31
	roma	84	0.40	0.21	0.13 - 1.40	51.96
	round	45	0.38	0.2	0.15 - 1.29	53.59
	grape	43	0.49	0.22	0.15 - 1.09	45.68
	cherry	32	0.40	0.14	0.12 - 1.77	36.91
Ascorbic acid (mg. 100g ⁻¹)	all	149	29.32	14.71	3.77 – 77.91	44.84
	roma	63	27.58	11.42	12.09 - 56.91	41.39
	round	35	24.6	9.13	11.30 - 41.65	37.12
	grape	27	33.85	22.59	3.78 – 77.91	52.10
	cherry	24	38.83	12.32	14.70 - 62.29	31.73
Citric acid (g. 100g ⁻¹)	all	149	0.54	0.21	0.11 - 1.10	39.28
	roma	63	0.54	0.22	0.11 - 0.97	41.99
	round	35	0.42	0.18	0.18 - 0.99	43.52
	grape	27	0.56	0.19	0.15 - 1.10	34.49
	cherry	24	0.64	0.14	0.42 - 1.04	21.99

N: number of samples; SD: Standard Deviation; CV: Coefficient of variation; TA: total acidity.

The balance between sugars, acids and aroma volatile compounds are responsible for the final flavour (Bertin & Génard, 2018; Davies & Hobson, 1981). The perception of fruit acidity is due mainly to citric acid level, which is one of the mainly organic acids in tomato. Besides taste, organic acids can be nutritionally important compounds, such as ascorbic acid (or vitamin C). The levels of ascorbic acid found in this work were wide, ranging from 3.8 to 77.9 mg/100g slightly above than found in similar works (Akpolat et al., 2020; Siddiqui, Chakraborty, Mishra, & Hazra, 2014). The average ascorbic acid content in all 4 tomatoes types was 29.3 mg 100 g⁻¹; these results are similar to Akpolat et al.(2020).

3.2. PLS prediction models based on NIR measurements.

Figure 2 shows the NIR spectra of all 4 tomato types (Roma, round, grape and cherry) that consist of overtones and combination bands of fundamental vibrations (Feng, Zhang, Adhikari, & Guo, 2018; Nicolai et al., 2007; Pavia, Kris, Lampman, & Vyvyan, 2010). Two strong NIR bands assigned to O-H absorbance are shown between 5200 – 5100 (first overtone) and 6600 – 7150 cm⁻¹ (O-H stretching) (Rodriguez-saona, Ayvaz, & Wehling, 2017). The bands from 7400 – 7000 and from 4350 – 4033 cm⁻¹ can be assigned to combination bands of C-H, typically from carbohydrates (Nicolai et al., 2007). The weak band close to 5555 cm⁻¹ has been assigned to the third C-H overtone and related to sugar contents.

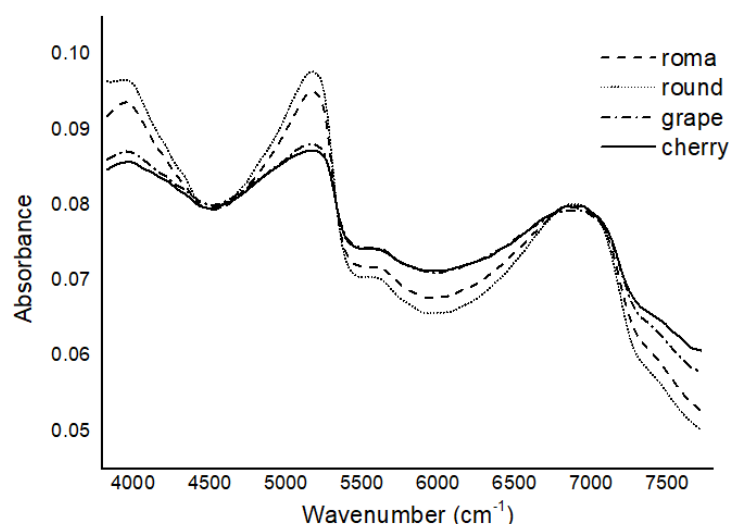


Figure 2. Normalized spectra of different types of tomato (Roma, round, grape, and cherry) collected using a portable NIR infrared (NeoSpectra-Module).

In **Figure 2** it is also possible to see the spectra pattern variation among the 4 different tomato types. Two separated groups were clustered, tomatoes with less sugar and more water content (Roma and round) and fruits with high sugar contents and less water (grape and cherry). Those groups have distinct spectrum pattern, mainly in the absorption bands related to O-H overtone (around 5000 cm^{-1}), C-H combinations and 3^{a} overtone ($4000 - 4400$ and around 5550 cm^{-1}) (Osborne, 2000). Those are characteristic regions for water and sugar compounds absorption (Nicolai et al., 2007; Pavia et al., 2010).

The absorption bands in the near-IR region are complex, with broad and overlapped bands. This overlap that can difficult the interpretation and requires multivariate analysis to understand and explain the set of spectra and identify its correlation with parameters of interest (Bureau, Cozzolino, & Clark, 2019; Rodriguez-saona et al., 2017). Using appropriate chemometric tools, preprocessing, and transformation these broad bands are especially useful for quantitative analysis. In **table 2** shows the statistical performance of the PLS models developed and each transformation used to develop these models.

Table 2. Statistical performance of the PLSR models developed.

Parameter	Preprocessing	N cal	F	SECV	r_{val}	N pred	SEP	r_{pre}
SSC ($^{\circ}$ BRIX)	Norm + 2° deriv (25) + OSC	250	4	0.52	0.91	69	0.49	0.87
Glucose (g. 100g^{-1})	Norm + 2° deriv + OSC	250	4	4.55	0.81	58	4.78	0.73
Fructose (g. 100g^{-1})	Norm + 2° deriv	250	6	2.11	0.86	58	1.97	0.91
TA (g citric acid. 100g^{-1})	2^{a} deriv (15)	150	8	0.07	0.57	50	0.09	0.41
Ascorbic acid (mg. 100g^{-1})	Norm + 2^{a} deriv (15)	100	6	3.78	0.82	46	4.09	0.81
Citric acid (g. 100g^{-1})	Norm + 2^{a} deriv (25)	100	6	0.06	0.87	46	0.07	0.86

SSC: soluble solid contents; TA: total acidity; N cal: number of samples in the calibration set; N pred: number of samples in the prediction set; F: factors; SECV: standard error of cross validation; r_{val} : correlation coefficient of cross validation; SEP: standard error of prediction; r_{pred} : correlation coefficient of prediction; Norm: normalization; Deriv: derivative; OSC: orthogonal signal correction.

SSC and fructose models showed the best performance, with higher correlation coefficients of prediction (0.87 and 0.91, respectively). The SEP for SSC was low (0.49°) for the wide range of concentrations (2.9° - 11.2° BRIX). Similar studies have reported standard error of prediction values of 0.57 (Castrignanò et al., 2019) and 0.20° (Feng et al., 2018) but for shorter SSC ranges compared to ours (4.20° – 8.30° and 5.50° – 7.40° BRIX, respectively). The SSC correlation coefficients found were higher than the results showed by Castrignanò et al (2019) evaluating the feasibility of a miniaturized Near-infrared spectrometer to determine the quality attributes of tomatoes ($r_{\text{pre}} = 0.75$). And similar to results found by Feng et al. (2018) by using a portable spectrometer to evaluate Cherry tomato quality.

The models for predicting total acidity that included all tomato types was not successful ($r_{\text{pred}} = 0.41$). In order to solve this problem, separate PLSR models were developed to each tomato type (**Table 3**). The spectra were preprocessed by normalization and second derivative and smooth using Savitsky-Golay with 35 datapoint window. Results showed that separate PLSR models improved the predictive ability compared to a single model for all 4 tomato types. Each calibration model had calibration correlation coefficient above 0.70, and low prediction errors (below $0.08\text{g citric acid } 100\text{g}^{-1}$).

Table 3: Statistical performance of Total acidity PLSR models developed separately to roma, round, grape, and cherry tomatoes.

Tomato type	N cal	F	SECV	r_{val}
roma	84	7	0.07	0.74
round	45	7	0.05	0.79
grape	43	5	0.08	0.80
cherry	32	7	0.08	0.70

N cal: number of samples in the calibration set; F: factors; SECV: standard error of cross validation; r_{val} : correlation coefficient of cross validation; r_{val} : correlation coefficient of cross validation.

The ascorbic and citric acids results (**Table 2**) demonstrate the applicability of portable spectrometer (Neospectra) to make accurate predictions, with low errors of $4.1\text{ mg } 100\text{g}^{-1}$ and $0.07\text{ g } 100\text{g}^{-1}$, respectively, in a few seconds without any sample

preparation. Our organic acids results were comparable to those obtained in other studies that used mid-IR spectroscopy. (Akpola et al., 2020; Wilkerson et al., 2013).

Correlation plots, besides statistical PLSR models results, are helpful to visualize the relationship between measured and NIR predicted quality parameters levels. **Figure 3** shows the correlation plots between reference values and predicted values obtained by using NIR spectra for all quality parameters in both steps (calibration and prediction) and regression vectors of PLSR models. The correlation plots show a high degree of linearity that indicates that the models are able to accurately predict the tomato quality attributes based on the NIR spectral data. The regression vectors information enables to identify the most important absorption bands related to each quality parameter, showing the bands more responsible for the prediction of quality parameter. The main responsible bands found were absorption bands around 5000 (O-H absorption region), 5555 (3^o C-H overtone), 6700 (O-H stretch), and 7300 cm⁻¹ (C-H combinations) are tightly related to absorption of molecules present in quality attributes in tomatoes (carbohydrates, organic acids and water) (Nicolai et al., 2007).

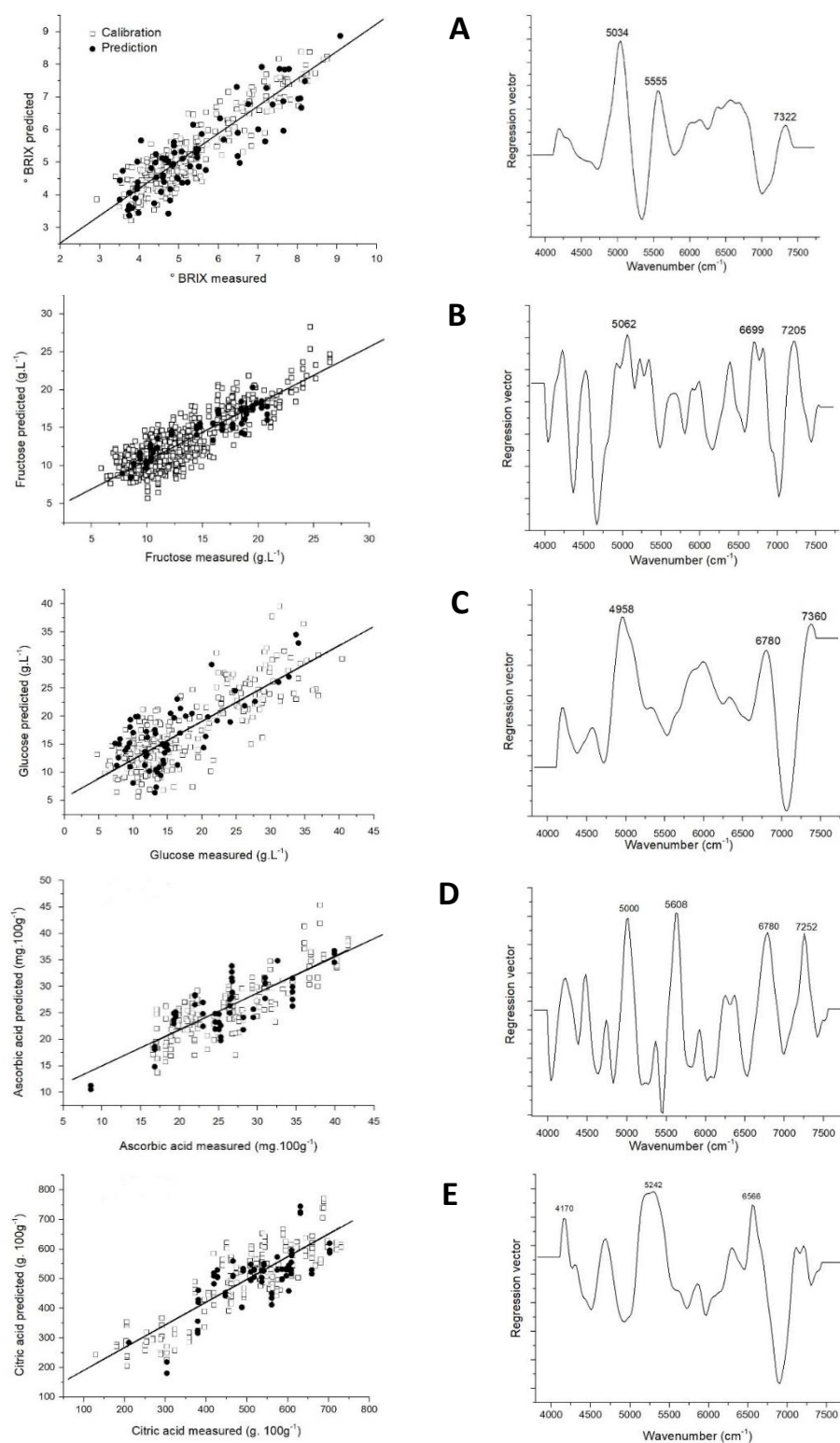


Figure 3. Scatter plot of PLSR models predicted and actual value and regression vector for SSC (A), fructose (B), glucose (C), ascorbic (D) and citric acid (E).

4. Conclusion

The strong correlation between spectral data and the reference values ($r_{\text{pred}} > 0.73$) as well as low standard errors found in this work, except for the total acidity PLSR models, evidence the potential of portable NIR spectrometer to determine tomato quality parameters. The non-invasive analyses performed in approximately 20 seconds shows that NIR portable spectrometer can be a useful tool to help farmers to make fast management decision (for example, about harvest time). Furthermore, it is possible to quantify all quality parameters, simultaneously, to different types of tomato using the same PLSR model. Also, it was shown to be able to do the measurement in the field without the need to transport the tomatoes to a laboratory.

Acknowledgments

This study was supported by CAPES institution providing the scholarship (process number 88881.187998/2018-01), CNPq - Research Productivity Fellowship processes 309043/2016-6 and 302866/2017-5 and Lipman Family Farms (Florida, US) by providing tomato samples.

References

- Akpolat, H., Barineau, M., Jackson, K. A., Aykas, D. P., & Rodriguez-saona, L. E. (2020). LWT - Food Science and Technology Portable infrared sensing technology for phenotyping chemical traits in fresh market tomatoes, *124*(February).
- Bai, Y., & Lindhout, P. (2007). Domestication and breeding of tomatoes: What have we gained and what can we gain in the future? *Annals of Botany*, *100*(5), 1085–1094. <https://doi.org/10.1093/aob/mcm150>
- Beghi, R., Spinardi, A., Bodria, L., & Mignani, I. (2013). Apples Nutraceutic Properties Evaluation Through a Visible and Near-Infrared Portable System, 2547–2554. <https://doi.org/10.1007/s11947-012-0824-7>
- Bertin, N., & Génard, M. (2018). Tomato quality as influenced by preharvest factors. *Scientia Horticulturae*, *233*(April 2017), 264–276. <https://doi.org/10.1016/j.scienta.2018.01.056>
- Blakey, R. J. (2016). Evaluation of avocado fruit maturity with a portable near-infrared

spectrometer. *Postharvest Biology and Technology*, 121, 101–105.
<https://doi.org/10.1016/j.postharvbio.2016.06.016>

Bureau, S., Cozzolino, D., & Clark, C. J. (2019). Contributions of Fourier-transform mid infrared (FT-MIR) spectroscopy to the study of fruit and vegetables: A review. *Postharvest Biology and Technology*, 148(October 2018), 1–14.
<https://doi.org/10.1016/j.postharvbio.2018.10.003>

Castrignanò, A., Buttafuoco, G., Malegori, C., Genorini, E., Iorio, R., Stipic, M., ... Venezia, A. (2019). Assessing the Feasibility of a Miniaturized Near-Infrared Spectrometer in Determining Quality Attributes of San Marzano Tomato. *Food Analytical Methods*, 12(7), 1497–1510. <https://doi.org/10.1007/s12161-019-01475-x>

Cirilli, M., Bellincontro, A., Urbani, S., Servili, M., Esposto, S., Mencarelli, F., & Muleo, R. (2016). On-field monitoring of fruit ripening evolution and quality parameters in olive mutants using a portable NIR-AOTF device. *Food Chemistry*, 199, 96–104. <https://doi.org/10.1016/j.foodchem.2015.11.129>

Costa, J. M., & Heuvelink, E. (2018). The global tomato industry. In E. Heuvelink (Ed.), *Tomatoes* (2nd ed., pp. 1–26). CABI.
<https://doi.org/https://doi.org/10.1079/9781780641935.0001>

Crocombe, R. A. (2018). Portable Spectroscopy. *Applied Spectroscopy*, 72(12), 1701–1751. <https://doi.org/10.1177/0003702818809719>

Davies, J., & Hobson, G. (1981). The constituents of tomato fruit - The influence of environment, nutrition and genotype. *Food Science and Nutrition*, 15(3), 205–208.

Erfan, M., Sabry, Y. M., Sakr, M., Mortada, B., Medhat, M., & Khalil, D. (2016). On-Chip Micro-Electro-Mechanical System Fourier Transform Infrared (MEMS FT-IR) Spectrometer-Based Gas Sensing. *Applied Spectroscopy*, 70(5), 897–904.
<https://doi.org/10.1177/0003702816638295>

FAO. (2020). Food and Agriculture Organization of the United Nations.

Feng, L., Zhang, M., Adhikari, B., & Guo, Z. (2018). Nondestructive Detection of Postharvest Quality of Cherry Tomatoes Using a Portable NIR Spectrometer and Chemometric Algorithms. *Food Analytical Methods*, 12(4), 914–925.

<https://doi.org/10.1007/s12161-018-01429-9>

Fernández-Espinosa, A. J. (2016). Combining PLS regression with portable NIR spectroscopy to on-line monitor quality parameters in intact olives for determining optimal harvesting time. *Talanta*, 148, 216–228. <https://doi.org/10.1016/j.talanta.2015.10.084>

Ibáñez, G., Valcárcel, M., Cebolla-Cornejo, J., & Roselló, S. (2019). FT-MIR determination of taste-related compounds in tomato: a high throughput phenotyping analysis for selection programs. *Journal of the Science of Food and Agriculture*, 99(11), 5140–5148. <https://doi.org/10.1002/jsfa.9760>

Kenda, A., Frank, A., Kraft, M., Tortschanoff, A., Sandner, T., Schenk, H., & Scherf, W. (2009). Compact High-Speed Spectrometers Based on MEMS Devices with Large Amplitude In-Plane Actuators. *Procedia Chemistry*, 1(1), 556–559. <https://doi.org/10.1016/j.proche.2009.07.139>

Liu, Y., Gao, R., Hao, Y., Sun, X., & Ouyang, A. (2012). Improvement of Near-Infrared Spectral Calibration Models for Brix Prediction in “Gannan” Navel Oranges by a Portable Near-Infrared Device. *Food and Bioprocess Technology*, 5(3), 1106–1112. <https://doi.org/10.1007/s11947-010-0449-7>

Marques, E. J. N., De Freitas, S. T., Pimentel, M. F., & Pasquini, C. (2016). Rapid and non-destructive determination of quality parameters in the “Tommy Atkins” mango using a novel handheld near infrared spectrometer. *Food Chemistry*, 197, 1207–1214. <https://doi.org/10.1016/j.foodchem.2015.11.080>

Nicolai, B. M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K. I., & Lammertyn, J. (2007). Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. *Postharvest Biology and Technology*, 46(2), 99–118. <https://doi.org/10.1016/j.postharvbio.2007.06.024>

Osborne, B. G. (2000). Near-Infrared Spectroscopy in Food Analysis. *Encyclopedia of Analytical Chemistry*, 1–14. <https://doi.org/10.1002/9780470027318.a1018>

Pavia, D. L., Kris, G. S., Lampman, G. M., & Vyvyan, J. R. (2010). *Introdução à espectroscopia*. (1ª). Cengage Learning.

- Pérez-Marín, D., Paz, P., Guerrero, J. E., Garrido-Varo, A., & Sánchez, M. T. (2010). Miniature handheld NIR sensor for the on-site non-destructive assessment of post-harvest quality and refrigerated storage behavior in plums. *Journal of Food Engineering*, 99(3), 294–302. <https://doi.org/10.1016/j.jfoodeng.2010.03.002>
- Pérez-marín, D., Torres, I., Entrenas, J., Vega, M., & Sánchez, M. (2019). Spectrochimica Acta Part A : Molecular and Biomolecular Spectroscopy Pre-harvest screening on-vine of spinach quality and safety using NIRS technology, 207(1258), 242–250.
- Reuhs, B. L., & Simsek, S. (2017). Nuclear Magnetic Resonance. In S. S. Nielsen (Ed.), *Food Analysis* (fifth). Cham: Springer International Publishing. <https://doi.org/10.1007/978-3-319-45776-5>
- Rodriguez-Saona, L. E., & Aykas, D. P. (2019). *Tomato Chemistry, Industrial Processing and Product Development*. (S. Porretta, Ed.), *Tomato Chemistry, Industrial Processing and Product Development*. Cambridge: Royal Society of Chemistry. <https://doi.org/10.1039/9781788016247>
- Rodriguez-saona, L. E., Aykas, D. P., Borba, K. R., & Urbina, A. U. (2020). Miniaturization of Optical Sensors and their Potential for High-Throughput Screening of Foods. *Current Opinion in Food Science*. <https://doi.org/10.1016/j.matdes.2020.108709>
- Rodriguez-saona, L. E., Ayvaz, H., & Wehling, R. L. (2017). Infrared and Raman Spectroscopy. In S. S. Nielsen (Ed.), *Food Analysis* (pp. 107–127). Springer International Publishing. <https://doi.org/10.1007/978-3-319-45776-5>
- Sánchez, M. T., De La Haba, M. J., Benítez-López, M., Fernández-Novales, J., Garrido-Varo, A., & Pérez-Marín, D. (2012). Non-destructive characterization and quality control of intact strawberries based on NIR spectral data. *Journal of Food Engineering*, 110(1), 102–108. <https://doi.org/10.1016/j.jfoodeng.2011.12.003>
- Sánchez, M. T., De la Haba, M. J., Guerrero, J. E., Garrido-Varo, A., & Pérez-Marín, D. (2011). Testing of a local approach for the prediction of quality parameters in intact nectarines using a portable NIRS instrument. *Postharvest Biology and Technology*, 60(2), 130–135. <https://doi.org/10.1016/j.postharvbio.2010.12.006>

Sánchez, M. T., Torres, I., de la Haba, M. J., Chamorro, A., Garrido-Varo, A., & Pérez-Marín, D. (2019). Rapid, simultaneous, and in situ authentication and quality assessment of intact bell peppers using near-infrared spectroscopy technology. *Journal of the Science of Food and Agriculture*, 99(4), 1613–1622. <https://doi.org/10.1002/jsfa.9342>

Sánchez, M., Torres, I., Haba, M. De, Chamorro, A., Garrido-varo, A., & Pérez-marín, D. (2018). Rapid , simultaneous , and in situ authentication and quality assessment of intact bell peppers using near-infrared spectroscopy technology, (July). <https://doi.org/10.1002/jsfa.9342>

Siddiqui, M. W., Chakraborty, I., Mishra, P., & Hazra, P. (2014). Bioactive attributes of tomatoes possessing dg, ogc, and rin genes. *Food and Function*, 5(5), 936–943. <https://doi.org/10.1039/c3fo60520e>

Sorak, D., Herberholz, L., Iwascek, S., Altinpinar, S., Pfeifer, F., & Siesler, H. W. (2012). New developments and applications of handheld raman, mid-infrared, and near-infrared spectrometers. *Applied Spectroscopy Reviews*, 47(2), 83–115. <https://doi.org/10.1080/05704928.2011.625748>

Torres, I., Pérez-Marín, D., De la Haba, M. J., & Sánchez, M. T. (2017). Developing universal models for the prediction of physical quality in citrus fruits analysed on-tree using portable NIRS sensors. *Biosystems Engineering*, 153, 140–148. <https://doi.org/10.1016/j.biosystemseng.2016.11.007>

USDA. (2020). Tomato grading visual aids. Retrieved December 10, 2015, from https://www.ams.usda.gov/sites/default/files/media/Tomato_Visual_Aids%5B1%5D.pdf

Wang, L., Sun, D.-W., Pu, H., & Cheng, J.-H. (2017). Quality analysis, classification, and authentication of liquid foods by near-infrared spectroscopy: A review of recent research developments. *Critical Reviews in Food Science and Nutrition*, 57(7), 1524–1538. <https://doi.org/10.1080/10408398.2015.1115954>

Wilkerson, E. D., Anthon, G. E., Barrett, D. M., Sayajon, G. F. G., Santos, A. M., & Rodriguez-Saona, L. E. (2013). Rapid assessment of quality parameters in processing tomatoes using hand-held and benchtop infrared spectrometers and multivariate

analysis. *Journal of Agricultural and Food Chemistry*.
<https://doi.org/10.1021/jf304968f>

Capítulo 2.

**Espectroscopia NIR e MIR na análise e classificação da qualidade
de tomates destinados ao processamento**

Espectroscopia NIR e MIR na análise e classificação da qualidade de tomates destinados ao processamento industrial

Karla R. Borba^a, Nayara Bomfin^b, Luiz A. Colnago^c e Marcos D. Ferreira^c.

^aDepartamento de Alimentos e Nutrição, Faculdade de Ciências farmacêuticas, Universidade Estadual Paulista (UNESP), 4801-902, Araraquara, SP, Brasil.

^bUniversidade Federal de São Carlos (UFSCar), São Carlos, SP, Brasil.

^cEMBRAPA Instrumentação, Caixa Postal 741, 13561-206, São Carlos, SP, Brasil.

RESUMO

Aproximadamente 30% do tomate produzido no Brasil é destinado ao processamento industrial. Para garantir que o produto final satisfaça as exigências e requisitos do mercado é necessário fazer o controle de atributos que são importantes para indústria de processamento como: teor sólidos solúveis (SS), pH e firmeza. Em geral, os métodos analíticos de qualidade convencionais são destrutivos, apresentam custos elevados com amostras e reagentes, são demorados e necessitam de mão de obra especializada. Para minimizar ou eliminar esses problemas tem sido proposto o uso de técnicas não invasivas como a espectroscopia de infravermelho próximo (NIR) e médio (MIR). Essas técnicas já têm sido estudadas para avaliar qualidade interna e externa de frutas e hortaliças. Desta forma, este trabalho teve como objetivo avaliar a utilização de NIR e MIR associadas a ferramentas quimiométricas na classificação e análise dos padrões de qualidade de tomates destinados para processamento. Para isso foram desenvolvidos modelos PLS de regressão (Regressão por mínimos quadrados parciais, do inglês *Partial Least Squares*) para a determinação de SS, pH e firmeza. Também foram desenvolvidos modelos PLS-DA (Mínimos Quadrados Parciais para Análise Discriminante, do inglês *Partial Least Squares for Discriminant Analysis*) de classificação de tomates de acordo com seu índice de maturação e firmeza. A técnica de MIR resultou em modelos PLS com maiores coeficientes de correlação de Pearson para SS e firmeza ($r = 0,79$ e $0,83$, respectivamente), os modelos PLS para predição de pH apresentaram baixos erros médios padrão (SEP igual ou abaixo de $0,07$) em todas as etapas (calibração e validação externa). A técnica de NIR resultou em predições menos precisas para SS e firmeza ($r = 0,71$ e $0,70$, respectivamente), e performance similar aos resultados de MIR para determinação de pH ($r = 0,68$ e $SEP = 0,07$). Em relação a classificação dos tomates de acordo com o índice de

maturação e firmeza, os modelos PLS-DA desenvolvidos com os espectros de MIR resultaram em acurácias maiores para a classificação. Com valores de 98,30% e 83,87% para índice de maturação e firmeza, respectivamente. A classificação por meio de espectros de NIR obteve acurácias acima de 75% (IM = 76,19% e firmeza = 83,88%). Os resultados mostraram que as técnicas espectroscópicas apresentam aplicabilidade como métodos rápidos e precisos de avaliação de qualidade de tomates destinados para o processamento.

Palavras-chave: Espectroscopia; Infravermelho; Quimiometria; PLS; PLS-DA; Análise de qualidade; indústria.

INTRODUÇÃO

O Brasil está entre os 10 maiores produtores mundiais de tomate (1), sendo que o estado de São Paulo é o segundo maior produtor no país, com aproximadamente 22,29% do total (1). No Brasil, 30% da produção de tomates é utilizada como matéria-prima para processamento industrial (2). Cultivares distintos são direcionados para as duas finalidades, o tomate rasteiro, ou tomate industrial, é destinado para o processamento, sendo matéria-prima dos mais diversos produtos atomatados (extratos, molhos, sucos e outros derivados) (3). Por sua vez, o tomate estaqueado é cultivado e abastece o mercado para consumo *in natura*.

A qualidade de um produto vegetal é importante tanto para tomates que serão consumidos *in natura* quanto para aqueles que serão processados, pois irá interferir diretamente no resultado final do produto assim como nos custos de produção. Seu controle garante que o produto final satisfaça as exigências e requisitos do mercado (4). Na indústria de processamento de tomate alguns atributos de qualidade, como a coloração, firmeza, consistência, acidez e o teor de sólidos solúveis são os mais importantes (5). A adição de uma etapa de classificação e exclusão de frutos impróprios, anteriormente ao processamento, pode resultar em um produto final de melhor qualidade sensorial, pois significaria melhoria na matéria prima utilizada. Isso porque, durante a pasteurização da polpa de tomate o produto é estabilizado no aspecto microbiológico, porém padrões de qualidade alterados pelos micro-organismos não são recuperados (6). Com técnicas rápidas e não destrutivas, como a utilização de espectroscopia, é possível uma etapa de classificação e controle de qualidade otimizados e o produto final apresentar maior valor agregado e qualidade.

Os métodos convencionais utilizados para a análise da qualidade em tomates, e demais produtos vegetais, são geralmente destrutivos. Esse tipo de metodologia destrutiva requer despesas mais elevadas com amostras e reagentes, além de demandar muito tempo e mão de

obra especializada para análise (7). Métodos não destrutivos, onde a manipulação da amostra é a mínima possível, são uma alternativa para analisar e garantir o controle de qualidade de frutas e hortaliças ao longo de um período, tornando possível o monitoramento durante transporte, armazenamento e comercialização do vegetal (7,8).

Técnicas não destrutivas, como a espectroscopia de infravermelho próximo (NIR) e médio (MIR), têm sido propostas para avaliar características representativas de qualidade em frutas e hortaliças, pois permitem rápida execução das medições, com potencial para estimar vários atributos de qualidade simultaneamente além da possibilidade de repetir as medições nas mesmas amostras (9–12). A espectroscopia é o estudo da interação da radiação eletromagnética com a matéria, essa radiação pode ser dividida em regiões de energia (13). A radiação de infravermelho corresponde à parte do espectro eletromagnético situada entre a região do visível e das micro-ondas, subdividido em infravermelho próximo ($4000 - 12500 \text{ cm}^{-1}$), infravermelho médio ($400 - 4000 \text{ cm}^{-1}$) e distante ($100 - 400 \text{ cm}^{-1}$) (13). O comprimento de onda de uma absorção depende das massas relativas dos átomos, das constantes de força das ligações e da geometria dos átomos (14).

Este trabalho teve como objetivo avaliar a utilização de técnicas de espectroscopia de infravermelho próximo e médio associadas a ferramentas quimiométricas na classificação e análise de padrões de qualidade em tomates destinados ao processamento industrial.

MATERIAL E MÉTODOS

O modelo vegetal utilizado foi o tomate (*Lycopersicon esculentum*) rasteiro destinado à indústria, originado de lotes direcionados a uma empresa de processamento de polpa de tomate no interior de São Paulo, coletados em três períodos distintos de colheita – agosto, setembro e outubro de 2017. Após amostragem, os frutos foram transportados até o laboratório de pós-colheita, Embrapa Instrumentação, São Carlos, onde foram selecionados, lavados e sanitizados. Tomates nos estádios verde, pintado e vermelho-intenso foram utilizados no desenvolvimento dos modelos quimiométricos. Essa diversidade de estádios, assim como a coleta dos frutos em períodos distintos, foi utilizada para aumentar a variabilidade entre as amostras, para que os modelos de predição e classificação desenvolvidos apresentassem maior robustez. e com isso desenvolver modelos representativos e aplicáveis para uma maior gama de variação. Análises físico-químicas e espectroscópicas foram realizadas sob as mesmas condições de temperatura e umidade relativa ($21^\circ \text{C} \pm 1$ e 60% UR).

Análises de referência

Massa fresca

Os frutos foram pesados individualmente em balança eletrônica Marte AS 2000C, com precisão de 0,01 g, os valores foram expressos em gramas.

Teor de Sólidos Solúveis

O teor de sólidos solúveis foi quantificado inserindo-se alíquota de 2-3 mL da polpa de tomate em refratômetro de bancada Atago RX-5000cx. Os resultados foram expressos em °Brix (AOAC, 1997 - método 932.12).

pH

A medida de pH foi realizada com a imersão do eletrodo do equipamento QX 1500 QUALSTRON em 30 mL da polpa de tomate extraída.

Coloração

A coloração foi realizada em colorímetro, Konica Minolta Sensing (Tokio, Japão), no sistema L, a* e b* (*color space*) proposto pela *Comission Internationale de l'Éclairage* (CIE). A casca dos frutos foi avaliada quanto à cor, com três repetições na região equatorial de cada fruto e posteriormente foi realizada a média das três repetições.

Firmeza

A firmeza foi avaliada utilizando-se o texturômetro digital de bancada TA.XT. Plus Texture Analyser® e os resultados foram expressos em Newton (N). A medição foi feita em triplicata em regiões equatoriais equidistantes.

Infravermelho próximo (NIR)

Os espectros de NIR foram adquiridos por absorbância em espectrômetro de infravermelho comercial, Spectrum 100N (Perkin-Elmer Corp, Norwalk, CT). Os sinais das amostras foram obtidos pela análise na região equatorial de cada tomate fresco em triplicata e, posteriormente, foi realizada a média dos três espectros. Os espectros foram coletados na região entre 4000-10000 cm^{-1} com uma resolução de 32 cm^{-1} e 32 scans cada para melhorar a relação sinal-ruído.

Infravermelho médio (MIR)

Os espectros de absorbância da região do infravermelho médio (MIR) foram coletados pelo espectrômetro Agilent Cary 630 FTIR na região entre 650-4000 cm^{-1} com uma resolução de 16 cm^{-1} e 64 scans cada para melhorar a relação sinal-ruído. As análises foram realizadas pela análise na região equatorial de cada fruto fresco em triplicata e, posteriormente, foi realizada a média dos três espectros.

Processamento dos dados

A análise dos espectros de NIR e MIR, o processamento e o desenvolvimento dos modelos de predição e classificação foram realizados pelos softwares Origin 9 OriginLab, Northampton, MA, USA) e Pirouette v.4.5 (Infometrix, Inc. Bothell – WA). Antes de desenvolver os modelos de predição e classificação, foi feita uma análise de componentes principais (PCA) nos conjuntos de espectros de NIR e MIR com a finalidade de análise exploratória dos dados. Depois disso, os espectros foram correlacionados com os resultados das análises físico-químicas de referência para a construção dos modelos de predição e classificação. PLS (Regressão por mínimos quadrados parciais, do inglês *Partial least squares*) e PLS-DA (Mínimos Quadrados Parciais para Análise Discriminante, do inglês *Partial Least Squares for Discriminant Analysis*) foram as ferramentas quimiométricas utilizadas. Durante o pré processamento, os espectros de NIR e MIR foram centrados na média, normalizados (0-1) e derivatizados (Savitzky – Golay) aplicando polinômio de segunda ordem (com 25 pontos).

As ferramentas quimiométricas PLS (mínimos quadrados parciais) e PLS-DA são técnicas supervisionadas utilizadas para predição e classificação, respectivamente (Ferreira, 2015). Tratando-se de modelos supervisionados, as variáveis dos espectros são correlacionadas com valores obtidos previamente por técnicas de referência e para a classificação são atribuídas classes distintas. Para isso, foram construídas matrizes onde os sinais analíticos (espectros) foram utilizados na matriz X, contendo as variáveis independentes. A matriz Y foi construída com os dados obtidos pelas análises químicas de referência, que são as variáveis dependentes no caso do PLS e as classes para PLS-DA. Durante a construção de modelos quimiométricos é importante que aconteça uma etapa de validação externa para a calibração desenvolvida, que precisa ser feita com amostras não conhecidas. Portanto, o conjunto de dados totais foi previamente dividido, de forma aleatória, entre dois grupos. Onde 70% dos dados formaram o grupo de calibração e validação interna (com teste de validação interna cruzada *leave-one-out*) e os outros 30% formaram o conjunto de validação externa (ou predição).

Alguns resultados são importantes para a avaliação da performance de cada modelo desenvolvido, possibilitando a avaliação da precisão e aplicabilidade, como por exemplo: coeficientes de correlação de Pearson (r) da calibração (r_{Cal}), validação (r_{Val}) e predição (r); erros quadráticos médios da calibração (RMSEC), validação (RMSEV) e predição (RMSEP); número de fatores utilizados; e acurácia no caso de modelos de classificação. Os modelos PLS com maior precisão são aqueles com coeficientes de correlação mais próximos de 1 e com

baixos erros quadráticos médios, assim como quanto mais alta a capacidade de acerto na classificação (acurácia em %) melhor o desempenho do modelo PLS-DA de classificação.

RESULTADOS E DISCUSSÃO

Caracterização dos tomates

Os resultados das análises físico-químicas de referência, Tabela 1, mostram a amplitude de variação de cada parâmetro. Para o desenvolvimento de modelos de predição e classificação é essencial que haja grande amplitude nos resultados, para que os modelos apresentem aplicabilidade, pois a regressão precisa ser calibrada com amostras que representem a variação real existente.

Tabela 1: Caracterização do conjunto amostral, valores mínimos, máximo, média e desvio padrão dos frutos de tomate utilizados no desenvolvimento dos modelos PLS e PLS-DA.

	Mínimo	Máximo	Média	Desvio padrão
Massa (g)	35,45	98,21	61,89	11,54
Diâmetro (mm)	38,80	73,77	47,86	4,38
Altura (mm)	42,59	73,45	57,21	4,56
pH	3,98	5,48	4,40	0,17
SST (°Brix)	2,91	6,12	4,71	0,45
Firmeza (N)	6,02	34,23	15,58	6,62
Coloração				
<i>L</i>*	21,57	58,97	44,93	6,56
<i>a</i>*	-14,16	29,17	9,78	0,99

SS: Sólidos Solúveis; L: luminosidade; *a*: parâmetro de cor do eixo CIE L*a*b* com variação de verde a vermelho.

A cor de uma fruta e/ou hortaliça é um parâmetro importante na agricultura, pois é um indicativo de qualidade e, por vezes, é utilizado como atributo de valoração do produto (16,17). Durante o processamento de tomate os lotes utilizados podem conter frutos em distintos estádios de maturação, como é possível observar nos valores máximos e mínimos dos parâmetros de qualidade avaliados. Esse fato pode resultar em um produto final com qualidade inferior àquela requerida pelos consumidores, principalmente em relação à coloração do extrato final. Produtos atomatados são mais aceitos pelos consumidores quando apresentam coloração vermelho intenso (18), com o aumento do uso de tomates no estágio verde para o processamento o resultado final será um produto com coloração marrom. Das análises de

coloração, que foram obtidas pelo sistema L , a e b , foram utilizados apenas os valores de L e a . Sendo o valor de a o mais importante para avaliar a qualidade e estágio de maturação de tomates, uma vez que esse mostra a variação de cor desde verde até vermelho. Nota-se que o conjunto de tomates utilizados apresentou grande variação em relação ao valor a .

Valores de pH e teor de sólidos solúveis são importantes durante o processamento de tomates, pois eles irão influenciar diretamente no rendimento e custo de processo uma vez que valores inadequados irão levar a necessidade de “manejo” durante as etapas da cadeia. O pH baixo, inferior a 4,5, irá impedir a proliferação de micro-organismos no produto, pHs acima deste valor irão requerer mais tempo de esterilização aumentando os gastos com energia durante o processamento (19). Quanto maior o teor de SS maior será o rendimento e menor o gasto de energia no processo de concentração da polpa (20). A cada 1° Brix a mais na matéria-prima há um incremento de 20% no rendimento final (2).

Transformações dos espectros

Os espectros de infravermelho foram submetidos a pré processamentos com a finalidade de corrigir problemas como deformação na linha de base, espalhamento de luz, presença de ruído e também com o objetivo de evidenciar bandas e obter o máximo de informação relevante para os parâmetros analisados. Dessa forma foram utilizados os métodos de centralização na média, normalização (0-1) e aplicação de primeira derivada, com alisamento de segunda ordem usando o método de Savitzky-Golay, com janela de 25. Na figura 1 estão apresentados os espectros de reflectância de infravermelho próximo (NIR) brutos (A) e pré processados (B). Nota-se que o tratamento dos espectros corrigiu o problema de deformação da linha de base, diminuiu a largura das linhas e consequentemente evidenciou bandas.

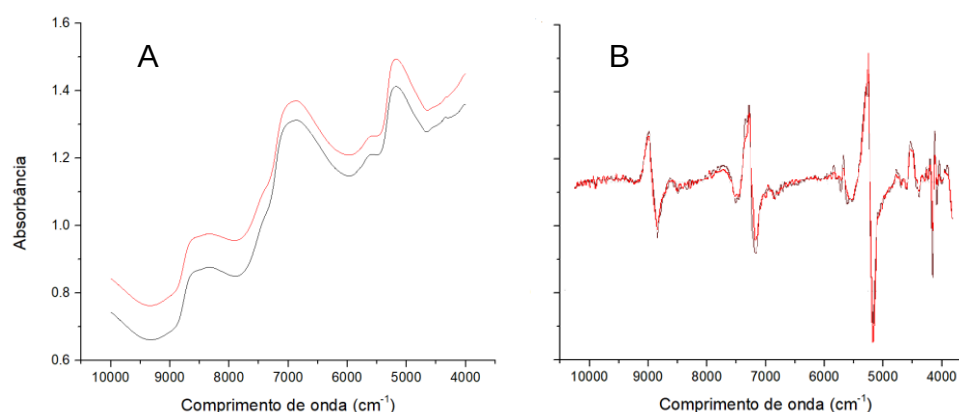


Figura 1: Espectros, de reflectância no infravermelho próximo - NIR, brutos (A) e após centralização na média, normalização (0-1) e primeira derivada, com alisamento de segunda

ordem usando o método de Savitzky-Golay, com janela de 25 pontos (B); (-----) 5,06 °Brix e (- - - - -) tomate com 4,12 °Brix.

A espectroscopia de NIR é caracterizada por espectros complexos com regiões de sobretons e combinações (7). A região entre 6500 e 7500 cm^{-1} é caracterizada pelo estiramento assimétrico da ligação O-H, no segundo *overtone* (sobretom) da água (7,21), entre 5100 e 5200 é a banda característica do primeiro *overtone* de O-H (22). Banda de C-H, tipicamente presentes em carboidratos, estão localizadas na faixa de 7400 – 7000 e 4350 – 4033 cm^{-1} (7).

Os espectros de reflectância no infravermelho médio (MIR) estão representados na Figura 2, onde estão apresentados espectros brutos (A) e aqueles que passaram por centralização na média, normalização (0-1) e primeira derivada, com alisamento de segunda ordem usando o método de Savitzky-Golay, com janela de 25 pontos (B). Assim como na figura anterior, a figura 2 apresenta espectros de tomates nas diferentes classes.

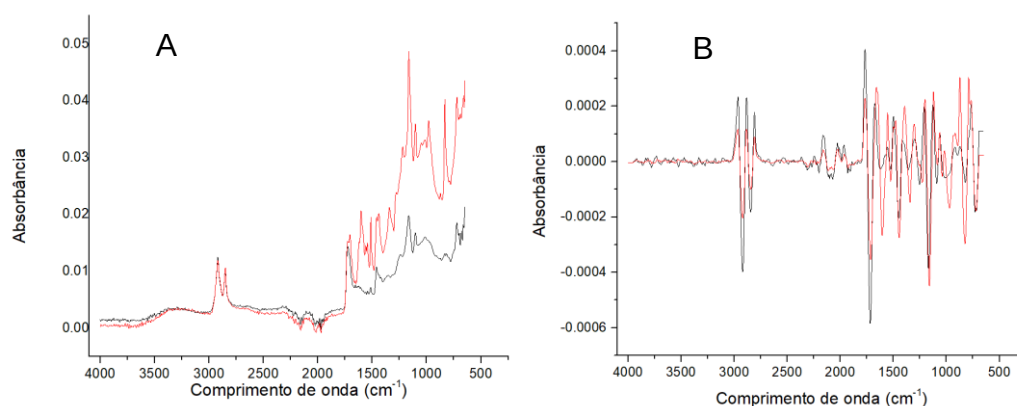


Figura 2: Espectros, de reflectância no infravermelho médio - MIR, brutos (A) e após centralização na média, normalização (0-1) e primeira derivada, com alisamento de segunda ordem usando o método de Savitzky-Golay, com janela de 25 pontos (B); (-----) 5,06 °Brix e (- - - - -) tomate com 4,12 °Brix.

Nos espectros de MIR foi realizada a seleção de variáveis, onde os intervalos de 3656 a 4000 cm^{-1} e 2777 a 1773 cm^{-1} foram excluídas. Nessas regiões não existem sinais importantes, somente ruídos que podem influenciar a correlação entre os sinais espectroscópicos com as medidas de referência (21,23). A região do espectro de MIR entre 3200 a 3600 cm^{-1} é a banda característica da absorção de grupos O-H (hidroxila), que estão presentes, em moléculas de ácidos orgânicos e água (21).

O intervalo entre 1790 e 1706 cm^{-1} é associado ao estiramento dos grupos C=O e C=C (23) e as bandas entre 1500 e 900 cm^{-1} são dominadas por grupos de estiramento C-C e C-O

(24). Mais especificamente, as regiões no intervalo de 1650 e 1850 cm^{-1} são associadas ao grupo carbonila e carboxila (13), encontrados nos aldeídos e cetonas, constituintes de moléculas de monossacarídeos (25). As bandas de absorção localizadas entre 3000 e 2750 cm^{-1} são características de moléculas com ligações C-H, esse tipo de ligação é característico de lipídeos (10).

Modelos de Predição de Parâmetros de qualidade

A tabela 2 apresenta os resultados das performances dos modelos de predição de SS, pH e firmeza desenvolvidos por meio de PLS. A escolha do número de fatores (ou variáveis latentes) a serem incluídos no modelo de regressão é uma etapa importante. Pois um modelo com um número muito grande de fatores pode estar modelando ruído e assim ser super ajustado, enquanto um modelo com poucos fatores pode estar deixando de lado informações importantes. Os números de fatores foram escolhidos por avaliação do SEV (erro médio padrão da validação), onde o menor erro padrão da validação foi o mais adequado (26).

Tabela 2: Resultados obtidos com os modelos PLS NIR e MIR para predição dos parâmetros de qualidade de tomates destinados à indústria na calibração, validação interna e externa.

Parâmetro	Método	Calibração e validação interna				Validação Externa		
		Fator	SEC	r_{cal}	SEV	r_{val}	SEP	r
SS (° BRIX)	NIR	2	0,09	0,96	0,09	0,95	0,29	0,71
	MIR	6	0,20	0,80	0,22	0,74	0,19	0,79
pH	NIR	3	0,07	0,71	0,08	0,66	0,07	0,68
	MIR	5	0,05	0,81	0,05	0,73	0,06	0,66
Firmeza	NIR	5	2,35	0,89	3,36	0,76	3,79	0,70
	MIR	2	2,63	0,90	2,68	0,90	3,38	0,83

NIR: Infravermelho próximo; MIR: Infravermelho médio; SEC: erro médio padrão da calibração; SEV: erro médio padrão da validação interna; SEP: erro médio padrão da predição; r_{cal} : Coeficiente de correlação de Pearson da calibração; r_{val} : Coeficiente de correlação de Pearson da validação interna; r: Coeficiente de correlação de Pearson da validação externa; SS: Sólidos solúveis.

Durante a modelagem PLS foi realizado o procedimento de validação cruzada (*cross-validation leave-one-out*) que consiste em uma validação interna para estimar a capacidade preditiva do modelo de calibração. Para avaliar a precisão aplicabilidade de um modelo PLS é necessário avaliar alguns índices como: erros médios da calibração, validação interna e externa; coeficientes de correlação de Pearson da calibração, validação interna e externa, número de fatores utilizados, dentre outros (27). Onde modelos com maior aplicabilidade são aqueles com menores erros e alta correlação. Em relação ao coeficiente de correlação de Pearson, tem-se que valores entre 0,50 e 0,70 são considerados moderados e modelos com

valores de coeficiente de correlação acima de 0,70 são considerados modelos com alta correlação entre os sinais e valores de referência (26).

Todos os modelos desenvolvidos, com exceção daqueles para determinar o pH das amostras, resultaram em boas correlações e baixos erros. Os modelos para a determinação de SS apresentaram ótima performance durante a etapa de calibração e validação cruzada, com altos coeficientes de Person principalmente o desenvolvido utilizando NIR ($r_{\text{val}} = 0,95$). Após a validação externa PLS-NIR apresentou um decréscimo na performance, porém ainda com correlação moderada ($r = 0,71$). Já o modelo para determinação de SS com a utilização do MIR se mostrou constante durante ambas as etapas (calibração/validação interna e validação externa), com valores similares de coeficiente de correlação ($r_{\text{val}} = 0,80$ e $r = 0,79$) e erros ($\text{SEV} = 0,22$ °Brix e $\text{SEP} = 0,19$ °Brix). Os resultados para predição de SS com espectros de MIR do presente estudo apresentaram melhor performance quando comparados com um estudo similar de Akpolat e colaboradores (2020), que utilizaram o mesmo método e desenvolveram modelos PLS de predição de SS em tomates com coeficientes de correlação entre 0,55 – 0,72.

O valor de pH no conjunto amostral variou entre 3.98 – 5.48 (tabela 1). Mesmo com modelos apresentando coeficientes de correlação abaixo de 0,70, os erros médios padrão em todas as etapas foram baixos (iguais ou abaixo de 0.07). Os erros encontrados nos modelos PLS para predição de pH estão abaixo do que em trabalhos similares com correlação acima de 0,80, por exemplo no estudo desenvolvido por He e colaboradores utilizando infravermelho VIS/NIR onde o menor erro quadrático médio de predição foi de 0,09.

Assim como o teor de sólidos solúveis, a firmeza (ou firmeza) de um fruto está relacionada ao grau de amadurecimento do mesmo (29,30). A técnica de MIR resultou novamente no modelo com maior coeficiente de correlação na determinação da firmeza do tomate, com $r = 0,83$, do que aquele desenvolvido com infravermelho próximo ($r = 0,70$). Porém, mesmo com moderada correlação PLS-NIR teve baixo valor de erro ($\text{SEP} = 3,79$ N). Esses resultados são similares com aqueles encontrados por Sirisomboon e colaboradores (2012), que estudando a aplicabilidade de espectroscopia de NIR para a predição de firmeza em tomates ‘Momotaro’, obtiveram correlação de 0,66 no atributo de firmeza deformação na ruptura.

Adicionalmente, os gráficos da relação entre os valores medidos pelos métodos de referência e os valores preditos através dos espectros de NIR e MIR estão representados nas figuras 3 e 4, respectivamente.

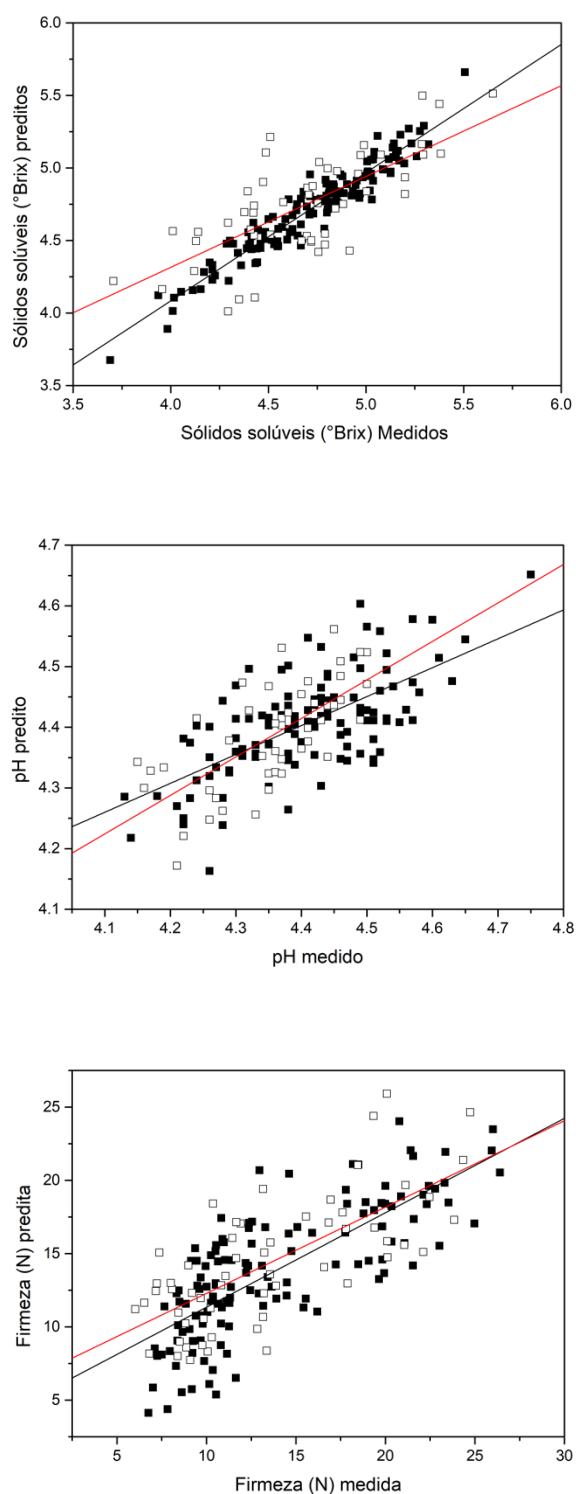


Figura 3: Gráficos dos valores medidos pelas análises de referência e preditos pelos espectros de NIR para SS, pH e firmeza em Newton. Quadrados cheios representam amostras do conjunto de calibração e vazios de validação externa. Reta de regressão preta: conjunto calibração e reta de regressão vermelha: conjunto validação externa.

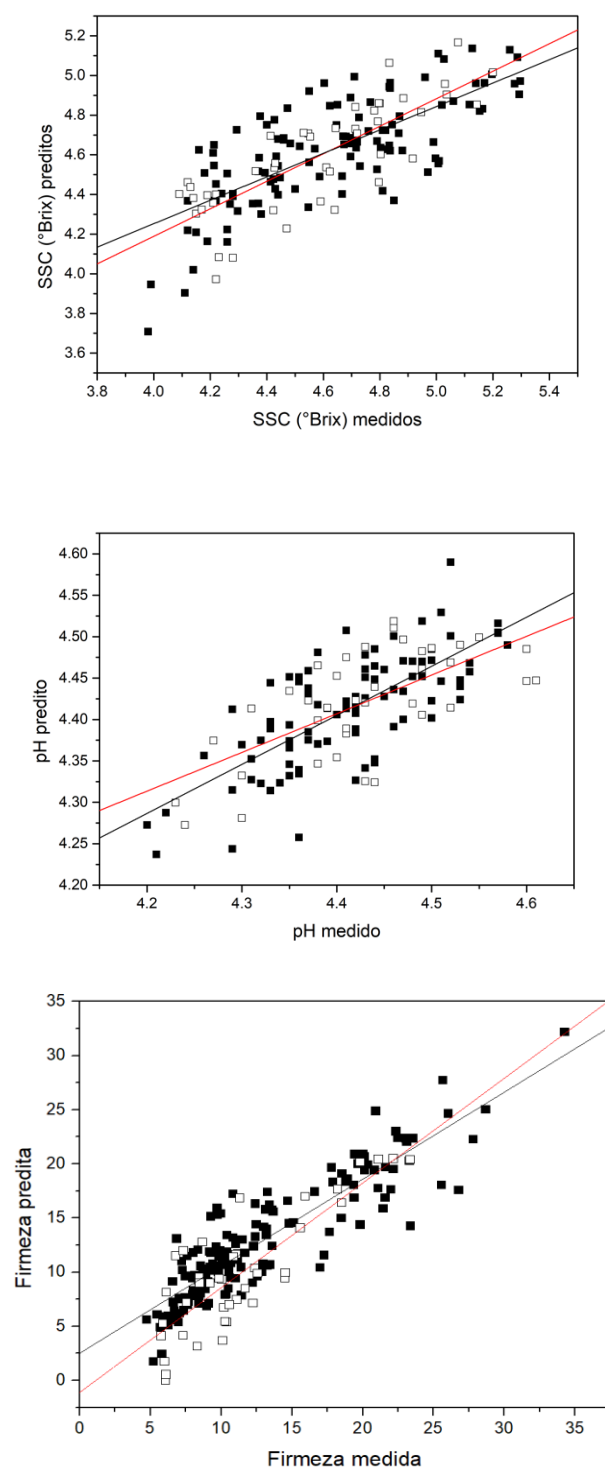


Figura 4: Gráficos dos valores medidos pelas análises de referência e preditos pelos espectros de MIR para SS, pH e firmeza. Quadrados cheios representam amostras do conjunto de calibração e vazios de validação externa. Reta de regressão preta: conjunto calibração e reta de regressão vermelha: conjunto validação externa.

Modelos de Classificação

Para os atributos índice de maturação e firmeza o conjunto de dados foi dividido em classes. As classes para maturação consistiam em: classe 1, tomates no estágio verde e pintando; e classe 2 com tomates vermelhos e vermelho intenso. A classificação de índice de maturação foi feita de acordo com padronização desenvolvida por cartilhas da USDA (Anexo 1). Em relação a firmeza as amostras foram separadas em: classe 1, tomates com firmeza igual ou acima de 15,58 N; e classe 2 frutos abaixo de 15,58 N.

Os modelos de classificação foram avaliados com relação a sua acurácia, de acordo com a equação (I). A acurácia é a medida mais simples da qualidade de um modelo de classificação e representa a porcentagem de amostras classificadas corretamente (31).

$$Acurácia (\%) = 100 - \left\{ \left(\frac{\sum y_1 + \sum y_2}{N} \right) \times 100 \right\} \text{ Eq. (I)}$$

Onde,

$\sum y_1$: soma dos erros de classificação da classe 1

$\sum y_2$: soma dos erros de classificação da classe 2

N : número amostral

Na tabela 3 estão apresentados os resultados dos modelos de classificação, nas etapas de calibração e validação externa, utilizando espectros de MIR e NIR. Para IM o modelo desenvolvido com infravermelho médio apresentou acurácias próximas de 100% de acerto, tanto na etapa de calibração quanto na validação (96,88% e 98,30%). Uma boa seleção de tomates, de acordo com seu grau de amadurecimento, possibilita que o produto final processado apresente características desejáveis pelo consumidor, como por exemplo a coloração vermelha intensa no molho de tomate. Mesmo resultando em uma menor porcentagem de acurácia, o modelo desenvolvido com infravermelho próximo ainda apresentou boa performance de classificação, como baixa porcentagem de erros na etapa de teste (validação externa) e com acurácia acima de 75%. Para a classificação de frutos com base na sua firmeza ambas técnicas resultaram em bons modelos, sendo que o modelo PLS-DA a partir de espectros de MIR apresentou maior acurácia (83,87%) do que aquele gerado com espectros de reflectância no infravermelho próximo – NIR (80,95%).

Tabela 3: Resultados dos modelos de classificação de tomates a partir do seu índice de maturação e firmeza (em Newtons), desenvolvidos por PLS-DA, utilizando espectroscopia de NIR e MIR.

		Etapa de Calibração + Validação interna					Etapa de Validação Externa				
		Classe 1		Classe 2		Acurácia (%)	Classe 1		Classe 2		Acurácia (%)
		n	Incorretas	n	Incorretas		n	Incorretas	n	Incorretas	
NIR	IM	34	10	110	11	85,42	27	10	36	5	76,19
	Firmeza (N)	95	10	49	16	81,94	30	5	33	7	80,95
MIR	IM	44	2	81	2	96,80	20	0	39	1	98,30
	Firmeza (N)	80	11	64	15	81,94	41	4	21	6	83,87

IM: índice de maturação; n: número amostral.

Com valores acima de 75% de acerto durante a classificação na validação externa, ambas as técnicas se mostraram aplicáveis em etapas que antecedem o processamento de tomates. Estes resultados são similares ao encontrados na literatura, Tiwari e colaboradores (2013), utilizando infravermelho próximo, desenvolveram modelos de classificação de tomates verdes com acurácia entre 76 e 84%. Em outro estudo, Sirisomboon e colaboradores (2012) avaliaram a aplicabilidade do NIR em classificar níveis de maturação de tomates ‘Momotaro’, e observaram baixa porcentagem de acerto nos modelos desenvolvidos com a ferramenta SIMCA, mas a análise de componentes principais mostrou distinção entre os estádios de maturação.

Utilizando os espectros de MIR, os quais resultaram em melhores modelos de classificação, foi feita a análise de componentes principais (PCA) com a finalidade de observar e explorar os padrões das amostras. Na figura 5 estão os gráficos de dispersão das amostras, no conjunto total de dados (calibração + validação externa).

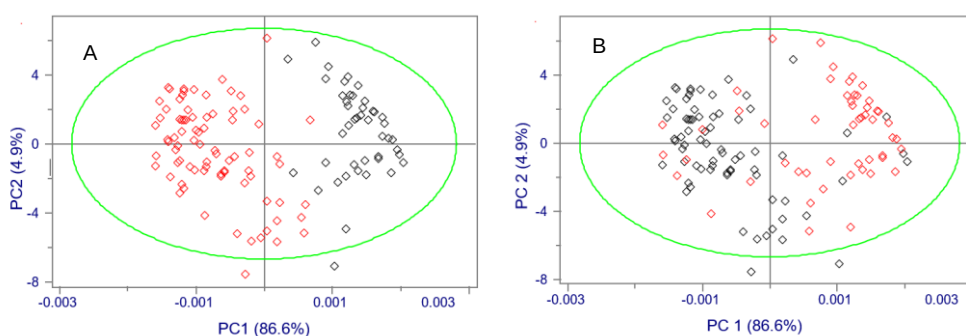


Figura 5: Gráfico de scores da PCA realizadas com os espectros de MIR, com classe IM (A) (preto = classe 1 e vermelhos = classe 2) e firmeza (B) (preto = classe abaixo 15,58 e vermelho = classe acima de 15,58).

A ferramenta quimiométrica PCA é uma técnica não supervisionada utilizada em geral para realizar uma análise exploratória dos dados (15). Método que transforma um conjunto de dados multivariados em outro de menor dimensão, denominados componentes principais (34). Observa-se que com apenas duas componentes principais foi possível explicar acima de 90% da variação dos dados (PC1: 86,6%; PC2: 4,9%). Cada variável espectral apresenta um peso, ou correlação, com a informação da variação do conjunto de amostra. Esses pesos, também chamado de *loadings*, estão apresentados na figura 6, nota-se que tanto a PC 1 quanto a PC 2 apresentam *loadings* em regiões similares do espectro de infravermelho médio. Essas regiões foram utilizadas tanto no desenvolvimento dos modelos de determinação de SS, pH e firmeza, quanto para criação dos modelos de classificação. E representam as regiões de bandas de absorção que apresentam maior correlação com as variáveis dependentes (resultados das análises de referência e/ou classes). A avaliação dos *loadings* pode ser uma ótima ferramenta para a seleção de variáveis no desenvolvimento dos modelos de predição e classificação.

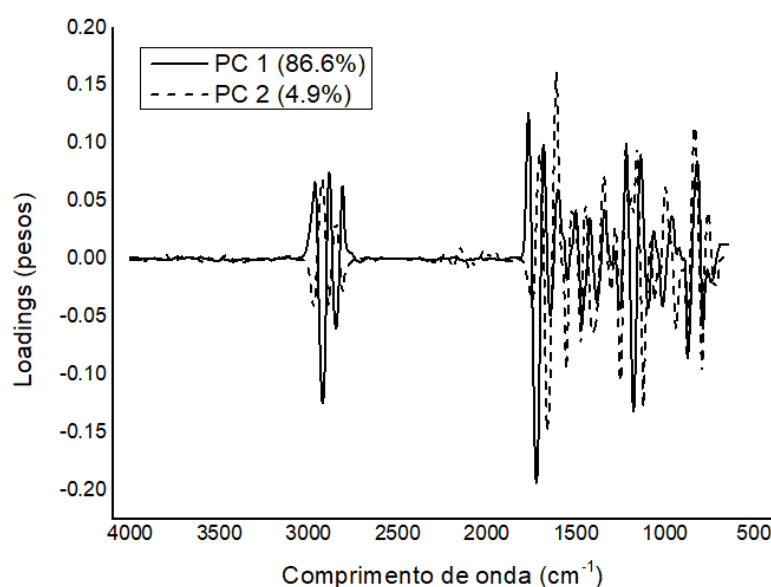


Figura 6: Gráfico dos pesos (*loadings*) mostrando a importância de cada variável do espectro de MIR para classificação da amostra em cada uma das classes.

CONCLUSÕES

O presente estudo mostrou que as técnicas de infravermelho próximo e médio tem potencial para serem utilizadas como métodos rotineiros de avaliação de qualidade de tomates destinados para o processamento. A espectroscopia de Infravermelho apresentou viabilidade de aplicação para a determinação de parâmetro SS, pH e firmeza. Sendo que a técnica de MIR resultou em modelos com coeficientes de correlação de Pearson maiores na determinação

Em relação aos modelos de classificação de tomates de acordo com o índice de maturação e firmeza, os modelos PLS-DA desenvolvidos com os espectros de MIR resultaram em acurácias maiores para a classificação. Com valores de 98,30% e 83,87% de acertos para índice de maturação e firmeza, respectivamente. Apesar de ter resultado em modelos de classificação com menor acurácia, quando comparado com a técnica de MIR, a espectroscopia de infravermelho próximo apresentou boas taxas de acerto (acima de 75%).

Agradecimentos

Este estudo contou com o apoio da instituição CAPES (processo nº 88881.187998 / 2018-01), bolsas de Produtividade em Pesquisa - CNPq (processos: 309043/2016-6 e 302866/2017-5) e auxílio FAPESP (n.º 2020 / 07017-0 e 2019/13656 -8).

REFERÊNCIAS

1. FAO. Food and Agriculture Organization of the United Nations. 2020.
2. Embrapa. Sistemas de Produção [Internet]. 2006. Available from: https://sistemasdeproducao.cnptia.embrapa.br/FontesHTML/Tomate/TomateIndustrial_2ed/index.htm
3. Bureau S, Boas AA de CV, Giovinazzo R, Jaillais B, Page D. Toward the implementation of mid-infrared spectroscopy along the processing chain to improve quality of the tomato based products. LWT [Internet]. 2020 Aug;130:109518. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0023643820305077>
4. Ścibisz I, Reich M, Bureau S, Gouble B, Causse M, Bertrand D, et al. Mid-infrared spectroscopy as a tool for rapid determination of internal quality parameters in tomato. Food Chem. 2011;125(4):1390–7.
5. Ayvaz H, Sierra-Cadavid A, Aykas DP, Mulqueeney B, Sullivan S, Rodriguez-Saona LE. Monitoring multicomponent quality traits in tomato juice using portable mid-infrared (MIR) spectroscopy and multivariate analysis. Food Control [Internet]. 2016 Aug;66(January):79–86. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0956713516300329>
6. Gould WA. Total Acidity and pH. Tomato Prod Process Technol [Internet]. 1992;345–52. Available from:

<http://linkinghub.elsevier.com/retrieve/pii/B9781845695996500281>

7. Nicolai BM, Beullens K, Bobelyn E, Peirs A, Saeys W, Theron KI, et al. Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. *Postharvest Biol Technol.* 2007 Nov;46(2):99–118.
8. Borba KR, Spricigo PC, Aykas DP, Mitsuyuki MC, Colnago LA, Ferreira MD. Non-invasive quantification of vitamin C, citric acid, and sugar in ‘Valência’ oranges using infrared spectroscopies. *J Food Sci Technol.* 2020;(Usda 2019).
9. Pérez-marín D, Torres I, Entrenas J, Vega M, Sánchez M. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* Pre-harvest screening on-vine of spinach quality and safety using NIRS technology. 2019;207(1258):242–50.
10. Nicolai BM, Defraeye T, De Ketelaere B, Herremans E, Hertog MLATM, Saeys W, et al. Nondestructive measurement of fruit and vegetable quality. Vol. 5, *Annual review of food science and technology.* 2014. 285–312 p.
11. De Oliveira GA, Castilhos F, Renard CMA-G, Bureau S. Comparison of NIR and MIR spectroscopic methods for determination of individual sugars, organic acids and carotenoids in passion fruit. *Food Res Int.* 2014 Jun;60:154–62.
12. Nascimento PAM, Carvalho LC de, Júnior LCC, Pereira FMV, Teixeira GH de A. Robust PLS models for soluble solids content and firmness determination in low chilling peach using near-infrared spectroscopy (NIR). *Postharvest Biol Technol.* 2016;111:345–51.
13. Dufour É. *Principles of Infrared Spectroscopy.* *Infrared Spectrosc Food Qual Anal Control.* 2009;
14. Subramanian A, Prabhakar V, Rodriguez-Saona L. *Analytical Methods: Infrared Spectroscopy in Dairy Analysis.* Reference Module in Food Science. 2016.
15. Ferreira MMC. *Qumiometria: Conceitos, Métodos e Apliações.* Editora Un. Campinas; 2015. 493 p.
16. Clément A, Dorais M, Vernon Ma. Multivariate Approach to the Measurement of Tomato Maturity and Gustatory Attributes and Their Rapid Assessment by Vis - NIR Spectroscopy. *J Agric Food Chem.* 2008;56:1538–44.
17. Ferreira MD, Spricigo PC. *Colorimetria – Princípios e aplicações na agricultura.* In: *Instrumentação Pós-Colheita em Frutas e Hortaliças.* Brasília – DF: Embrapa; 2017.

18. Gould WA. Color and Color Measurement. Tomato Prod Process Technol [Internet]. 1992;297–312. Available from: <http://linkinghub.elsevier.com/retrieve/pii/B9781845695996500256>
19. Monteiro CS, Balbi ME, Miguel OG, Penteado PTP da S, Haracemiv SMC. QUALIDADE NUTRICIONAL E ANTIOXIDANTE DO TOMATE “ TIPO ITALIANO .” 2008;25–31.
20. Gliordano LB de, Ribeiro CS. Origem. Botânica e Composição Química do fruto. Tomate para processamento industrial. 2000;12–7.
21. Sun D-W, Schulz H, Baranska M. Infrared Spectroscopy for Food Quality Analysis and Control. Infrared Spectroscopy for Food Quality Analysis and Control. 2009. 321–353 p.
22. Ayvaz H, Rodriguez-Saona L, L. Wehling R. Infrared and Raman Spectroscopy. 2017. p. 129–50.
23. Yang H, Irudayaraj J. Rapid determination of vitamin C by NIR, MIR and FT-Raman techniques. J Pharm Pharmacol. 2002 Sep;54(9):1247–55.
24. Pavia DL, Kris GS, Lampman GM, Vyvyan JR. Introdução à espectroscopia. 1ª. Cengage Learning; 2010. 716 p.
25. Nelson DL, Cox M. Lehninger – Princípios de Bioquímica. 3ª. São Paulo: Sarvier; 2002. 1009 p.
26. Kowalski BR, Seasholtz MB. Recent developments in multivariate calibration. J Chemom. 1991;5:129–45.
27. Aykas DP, Karaman AD, Keser B, Rodriguez-Saona L. Non-targeted authentication approach for extra virgin olive oil. Foods. 2020;9(2):1–17.
28. Akpolat H, Barineau M, Jackson KA, Aykas DP, Rodriguez-saona LE. LWT - Food Science and Technology Portable infrared sensing technology for phenotyping chemical traits in fresh market tomatoes. 2020;124(February).
29. Chitarra MIF, Chitarra AB. Pós-colheita de frutas e hortaliças: fisiologia e manuseio. 2ª EDIÇÃO. Lavras: FAEPE, editor. 2005; 2005. 783 p.
30. Kader AA. Postharvest Technology of Horticultural Crops. trird. University of California. Agriculture ans Natural Resources; 2002.

31. Ballabio D, Todeschini R. Multivariate Classification for Qualitative Analysis. *Infrared Spectrosc Food Qual Anal Control*. 2009;83–104.
32. Tiwari G, Slaughter DC, Cantwell M. Nondestructive maturity determination in green tomatoes using a handheld visible and near infrared instrument. *Postharvest Biol Technol*. 2013 Dec;86:221–9.
33. Sirisomboon P, Tanaka M, Kojima T, Williams P. Nondestructive estimation of maturity and textural properties on tomato “Momotaro” by near infrared spectroscopy. *J Food Eng*. 2012;112(3):218–26.
34. Varella CAA. *Análise Multivariada Aplicada as Ciências Agrárias: Análise de componentes principais*. Universidade Federal do Rio de Janeiro (UFRJ); 2008.

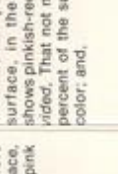
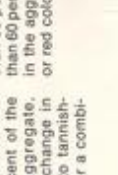
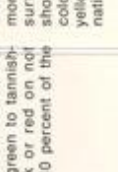
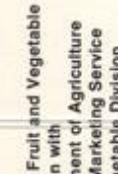
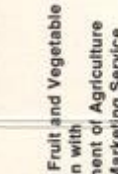
TOMATOES

UNITED STATES STANDARDS FOR GRADES OF FRESH

REQUIREMENTS IN

United Fresh Fruit and Vegetable Association
in cooperation with
U. S. Department of Agriculture
Agricultural Marketing Service
Fruit and Vegetable Division
U.S.D.A. Visual Aid TM-L-1; February '75
The John Henry Company
P.O. Box 1410, Lansing, Mich. 48904



(1) "Green" means that the surface of the tomato is completely green in color. The shade of green color may vary from light to dark;	    GREEN 	    BREAKERS 	    TURNING 	    PINK 	    LIGHT RED 	    RED 	(2) "Breakers" means that there is a definite break in color from green to tannish-yellow, pink or red on not more than 10 percent of the surface; (3) "Turning" means that more than 10 percent but not more than 30 percent of the surface, in the aggregate, shows a definite change in color from green to tannish-yellow, pink, red, or a combination thereof; (4) "Pink" means that more than 30 percent but not more than 60 percent of the surface, in the aggregate, shows pink or red color; (5) "Light red" means that more than 60 percent of the surface, in the aggregate, shows pinkish-red or red; PROVIDED, That not more than 90 percent of the surface is red color; and, (6) "Red" means that more than 90 percent of the surface, in the aggregate, shows red color.
---	---	---	---	--	---	---	--

The above photographs are only guides illustrating the shade and percentage of surface color specified for each of the color terms. These photographs do not necessarily depict absolute limits of minimum or maximum shades and/or percentage of color required for each term.

Capítulo 3.

**Selection of industrial tomatoes using TD-NMR data and
computational classification methods.**

Artigo em submissão na revista Industrial Crops and Products

Selection of industrial tomatoes using TD-NMR data and computational classification methods.

Karla R. Borba^a, Fernanda C.A. Oldoni^a, Tatiana Monaretto^{b,c}, Luiz A. Colnago^b and Marcos D. Ferreira^b

^aDepartment of Food and Nutrition, School of Pharmaceutical Sciences, São Paulo State University-UNESP, Araraquara – Jaú, Km 1, 14801-902 - Araraquara-SP, Brazil.

^bEmbrapa Instrumentation, XV de Novembro, 1452, 13561-206 - São Carlos-SP, Brazil.

^c São Carlos Institute of Chemistry, São Paulo University, Trabalhador São Carlense Avenue 400, São Carlos-SP, 13566-590.

*Corresponding author < borbakr@gmail.com >

Highlights

- TD-NMR technique is a potential high-throughput screening application before the industry processing.
- Computational methods associated with TD-NMR relaxometry can be highly applied to classify unripened, low SSC, and defected tomatoes.
- Classification models great accuracies: Maturity index: 97% (SVM); SSC: 100% (SVM and kNN); presence of defects: 90% (PLS-DA).

Abstract

The tomato processing chain has a world economic relevance for the food industry and the agribusiness sector, providing ready-to-eat products and raw material (mainly tomato paste) for other production chains (sauces, juices, ketchup, and other derivatives). The final product quality is depending on control of some fruit attributes, such as color (related to the ripeness index), soluble solids content (SSC), and defects (mostly rot and immature tomatoes). The aim of this study was to develop accurate and nondestructive classification models according to the tomato maturation stage, SSC, and presence of defects using TD-NMR (Time domain – Nuclear Magnetic Resonance) associated with SIMCA (Soft Independent Modeling by Class Analogy), PLS-DA (Discriminant analysis by partial least squares classification), SVM (Support Vector Machine), and kNN (k-nearest neighbors). A total of 132 tomatoes were collected in a tomato processing facility, divided into different classes, for maturity (according USDA maturity classification, 2020) and defects, and analyzed using a low-field NMR spectrometer. Each class showed different decay times, for example, green tomatoes showed a short decay signal than red tomatoes. Mainly, due to the relaxation signal is related to the mobility of the water in different vegetable tissue compartments. Classification models resulted in great accuracy performances, the best accuracy for each classification were: Maturity index: 97% (SVM); SSC: 100% (SVM and kNN); presence of defects: 90% (PLS-DA). These results show that CPMG decays associated with chemometrics (SIMCA and PLS-DA) and machine learning algorithms (SVM and kNN) can be used in the tomato processing industry to classify samples according to maturation index, SSC, and defects presence. These classification models showed the potential of TD-NMR technique in a high-throughput screening application before the processing.

Keywords: NMR; tomato processing; relaxation time; PLS; PLS-DA; SIMCA; SVM; kNN.

1. Introduction

Tomato is the second most-produced and consumed vegetable in the world. Its global production in 2019 was approximately 197 million tons (FAO, 2020). Brazil is the ninth-largest tomato producing country, representing almost 2.5% of world production, with 4.1 million tons in 2018 (FAO, 2020). The industrial tomato production chain has economic relevance for the food industry and agribusiness sector, providing ready-to-eat products and inputs for other production chains (sauces, juices, and other derivatives) (Bureau, Boas, Giovinazzo, Jaillais, & Page, 2020).

During food industry processing, quality control of the raw material is essential for consumer acceptance and for the final industry yield. Control of some quality attributes of tomatoes, such as color (related to the ripeness index), soluble solids content, and defects (mainly rot and immature tomatoes) are important to achieve the final quality required by market (Ayvaz et al., 2016; Ścibisz et al., 2011). Most of industries use manual resources for a prior selection of their raw material, for example, manual removal of inappropriate tomatoes on the production running machine. However, this strategy is affected by the high speed and amount of tomatoes that pass through the system during processing. Reducing labor requirements and efficiently sorting tomatoes based on their properties are alternative and effective methods that can significantly increase profitability at the processing tomatoes sector (Tu, Choi, McCarthy, & McCarthy, 2007). Implementation of an automatic step to classify and exclude unsuitable fruits would result in greater efficiency and reliability than the usual methods in use.

Products derived from tomatoes are more accepted by consumers when they have an intense red color. Therefore, processing tomatoes must be selected according to their stage of ripeness. If during processing, the percentage of green tomatoes is large, their pulp or final tomato paste will show a brownish color, which causes a bad impact on consumers (Gould, 1992). In addition, more mature tomatoes (red ripe stage) results in high levels of carotenoids, more specifically lycopene, which increases the nutritional value of the final product (Gould, 1992). Besides that, ripeness stage also influences the peelability of tomatoes; ripe tomatoes have easy removal of its peel, which is one of the crucial industrial processing steps (Barrett, Garcia, & Miyao, 2006).

Another quality parameter that must be controlled throughout the production chain is the soluble solids content (SSC), which is also related to the degree of ripeness (Ayvaz et al., 2016; Bureau et al., 2020). The SSC is directly associated with final yield, because to reach a certain quantity of final concentrated product with high levels of SSC will be required a small amount of fruit and time to water evaporation (Wilkerson et al., 2013). According to studies done by Embrapa (2006), for each 1 ° BRIX of increase in the raw material there is an increase of 20% in the industrial yield.

The presence of defects directly affects the quality of processed tomatoes. Defects can be classified as serious as rotten fruits or as light defects as superficial damage, spots, deformation, and immature tomatoes (Barrett et al., 2006). Several factors may cause defects such as bacterial

disease, attack of insects, mold growth, advanced maturity, inadequate plant nutrition/fertilization, or mishandling during harvesting and/or transportation (Barrett et al., 2006). Some of these defects are more detrimental during processing, such as rotten and immature tomatoes for example, while others, like spots and deformation, are lighter. Also, some defects may be cosmetic, affecting only tomato appearance, while others may affect quality aspects such as the presence of a heavy load of mold (Barrett et al., 2006; Barrett, Garcia, & Wayne, 1998).

Several new instrumental techniques had been increasingly applied in the food industry, especially for quality inspection and defect sorting sectors. Machine vision (Dhakshina Kumar, Esakkirajan, Bama, & Keerthiveena, 2020), image processing (Arakeri & Lakshmana, 2016; Ireri, Belal, Okind, Makange, & Ji, 2019; Tonguç & Yakut, 2009), time-domain nuclear magnetic resonance (TD-NMR), spectroscopy (Fan & Zhang, 2019) and infrared spectroscopy (Nicolai et al., 2007) are some of the current techniques. All those techniques result in a complex output data, with a thousand variables (Pasquini, 2018). To solve this issue several computational classification methods are used, such as PLS-DA (Discriminant analysis by partial least squares classification), SIMCA (Soft Independent Modeling by Class Analogy), SVM (Support Vector Machine), and kNN (k-nearest neighbors).

TD-NMR relaxometry and diffusimetry, using bench top, low field instruments have been widely used for quantitative and qualitative food analyses (Bizzani et al., 2020; Carvalho, Verbi Pereira, Morais, Lima, & Teixeira, 2019; Flores, Colnago, Ferreira, & Spoto, 2016; Kirtil, Cikrikci, McCarthy, & Oztop, 2017; Poliana M. Santos et al., 2014). Time-domain NMR has been used to determine ripeness, defects, and quality of intact fruits and vegetables such as plums (Pereira, Carvalho, Cabeça, & Colnago, 2013), orange (Bizzani, Flores, Colnago, & Ferreira, 2017), mangoes (Bizzani et al., 2020), bananas (Raffo, Gianferri, Barbieri, & Brosio, 2005; Ribeiro et al., 2010), nuts (Carvalho et al., 2019), and cucumber (Dong et al., 2017). The major advantages of TD-NMR when compared with others physical methods are: it is a non-destructive and non-invasive technique that preserve the sample integrity; analyze the whole sample content (not surface analysis), simple and rapid measurements and has high-throughput screening potential, and the ability to perform direct analysis without sample preparations (Colnago, Engelsberg, Souza, & Barbosa, 2007; Pereira et al., 2013; Prestes et al., 2007).

Therefore, the potential of using TD-NMR relaxometry associated with SIMCA, PLS-DA, SVM, and kNN algorithms, to develop accurate classification models according to the tomato maturation stage, SSC, and presence of defects, was studied.

2. Material and methods

Samples

A total of 132 tomatoes in different ripening stages, with and without defects were collected in a tomato processing facility at the São Paulo state, Brazil. The samples were collected from three different harvests times during the 2017 season.

Maturity and defects clusters

For the maturity classification two classes were developed, according to United States Standards for Grades of Fresh Tomatoes (USDA, 2020), class 1 named ‘green’ with only green tomatoes and class 2, ‘red’, that included light red and red-ripe tomatoes. The classification of fruit defects was also based on two, as follow, the class ‘defect’ presented some serious defect, such as rotten, overly ripened, or immature; and class ‘no defect’, free of any of these serious issues (Barrett et al., 2006).

Solids Soluble Content (SSC) quantification

The SSC was determined using a digital refractometer ATAGO (ATAGO, Bellevue, WA) at 20 °C with $\pm 0.1\%$, an aliquot of 2 mL of extracted juice was placed to refractometer detector and the results were expressed as °Brix. Tomato samples were separated into two groups, group 1 with SSC below or equal 4.5 °Brix and group 2 tomatoes that presented SSC above 4.5 °Brix.

Time-domain nuclear magnetic resonance (TD-NMR)

TD-NMR measurements of intact tomatoes were carried out on an SLK-MRI-1400 relaxometer (Spinlock Magnetic Resonance Solution, Cordoba, Argentina) equipped with a 0.23 T (9 MHz for ^1H) Halbach permanent magnet and a 10 cm diameter bore. Carr–Purcell–Meiboom–Gill (CPMG) pulse sequence (De Andrade et al., 2011) was used for ^1H transverse relaxation time measurement (T_2), with 90° and 180° pulses of 32 and 64 μs , echo time (τ) of 940 μs , 3000 echoes, 8 scans, and a recycle delay of 5 s. Analysis were performed at 21 ± 1 °C. The average CPMG signals, shown in Fig. 1, were obtained by the arithmetic mean of three normalized (from 1 to 0) CPMG signals from to the same tomatoes class.

Data processing

The TD-NMR data were preprocessed, using Origin 8.1 software (Origin Lab, Northampton, MA, USA), through normalization by the maximum (0-1), smothered by a second-order polynomial Savitzky-Golay algorithm with 25-point window to reduce eventual noise, and mean centered.

The Inverse Laplace Transform (ILT) was implemented in the average CPMG signals using the Tikonov regularization (Tikhonov & Arsenin, 1977). The optimal regularization parameter (α) was determined by the BRD (Butler-Reeds-Dowson) method (Mitchell, Chandrasekera, & Gladden, 2012).

Multivariate analysis and machine learning algorithms

PCA (Principal Components Analysis) was applied using TD-NMR CPMG signals as an exploratory tool. The SIMCA (Soft Independent Modeling by Class Analogy) and PLS-DA (Discriminant analysis by partial least squares classification) models were built using Pirouette v.4.5 software (Infometrix, Bothell, WA, USA). The preprocessed data were divided, randomly, into two sets, with 70% used for calibration (training) and 30% used for validation (testing). The performance of developed classification models was evaluated by precision, sensitivity, specificity, and accuracy.

SVM (Support Vector Machine) were applied to the TD-NMR decay using Python 3.7 in the Visual Studio Code version 1.46.1 development environment. The data structure was organized, analyzed, and evaluated using the Panda 0.18.1 and Numpy 1.11.1. The auxiliary package used to perform the SVM calculations was Scikit-learn 0.19.1 and graphs were generated with Matplotlib 2.2.3. The SVM classifier was optimized for the radial basis function (RBF) and was chosen based on previous performance tests. The k-nearest neighbors (kNN) algorithm, with $k = 3$. To SVM and kNN analysis, the data were split into training (80%) and test (20%) set, and the accuracies (and further performance factors) of both sets were evaluated.

3. Results and discussion

3.1 TD-NMR relaxation signals

Fig. 1 shows the decays of average CPMG signals from tomatoes samples classified by color or ripening stage (green: unripe, or red: ripe), SSC content ($4.5 \leq \text{Brix} < 4.5$), and with or no defect. The CPMG decay is governed by transverse relaxation time (T_2) that shows the tomatoes samples can be arrange in two groups: the ones indicated by the green line, with high SSC and no defect tomatoes, show longer CPMG decays signals than the ones on the red line, that have low SSC and defects presence.

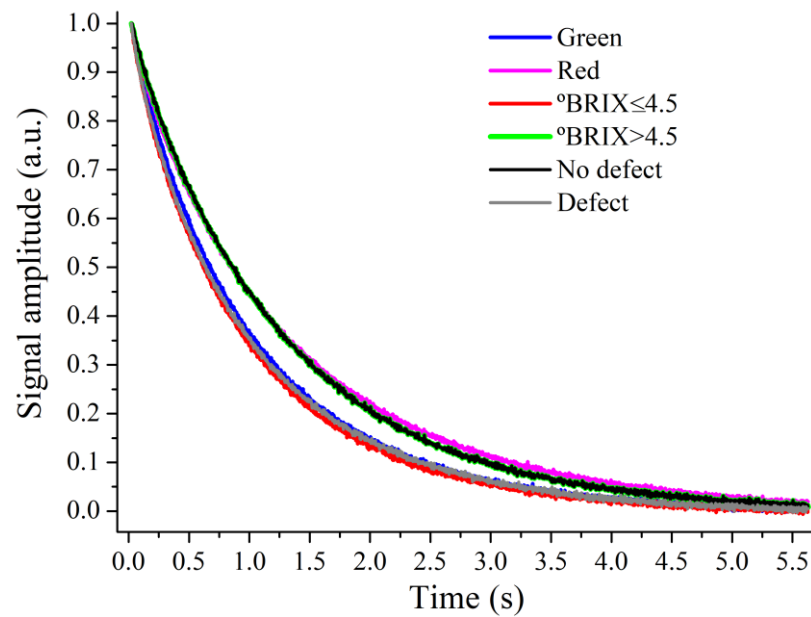


Figure 1. Average CPMG decay signals of tomatoes from the three classes: green/unripe (blue), red/ripe (pink), $^{\circ}\text{Brix} \leq 4.5$ (red) and $^{\circ}\text{Brix} > 4.5$ (green), and defect (grey) and no defects (black).

Fig. 2, shows the distribution of the transverse relaxation time obtained by employing the ILT algorithm to the average CPMG signals from Fig.1 (Monaretto et al., 2019). The T_2 distribution time of the tomatoes, show in Fig. 2, are composed of three relaxation times, as it is commonly observed in fruits (Ribeiro et al., 2010; Hills, 1999). T_2 values and respective areas are in the Table 1. All samples show a strong peak at approximately 1 s (T_{23}), which represent more than 95% of the water content in the tomatoes. These peaks have been assigned to the water in the vacuoles or large porous (Ribeiro et al., 2010; Hills, 1999). The other two small peaks are the T_{22} (from 0.130 to 0.455 s) and T_{21} (from 0.021 to 0.083 s). The T_{22} has been assigned to the water in

the cytoplasm and T_{21} water in cells walls and attached to a large structure such as starch grain (Ribeiro et al., 2010).

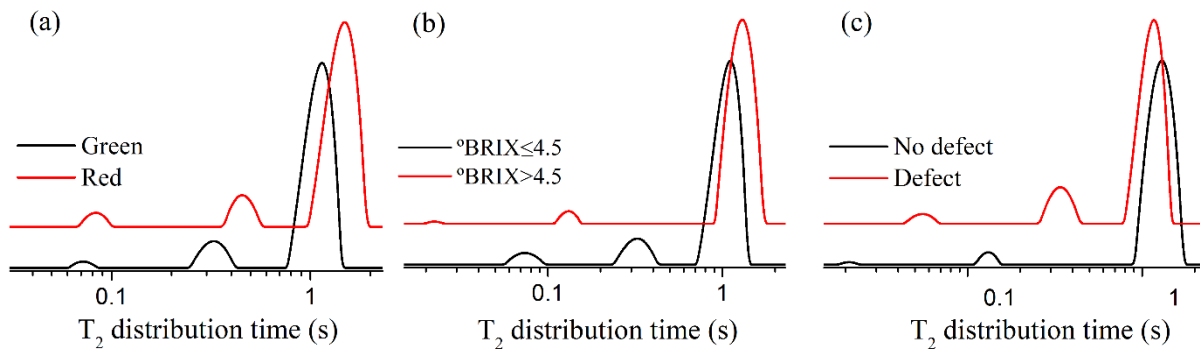


Figure 2. T_2 distribution time of average CPMG signals obtained from each class of tomatoes: (a) green and red; (b) $^{\circ}\text{Brix} \leq 4.5$ and $^{\circ}\text{Brix} > 4.5$; and (c) no defect and defect.

The three T_2 of the red tomatoes have the highest values of all tomatoes class studied, as shown in Fig. 2 and Table 1. The T_2 of water in fruits, like tomatoes, is dependent on water dynamics that depends on the food microstructure (pore size), water content and solutes concentration (Hills, 1999; Ishida, Koizumi, & Kano, 1994). After the ripening process, the total amount of water with high mobility increases. Red-ripe tomatoes show large water content in the outer wall pericarp, dissepiments, and columella (Hills & Clark, 2003; Ishida et al., 1994; Ishida, Kobayashi, Koizumi, & Kano, 1989). One of the most important tomato changes occurs during the ripening of green to pink tomato, when occurs the liquefaction of locular tissues (Hills & Clark, 2003). All these ripening modifications are useful, and observed, in the TD-NMR signals and can be used to classify tomatoes according to its maturity.

The green tomato signal showed a shorter decay than the red one (Fig. 1), which reflects the different water mobility in different vegetable tissue compartments (Ishida et al., 1989). In green tomatoes, the large amounts of water are in seeds and seed envelopes, and most of that water has a restricted mobility (Zhang, McCarthy, & McCarthy, 2016). The tomatoes with no defect have the longer and the larger relative area of T_{23} than tomatoes with defect, as is possible to see in Fig. 2 (c) and Table 1. A similar effect has also been observed in banana (Ribeiro et al., 2010). The defect may cause damage in some vacuoles and the vacuolar water can reach the cytoplasm as the area of T_{22} increases.

Table 1 and Fig. 2 (b) shows that tomatoes with higher °Brix have higher T_{23} value than tomatoes with lower °Brix. Normally, fruits, with much higher °Brix than tomatoes, like oranges (Flores et al., 2016) and plums (Pereira et al., 2013), the T_2 has an inverse correlation with °Brix. The increased viscosity of the water due to the high concentration of sugars, which consequently reduces the T_2 values, has explained this subject. In tomatoes, the °Brix and T_2 seems to have no inverse correlation may be because tomatoes have low sugar concentration. The T_2 variation in tomatoes may be related to the changes in cellular structure on the ripening process, once a higher °Brix value is observed in ripe tomatoes.

Table 1: T_2 values and respective areas for each tomato class (maturation, SSC, and defects).

TOMATOES CLASS		T_2 (S)	RELATIVE AREA %
Green	T_{21}	0.072	0.3
	T_{22}	0.322	3.2
	T_{23}	1.16	96.5
Red	T_{21}	0.083	0.4
	T_{22}	0.455	3.3
	T_{23}	1.48	96.3
°Brix ≤ 4.5	T_{21}	0.075	0.5
	T_{22}	0.322	3.4
	T_{23}	1.130	96.1
°Brix > 4.5	T_{21}	0.022	0.1
	T_{22}	0.130	0.3
	T_{23}	1.290	99.6
No defect	T_{21}	0.021	0.1
	T_{22}	0.130	0.3
	T_{23}	1.29	99.6
Defect	T_{21}	0.055	0.4
	T_{22}	0.345	4.7
	T_{23}	1.16	94.9

The difference between CPMG decays of these different classes were used to build classification models, using chemometrics methods and machine learning algorithms.

3.2 PCA

Figure 3 shows the PCA score plot of CPMG decays for maturation, SSC, and defect data sets. The principal component 1 (PC1) and 2 (PC2) were responsible to describe 98,4% of the

NMR data variance for maturation set (fig. 1A). It is possible to observe class discrimination between ripe (red points) and unripe (black points) tomatoes. PCA from SSC data set (fig. 3B) also shows PC1 and PC2 explaining around 98% of the total variance, and tomatoes content high SSC values (red spots) are localized mostly in the positive PC1 and all samples from class with SSC above 4.5°Brix are in negative PC1. PC1 for defects tomatoes data set also was capable to discriminate classes for almost all samples, positive portion showed only tomatoes free of any defect and negative PC1 portion presented all samples with defects (fig.3C). PCA, as an unsupervised exploratory analysis, showed that CMPG decay signals are suitable for classes separating using only 2 principal components.

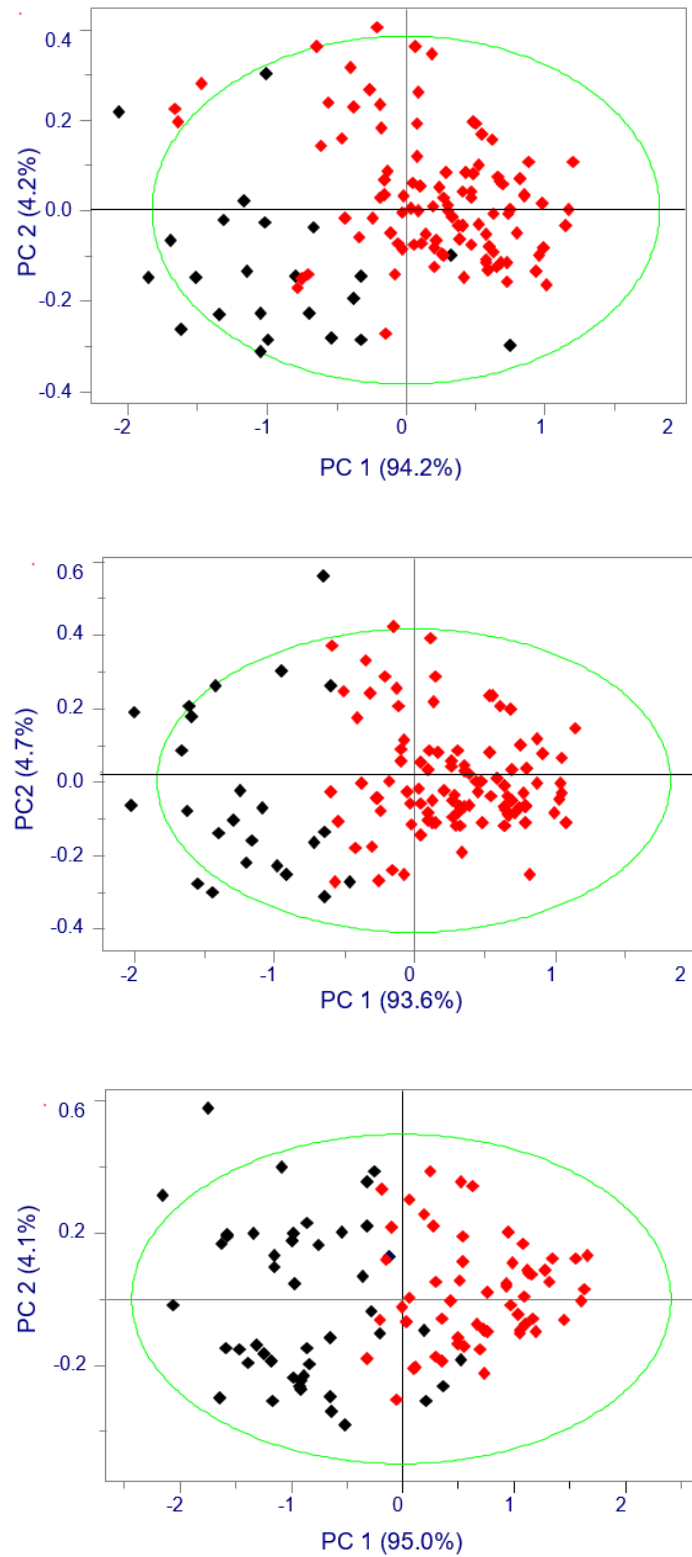


Figure 3: Principal components analysis (PCA) of Carr-Purcell-Meiboom-Gill decay data for (A) maturation (black: (green)unripe and red: (red) ripe tomatoes), (B) SSC (black: $^{\circ}\text{Brix} \leq 4.5$ and red: $^{\circ}\text{Brix} > 4.5$), and (C) defect (black: defects and red: no defects) data sets.

3.3 Classification models

To discriminate between the different classes of tomato maturation, SSC and presence of defects the supervised chemometric tools SIMCA and PLS-DA and machine learning algorithms SVM and kNN were carried out.

Figure 4 shows the 3D plots and decision boundaries of TD-NMR data for tomato samples separated into two maturity index classes. SIMCA model showed that the two classes were independent of one another, requiring three PCs to explain 99% of the total variance within groups. As a supervised chemometric method for classification, the training data set is assigning to classes and then establishes a principal component analysis (PCA) model for each class, and the corresponding models are developed individually (Santos, Pereira-Filho, & Rodriguez-Saona, 2013).

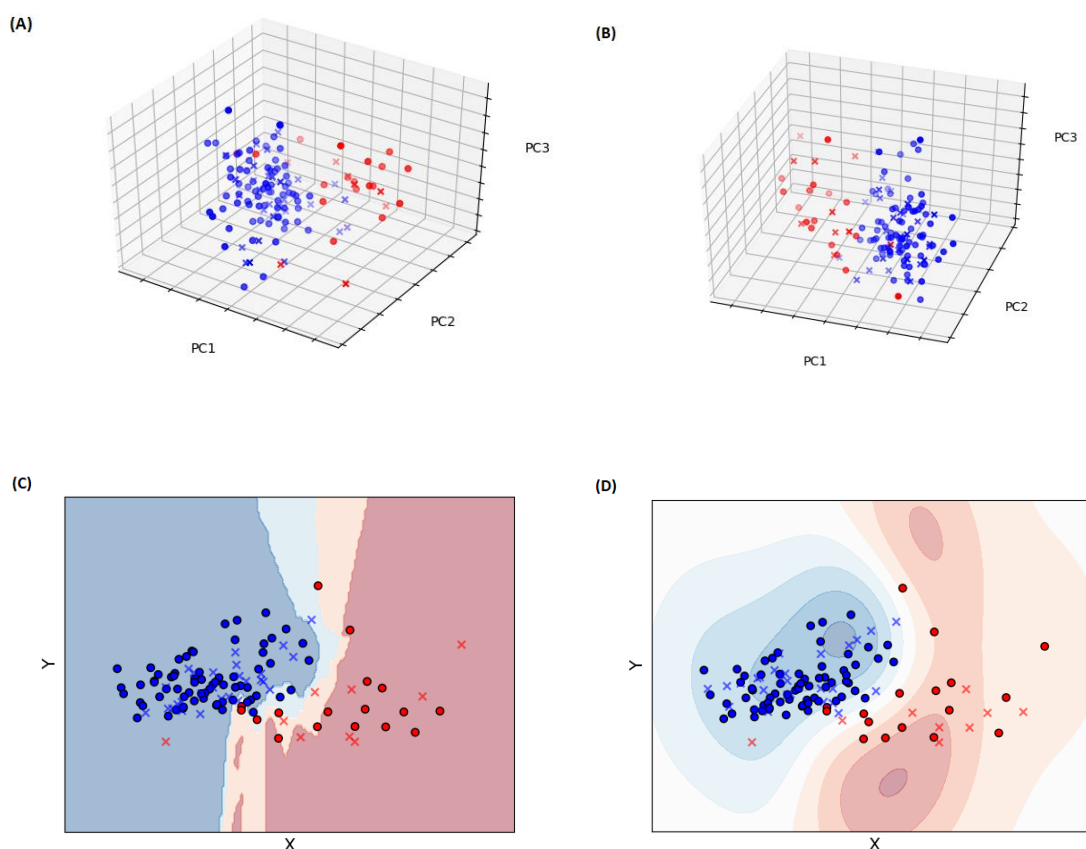


Figure 4. (A) SIMCA, (B) PLS-DA 3D projection plots of TD-NMR data for different maturity index of tomatoes, (C) SVM computed decision boundaries in the SVM model (rbf with $C = 10$), and (D) computed decision boundaries in the kNN model ($k=3$). Circles means calibration set and

crosses validation or test set; Red symbols are the green (unripe) tomatoes group and blue are red (ripe) tomatoes samples.

In table 2 are presented the performance metrics of each model during the training process, SIMCA classification obtained 100% of precision. All classification approaches resulted in great models, with a capacity of classification above 90%. PLS-DA PC1, 2 and 3 corresponded to 93.87%, 5.22%, and 0.28%. Only two principal components were able to explain 99% of the variance, which indicates that the model should be robust and minimizes overfitting (Aykas, Karaman, Keser, & Rodriguez-Saona, 2020). PLS-DA is a supervised algorithm based on the analytical data (TD-NMR signals in this case) and established categories relationship. this algorithm seeks to maximize covariance between the two types of variables (Amodio, Ceglie, Chaudhry, Piazzolla, & Colelli, 2017).

Table 2. Summary of the maturation index, SSC, and defects classification models for training data set performed using SIMCA, PLS-DA, SVM and kNN.

Attribute	Model		Classes		Precision	Sensitivity	Specificity	Accuracy
Maturity	SIMCA	Green	Green	Red				
			16	0	1.00	1.00	1.00	
	PLS-DA	Red	0	76	1.00	1.00	1.00	1.00
		Green	13	3	1.00	0.81	1.00	
	SVM	Red	0	74	0.96	1.00	0.81	0.94
		Green	15	2	1.00	0.88	1.00	
	KNN	Red	0	67	0.97	1.00	0.88	0.98
		Green	15	2	1.00	0.88	1.00	
		Red	0	67	0.97	1.00	0.88	0.98
SSC	SIMCA	Group 1	Group 1	Group 2				
			13	2	1.00	0.87	1.00	
	PLS-DA	Group 2	0	72	0.97	1.00	0.87	0.98
		Group 1	11	3	1.00	0.79	1.00	
	SVM (rbf)	Group 2	0	72	0.96	1.00	0.79	0.97
		Group 1	16	3	0.94	0.84	0.99	
	KNN	Group 2	1	68	0.96	0.99	0.84	0.95
		Group 1	16	3	0.94	0.84	0.99	
		Group 2	1	68	0.96	0.99	0.84	0.95
Defects	SIMCA	Defect	Defect	No defect				
			31	0	1.00	1.00	1.00	
	PLS-DA	No defect	0	44	1.00	1.00	1.00	1.00
		Defect	24	6	0.86	0.80	0.91	
	SVM	No defect	4	40	0.87	0.91	0.80	0.86
		Defect	37	1	0.9	0.97	0.89	
	KNN	No defect	4	31	0.97	0.89	0.97	0.93
		Defect	36	2	0.97	0.95	0.97	
		No defect	1	34	0.94	0.97	0.95	0.96

SIMCA: Soft Independent Modeling by Class Analogy; PLS-DA: Discriminant analysis by partial least squares classification; SVM: Support Vector Machine (radial basis function); and kNN: k-nearest neighbors' algorithm.

SVM and kNN algorithms presented higher accurate models to classify green tomatoes, with 97 and 94% accuracy, respectively, in the test set and showed lower misclassifications (1 and 2 wrong matches, respectively) (Table 3). The k-nearest neighbor algorithm is an object classification technique based on closest training examples in the problem space, kNN algorithm

is amongst the simplest of all machine learning algorithms (Raikwal & Saxena, 2012), the object (sample) is assigned to the class most common amongst its K nearest neighbors.

Table 3. Summary of the maturation index, SSC, and defects classification models for validation (test) data set performed using SIMCA, PLS-DA, SVM and kNN.

Attribute	Model		Classes		Precision	Sensitivity	Specificity	Accuracy
Maturity	SIMCA	Green	4	4	0.80	0.50	0.97	0.88
		Red	1	31	0.89	0.97	0.50	
	PLS-DA	Green	6	2	0.60	0.75	0.88	0.85
		Red	4	28	0.93	0.88	0.75	
	SVM (rbf)	Green	8	1	1.00	0.89	1.00	0.97
		Red	0	27	0.96	1.00	0.89	
	KNN	Green	8	1	0.89	0.89	0.96	0.94
		Red	1	26	0.96	0.96	0.89	
SSC	SIMCA	Group 1	4	5	1.00	0.44	1.00	0.87
		Group 2	0	30	0.86	1.00	0.44	
	PLS-DA	Group 1	7	2	1.00	0.78	1.00	0.90
		Group 2	0	28	0.93	1.00	0.78	
	SVM (rbf)	Group 1	5	0	1.00	1.00	1.00	1.00
		Group 2	0	33	1.00	1.00	1.00	
	KNN	Group 1	5	0	1.00	1.00	1.00	1.00
		Group 2	0	33	1.00	1.00	1.00	
Defects	SIMCA	Defect	14	5	0.93	0.74	0.92	0.81
		No defect	1	11	0.69	0.92	0.74	
	PLS-DA	Defect	16	3	1.00	0.84	1.00	0.90
		No defect	0	12	0.80	1.00	0.84	
	SVM (rbf)	Defect	11	0	0.73	1.00	0.81	0.88
		No defect	4	17	1.00	0.81	1.00	
	KNN	Defect	11	0	0.73	1.00	0.81	0.88
		No defect	4	17	1.00	0.81	1.00	

SIMCA: Soft Independent Modeling by Class Analogy; PLS-DA: Discriminant analysis by partial least squares classification; SVM (rbf): Support Vector Machine (radial basis function); and kNN: k-nearest neighbors' algorithm; Group 1: $SSC \leq 4.5$ °Brix; Group 2: $SSC > 4.5$ °Brix.

The support vector machine is a linear or non-linear classifier, which is a mathematical function that can distinguish two different kinds of objects (Raikwal & Saxena, 2012). SVM

objective is to find a hyperplane in a N-dimensional space that distinctly classifies the data points (Jaganathan, Badawy, Baili, & Changalasetty, 2016), basically, the original data set is compressed to support vector set, then new knowledge is gained by learning to use the subset and also the rules decided by support vector are given (Liu, Lv, Liu, & Shi, 2010). Maturity index models results showed the TD-NMR, associated to chemometrics tools and machine learning, is a powerful alternative to separate green tomato samples before tomato processing starts, the purpose of this models is to be a rapid, qualitative, and accurate method for use in industry to ensure the final quality of tomato products.

Description of classification models according to SSC is presented in table 2 (training) and table 3 (test). During the training, all models showed high accuracy levels (between 95-98%) and 2 - 4 misclassification samples. Even with a low SSC amplitude (3.51 - 5.80 °Brix) was possible to visualize the clustering between the two classes (Figure 5). Computed decision boundaries help to detect the clustering of the samples in the new hyperplane developed through the algorithms. Only a few intermixed samples were observed, several numbers of samples intermixed in another class may cause the decision boundary too complex and may be harmful to improve the precision of classifier in SVM and kNN models (Jaganathan et al., 2016). During tomato processing, producers search for tomatoes presenting SSC above 4.5 °Brix because high SSC content results in a large yield.

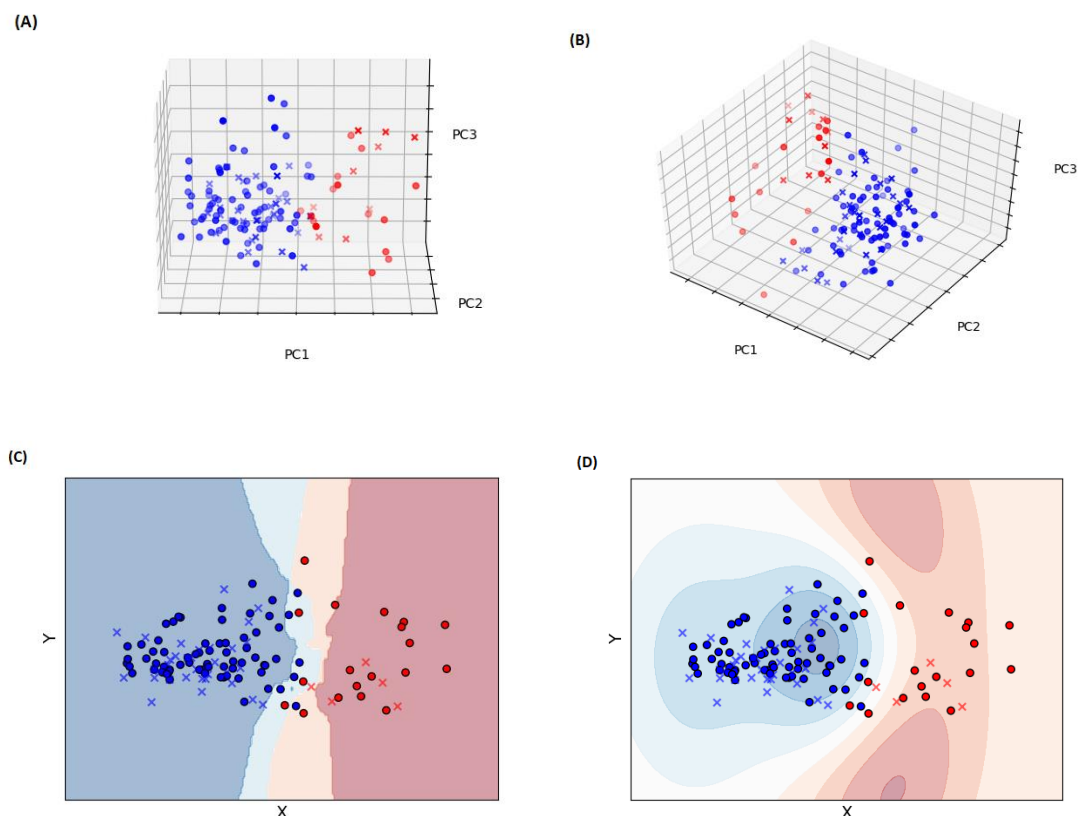


Figure 5. (A) SIMCA, (B) PLS-DA 3D projection plots of TD-NMR data for SCC of tomatoes, (C) SVM computed decision boundaries in the SVM model (rbf with $C = 10$), and (D) computed decision boundaries in the kNN model ($k=3$). Circles means calibration set and crosses validation or test set; Red symbols are group 1 and blue group 2 of SSC.

SIMCA resulted in more accurate SSC classification during the training set (accuracy = 98%) and only 2 tomato samples from group 1 were misclassified in group (table 2). In the PLS-DA the first three principal components explained 96% of the total TD-NMR signal decay variance (PC1 94.2%, PC2 3.1%, and PC3 0.6%). Training with by SVM and kNN algorithms, with 70% of the NMR signals, resulted in equal performance (accuracy = 95%). On the other hand, in the test set, SVM and kNN had better performance than SIMCA and PLS-DA (table 3). These models correctly classified all samples (accuracy = 100%).

The use of tomatoes with defects in the industry processing can directly affect the final product. The classification models built in this work showed the feasibility of TD-NMR use as a tool to select good and defected tomato fruits. The performance of each developed model, during the test set, is summarized in table 3. All algorithms resulted in good models, with the percentage

of correct answers above 80%. PLS-DA model, using 3 principal components, explained almost all the sample variation (PC1 95.4%, PC2 3.8%, and PC3 0.3%). Some samples were misclassified, mainly during the training set, it may be because some samples presented less degree of defects. So, the TD-NMR signals were similar to the defect's free tomatoes, it can be seen by the SIMCA and PLS-DA 3D projection plots and SVM and kNN decision boundaries graphics (figure 6).

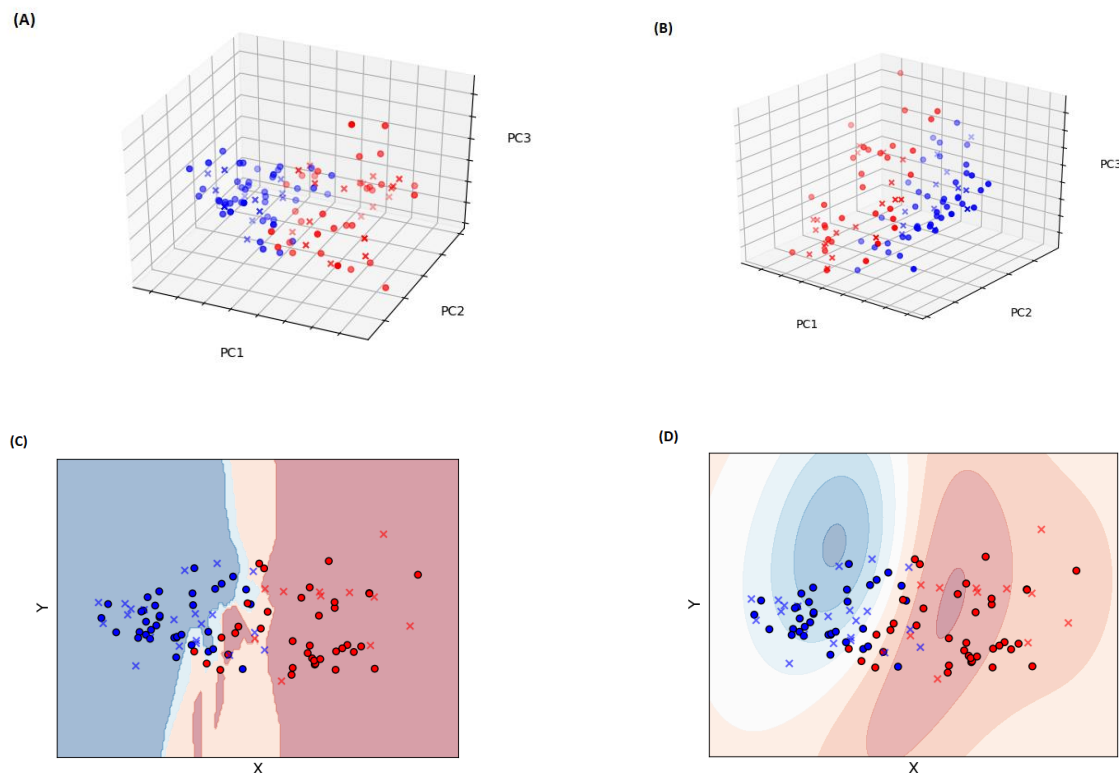


Figure 6. (a) SIMCA, (b) PLS-DA 3D projection plots of TD-NMR data for tomatoes defects, (c) SVM computed decision boundaries in the SVM model (rbf with $C = 10$), and (d) computed decision boundaries in the kNN model ($k=3$). Circles means calibration set and crosses validation or test set; Blue symbols are defected tomatoes and red tomatoes without defects.

4. Conclusion

With these results, we concluded that CPMG decays associated with chemometrics (SIMCA and PLS-DA) and machine learning algorithms (SVM and kNN) could be successfully used in the tomato processing industry to classify samples according to maturation index, SSC, and defects presence. These models showed the potential of TD-NMR technique in a high-throughput screening application before the processing.

Acknowledgments

This study was supported by CAPES institution providing the scholarship (process number 88881.187998/2018-01) and CNPq - Research Productivity Fellowship processes 309043/2016-6 and 302866/2017-5 and FAPESP grant numbers 2020/07017-0 and 2019/13656-8.

References

- Amodio, M. L., Ceglie, F., Chaudhry, M. M. A., Piazzolla, F., & Colelli, G. (2017). Potential of NIR spectroscopy for predicting internal quality and discriminating among strawberry fruits from different production systems. *Postharvest Biology and Technology*, 125, 112–121. <https://doi.org/10.1016/j.postharvbio.2016.11.013>
- Arakeri, M. P., & Lakshmana. (2016). Computer Vision Based Fruit Grading System for Quality Evaluation of Tomato in Agriculture industry. *Procedia Computer Science*. <https://doi.org/10.1016/j.procs.2016.03.055>
- Aykas, D. P., Karaman, A. D., Keser, B., & Rodriguez-Saona, L. (2020). Non-targeted authentication approach for extra virgin olive oil. *Foods*, 9(2), 1–17. <https://doi.org/10.3390/foods9020221>
- Ayvaz, H., Sierra-Cadavid, A., Aykas, D. P., Mulqueeney, B., Sullivan, S., & Rodriguez-Saona, L. E. (2016). Monitoring multicomponent quality traits in tomato juice using portable mid-infrared (MIR) spectroscopy and multivariate analysis. *Food Control*, 66(January), 79–86.

<https://doi.org/10.1016/j.foodcont.2016.01.031>

- Barrett, D. M., Garcia, E., & Miyao, G. (2006). DEFECTS AND PEELABILITY OF PROCESSING TOMATOES. *Journal of Food Processing and Preservation*, 30(1), 37–45. <https://doi.org/10.1111/j.1745-4549.2005.00045.x>
- Barrett, D. M., Garcia, E., & Wayne, J. E. (1998). Textural modification of processing tomatoes. *Critical Reviews in Food Science and Nutrition*, 38(3), 173–258. <https://doi.org/10.1080/10408699891274192>
- Bizzani, M., Flores, D. W. M., Colnago, L. A., & Ferreira, M. D. (2017). Non-invasive spectroscopic methods to estimate orange firmness, peel thickness, and total pectin content. *Microchemical Journal*, 133, 168–174. <https://doi.org/10.1016/j.microc.2017.03.039>
- Bizzani, M., William Menezes Flores, D., Bueno Moraes, T., Alberto Colnago, L., David Ferreira, M., & Helena Fillet Spoto, M. (2020). Non-invasive detection of internal flesh breakdown in intact Palmer mangoes using time-domain nuclear magnetic resonance relaxometry. *Microchemical Journal*. <https://doi.org/10.1016/j.microc.2020.105208>
- Bureau, S., Boas, A. A. de C. V., Giovinazzo, R., Jaillais, B., & Page, D. (2020). Toward the implementation of mid-infrared spectroscopy along the processing chain to improve quality of the tomato based products. *LWT*, 130, 109518. <https://doi.org/10.1016/j.lwt.2020.109518>
- Carvalho, L. C. de, Verbi Pereira, F. M., Morais, C. de L. M. de, Lima, K. M. G. de, & Teixeira, G. H. de A. (2019). Assessment of macadamia kernel quality defects by means of near infrared spectroscopy (NIRS) and nuclear magnetic resonance (NMR). *Food Control*, 106, 106695. <https://doi.org/10.1016/j.foodcont.2019.06.021>
- Colnago, L. A., Engelsberg, M., Souza, A., & Barbosa, L. (2007). High-Throughput, Non-Destructive Determination of Oil Content in Intact Seeds by Continuous Wave-Free Precession NMR, 79(3), 1271–1274.
- Dhakshina Kumar, S., Esakkirajan, S., Bama, S., & Keerthiveena, B. (2020). A microcontroller based machine vision approach for tomato grading and sorting using SVM classifier. *Microprocessors and Microsystems*, 76. <https://doi.org/10.1016/j.micpro.2020.103090>
- Dong, X., Li, Y., Li, Y., Song, L., Cheng, S., Li, D., ... Tan, M. (2017). Combination of NMR and MRI Techniques for Non-invasive Assessment of Sea Cucumber (*Stichopus japonicas*)

- Tenderization During Low-Temperature Heating Process. *Food Analytical Methods*, 10(7), 2207–2216. <https://doi.org/10.1007/s12161-016-0770-5>
- Embrapa. (2006). Sistemas de Produção. Retrieved from https://sistemasdeproducao.cnptia.embrapa.br/FontesHTML/Tomate/TomateIndustrial_2ed/index.htm
- Fan, K., & Zhang, M. (2019). Recent developments in the food quality detected by non-invasive nuclear magnetic resonance technology. *Critical Reviews in Food Science and Nutrition* (Vol. 59). <https://doi.org/10.1080/10408398.2018.1441124>
- FAO. (2020). Food and Agriculture Organization of the United Nations.
- Flores, D. W. M., Colnago, L. A., Ferreira, M. D., & Spoto, M. H. F. (2016). Prediction of Orange juice sensorial attributes from intact fruits by TD-NMR. *Microchemical Journal*, 128, 113–117. <https://doi.org/10.1016/j.microc.2016.04.009>
- Gould, W. A. (1992a). Color and Color Measurement. *Tomato Production, Processing and Technology*, 297–312. <https://doi.org/10.1533/9781845696146.3.297>
- Gould, W. A. (1992b). Tomato Pulp and Paste Manufacture. *Tomato Production, Processing and Technology*, 219–231. <https://doi.org/10.1533/B978-1-84569-599-6.50018-9>
- Hills, B. (1999). Food Science, Applications of NMR Spectroscopy. *Encyclopedia of Spectroscopy and Spectrometry*, 593–601. <https://doi.org/10.1006/rwsp.2000.0366>
- Hills, B. P., & Clark, C. J. (2003). Quality Assessment of Horticultural Products by NMR. *Annual Reports on NMR Spectroscopy*, 50, 75–120. [https://doi.org/10.1016/S0066-4103\(03\)50002-3](https://doi.org/10.1016/S0066-4103(03)50002-3)
- Ireri, D., Belal, El., Okind, C., MAkange, N., & Ji, C. (2019). A computer vision system for defect discrimination and grading in tomatoes using machine learning and image processing. *Artificial Intelligence in Agriculture*, 6, 28–37. <https://doi.org/https://doi.org/101016/j.aiia.2019.06.001>
- Ishida, N., Koizumi, M., & Kano, H. (1994). Ontogenetic changes in water in cherry tomato fruits measured by nuclear magnetic resonance imaging. *Scientia Horticulturae*, 57(4), 335–346. [https://doi.org/10.1016/0304-4238\(94\)90116-3](https://doi.org/10.1016/0304-4238(94)90116-3)
- Ishida, Nobuaki, Kobayashi, T., Koizumi, M., & Kano, H. (1989). ¹H-Nmr Imaging of Tomato

- Fruits. *Agricultural and Biological Chemistry*, 53(9), 2363–2367. <https://doi.org/10.1080/00021369.1989.10869666>
- Jaganathan, S. C. B., Badawy, A., Baili, J., & Changalasetty, S. B. (2016). A Hybrid GA/kNN/SVM Algorithm for Classification of data. *Biohouse Journal of Computer Science*, 2(2).
- Kirtil, E., Cikrikci, S., McCarthy, M. J., & Oztop, M. H. (2017). Recent advances in time domain NMR & MRI sensors and their food applications. *Current Opinion in Food Science*, 17, 9–15. <https://doi.org/10.1016/j.cofs.2017.07.005>
- Liu, Z., Lv, X., Liu, K., & Shi, S. (2010). Study on SVM compared with the other text classification methods. *2nd International Workshop on Education Technology and Computer Science, ETCS 2010, 1*, 219–222. <https://doi.org/10.1109/ETCS.2010.248>
- Mitchell, J., Chandrasekera, T. C., & Gladden, L. F. (2012). Numerical estimation of relaxation and diffusion distributions in two dimensions. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 62, 34–50. <https://doi.org/10.1016/j.pnmrs.2011.07.002>
- Monaretto, T., Souza, A., Moraes, T. B., Bertucci-Neto, V., Rondeau-Mouro, C., & Colnago, L. A. (2019). Enhancing signal-to-noise ratio and resolution in low-field NMR relaxation measurements using post-acquisition digital filters. *Magnetic Resonance in Chemistry*, 57(9), 616–625. <https://doi.org/10.1002/mrc.4806>
- Nicolai, B. M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K. I., & Lammertyn, J. (2007). Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. *Postharvest Biology and Technology*, 46(2), 99–118. <https://doi.org/10.1016/j.postharvbio.2007.06.024>
- Pasquini, C. (2018). Near infrared spectroscopy: A mature analytical technique with new perspectives – A review. *Analytica Chimica Acta*, 1026, 8–36. <https://doi.org/10.1016/j.aca.2018.04.004>
- Pereira, F. M. V., Carvalho, A. de S., Cabeça, L. F., & Colnago, L. A. (2013). Classification of intact fresh plums according to sweetness using time-domain nuclear magnetic resonance and chemometrics. *Microchemical Journal*, 108, 14–17. <https://doi.org/10.1016/j.microc.2012.12.003>

- Prestes, R. A., Colnago, L. A., Forato, L. A., Vizzotto, L., Novotny, E. H., & Carrilho, E. (2007). A rapid and automated low resolution NMR method to analyze oil quality in intact oilseeds. *Analytica Chimica Acta*, 596(2), 325–329. <https://doi.org/10.1016/j.aca.2007.06.022>
- Raffo, A., Gianferri, R., Barbieri, R., & Brosio, E. (2005). Ripening of banana fruit monitored by water relaxation and diffusion ^1H -NMR measurements. *Food Chemistry*, 89(1), 149–158. <https://doi.org/10.1016/j.foodchem.2004.02.024>
- Raikwal, J. S., & Saxena, K. (2012). Performance Evaluation of SVM and K-Nearest Neighbor Algorithm over Medical Data set. *International Journal of Computer Applications*, 50(14), 35–39. <https://doi.org/10.5120/7842-1055>
- Ribeiro, F. Z., Marconcini, L. V., de Toledo, I. B., de Vasconcellos Azeredo, R. B., Barbosa, L. L., & Colnago, L. A. (2010). Nuclear magnetic resonance water relaxation time changes in bananas during ripening: A new mechanism. *Journal of the Science of Food and Agriculture*, 90(12), 2052–2057. <https://doi.org/10.1002/jsfa.4051>
- Santos, P. M., Pereira-Filho, E. R., & Rodriguez-Saona, L. E. (2013). Rapid detection and quantification of milk adulteration using infrared microspectroscopy and chemometrics analysis. *Food Chemistry*. <https://doi.org/10.1016/j.foodchem.2012.10.024>
- Santos, Poliana M., Corrêa, C. C., Forato, L. A., Tullio, R. R., Cruz, G. M., & Colnago, L. A. (2014). A fast and non-destructive method to discriminate beef samples using TD-NMR. *Food Control*, 38(1), 204–208. <https://doi.org/10.1016/j.foodcont.2013.10.026>
- Ścibisz, I., Reich, M., Bureau, S., Gouble, B., Causse, M., Bertrand, D., & Renard, C. M. G. C. (2011). Mid-infrared spectroscopy as a tool for rapid determination of internal quality parameters in tomato. *Food Chemistry*, 125(4), 1390–1397. <https://doi.org/10.1016/j.foodchem.2010.10.012>
- Tikhonov, A. N., & Arsenin, V. Y. (1977). *Solutions of ill-posed problems*. (F. John, Ed.) (xiii). New York: Halsted Press.
- Tonguç, G., & Yakut, A. K. (2009). Fruit Grading Using Digital Image Processing Techniques. *Journal of Agricultural Machinery Science*, 5(1), 93–101.
- Tu, S. S., Choi, Y. J., McCarthy, M. J., & McCarthy, K. L. (2007). Tomato quality evaluation by peak force and NMR spin-spin relaxation time. *Postharvest Biology and Technology*, 44(2),

157–164. <https://doi.org/10.1016/j.postharvbio.2006.11.011>

USDA. (2020). Tomato grading visual aids. Retrieved December 10, 2015, from https://www.ams.usda.gov/sites/default/files/media/Tomato_Visual_Aids%5B1%5D.pdf

Wilkerson, E. D., Anthon, G. E., Barrett, D. M., Sayajon, G. F. G., Santos, A. M., & Rodriguez-Saona, L. E. (2013). Rapid assessment of quality parameters in processing tomatoes using hand-held and benchtop infrared spectrometers and multivariate analysis. *Journal of Agricultural and Food Chemistry*. <https://doi.org/10.1021/jf304968f>

Zhang, L., McCarthy, K. L., & McCarthy, M. J. (2016). Study of tomato maturity using nuclear magnetic resonance relaxometry and imaging. *Acta Horticulturae*, 1119, 313–320. <https://doi.org/10.17660/ActaHortic.2016.1119.43>

Capítulo 4.

**Rapid quality control of tomato paste based on a miniaturized NIR
spectroscopy system.**

Rapid quality control of tomato paste based on a miniaturized NIR spectroscopy system

Karla R. Borba^{a*}, Didem P. Aykas^{b, c}, Luiz A. Colnago^d, Marcos D. Ferreira^d and Luis E. Rodriguez-Saona^b.

^aDepartment of Food and Nutrition, School of Pharmaceutical Sciences, São Paulo State University-UNESP, Araraquara – Jaú, Km 1, 14801-902 - Araraquara-SP, Brazil.

^b Department of Food Science and Technology, The Ohio State University, 100 Parker Food Science and Technology Building, 2015 Fyffe Road, Columbus, Ohio – USA.

^cDepartment of Food Engineering, Faculty of Engineering, Adnan Menderes University, Aydin, Turkey.

^d Embrapa Instrumentation, XV de Novembro, 1452, 13561-206 - São Carlos-SP, Brazil.

*Corresponding author < borbakr@gmail.com >

ABSTRACT

A portable/handheld Near-infrared (NIR) spectrometer (Neospectra, Si-ware) with a spectral range of 1350 – 2500 nm, equipped with a micro-electro-mechanical system (MEMs) was used in combination with partial least square regression (PLS) method to develop simultaneous prediction models of multiple quality traits in tomato paste without sample preparation. A total of 240 samples were obtained from 4 different leading tomato paste processors in California, USA, in 2019. The reference values of natural tomato soluble solids (NTSS), pH, titratable acidity (TA), Bostwick consistency, and serum viscosity were obtained using official methods. NIR spectra were directly collected on the portable spectrometer and PLS was used to develop calibration and validation models. The models were evaluated with a randomly selected external validation set. PLS models showed high correlation (r) and low standard errors of prediction (SEP) for almost all attributes. r for NTSS, pH, Bostwick consistency, Serum viscosity, and TA were: 0.98, 0.77, 0.94, 0.93, and 0.81 and SEP were = 0.58 °Brix, 0.02, 0.06 cm, 0.09 log cSt, and 0.11 % citric, respectively. These results show that a portable NIR based on MEMs technology could be an ideal on-line and in-line routine quality control tool for tomato paste and its related product evaluation, which would provide the tomato industry with accurate results in short time and lower cost.

Keywords: Tomato paste, MEMs, NIR, PLS, infrared, quality control.

1. Introduction

Tomato (*Solanum lycopersicum*) is consumed fresh as well as in the processed tomato products such as juice, tomato puree, tomato sauce, ketchup, and tomato paste (1,2). Tomato paste is the majority portion of consumption, in volume and commercial value (1,3), and its widely acceptance around the world is becoming high due to beneficial health effects (4). In 2018/2019 the global exports of tomato paste generated more than 2.93 billion dollars of the 6 billion generated by the global trade of tomato products (5). The United States are one of the major tomato processing producers (3).

Tomato paste is generally used as an intermediate, or base, product with water and other ingredients to be reconstituted into several final tomato products (6). The current competitive scenario in processing foods has been required for the sector to produce high quality products, for this reason, the research, and development of fast, accurate and non-destructive tools that are capable of evaluating products in-line is important and can be a great differential (7,8). Paste quality can be influenced by factors such as cultivar, maturation, and processing conditions (9). Wherefore, it is essential to evaluate and maintain quality during tomato paste processing. Tomato rheological properties (determining whether they are more or less viscous), color (preferred as deep red as possible), and still their taste and aroma are crucial quality attributes of tomato products to be ensured to reach the acceptance standards (1,10). Soluble solids (SS), pH, acidity, consistency, viscosity, and color are important tomato quality parameters to control during tomato processing chain (6,11).

Several methods had been developed and used to ensure tomato paste quality during industrial processing, most of them show disadvantages such as time-consuming, sample preparation, many steps in the analyzes, hazardous solvents required, expensive equipment, and high level of training for employees, and also their applications are limited to laboratory environments (12). Near infrared spectroscopy (NIR) is an alternative method that can determine several quality parameters, simultaneously, with no sample preparation, no complex training staff required to operate the spectrometer, no laboratory environment need, and results achieving are fast (5 seconds). The spectral information measured in the NIR (800 – 2500 nm) region originates from the excitations of higher quanta transitions, mostly overtones and binary combinations (13–16).

The spectrometer miniaturization brings even more advantages and has revolutionized the food industry by allowing on-site, in-line and real-time monitoring of food products and production processes to ensure quality and safety by using a palm-size equipment with low-cost and with spectral performances comparable to their benchtop laboratory counterparts (13,17,18). Spectrometer miniaturization had become possible due to the successfully widespread development in micro-electromechanical systems (MEMS) production, optical components, wavelength selectors and detectors, and solid-state lasers (13). MEMS combines mechanical parts, sensors, actuators, and electronics components on the same substrate that are developed with inherited technology from the semiconductor microfabrication (19). MEMS chips are etched using batch processes, that allows a fast and low-cost production. All these technologies advances enable that bulky Michelson interferometer with discrete optical, mechanical, and electrical components were exchanged for single-chip Michelson interferometer with monolithic opto-electro-mechanical structure (Si-Ware Systems, 2019).

Portable infrared has been evaluated and applied for quality control of agricultural products, Baqueta et al. (2020) evaluated the use of a portable near-infrared (NIR) spectroscopy as a fast alternative of conventional methods, suitable for in-line use for quality control in the industrial coffee processing. The results showed predictive capacity and high sensitivity in determining granulometry, moisture content and infusion time. Xiao et al (2017) found in a comparison between a benchtop and a portable NIR spectrometer to determine SSC in grape berry, that the portable NIR performed better. A different portable NIR was used, in parallel to a benchtop device, to evaluate acerola fruit quality (23), the miniaturized spectrometer demonstrated to be a robust device for the application of spectroscopy in field, even though the reduced number of wavelengths.

Therefore, the aim of this study was to develop robust and accurate prediction multivariate models using a portable (palm-size) NIR spectrometer data to ensure quality on tomato paste process.

2. Materials and Methods

Tomato paste samples

A total of 240 tomato paste samples were provided by four different major tomato processors in California, USA in 2019 harvest season between June-September.

Reference Analyses

The reference values for natural tomato soluble solids (NTSS, expressed in ° Brix), pH, titratable acidity (% citric acid), Bostwick consistency (cm), and serum viscosity (cSt), were measured, by official methods, and provided by the quality assurance department of each tomato processor company.

NIR infrared Analysis

The near infrared analysis was carried out using a portable spectrometer (NeoSpectra – Micro, Sci-ware Systems, Cairo, Egypt) based on Fourier Transform Infrared (FTIR) and MEMs technology. Spectra were collected in the 1300 – 2500 nm range, 16 nm resolution, and 5 seconds scan. Tomato paste sample was directly applied to the detector area and covered with a metal device trying to avoid the light scattering. The spectral collection was done in triplicate for each sample, and it was carried out in California at each tomato processor's quality assurance laboratory, right after the paste production, to minimize any change in the reference values.

PLS models

To build the tomato paste quality prediction models, the chemometric tool partial least squares regression (PLS) were used in Pirouette software (Infometrix, Bothell, WA, USA). PLS models were built by using the reference values as dependent variables (matrix Y) and the NIR spectral data as independent variables (matrix X). Two sets were, randomly, split to perform the calibration and external validation (or prediction) steps, with 70% and 30% of the total data, respectively. In the calibration step cross-validation, by using leave-one-out algorithm, was applied to do an internal validation. Using this algorithm each spectrum is dropped from the data set, and the remaining spectra are used for prediction of left-out data point until every spectrum is left-out once.

Additionally, statistical results were evaluated using models performance such as the correlation coefficient of cross validation (r_v), the correlation coefficient of prediction (r_p), the standard error of cross validation (SECV), and the standard error of prediction (SEP) that shows the magnitude of error expected when the concentration of the independent/unknown samples are predicted, and number of factors.

3. Results and Discussion

3.1 Reference values in tomato paste samples

The ranges of reference compositional values for quality traits, available in table 1, obtained in this study were similar to those reported in the literature (6,24–28). Soluble solids content (SSC), commonly called Natural Tomato Soluble Solids (NTSS), are a quality and yield indicator and it is economical essential in tomato processing (11). Some companies are using a price-increase according to SSC to encourage the incoming of high SSC tomatoes in the factory (29), because higher levels of NTSS mean that less amount of tomato is required to concentrate the paste and the evaporation time is also minor. According to USDA regulamentation, tomato paste NTSS should not be less than 24 °Brix and all tested tomato paste samples were above that limit, ranging from 25.5 to 38 °Brix (**table 1**). Typically, the finished paste has a final NTSS of 28 – 31°Brix (30). Therefore, all samples were within the recommended levels.

Moreover, soluble solids jointly with pH and acidity parameters are associated to tomato paste flavor (29,31). Low acidic levels of tomatoes (pH 4,4 - 4.6) are important during processing, because decrease thermal treatments to improve microbiological food safety, avoiding bacterial growth and spore dissemination (32). Industrial tomato processors in California targets a range of pH 4.30 ± 0.30 in order to ensure a safety margin (32), the pH levels of tomato paste found in this research was from 4.19 to 4.48 (**table 1**), in accepted values regarding to prevent microbiology issues.

Consistency and viscosity are related to texture and consequently to consumer acceptance of final product (11). The particularity of each one is that consistency is associated to insoluble solids (pulp) in the tomato products and viscosity is the result of soluble solids, mainly pectins that is one of the most significant viscosity defining compounds in tomato products (33). The range of Bostwick consistency (centimeters) found was larger, 1.00 – 7.70 (**table 1**), than results reported by Barret and others (1998) (3.10 – 5.50 cm). The critical processing steps, that influence consistency and viscosity, are breaking temperature (i.e. temperature at which fruits are crushed and initially heated) and progressive juice concentration influenced by thermal treatment under vacuum (29), therefore it is an important parameter to control during processing in order to adjust or modify processing parameters to reach the desirable value.

The viscosity of tomato products is related to different compositional characteristics such as juice apparent viscosity, serum viscosity and pulp content, defined during the concentration and reconstitution processes where pectin is irreversibly damaged (35). In this study the values of serum viscosity were between 1.83 and 2.89 log cSt (**table 1**), the viscosity of the serum, the soluble fraction of the tomato juice after removal of insoluble material by centrifugation, can be determined using a Cannon-Fenske-type viscometer. This serum viscosity is the result of the solutes present, particularly the polymeric material, which in tomato juice consists mostly of pectin.

Table 1. Tomato paste quality parameters values.

Quality parameter	Range	Mean	SD
NTSS (°Brix)	25.0 - 38	29.87	2.8
pH	4.19 - 4.48	4.38	0.04
Bostwick Consistency (cm)	1.00 - 7.70	3.27	1.53
Serum Viscosity (log cSt)	1.83 - 2.89	2.42	0.24
Total acids	1.03 - 2.27	1.58	0.2

SD: standard derivation; NTSS: Natural Tomato Soluble solids.

3.2 NIR spectrum

Figure 1 shows the original spectrum of absorbance, collected by the NeoSpectra-micro portable spectrometer, for a representative tomato paste sample, spectrum ranges covered first and second overtones and combination band region. The problem with most of portable devices is that they have a short range and, generally, it does not include important overtones regions. Variables from 1893 to 1977 nm were excluded during the model's development because in this range samples had been saturated the detector; concentration was higher than the detection limit.

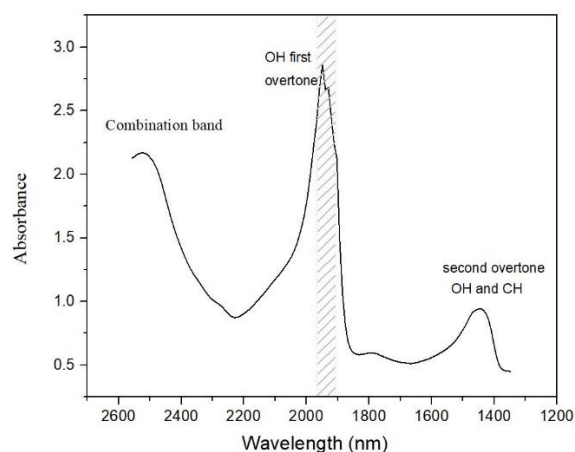


Figure 1. Representative raw near infrared absorption spectrum of tomato paste samples in wavelength range 1350 – 2500 nm, obtained using NeoSpectra – Micro (Si-Ware Systems).

Before to build PLS models based on NIR spectra data are often preprocessed, because it can contain systematic variation that is unrelated to the response y (reference values) (36). Different pre-processes were used to correct systematic and other problems such as baseline shifts, light scattering, and to highlight some bands in the overtones. The best pre-processes selected are in **table 1**, as well as the performance results of each quality parameter PLS model. Orthogonal signal correction (OSC) were well applied in all models, this consists in a mathematic filter that removes from the spectra matrix only the part that definitely is unrelated to Y (36,37). In other words, OSC removes components in the X block that are orthogonal to the

Y block, so the OSC-corrected X block should then produce models more parsimonious and/ result in lower predictions errors (Infometrix, Bothell, WA, USA).

PLS models

The differences in the total number of samples among the quality parameters are due to the exclusion of samples with high leverage and/or studentized residual, that were identified as an outlier, as well as unpattern spectra samples previously detected. As can be seen in tables 2 and 3, the variation in the range between calibration and validation sets were very similar, respectively. All quality traits PLS models resulted in correlation coefficients values up to 0.80, except for pH ($r_p = 0.77$) (Table 3). In general, a model with a higher correlation coefficient and lower errors represents a model with a better prediction accuracy (11). Even with the lowest r_p , pH model resulted really low errors (SEC = 0.02) (Table 2), when compared with other studies in the literature our pH errors are lower (11). For NTSS attribute, PLS model was built using a total of 236 samples, and otherwise, results showed a high correlation between spectral signals and reference value. With high correlation coefficients (0.98), both in calibration and validation set (**tables 2 and 3**, respectively).

Table 2. Statistical performances of the tomato quality PLSR models developed, during the calibration step.

Quality parameter	preprocessing	Range	N	F	SEC	r_{cal}	SECV	r_v
NTSS (°Brix)	Norm + SD (25) + OSC (1)	25.5 - 38	171	5	0.53	0.98	0.54	0.9
pH	SNV + OSC (1)	4.19 - 4.48	142	3	0.02	0.88	0.02	0.8
Bostwick (cm)	Norm + SD (25) + OSC (1)	1.00 - 7.70	150	4	0.53	0.92	0.54	0.9
Serum Viscosity (log cSt)	SNV + OSC (1)	1.83 - 2.89	135	8	0.07	0.95	0.07	0.9
Titrateable Acidity (% Citric)	SNV + OSC (1)	1.24 - 2.27	160	4	0.09	0.88	0.09	0.8

N: number of samples; F: factors; SEC: standard error of calibration; SECV: standard error of cross validation; r_{cal} : correlation coefficient of calibration; r_{val} : correlation coefficient of cross validation; NTSS: Natural Tomato Soluble solids; Norm: normalization (0-1); SD: second derivative; OSC: Orthogonal Signal Correction; SNV: Standard Normal variate.

Texture properties, Bostwick consistency and serum viscosity (crucial parameters to guarantee the consumer acceptance), exhibiting good predictive ability ($r_p \geq 0.93$). Bostwick consistency models showed best statistical performance values ($r_p = 0.94$) (**Table 3**) than those found by Zhang and others (2014) using guided microwave spectroscopy for in-line measurement of tomato paste properties ($r_{cv} = 0.88$). Serum viscosity, $r_p = 0.93$ and SEP = 0.09, showed best accuracy PLS model than that already been found in similar studies (11) using infrared spectroscopy.

Table 3. Statistical performances of the PLSR models developed in validation step.

Quality parameter	Range	N	SEP	r_p
NTSS (°Brix)	25.0 - 37.7	65	0.58	0.98
pH	4.25 - 4.45	62	0.02	0.77
Bostwick Consistency (cm)	1.00 - 7.70	57	0.6	0.94
Serum Viscosity (log cSt)	2.02 - 2.89	45	0.09	0.93
Titrateable Acidity (% Citric)	1.30 - 2.10	59	0.11	0.81

N: number of samples; SEP: standard error of prediction (external validation); r_p : correlation coefficient of prediction; NTSS: Natural Total Soluble solids.

Figure 2 shows the correlation plots between reference values and predicted NIR portable spectroscopy levels of different quality attributes during calibration and validation steps. The solid line is the regression line standing for the ideal correlation between the predicted and measured values. Dispersion of points near to the calibration and validation regression lines (white and black squares, respectively) are important and therefore, shows an excellent assessment performance of PLS models (38).

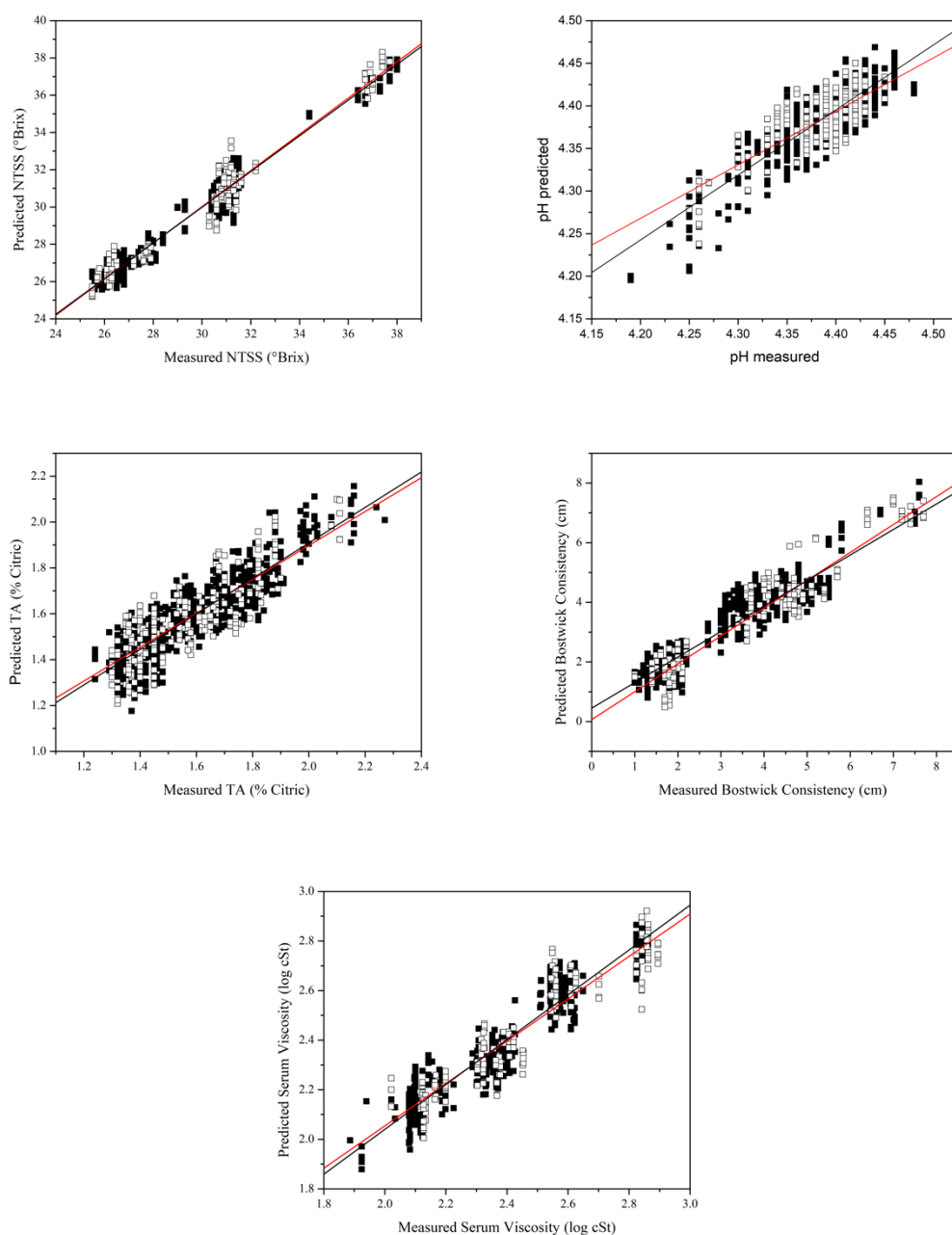


Figure 2. PLSR correlation plots for different quality traits in tomato paste using portable NIR unit equipped (white and black squares represent calibration and validation set samples, respectively; black regression line: calibration set and red regression line: external validation set).

Figure 3 shows loadings from each PLS model, the use of loadings plots enabled the identification of spectral portions that are most important for prediction (12). Wherefore loadings are generally used to identify the specific areas of the infrared spectra that were related with highest variation in the calibration set which areas of the IR spectra were more related to sample variation (11,12). The region around 2000 nm is the characteristic region of OH absorption (combination bands) (16,39), associated with water and organic acids. NTSS, TA, Bostwick, and serum viscosity loading vectors shows these bands as important

spectra regions. Furthermore, spectra variables around 1800 nm were also identified as important for prediction of NTSS, (Figure 3 a, c, d and e) it's the C-H third overtone one important region to predict carbohydrates (sugars) (39). Another important band is localized around 1470, the O-H stretch region.

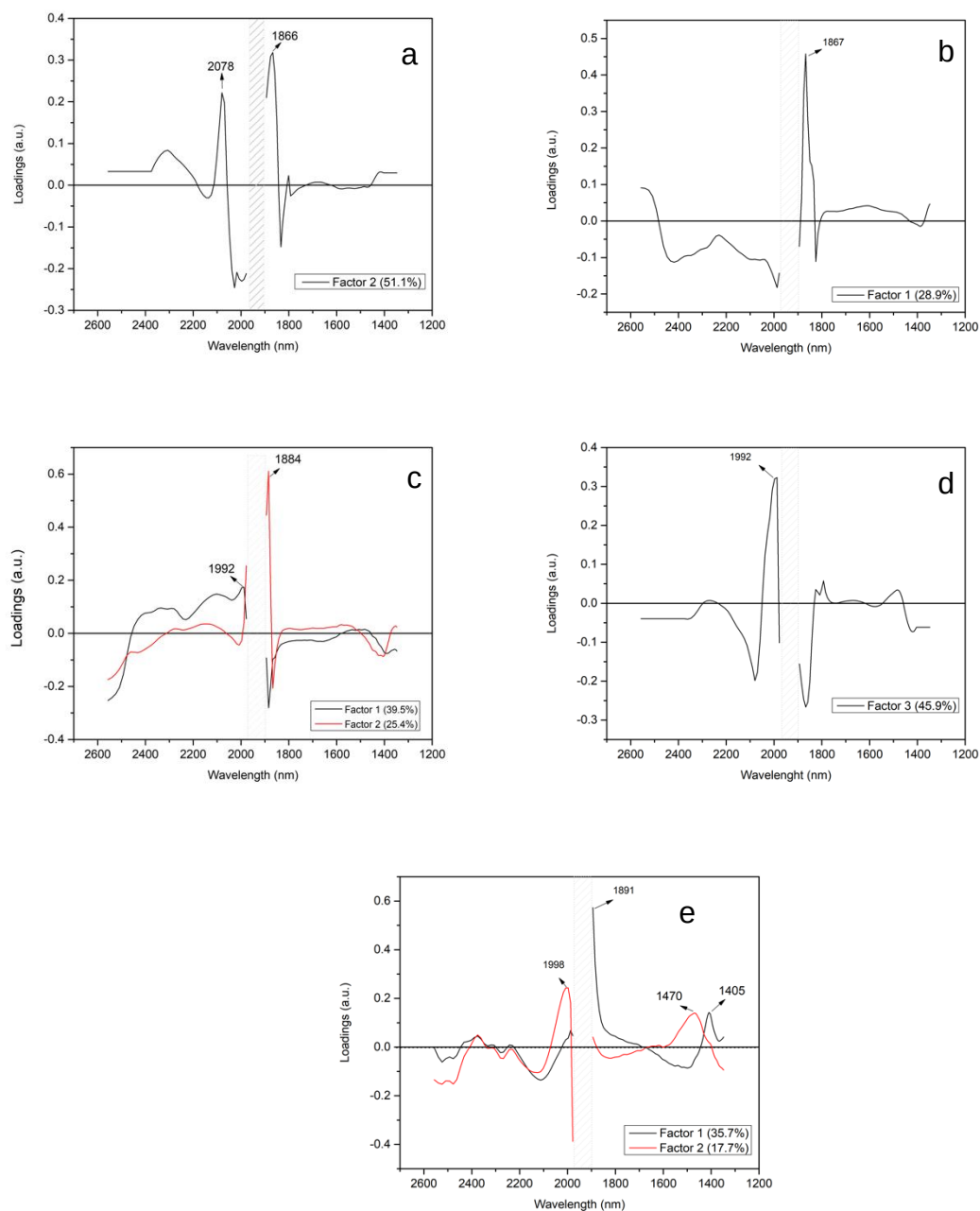


Figure 3: PLS loadings vectors for NTSS (a), pH (b), TA (c), Bostwick(d), and serum viscosity(e). Variables between 1893 to 1777 nm were excluded during the models development.

4. Conclusion

Our results showed prediction models with strong correlation between spectral data and the reference values ($r_p \geq 0.93$; except for pH, 0.77, and TA, 0.81) and very low errors value for each quality parameter. The developed calibration models were validated using an external validation set, and the robustness of the models was confirmed by obtaining similar SECV and SEP values. These findings show that the combination of a portable near infrared sensor with PLS modeling is a great alternative method to simultaneous prediction of multiple quality traits in tomato paste with no sample preparation. The implementation of Portable NIR based on MEMs tools throughout the production regarding quality targets could be an ideal tool for routine in-plant assessment of the tomato paste quality and its related products, which would provide the tomato industry with accurate results in short time and low cost.

Acknowledgements

This study was supported by CAPES institution providing the scholarship (process number 88881.187998/2018-01), CNPq - Research Productivity Fellowship processes 309043/2016-6 and 302866/2017-5 and the authors would like to thank the California League of Food Processors for their financial support through a research grant (Grant number: TR 4-19 GRT00056473) and for providing tomato paste samples.

References

1. Bureau S, Boas AA de CV, Giovinnazzo R, Jaillais B, Page D. Toward the implementation of mid-infrared spectroscopy along the processing chain to improve quality of the tomato based products. LWT [Internet]. 2020 Aug;130:109518. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0023643820305077>
2. Liu C, Hao G, Su M, Chen Y, Zheng L. Potential of multispectral imaging combined with chemometric methods for rapid detection of sucrose adulteration in tomato paste. J Food Compos Anal. 2017;215:78–83.
3. Mashad HM El, Zhao L, Zhang R, Pan Z. Tomato. In: Integrated Processing Technologies for Food and Agricultural By-Products [Internet]. 1st ed. Elsevier; 2019. p. 452. Available from: <https://linkinghub.elsevier.com/retrieve/pii/C20170009011>
4. Burton-freeman B, Talbot J, Park E, Krishnankutty S, Edirisinghe I. Protective activity of processed tomato products on postprandial oxidation and inflammation: A clinical trial in healthy weight men and women. 2012;622–31.
5. The Tomato News Conference. Tomato News [Internet]. 2019. Available from: http://www.tomatonews.com/en/the-tomato-news-conference_10_49.html
6. Zhang L, Schultz MA, Cash R, Barrett DM, McCarthy MJ. Determination of quality parameters of tomato paste using guided microwave spectroscopy. Vol. 40, Food Control. 2014. p. 214–23.

7. Cortés V, Blasco J, Aleixos N, Cubero S, Talens P. Monitoring strategies for quality control of agricultural products using visible and near-infrared spectroscopy: A review. *Trends Food Sci Technol* [Internet]. 2019 Mar;85(July 2018):138–48. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0924224418304825>
8. Porep JU, Kammerer DR, Carle R. On-line application of near infrared (NIR) spectroscopy in food production. *Trends Food Sci Technol* [Internet]. 2015 Dec;46(2):211–30. Available from: <https://doi.org/10.1016/j.tifs.2015.10.002>
9. Shobo BA, Bodunde JG, Makinde EA, Olowe VIO, Ogunleye MT. Paste Quality of Tomato (*Lycopersicon esculentum* Mill) as Influenced by Variety and Organic Fertilizer. *J Org Agric Enviroment*. 2018;6.
10. Berta M, Wiklund J, Kotz R. Correlation between in-line measurements of tomato ketchup shear viscosity and extensional viscosity. 2016;173:8–14.
11. Ayvaz H, Sierra-Cadavid A, Aykas DP, Mulqueeney B, Sullivan S, Rodriguez-Saona LE. Monitoring multicomponent quality traits in tomato juice using portable mid-infrared (MIR) spectroscopy and multivariate analysis. *Food Control* [Internet]. 2016 Aug;66(January):79–86. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0956713516300329>
12. Wilkerson ED, Anthon GE, Barrett DM, Sayajon GFG, Santos AM, Rodriguez-Saona LE. Rapid assessment of quality parameters in processing tomatoes using hand-held and benchtop infrared spectrometers and multivariate analysis. *J Agric Food Chem*. 2013;
13. Rodriguez-saona LE, Aykas DP, Borba KR, Urbina AU. Miniaturization of Optical Sensors and their Potential for High-Throughput Screening of Foods. *Curr Opin Food Sci* [Internet]. 2020; Available from: <https://doi.org/10.1016/j.matdes.2020.108709>
14. Pasquini C. Near infrared spectroscopy: A mature analytical technique with new perspectives – A review. *Anal Chim Acta* [Internet]. 2018;1026:8–36. Available from: <https://doi.org/10.1016/j.aca.2018.04.004>
15. Sánchez M-T, De la Haba M-J, Serrano I, Pérez-Marín D. Application of NIRS for Nondestructive Measurement of Quality Parameters in Intact Oranges During On-Tree Ripening and at Harvest. *Food Anal Methods*. 2012 Aug 17;6(3):826–37.
16. Nicolai BM, Beullens K, Bobelyn E, Peirs A, Saeys W, Theron KI, et al. Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. *Postharvest Biol Technol*. 2007 Nov;46(2):99–118.
17. Sánchez MT, De La Haba MJ, Benítez-López M, Fernández-Novales J, Garrido-Varo A, Pérez-Marín D. Non-destructive characterization and quality control of intact strawberries based on NIR spectral data. *J Food Eng*. 2012;110(1):102–8.
18. Sorak D, Herberholz L, Iwascek S, Altinpinar S, Pfeifer F, Siesler HW. New developments and applications of handheld raman, mid-infrared, and near-infrared spectrometers. *Appl Spectrosc Rev*. 2012;47(2):83–115.
19. Crocombe RA. Portable Spectroscopy. *Appl Spectrosc*. 2018;72(12):1701–51.
20. Si-Ware S. The “Neo” way for FT-IR [Internet]. 2019. Available from: <https://www.neospectra.com/how-it-works/mems-ft-ir/>
21. Baqueta MR, Coqueiro A, Março PH, Valderrama P. Quality Control Parameters in the Roasted Coffee Industry: a Proposal by Using MicroNIR Spectroscopy and Multivariate Calibration. *Food Anal Methods*. 2020;13(1):50–60.

22. Xiao H, Sun K, Sun Y, Wei K, Tu K, Pan L. Comparison of benchtop fourier-transform (FT) and portable grating scanning spectrometers for determination of total soluble solid contents in single grape berry (*Vitis vinifera* L.) and calibration transfer. *Sensors (Switzerland)*. 2017;17(11):1–14.
23. Malegori C, José E, Marques N, Tonetto S, Freitas D, Pimentel F, et al. Talanta Comparing the analytical performances of Micro-NIR and FT-NIR spectrometers in the evaluation of acerola fruit quality, using PLS and SVM regression algorithms. 2017;165(December 2016):112–6.
24. Capanoglu E, Beekwilder J, Boyacioglu D, Hall R, De Vos R. Changes in antioxidant and metabolite profiles during production of tomato paste. *J Agric Food Chem*. 2008;56(3):964–73.
25. Carmen Sánchez M, Valencia C, Gallegos C, Ciruelos A, Latorre A. Influence of processing on the rheological properties of tomato paste. *J Sci Food Agric*. 2002;82(9):990–7.
26. Deda MS, Bloukas JG, Fista GA. Effect of tomato paste and nitrite level on processing and quality characteristics of frankfurters. *Meat Sci*. 2007;76(3):501–8.
27. Kadakal Ç, Artik N. Degradation kinetics of ergosterol in tomato paste serum. *Eur Food Res Technol*. 2008;227(3):683–8.
28. Shatta AAB. Impact of Processing Steps on Physicochemical and Rheological Properties of Tomato Paste (Cold-Break). *MOJ Food Process Technol*. 2017;5(2):263–71.
29. Bureau S, Giovinazzo R, Page D. Quality improvement of the tomato-based products using infrared tool along the production chain. 2019;169–74.
30. Anthon GE, Barrett DM. Changes in tomato paste during storage and the effects of heating on consistency of reconstituted tomato paste. *J Texture Stud*. 2010;41(3):262–78.
31. Zhang L, Schultz MA, Cash R, Barrett DM, McCarthy MJ. Determination of quality parameters of tomato paste using guided microwave spectroscopy. *Food Control [Internet]*. 2014;40(1):214–23. Available from: <http://dx.doi.org/10.1016/j.foodcont.2013.12.008>
32. Anthon GE, Barrett DM. Pectin methylesterase activity and other factors affecting pH and titratable acidity in processing tomatoes. *Food Chem [Internet]*. 2012;132(2):915–20. Available from: <http://dx.doi.org/10.1016/j.foodchem.2011.11.066>
33. Anthon GE, Diaz J V., Barret DM. Changes in Pectins and Product Consistency during the Concentration of Tomato Juice to Paste. *J Agric Food Chem*. 2008;56:7100–5.
34. Barrett DM, Garcia E, Wayne JE. Textural modification of processing tomatoes. *Crit Rev Food Sci Nutr*. 1998;38(3):173–258.
35. Beresovsky N, Kopelman IJ, Mizrahi S. THE ROLE OF PULP INTERPARTICLE INTERACTION IN DETERMINING TOMATO JUICE VISCOSITY. *J Food Process Preserv*. 1995;19:133–46.
36. Wold S, Antti H, Lindgren F, Öhman J. Orthogonal signal correction of near-infrared spectra. *Chemom Intell Lab Syst*. 1998;44(1–2):175–85.
37. Harrington PDB, Kister J, Artaud J, Dupuy N, Ce P. Automated Principal Component-Based Orthogonal Signal Correction Applied to Fused Near Infrared - Mid-Infrared Spectra of French Olive Oils. *Anal Chem*. 2009;81(17):7160–9.
38. Baranska M, Schütze W, Schulz H. Determination of lycopene and β -carotene content in tomato fruits and related products: Comparison of FT-Raman, ATR-IR, and NIR spectroscopy. *Anal Chem*. 2006;
39. Ribeiro JS, Ferreira MMC, Salva TJG. Chemometric models for the quantitative descriptive sensory analysis of Arabica coffee beverages using near infrared spectroscopy. *Talanta [Internet]*.

2011;83(5):1352–8. Available from: <http://dx.doi.org/10.1016/j.talanta.2010.11.001>

Considerações Finais

Este estudo resultou na comprovação de que as técnicas espectroscópicas de infravermelho próximo, médio e de ressonância magnética nuclear juntamente com ferramentas quimiométricas e algoritmos de aprendizagem de máquina, podem ser uma forma alternativa e rápida para analisar parâmetros de qualidade em tomates e seus produtos derivados.

Os espectrômetros NIR portáteis, NeoSpectra e NeoSpectra – Micro, utilizados nos capítulos 1 e 4 resultaram em performances similares ou melhores quando comparados com equipamentos de bancada. Trazendo vantagens como valor muito mais baixo e acessível a pequenas empresas e produtores, análises rápidas com alta acurácia e avaliação de diversos atributos de qualidade simultaneamente.

A aplicação de RMN-DT na classificação de tomates de acordo com seu índice de maturação, teor de sólidos solúveis e firmeza representará um grande ganho de valor agregado na indústria de processamento de tomates. Pois com um melhor controle da sua matéria prima esse setor pode obter produtos finais de melhor qualidade e diminuir gastos durante a produção.

Portanto, fica evidente que a implementação de métodos espectroscópicos durante toda a cadeia de produção e processamento de tomates pode trazer grandes avanços e melhorias no setor agroindustrial e alimentício.

Referências

1. Bai Y, Lindhout P. Domestication and breeding of tomatoes: what have we gained and what can we gain in the future? *Ann Bot.* 2007;100(5):1085-94. doi: 10.1093/aob/mcm150.
2. Pedro AM, Ferreira MM. Nondestructive determination of solids and carotenoids in tomato products by near-infrared spectroscopy and multivariate calibration. *Anal Chem.* 2005;77(8):2505-11. doi: 10.1021/ac048651r.
3. Chitarra MIF, Chitarra AB. Pós-colheita de frutas e hortaliças : fisiologia e manuseio. 2. ed. rev. amp. Lavras: Universidade Federal de Lavras, 2005.
4. Kader AA. Quality and Safety Factors: Definition and Evaluation for Fresh Horticultural Crops. In: Kader AA. editor. *Postharvest Technology of Horticultural Crops*. 3rd. ed. California: University of California, Division of Agriculture and Natural Resources, 2002. p. 535.
5. Ajayi AA, Olasehinde IG. Studies on the pH and protein content of tomato (*Lycopersicon esculentum* Mill) fruits deteriorated by *Aspergillus niger*. *Sci Res Essays.* 2009;4(3):185–7.
6. Macleod RF, Kader AA, Morris LL. Damage to fresh tomatoes can be reduced. *California Agriculture* 1976;30(12):10-12.
7. Monteiro CS, Balbi ME, Miguel OG, Penteado PTP da S, Haracemiv SMC. Qualidade nutricional e antioxidante do tomate “tipo italiano”. *Alimentos e Nutrição.* 2008;19(1):25-31.
8. FAO. Food and Agriculture Organization of the United Nations.

- [Internet]. 2020. Acessado em 28/08/2020.
<http://www.fao.org/faostat/en/#data/QC>.
9. Embrapa Hortaliças. Cultivo de Tomate para Industrialização. [Internet]. 2. ed. Brasília, DF: Embrapa; 2006. (Sistemas de Produção, 1) Available from:
https://sistemasdeproducao.cnptia.embrapa.br/FontesHTML/Tomate/TomateIndustrial_2ed/index.htm.
 10. Bai Y, Lindhout P. Domestication and breeding of tomatoes: What have we gained and what can we gain in the future? *Ann Bot*. 2007;100(5):1085–94.
 11. Ibáñez G, Valcárcel M, Cebolla-Cornejo J, Roselló S. FT-MIR determination of taste-related compounds in tomato: a high throughput phenotyping analysis for selection programs. *J Sci Food Agric*. 2019;99(11):5140–8.
 12. Ayvaz H, Sierra-Cadavid A, Aykas DP, Mulqueeney B, Sullivan S, Rodriguez-Saona LE. Monitoring multicomponent quality traits in tomato juice using portable mid-infrared (MIR) spectroscopy and multivariate analysis. *Food Control* [Internet]. 2016 Aug;66(January):79–86. Available from:
<https://linkinghub.elsevier.com/retrieve/pii/S0956713516300329>
 13. Gould WA. Color and Color Measurement. *Tomato Prod Process Technol* [Internet]. 1992;297–312. Available from:
<http://linkinghub.elsevier.com/retrieve/pii/B9781845695996500256>
 14. Agarwal S, Rao A V. Tomato lycopene and its role in human health and

- chronic diseases. CMAJ [Internet]. 2000;163(6):739–44. Available from: <http://www.cmaj.ca/content/163/6/739.abstract>
15. Fabrbri ADT. Estudo da radiação ionizante em tomates in natura (*Lycopersicum esculentum* mill) e no teor de licopeno do molho. [Dissertação]. São Paulo, Instituto de Pesquisas Energéticas e Nucleares, Tecnologia Nuclear - Aplicações. USP; 2009.
 16. Berta M, Wiklund J, Kotz R. Correlation between in-line measurements of tomato ketchup shear viscosity and extensional viscosity. 2016;173:8–14.
 17. Ishida N, Koizumi M, Kano H. Ontogenetic changes in water in cherry tomato fruits measured by nuclear magnetic resonance imaging. Sci Hortic (Amsterdam). 1994;57(4):335–46.
 18. Bureau S, Arbex de Castro Vilas Boas A, Giovinnazzo R, Jaillais B, Page D. Toward the implementation of mid-infrared spectroscopy along the processing chain to improve quality of the tomato based products. LWT [Internet]. 2020 Aug;130:109518. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0023643820305077>.
 19. Gliordano LB de, Ribeiro CS. Origem. Botânica e Composição Química do fruto. Tomate para processamento industrial. 2000;12–7.
 20. Pedro AMK, Ferreira MMC. Simultaneously calibrating solids , sugars and acidity of tomato products using PLS2 and NIR spectroscopy. 2007;595:221–7.
 21. Tavares CA, Rodriguez-Amaya DB. Carotenoid Composition of Brazilian Tomatoes and Tomato Products. LWT - Food Sci Technol.

- 1993;27(3):219-24.
22. AOAC. Official Methods of analysis of the Association of Analytical Chemists International. 14th. ed. Washington, DC: AOAC; 1992.
 23. Nicolai BM, Defraeye T, De Ketelaere B, Herremans E, Hertog MLATM, Saeys W, et al. Nondestructive measurement of fruit and vegetable quality. Vol. 5, Annual review of food science and technology. 2014; v. 5:285-312. <https://doi.org/10.1146/annurev-food-030713-092410>.
 24. Sun D-W, Schulz H, Baranska M. Fruits and vegetables. In: Sun Da-Wen, editor. Infrared Spectroscopy for Food Quality Analysis and Control. 2009. Chapter 12:321-53. <https://doi.org/10.1016/B978-0-12-374136-3.00012-2>.
 25. Sun D-W, Huang Y, Rasco BA, Cavinato AG. Infrared Spectroscopy for Food Quality Analysis and Control. Infrared Spectrosc Food Qual Anal Control. 2009;355–75.
 26. Huang H, Yu H, Xu H, Ying Y. Near infrared spectroscopy for on/in-line monitoring of quality in foods and beverages: A review. J Food Eng. 2008;87(3):303–13.
 27. Colnago LA, Engelsberg M, Souza A, Barbosa L. High-Throughput, Non-Destructive Determination of Oil Content in Intact Seeds by Continuous Wave-Free Precession NMR. 2007;79(3):1271–4.
 28. Pereira FMV, Carvalho A de S, Cabeça LF, Colnago LA. Classification of intact fresh plums according to sweetness using time-domain nuclear magnetic resonance and chemometrics. Microchem J [Internet]. 2013 May;108:14–7.

Available

from:

<http://dx.doi.org/10.1016/j.microc.2012.12.003>

29. Dufour É. Principles of Infrared Spectroscopy. In: Sun Da-Wen. editor. Infrared Spectroscopy for Food Quality Analysis and Control. San Diego, CA: Academic Press; 2009;3-27. <https://doi.org/10.1016/B978-0-12-374136-3.00001-8>
30. Skoog DA, Holler FJ, Nieman TA. Princípios de Análise Instrumental, 5a. ed. São Paulo: Bookman; 2002.
31. Subramanian A, Prabhakar V, Rodriguez-Saona L. Analytical Methods: Infrared Spectroscopy in Dairy Analysis. Reference Module in Food Science. Elsevier; 2016. ISBN 9780081005965. Available from: <http://www.sciencedirect.com/science/article/pii/B9780081005965006028>.
32. Bellon-Maurel V, McBratney A. Near-infrared (NIR) and mid-infrared (MIR) spectroscopic techniques for assessing the amount of carbon stock in soils – Critical review and research perspectives. Soil Biol Biochem. 2011 Jul;43(7):1398–410.
33. Nicolai BM, Beullens K, Bobelyn E, Peirs A, Saeys W, Theron KI, et al. Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. Postharvest Biol Technol. 2007 Nov;46(2):99–118.
34. Holl JM. Basic Atomic and Molecular Spectroscopy. (Tutorial Chemistry Texts): Volume 11. Cambridge: Royal Society of Chemistry; 2002. ISBN-13: 978-0854046676.
35. Pasikatan MC, Steele JL, Spillman CK, Haque E. Near Infrared

- Reflectance Spectroscopy for Online Particle Size Analysis of Powders and Ground Materials. *J Near Infrared Spectrosc.* 2001;9(3):153–64.
36. Madeira CRC, Piantavini MS, Trindade ACLB. Aplicação da espectrofotometria na região do infravermelho associado a calibração multivariada no controle de qualidade de fármacos – Revisão. *Visão Acadêmica*, Curitiba, 2015;16(1):179-97. <https://revistas.ufpr.br/academica/article/view/40358/26375>.
 37. Caramês ETS, Alamar PD, Poppi RJ, Pallone JAL. Quality control of cashew apple and guava nectar by near infrared spectroscopy. *J Food Compos Anal* [Internet]. 2016;56:41–6. Available from: <http://dx.doi.org/10.1016/j.jfca.2016.12.002>
 38. Mariani NCT, Da Costa RC, De Lima KMG, Nardini V, Cunha J??nior LC, Teixeira GHDA. Predicting soluble solid content in intact jaboticaba [*Myrciaria jaboticaba* (Vell.) O. Berg] fruit using near-infrared spectroscopy and chemometrics. *Food Chem.* 2014;159:458–62.
 39. Snyder AB, Sweeney CFCF, Rodriguez-saona LE, Giusti MM. Rapid authentication of concord juice concentration in a grape juice blend using Fourier-Transform infrared spectroscopy and chemometric analysis. *Food Chem* [Internet]. 2014 Mar;147:295–301. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0308814613014052>
 40. Limm W, Karunathilaka SR, Yakes BJ, Mossoba M. A portable mid-infrared spectrometer and a non-targeted chemometric approach for the rapid screening of economically motivated adulteration of milk powder. *Int Dairy J.* 2018;85:177–83.

41. De Oliveira GA, Castilhos F, Renard CMA-G, Bureau S. Comparison of NIR and MIR spectroscopic methods for determination of individual sugars, organic acids and carotenoids in passion fruit. *Food Res Int.* 2014 Jun;60:154–62.
42. Bizzani M, Flores DWM, Colnago LA, Ferreira MD. Non-invasive spectroscopic methods to estimate orange firmness, peel thickness, and total pectin content. *Microchem J.* 2017;133:168–74.
43. Dall' Acqua YG, Cunha Júnior LC, Nardini V, Lopes VG, Pessoa JD da C, de Almeida Teixeira GH. Discrimination of *Euterpe oleracea* Mart. (Açaí) and *Euterpe edulis* Mart. (Juçara) Intact Fruit Using Near-Infrared (NIR) Spectroscopy and Linear Discriminant Analysis. *J Food Process Preserv.* 2015;39(6):2856–65.
44. Maniwaru P, Nakano K, Boonyakiat D, Ohashi S, Hiroi M, Tohyama T. The use of visible and near infrared spectroscopy for evaluating passion fruit postharvest quality. *J Food Eng.* 2014 Dec;143:33–43.
45. Cayuela JA, Weiland C. Intact orange quality prediction with two portable NIR spectrometers. *Postharvest Biol Technol.* 2010;58(2):113–20.
46. Cunha Júnior LC, Teixeira GH de A, Nardini V, Walsh KB. Quality evaluation of intact açaí and juçara fruit by means of near infrared spectroscopy. *Postharvest Biol Technol* [Internet]. 2016;112:64–74. Available from: <http://dx.doi.org/10.1016/j.postharvbio.2015.10.001>
47. Pérez-marín D, Torres I, Entrenas J, Vega M, Sánchez M. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* Pre-harvest screening on-vine of spinach quality and safety using NIRS

- technology. 2019;207(1258):242–50.
48. Nascimento PAM, Carvalho LC de, Júnior LCC, Pereira FMV, Teixeira GH de A. Robust PLS models for soluble solids content and firmness determination in low chilling peach using near-infrared spectroscopy (NIR). *Postharvest Biol Technol*. 2016;111:345–51.
 49. Betemps DL, Fachinello JC, Galarça SP, Machado NP, Remorini D, Massai R, et al. Espectroscopia do visível e infra-vermelho próximo para estimar sólidos solúveis e firmeza de polpa em função da época de colheita em pêssegos. *Semin Ciências Agrárias*. 2014 Jun 23;35(3):1257.
 50. Cayuela Sánchez JA. Citrus Internal Quality Predictions by NIR Spectroscopy. *Glob Sci Books*. 2008;(1):77–82.
 51. Gómez AH, He Y, Pereira AG. Non-destructive measurement of acidity, soluble solids and firmness of Satsuma mandarin using Vis/NIR-spectroscopy techniques. *J Food Eng*. 2006 Nov;77(2):313–9.
 52. Guthrie JA, Walsh KB, Reid DJ, Liebenberg CJ. Assessment of internal quality attributes of mandarin fruit. 1. NIR calibration model development. *Aust J Agric Res*. 2005;56(4):405.
 53. Kawano S, Fujiwara T. Determination Content in Satsuma Mandarin Near Infrared (NIR) Transmittance. *J Japanese Soc Hortic Sci*. 1993;62(2):465–70.
 54. Clément A, Dorais M, Vernon M. Nondestructive Measurement of Fresh Tomato Lycopene Content and Other Physicochemical Characteristics Using Visible - NIR Spectroscopy. *J Agric Food Chem*. 2008;56:9813–

- 8.
55. He Y, Zhang Y, Pereira AG, Gómez AH, Wang J. Nondestructive Determination of Tomato Fruit Quality Characteristics Using Vis / NIR Spectroscopy Technique. *Int J Inf Technol*. 2005;11(11):97-108.
56. Shiroma-Kian C, De Nardo T, Halim Y, Francis D, Rodriguez-Saona LE. Rapid and simultaneous determination of lycopene and beta-carotene contents in tomato juice by infrared spectroscopy. *J Agric Food Chem*. 2009 Feb 25;57(4):1105–12.
57. Pavia DL, Kris GS, Lampman GM, Vyvyan JR. Introdução à espectroscopia. 1ª. Cengage Learning; 2010. 716 p.
58. Ferreira DS. Aplicação de espectroscopia no infravermelho e análise multivariada para previsão de parâmetros de qualidade em soja e quinoa. [Tese]. Campinas: Universidade Estadual de Campinas; 2013.
59. Correa C. Avaliação do uso de espectroscopia no infravermelho para controle e garantia da qualidade da mistura biodiesel/diesel usando cartas de controle multivariadas. [Trabalho de Conclusão de Graduação]. Porto Alegre: Universidade Federal do Rio Grande do Sul; 2012.
60. Santos PM, Pereira-Filho ER, Rodriguez-Saona LE. Rapid detection and quantification of milk adulteration using infrared microspectroscopy and chemometrics analysis. *Food Chem*. 2013; May 1;138(1):19-24.
61. Brás LP, Bernardino SA, Lopes JA, Menezes JC. Multiblock PLS as an approach to compare and combine NIR and MIR spectra in calibrations of soybean flour. *Chemom Intell Lab Syst*. 2005;

62. Ribeiro FZ, Marconcini LV, de Toledo IB, de Vasconcellos Azeredo RB, Barbosa LL, Colnago LA. Nuclear magnetic resonance water relaxation time changes in bananas during ripening: A new mechanism. *J Sci Food Agric.* 2010;90(12):2052–7.
63. Colnago LAC, Forato LA, Bernardes Filho R. Análise da Qualidade de Frutas por Ressonância Magnética Nuclear. In: Ferreira MD. editor. *Tecnologias pós-colheita em frutas e hortaliças.* São Carlos, SP: Embrapa Instrumentação, 2011. p. 286.
64. Santos PM dos. Desenvolvimento da técnica de precessão livre no estado estacionário para aumento da razão sinal ruído em espectros de RMN de alta resolução. [Dissertação]. São Carlos: Universidade de São Paulo; 2009.
65. Santos PM, Vinicius F, Kock C, Santos MS, Lobo CMS, Carvalho AS, et al. Non - Invasive Detection of Adulterated Olive Oil in Full Bottles Using Time - Domain NMR Relaxometry. *J Braz Chem Soc.* 2017;28(2):385-90.
66. Duynhoven J Van, Voda A, Witek M, As H Van. Chapter 3 - Time-Domain NMR Applied to Food Products [Internet]. 1st ed. Vol. 69, *Annual Reports on NMR Spectroscopy.* Elsevier Ltd.; 2010. 145–197 p. Available from: [http://dx.doi.org/10.1016/S0066-4103\(10\)69003-5](http://dx.doi.org/10.1016/S0066-4103(10)69003-5)
67. Bizzani M, Menezesflores DW, Colnago LA, Ferreira MD. Monitoring of soluble pectin content in orange juice by means of MIR and TD-NMR spectroscopy combined with machine learning. *Food Chem* [Internet]. 2020; Dec;332:127383. Available from:

<https://doi.org/10.1016/j.foodchem.2020.127383>

68. Flores DWM, Colnago LA, Ferreira MD, Spoto MHF. Prediction of Orange juice sensorial attributes from intact fruits by TD-NMR. *Microchem J.* 2016;128:113–7.
69. Bizzani M, William Menezes Flores D, Bueno Moraes T, Alberto Colnago L, David Ferreira M, Helena Fillet Spoto M. Non-invasive detection of internal flesh breakdown in intact Palmer mangoes using time-domain nuclear magnetic resonance relaxometry. *Microchemical Journal.* 2020;158: 105208.
<https://www.sciencedirect.com/science/article/pii/S0026265X20317136?via%3Dihub>.
70. Medeiros AM de. Determinação de sólidos solúveis totais em sucos de laranja comerciais utilizando espectroscopia e calibração multivariada. [trabalho de conclusão de Curso TCC] Campina Grande, PB: Universidade Estadual da Paraíba; 2014.
71. Ferreira MMC. *Qumiometria: Conceitos, Métodos e Apliações.* Campinas: Editora Un. Campinas; 2015. 493 p.
72. Varella CAA. *Análise Multivariada Aplicada as Ciências Agrárias: Análise de componentes principais.* Seropedica, RJ: Universidade Federal do Rio de Janeiro (UFRJ); 2008.
73. Geladi P, Kowalski BR. Partial Least-Squares Regression - a Tutorial. *Anal Chim Acta.* 1986;185:1–17.
74. Beebe KR, Randy JP, Seasholtz MB. *Chemometrics: A practical guide.* New York; John Wiley & Sons; 1998. (Wiley-Interscience Series on

Laboratory Automation).

75. Lavine BK. Chemometrics. Anal Chem. 1998 Jun;70(12):209–28.
76. Barker M, Rayens W. Partial least squares for discrimination. J Chemom. 2003;17(3):166–73.
77. Varmuza K, Filzmoser P. Introduction to Multivariate Statistical Analysis in Chemometrics. Boca Raton, Fl.: CRC Press; 2009. (Engineering & Technology, Physical Sciences).
78. Lima GF, Andrade SAC, da Silva VH, Honorato FA. Multivariate Classification of UHT Milk as to the Presence of Lactose Using Benchtop and Portable NIR Spectrometers. Food Anal Methods [Internet]. 2018 Oct 20;11(10):2699–706. Available from: <http://link.springer.com/10.1007/s12161-018-1253-7>
79. Almeida FG. Automação de classificador SVM para aplicação em projetos de consultoria de gestão. [Dissertação]. Brasília: Universidade de Brasília; 2019.
80. Lorena AC, Carvalho ACPLF. Introdução às Máquinas de Vetores Suporte (Support Vector Machines). [Relatório Técnico]. São Carlos; 2003AD. 66p.
81. Mitchell TM. Machine Learning, Hardcover; Blacklick, Ohio, U.s.a.: McGraw-Hill Science/Engineering/Math. 1997. ISBN 9780070428072.
82. Marsland S. Machine Learning: An Algorithmic Perspective. 2nd. ed. Herbrich R, Graepel T, editors. Boca Raton, Fl.: CRC Press; Chapman & Hall Book; 2015. (Chapman & Hall/CRC, Machine Learning & Pattern Recognition Series).

83. Yin Z, Liu J, Krueger M, Gao H. Introduction of SVM algorithms and recent applications about fault diagnosis and other aspects. In: 2015 IEEE 13th International Conference on Industrial Informatics (INDIN), Cambridge, 2015:550–555.
<https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7281793>.
84. Liu Z, Lv X, Liu K, Shi S. Study on SVM compared with the other text classification methods. 2nd Int Work Educ Technol Comput Sci ETCS 2010. 2010;1:219–22.
85. Hellicar AD, Smith D, Engelke U, Henshall J. A Hierarchical Learning Approach to Calibrate Allele Frequencies for SNP Based Genotyping of DNA Pools. 2014 International Joint Conference on Neural Networks (IJCNN). (6-11 July 2014, Beijing, China).
<https://ieeexplore.ieee.org/document/6889697>
86. Hassan UK, Nawi NM, Kasim S. Classify a Protein Domain using Sigmoid Support Vector Machine. 2014 International Conference on Information Science & Applications (ICISA). (6-7 May 2014, Seoul, South Korea). <https://ieeexplore.ieee.org/document/6847375>.
87. Belahcene M, Chouchane A, Mokhtari N. 2D and 3D face recognition based on IPC detection and patch of interest regions. 2014 International Conference on Connected Vehicles and Expo (ICCVE), Vienna, 2014: 627-628. <https://ieeexplore.ieee.org/document/7297624>.
88. Appana DK, Prosvirin A. Reliable fault diagnosis of bearings with varying rotational speeds using envelope spectrum and convolution neural networks. Soft Comput [Internet]. 2018;22(20):6719–29. Available from:

<https://doi.org/10.1007/s00500-018-3256-0>

89. Demetgul M. Fault diagnosis on production systems with support vector machine and decision trees algorithms. *The International Journal of Advanced Manufacturing Technology*. 2013;67(9-12):2183-94. <https://link.springer.com/article/10.1007%2Fs00170-012-4639-5>.
90. Teye E, Amuah CLY, McGrath T, Elliott C. Innovative and rapid analysis for rice authenticity using hand-held NIR spectrometry and chemometrics. *Spectrochim Acta - Part A Mol Biomol Spectrosc*. 2019;147–54.
91. Gao X, Lu T, Liu P, Lu Q. A soil moisture classification model based on SVM used in agricultural WSN. *2014 IEEE 7th Joint International Information Technology and Artificial Intelligence Conference*. 2014:432-436.
92. Zheng W, Fu X, Ying Y. Chemometrics and Intelligent Laboratory Systems Spectroscopy-based food classification with extreme learning machine. *Chemom Intell Lab Syst [Internet]*. 2014;139:42–7. Available from: <http://dx.doi.org/10.1016/j.chemolab.2014.09.015>
93. Zhang H, Berg AC, Maire M, Malik J. SVM-KNN : Discriminative Nearest Neighbor Classification for Visual Category. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*. 2006. p. 35.
94. Raikwal JS, Saxena K. Performance Evaluation of SVM and K-Nearest Neighbor Algorithm over Medical Data set. *Int J Comput Appl*. 2012;50(14):35–9.

95. Lebrini Y, Boudhar A, Lionboui RHH, Arrach LER. Identifying Agricultural Systems Using SVM Classification Approach Based on Phenological Metrics in a Semi - arid Region of Morocco. *Earth Syst Environ* [Internet]. 2019;3(2):277–88. Available from: <https://doi.org/10.1007/s41748-019-00106-z>

96. Gerhardt N, Schwolow S, Rohn S, Pérez-Cacho PR, Galán-Soldevilla H, Arce L, et al. Quality assessment of olive oils based on temperature-ramped HS-GC-IMS and sensory evaluation: Comparison of different processing approaches by LDA, kNN, and SVM. *Food Chem* [Internet]. 2018;11(95). Available from: <https://doi.org/10.1016/j.foodchem.2018.11.095>

97. Krithika N, Sevarani AG. An Individual Grape Leaf Disease Identification Using Leaf Skeletons and KNN Classification. In: 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS). 2017.