



Quantificação de açúcares com uma língua eletrónica: calibração multivariada com seleção de sensores

Vinicius da Costa Arca

*Dissertação apresentada à Escola Superior Agrária de Bragança
para obtenção do Grau de Mestre em Qualidade e Segurança Alimentar*

Orientado por
Professor Dr. Luís Avelino Guimarães Dias
Professor Dr. Evandro Bona

Esta dissertação não inclui as críticas e sugestões feitas pelo Júri

Bragança
2016



Quantificação de açúcares com uma língua eletrónica: calibração multivariada com seleção de sensores

Vinicius da Costa Arca

Dissertação apresentada à Escola Superior Agrária de Bragança para efeito da obtenção do Grau de Mestre em Qualidade e Segurança Alimentar através do acordo de Dupla Diplomação com a Universidade Tecnológica Federal do Paraná.

Orientado por:
Professor Dr. Luís Avelino Guimarães Dias
Professor Dr. Evandro Bona

**Bragança
2016**

Aos meus pais, Claudemir e Sonia

AGRADECIMENTOS

Ao meu orientador, Professor Doutor Luís Dias, pelos conhecimentos transmitidos, disponibilidade, ajuda, experiência e presença ao longo de todas as etapas do trabalho.

Ao meu co-orientador e exemplo de pesquisador, Professor Doutor Evandro Bona, pela minha inserção na área científica e oportunidades de aprendizagem proporcionadas desde o início da graduação.

Ao Eng. Jorge Sá Morais, pelo apoio e valiosos ensinamentos nas análises cromatográficas.

À Escola Superior Agrária do Instituto Politécnico de Bragança e ao Departamento de Alimentos da Universidade Tecnológica Federal do Paraná, pela oportunidade de orgulhosamente fazer parte da comunidade acadêmica e desenvolver este trabalho.

Aos professores que me acompanharam ao longo da minha jornada e que contribuíram não só com a minha formação acadêmica, mas também pessoal.

Aos amigos e colegas de Campo Mourão e Bragança, pela parceria e apoio enquanto estive longe de casa.

À minha família, pela formação do meu caráter e transmissão de valores, em especial aos meus pais, pelo incentivo e apoio incondicional aos estudos desde criança.

A todos manifesto a minha gratidão pela contribuição para a realização desta Tese de Mestrado.

RESUMO

Este trabalho incide na análise dos açúcares majoritários nos alimentos (glucose, frutose e sacarose) com uma língua eletrónica potenciométrica através de calibração multivariada com seleção de sensores. A análise destes compostos permite contribuir para a avaliação do impacto dos açúcares na saúde e seu efeito fisiológico, além de permitir relacionar atributos sensoriais e atuar no controlo de qualidade e autenticidade dos alimentos. Embora existam diversas metodologias analíticas usadas rotineiramente na identificação e quantificação dos açúcares nos alimentos, em geral, estes métodos apresentam diversas desvantagens, tais como lentidão das análises, consumo elevado de reagentes químicos e necessidade de pré-tratamentos destrutivos das amostras. Por isso se decidiu aplicar uma língua eletrónica potenciométrica, construída com sensores poliméricos selecionados considerando as sensibilidades aos açúcares obtidas em trabalhos anteriores, na análise dos açúcares nos alimentos, visando estabelecer uma metodologia analítica e procedimentos matemáticos para quantificação destes compostos. Para este propósito foram realizadas análises em soluções padrão de misturas ternárias dos açúcares em diferentes níveis de concentração e em soluções de dissoluções de amostras de mel, que foram previamente analisadas em HPLC para se determinar as concentrações de referência dos açúcares. Foi então feita uma análise exploratória dos dados visando-se remover sensores ou observações discordantes através da realização de uma análise de componentes principais. Em seguida, foram construídos modelos de regressão linear múltipla com seleção de variáveis usando o algoritmo *stepwise* e foi verificado que embora fosse possível estabelecer uma boa relação entre as respostas dos sensores e as concentrações dos açúcares, os modelos não apresentavam desempenho de previsão satisfatório em dados de grupo de teste. Dessa forma, visando contornar este problema, novas abordagens foram testadas através da construção e otimização dos parâmetros de um algoritmo genético para seleção de variáveis que pudesse ser aplicado às diversas ferramentas de regressão, entre elas a regressão pelo método dos mínimos quadrados parciais. Foram obtidos bons resultados de previsão para os modelos obtidos com o método dos mínimos quadrados parciais aliado ao algoritmo genético, tanto para as soluções padrão quanto para as soluções de mel, com R^2_{ajustado} acima de 0,99 e RMSE inferior a 0,5 obtidos da relação linear entre os valores previstos e experimentais usando dados dos grupos de teste. O sistema de multi-sensores construído se mostrou uma ferramenta adequada para a análise dos

açúcares, quando presentes em concentrações maioritárias, e alternativa a métodos instrumentais de referência, como o HPLC, por reduzir o tempo da análise e o valor monetário da análise, bem como, ter um preparo mínimo das amostras e eliminar produtos finais poluentes.

Palavras-chave: Língua eletrônica, calibração multivariada, açúcares, regressão linear múltipla, mínimos quadrados parciais, seleção de variáveis, algoritmo *stepwise*, algoritmo genético.

ABSTRACT

This work focuses on the analysis of the major sugars in foods (glucose, fructose and sucrose) with a potentiometric electronic tongue through multivariate calibration with sensors subset selection. The analysis of these compounds contributes to the assessment of the impact of sugars on health and its physiological effect; also, allows to relate sensory attributes and act on quality control and authenticity of the food. Although there are various analytical methods routinely used in the identification and quantification of sugars in foods, in general these methods have several disadvantages such as, slowness of the analysis, high consumption of chemicals and the need for destructive pretreatments of samples. Therefore, it was decided to apply a potentiometric electronic tongue, built with polymeric sensors selected considering the sensitivities to sugar obtained in previous studies, on the analysis of sugars in foods aiming to establish an analytical methodology and mathematical procedures to quantify these compounds. For this purpose, standard solutions of ternary mixtures of sugars and solutions of honey (previously analyzed by HPLC to determine the reference concentrations of sugars) at different levels of concentration were analysed. Hence, an exploratory data analysis aiming to remove sensors or discordant observations was made by performing a principal component analysis. Then multiple linear regression models were built with variable selection by using the stepwise algorithm and it was found that although it was possible to establish a good relation between the sensors responses and the sugars concentrations, the models did not show satisfactory predictive performance in test group data. Thus, in order to overcome this problem, new approaches were tested through the construction and optimization of the parameters of a genetic algorithm for selection of variables that could be applied to various regression techniques, including the regression with the method of partial least squares. Good predictive results were obtained for the partial least squares models combined with genetic algorithm for both standard solutions and for honey solutions, with values of R^2_{adjusted} above 0.99 and RMSE less than 0.5, obtained from the linear relationship between the predicted and experimental values using data of test groups. The multisensor system built proved to be a suitable tool for the analysis of sugars, when present in majority concentrations, and an alternative to analytical instrumental methods of reference, such as HPLC, by reducing the time of analysis and its monetary value, as well as, having a minimal sample preparation and eliminate pollutants.

Key-words: Electronic tongue, multivariate calibration, sugars, multiple linear regression, partial least squares, variable selection, stepwise algorithm, genetic algorithm.

ÍNDICE GERAL

PREÂMBULO	xiv
a) Gênese do trabalho	xvi
b) Objetivos	xvii
c) Estrutura do trabalho	xvii
1. REVISÃO BIBLIOGRÁFICA	1
1.1. Introdução	3
1.2. Açúcares nos alimentos	3
1.2.1. Açúcares na saúde e impacto sensorial	5
1.2.2. Metodologias analíticas	7
1.3. Língua eletrônica	8
1.3.1. Tipos de línguas eletrônicas	10
1.3.2. Aplicações das línguas eletrônicas	11
1.4. Principais técnicas de tratamento e interpretação de sinais	13
1.4.1. Análise de componentes principais	14
1.4.2. Regressão linear múltipla	15
1.4.3. Mínimos quadrados parciais	17
1.4.4. Algoritmo genético	18
2. MATERIAIS E MÉTODOS	21
2.1. Reagentes e amostras	23
2.2. Análises com HPLC	24
2.2.1. Preparo das soluções	24
2.2.2. Construção das curvas de calibração	25
2.3. Análises com língua eletrônica	25
2.3.1. Montagem do sistema de multi-sensores	25
2.3.2. Sensores químicos	26
2.3.3. Equipamento	26
2.3.4. Metodologia de análise	27
2.4. Análise de dados	31
2.4.1. Modelos de regressão linear múltipla	31
2.4.2. Modelos com método dos mínimos quadrados parciais	32
3. RESULTADOS E DISCUSSÕES	33
3.1. Análises por HPLC	35

3.2.	Análises com língua eletrônica.....	39
3.3.	Modelos multivariados de estimação e previsão	43
3.3.1.	Modelos MLR de estimação usando algoritmo <i>stepwise</i>	43
3.3.2.	Separação dos dados	46
3.3.3.	Modelos de previsão usando o algoritmo <i>stepwise</i>	47
3.4.	Modelos com seleção de variáveis usando algoritmo genético.....	49
4.	CONCLUSÕES.....	59
	REFERÊNCIAS BIBLIOGRÁFICAS	65

ÍNDICE DE TABELAS

Tabela 1. Reagentes utilizados nas análises em HPLC e com a língua eletrónica.....	23
Tabela 2. Descrição das amostras de méis	23
Tabela 3. Ordem das membranas no sistema de multi-sensores	27
Tabela 4. Níveis e concentrações dos açúcares nas soluções padrão com fundo iônico de KCl 1M	29
Tabela 5. Intervalos de concentrações e parâmetros de calibração para análise dos açúcares em HPLC.....	36
Tabela 6. Avaliação da repetibilidade e exatidão das soluções de controlo de qualidade	37
Tabela 7. Caracterização dos açúcares nas amostras de mel.....	37
Tabela 8. Modelos de estimação de MLR obtidos usando algoritmo <i>stepwise</i>	44
Tabela 9. Separação dos dados em conjuntos de treino e teste para construção dos modelos de regressão	47
Tabela 10. Modelos de MLR obtidos usando algoritmo <i>stepwise</i> e resultados da previsão para os dados de teste	48
Tabela 11. Modelos de PLS obtidos usando algoritmo genético e resultados da previsão para os dados de teste.....	52

ÍNDICE DE FIGURAS

Figura 1. Estruturas da glucose (α -D-Glicopirranose) e frutose (α -D-Frutofuranose).....	4
Figura 2. Ligação glicosídica α -1,2 entre uma molécula de glucose e frutose para formação de uma molécula de sacarose	4
Figura 3. Ligações α -1,4 e α -1,6 na estrutura do amido.....	5
Figura 4. Representação do funcionamento de um sensor em uma língua eletrónica.....	9
Figura 5. Decomposição da matriz original de dados (X) nas matrizes de <i>scores</i> (T) e <i>loadings</i> (P) através de PCA	15
Figura 6. Decomposição das matrizes de variáveis preditoras (X) e variáveis resposta (Y) através de PLS	17
Figura 7. Modelo generalizado do algoritmo genético.....	19
Figura 8. Esquema da montagem do sistema de multi-sensores	28
Figura 9. Desenho experimental ortogonal utilizado no preparo das soluções padrão de misturas de glucose, frutose e sacarose.....	30
Figura 10. Cromatogramas das amostras de mel.....	35
Figura 11. Curvas de calibração para sacarose, glucose e frutose obtidas por HPLC....	36
Figura 12. Concentração relativa dos açúcares nas soluções de méis para análises com língua eletrónica	39
Figura 13. Variação da intensidade do sinal potenciométrico da língua electrónica em função do tempo para uma solução de mel	40
Figura 14. Perfis dos sinais potenciométricos das soluções padrão e de mel com e sem fundo iônico de KCl 1M.....	41
Figura 15. Extremos e quartis dos sinais potenciométricos das soluções padrão e de mel com e sem fundo iônico de KCl 1M.....	42
Figura 16. PCA para soluções padrão e de mel medidas simultaneamente	43
Figura 17. Comparações das concentrações experimentais e estimadas de glucose usando os modelos de estimação de MLR obtidos usando algoritmo <i>stepwise</i>	45
Figura 18. Exemplo de seleção de variáveis utilizando o algoritmo de Kennard & Stone	46
Figura 19. Resultados de previsão para os dados de teste usando os modelos PLS para soluções sem fundo iônico	53
Figura 20. Resultados de previsão para os dados de teste usando os modelos PLS para soluções com fundo iônico de KCl 1M.....	54

Figura 21. Estabilização da função de desempenho para o GA obtida no modelo E.....	55
Figura 22. Frequência da presença dos sensores na construção dos modelos PLS com seleção de variáveis através do GA.....	56

LISTA DE ABREVIATURAS

AT – Açúcares totais

DOT – Limiar de detecção sensorial (abreviatura do inglês, *dose-over-threshold*)

E-tongue – Língua eletrônica (abreviatura do inglês, *electronic tongue*)

FI_{free} – Razão de intolerância à frutose livre (abreviatura do inglês, *free fructose intolerance ratio*)

FI_{total} – Razão de intolerância à frutose total (abreviatura do inglês, *total fructose intolerance ratio*)

GA – Algoritmo genético (abreviatura do inglês, *genetic algorithm*)

GC – Cromatografia gasosa (abreviatura do inglês, *gas chromatography*)

GI – Índice glicêmico (abreviatura do inglês, *glycemic index*)

GL – Carga glicêmica (abreviatura do inglês, *glycemic load*)

HPLC – Cromatografia líquida de alta eficiência (abreviatura do inglês, *high-performance liquid chromatography*)

MLR – Regressão linear múltipla (abreviatura do inglês, *multiple linear regression*)

MW – Massa molar (abreviatura do inglês, *molar weight*)

PCA – Análise de componentes principais (abreviatura do inglês, *principal component analysis*)

PCR – Regressão por componentes principais (abreviatura do inglês, *principal component regression*)

PLS – Mínimos quadrados parciais (abreviatura do inglês, *partial least squares*)

p_{crossover} – probabilidade de cruzamento

p_{mutation} – probabilidade de mutação

PVC – Policloreto de vinila (abreviatura do inglês, *polyvinyl chloride*)

R² – Coeficiente de determinação

R²_{ajustado} – Coeficiente de determinação ajustado

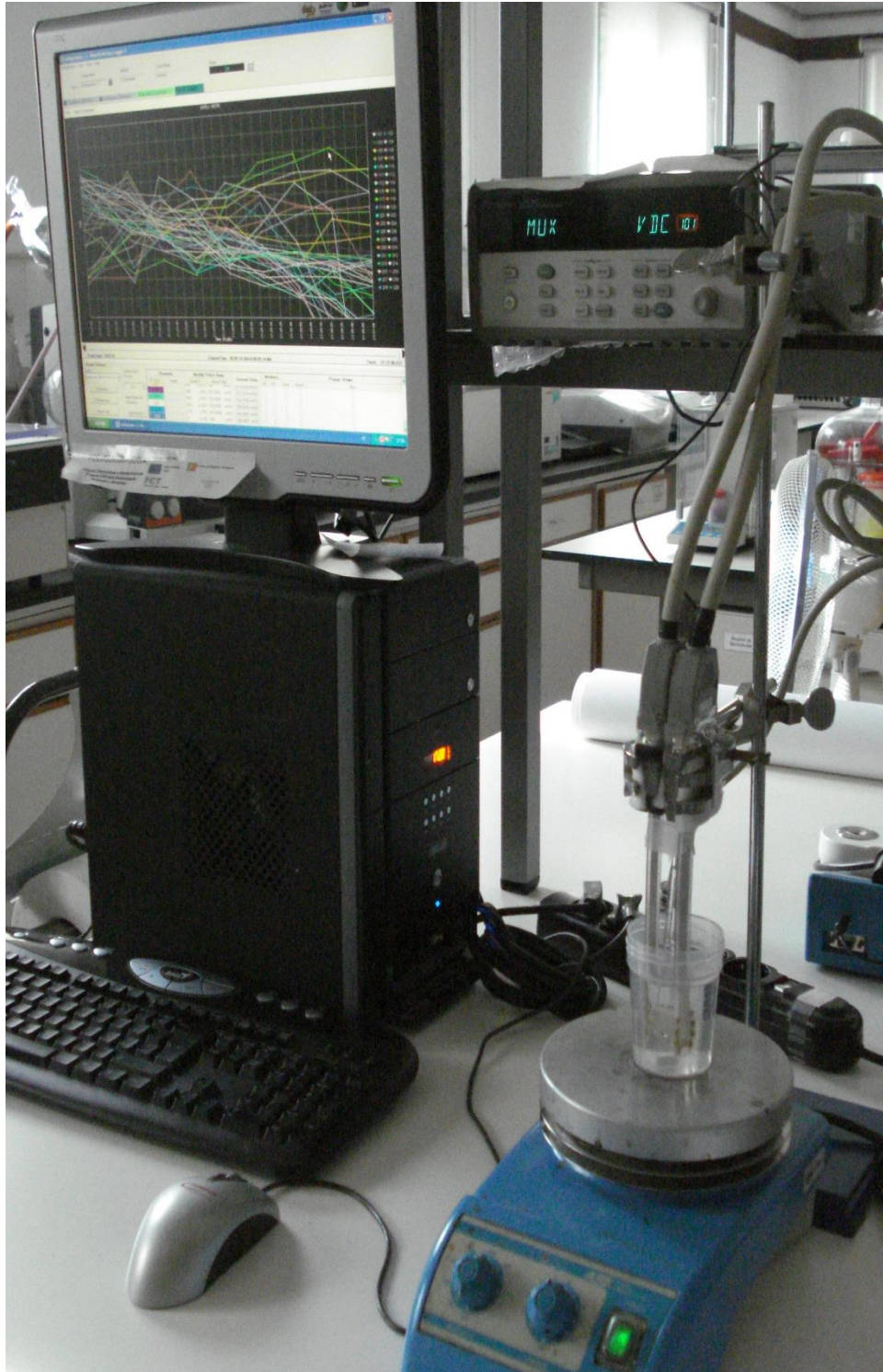
RMSE – Raíz do erro quadrático médio (abreviatura do inglês, *root mean squared error*)

RNA – Redes neurais artificiais

Sac.eq – Índice de doçura em sacarose equivalente

SE – Erro padrão (abreviatura do inglês, *standard error*)

PREÂMBULO



a) Gênese do trabalho

Os açúcares atuam de várias formas nos alimentos, e suas propriedades estão diretamente relacionadas com sua concentração e estrutura química. O tipo a quantidade desses compostos estão ligados a importantes questões de saúde e podem influenciar sensorialmente as características dos produtos alimentares. Portanto, considerando que este seja um fator que pode influenciar a escolha dos consumidores, além dos aspectos legais de autenticidade, nutricionais e de controlo de qualidade, é importante a quantificação da concentração dos açúcares nos alimentos.

Os métodos mais clássicos, como os de Lane-Eynon, Munson-Walker e Somogyi-Nelson, ou mesmo metodologias instrumentais modernas, como cromatografia gasosa e cromatografia líquida de alta eficiência ou métodos espectrofotométricos, apresentam diversas desvantagens, tais como lentidão das análises, consumo de reagentes químicos e necessidade de pré-tratamentos, geralmente destrutivos, das amostras. Estes fatores, aliados ao crescente apelo da química verde, impulsionam o desenvolvimento de novas metodologias confiáveis para este propósito. Entre elas, a língua eletrónica é uma ferramenta analítica que tem sido usada em estudos qualitativos e quantitativos na área de alimentos.

A língua eletrónica é composta por um sistema de multi-sensores químicos não específicos, de baixa seletividade e sensibilidade cruzada para diferentes espécies em solução que fornece um perfil global de sinais representativo da amostra analisada, do qual é possível extrair tanto informações qualitativas quanto quantitativas. Entretanto, uma vez que os sensores podem responder a diversos analitos em uma mesma leitura, e devido à elevada quantidade de dados obtidos através das análises, é necessário recorrer a técnicas multivariadas estatísticas para o tratamento dos dados. Entre as ferramentas multivariadas para estudos quantitativos, destacam-se técnicas de regressão, como a regressão linear múltipla e mínimos quadrados parciais, aliadas a algoritmos de seleção de variáveis, entre eles o algoritmo genético.

Foi nesse âmbito que se decidiu investigar a capacidade da utilização de uma língua eletrónica potenciométrica como metodologia alternativa para quantificação dos açúcares nos alimentos através da obtenção de modelos de previsão e seleção de variáveis usando o algoritmo genético.

b) Objetivos

O **objetivo geral** deste trabalho foi aplicar uma língua eletrônica potenciométrica na análise dos açúcares majoritários nos alimentos (glucose, frutose e sacarose) e açúcares totais em soluções padrão de mistura dos açúcares referidos e amostras de méis, visando estabelecer uma metodologia analítica e procedimentos matemáticos para quantificação destes compostos.

Os **objetivos específicos** para isto foram:

- ∅ Construir uma língua eletrônica potenciométrica com membranas poliméricas de policloreto de vinilo e diferentes combinações de plastificantes e aditivos;
- ∅ Determinar a concentração de glucose, frutose e sacarose em amostras de méis utilizando cromatografia líquida de alta eficiência (HPLC) através da construção de curvas de calibração com soluções padrão para estes compostos;
- ∅ Preparar diluições das amostras e soluções padrão em desenho experimental ortogonal na ausência e presença de fundo iônico;
- ∅ Analisar as soluções de méis com a língua eletrônica em paralelo com as soluções padrão de misturas dos açúcares;
- ∅ Realizar uma análise exploratória dos dados obtidos com a língua eletrônica;
- ∅ Utilizar técnicas de seleção de variáveis para otimizar a construção de modelos de regressão que relacionem os sinais potenciométricos da língua eletrônica com a concentração dos açúcares das soluções;
- ∅ Construir modelos de estimação e previsão utilizando técnicas de regressão multivariada.

c) Estrutura do trabalho

Este trabalho inicia-se com uma **REVISÃO BIBLIOGRÁFICA**, constituída primeiramente por uma caracterização dos principais açúcares presentes nos alimentos, suas funções e relevância do ponto de vista sensorial e impacto na saúde. Em seguida se apresentam os tipos de língua eletrônica, seu princípio de funcionamento e estudos já realizados utilizando esta ferramenta. Por fim, a secção é finalizada com breves descrições das principais ferramentas de análise e interpretação de dados utilizadas ao longo do trabalho.

Em seguida, são apresentados os **MATERIAIS E MÉTODOS**, onde faz-se a descrição dos reagentes, amostras, equipamentos e outros materiais. Nesta secção é apresentado o procedimento e desenho experimental de preparo das soluções padrão e de amostras para as análises com a língua eletrónica e por HPLC. Descreve-se o sistema de multi-sensores, os sensores químicos e o procedimento de análise com o mesmo. A secção é finalizada com a metodologia de tratamento e interpretação dos sinais adotada.

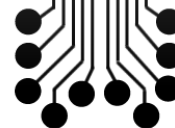
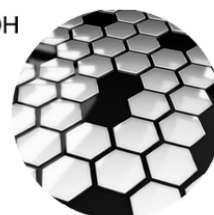
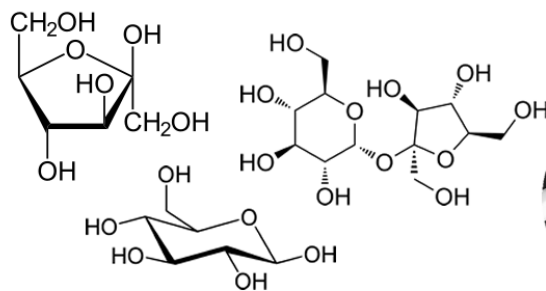
No terceiro capítulo, **RESULTADOS E DISCUSSÕES**, apresentam-se primeiramente os resultados das análises por HPLC e as curvas de calibração obtidas. Em seguida são apresentados os dados das análises com a língua eletrónica, os perfis e a análise exploratória dos dados. Ao fim do capítulo são apresentados os resultados da construção dos modelos de estimação e previsão para as soluções padrão e de mel.

No quarto capítulo, **CONCLUSÕES**, estão apresentadas as conclusões mais relevantes obtidas através da análise e discussões dos resultados e, uma verificação dos objetivos estabelecidos para realização do trabalho, além de sugestões para trabalhos futuros.

Por fim, são apresentadas as **REFERÊNCIAS BIBLIOGRÁFICAS** utilizadas.



1. REVISÃO BIBLIOGRÁFICA



1.1. Introdução

Os carboidratos são os nutrientes base da nutrição humana e constituem as biomoléculas mais abundantes da natureza. São os nutrientes mais consumidos seja na forma de açucarados naturais como mel e frutas; de açucarados propriamente ditos como o açúcar comercial; de alimentos elaborados à base de açúcar como sorvetes, biscoitos e chocolates; de amiláceos como arroz, milho, mandioca e batata e suas farinhas ou derivados como pães, massas, biscoitos ou cervejas [1,2].

Nos alimentos, os açúcares além de serem uma importante fonte energética, atuam como agentes de sabor (conferem doçura), textura (formam gomas e constituem a base da matriz dos alimentos) e escurecimento (sofrem reações de Maillard e caramelização). Também são usados como fonte de alimentação para microrganismos em produtos que passam por processos fermentativos ou como aditivos em produtos industrializados [2].

1.2. Açúcares nos alimentos

Existem três principais classes de carboidratos: os monossacarídeos, oligossacarídeos e polissacarídeos. Este estudo incide nos três açúcares mais presentes nos alimentos, os monossacarídeos glucose e frutose e o dissacarídeo sacarose. Os monossacarídeos, ou açúcares simples, consistem em uma única unidade de polihidroxil aldeído ou cetona, com tendência a apresentar conformação cíclica se possuir mais de quatro carbonos em sua estrutura. Os monossacarídeos mais abundantes na natureza (presentes em grande quantidade em produtos como mel, frutas e sumos), são a aldohexose D-glucose (piranose de seis carbonos), também referida como dextrose devido à sua rotação específica para direita (+52,7°), e a cetohexose D-frutose (furanose de seis carbonos), referida por levulose devido à sua rotação específica para esquerda (-92,4°) [2,3]. Na Figura 1 mostram-se as estruturas químicas da glucose e frutose.

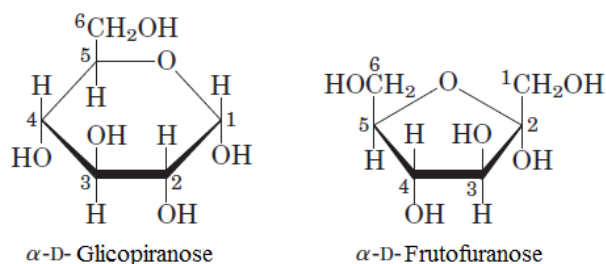


Figura 1. Estruturas da glucose (α -D-Glicopiranosse) e frutose (α -D-Frutofuranose)

Glucose, frutose e outros açúcares, por possuírem grupos aldeídos (possuem o grupo carbonilo na extremidade da cadeia de carbono) e cetonas (possuem o grupo carbonilo no meio da cadeia de carbono) livres na sua estrutura, são capazes de reduzir íons férricos (Fe^{3+}) ou cúpricos (Cu^{2+}) (teste qualitativo baseado na reação de Fehling) e, por isso, chamados açúcares redutores. Medindo-se a quantidade de agente oxidante reduzido pela solução contendo o açúcar é possível estimar a concentração desses compostos [1].

Os oligossacarídeos consistem em pequenas cadeias de unidades de monossacarídeos, ou resíduos de polissacarídeos, unidas por ligações glicosídicas. Os oligossacarídeos mais comuns são os dissacarídeos, compostos por duas unidades de monossacarídeos. O dissacarídeo mais abundante nos alimentos é a sacarose, constituída por uma unidade de D-glucose e uma de D-frutose unidas por uma ligação glicosídica α -1,2, segundo a equação da reação química apresentada na Figura 2. Outro dissacarídeo importante para a alimentação humana é a lactose, um dissacarídeo constituído por unidades de galactose e glucose (epímeros em relação ao quarto átomo de carbono) unidos através de ligação glicosídica β -1,4, o açúcar característico do leite [1].

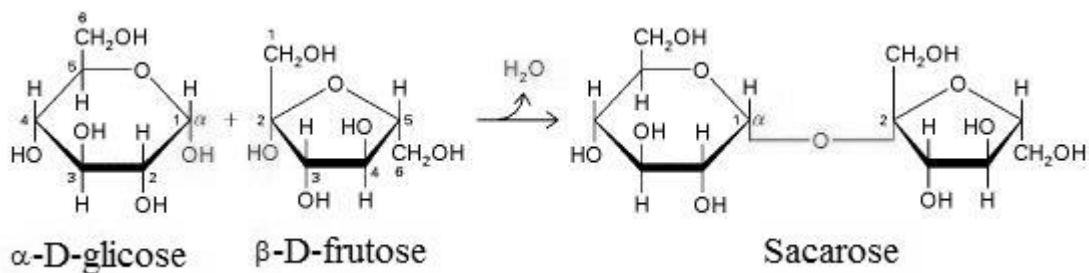


Figura 2. Ligação glicosídica α -1,2 entre uma molécula de glucose e frutose para formação de uma molécula de sacarose

Os polissacarídeos são polímeros que contém mais de 20 unidades de monossacarídeos. A celulose, o polissacarídeo mais abundante na natureza, apresenta uma estrutura linear composta apenas por unidades de glucose unidas por ligações β -1,4, as quais não são rompidas pelas enzimas digestivas do trato gastrointestinal humano. Entretanto, outros polissacarídeos também formados por unidades de glucose, como o glicogênio (polissacarídeo de reserva energética presente em células animais) e o amido (polissacarídeo de reserva energética presente em células vegetais), que possuem seções lineares (amilose) unidas por ligações α -1,4 e seções ramificadas (amilopectina) unidas por ligações α -1,6 (Figura 3), podem ser reduzidas a unidades de glucose por ação enzimática do sistema digestivo humano [1,2].

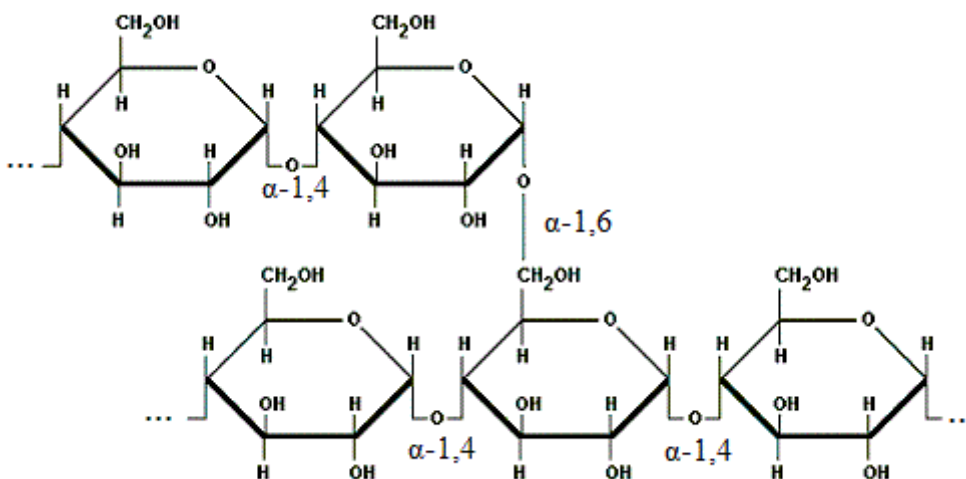


Figura 3. Ligações α -1,4 e α -1,6 na estrutura do amido

1.2.1. Açúcares na saúde e impacto sensorial

O consumo de alimentos açucarados, entre eles principalmente bebidas carbonatadas não alcoólicas, tais como refrigerantes, e bebidas de frutas, como néctares e sumos, teve um elevado crescimento nos últimos 30 anos. Embora os consumidores se refiram a essas bebidas, em especial as de fruta, como saudáveis por serem fonte de vitaminas e antioxidantes pelo seu conteúdo em sumo de frutas, seu consumo exagerado pode levar a graves problemas de saúde devido o alto teor de açúcares [4]. Estudos mostram que dietas ricas em frutose podem ser altamente prejudiciais à saúde, induzindo à obesidade, diabetes, dislipidemia e resistência à insulina [5] e que os efeitos da má absorção a

frutose podem ser minimizados através da ingestão simultânea de glucose em proporção igual ou superior [6].

Dessa forma, é conveniente avaliar estes atributos relacionados com a saúde através índices de fácil entendimento, uma vez que podem permitir uma melhor compreensão da relação entre os efeitos fisiológicos de alimentos ricos em carboidratos e a saúde humana [7]. O cálculo de índices de saúde importantes, tais como a carga glicêmica (*glycemic load*, GL), expressa pela equação (1), a razão de intolerância à frutose livre (*fructose intolerance ratio*, FI_{free}), definida pela equação (2), e à total (FI_{total}), usando a equação (3) – uma vez que uma molécula de sacarose pode gerar uma de glucose e outra de frutose e, estas podem apresentar efeitos metabólicos diferentes – pode ser realizado conhecendo-se o conteúdo e a proporção dos açúcares na amostra [4,7].

$$GL = \frac{GI_{sacarose} \times [Sacarose, gL^{-1}] + GI_{glucose} \times [Glucose, gL^{-1}] + GI_{frutose} \times [Frutose, gL^{-1}]}{100 \times (SS, L)} \quad (1)$$

Onde,

GI é o índice glicêmico médio de cada açúcar puro ($GI_{glucose} = 99 \pm 7$; $GI_{frutose} = 19 \pm 2$; $GI_{sacarose} = 68 \pm 5$) e SS é o volume (L) da amostra.

$$FI_{free} = \frac{[Glucose, gL^{-1}]}{[Frutose, gL^{-1}]} \quad (2)$$

$$FI_{total} = \frac{[Glucose, gL^{-1}] + [Sacarose, gL^{-1}] \times MW_{glucose} / MW_{sacarose}}{[Frutose, gL^{-1}] + [Sacarose, gL^{-1}] \times MW_{frutose} / MW_{sacarose}} \quad (3)$$

Onde, MW é a massa molar.

O índice glicêmico é uma medida de rapidez do aumento do açúcar no sangue (glucose) depois de comer um alimento ou um produto específico. A glucose é o açúcar que dá maior resposta de açúcar no sangue e, a sacarose tem um GI de 68 ± 5 , próximo do ponto médio entre a glucose e a frutose. Foi devido ao seu baixo GI que a frutose foi inicialmente promovida como um açúcar mais saudável do que a sacarose para os diabéticos. Mas, como já referido, trabalhos recentes sugerem que a frutose provoca efeitos indesejáveis na saúde humana. O índice de intolerância à frutose (FI) é

importante para as pessoas que sofrem de má absorção de frutose. Um alimento com a concentração de glucose superior ao da frutose tem o FI maior do que a unidade e, por isso, é tolerável para pessoas sensíveis à frutose [6].

Sendo assim, considerando que o impacto desses parâmetros nutricionais pode influenciar a escolha dos consumidores, é importante a quantificação da concentração dos açúcares majoritários nos alimentos (glucose, frutose e sacarose), bem como conhecer sua contribuição no impacto sensorial, uma vez que os açúcares podem realçar a percepção humana de alguns sabores e que a doçura está associada com proporção relativa de monossacarídeos individuais e com a relação entre o conteúdo de sacarose frente o conteúdo total de açúcares [7,8].

O realce de sabor doce está relacionado com o perfil de açúcares presente no alimento, ou seja, com a contribuição dos perfis de percepção da doçura de cada açúcar. A percepção de doçura do consumidor pode ser diretamente relacionada com o teor total de açúcares através do conceito de índice de doçura (*sweetness index*) em sacarose equivalente (*Sac.eq*), expresso através da equação (4). O impacto do conteúdo de cada açúcar individual na doçura pode ser avaliado através da relação entre a concentração de cada açúcar na amostra e seu respectivo limiar de detecção sensorial (*dose-over-threshold, DOT*) em água obtido através da literatura. Por definição, valores de DOT superiores a 1 indicam uma influência significativa no sabor [9].

$$\mathbf{Sac.eq = 1 \times [Sacarose] + 0.74 \times [Glucose] + 1.73 [Frutose]} \quad (4)$$

A expressão (4) mostra que a frutose tem um sabor mais acentuado de doçura e, por isso, mais facilmente detectado pelo consumidor. A sacarose confere uma doçura mais agradável para alimentos e bebidas devido ao desenvolvimento mais lento do sabor doce e da sua decomposição lenta.

1.2.2. Metodologias analíticas

A análise de alimentos é importante de forma a obter informação sobre a sua composição química, estrutura, propriedades físico-químicas e atributos sensoriais. Esta informação permite contribuir para a produção de alimentos que sejam nutricionais e com atributos desejáveis para o consumidor, considerando também o aspeto

informativo, ao nível dos rótulos, para ajudá-lo na sua dieta [10]. A informação analítica também é usada no controlo de qualidade e na autenticidade dos alimentos, para garantir que os consumidores não sejam vítimas de fraude. Por isso, é importante desenvolver técnicas analíticas rápidas, simples e de baixo custo para a análise de alimentos na sua composição, como por exemplo, ao nível dos açúcares [11].

A determinação de açúcares totais e de açúcares redutores nos alimentos é normalmente feita através de métodos pouco seletivos, fundamentados em sua maioria na redução de íons cobre em soluções alcalinas, na desidratação dos açúcares por uso de ácidos concentrados e posterior coloração com compostos orgânicos, ou redução simples e formação de compostos com coloração mensurável, como os métodos de Lane-Eynon, Munson-Walker, e Somogyi-Nelson [12–14]. Existem também uma série de metodologias instrumentais modernas capazes de fornecer o conteúdo total de açúcares ou a concentração específica de carboidratos, tais como cromatografia gasosa (GC), cromatografia líquida de alta eficiência (HPLC), análises enzimáticas, métodos eletroquímicos e espectrométricos [15,16]. A maioria desses métodos, entretanto, apresentam diversas desvantagens, como lentidão das análises, consumo de reagentes químicos e necessidade de pré-tratamentos das amostras, usando geralmente métodos destrutivos. Estes fatores, aliados ao crescente apelo da química verde, impulsionam o desenvolvimento de novas metodologias confiáveis para quantificação de açúcares nos alimentos [17]. É neste âmbito que se considera que as línguas eletrónicas poderão ser uma ferramenta analítica alternativa viável e com vantagens que a seguir se descreve.

1.3. Língua eletrónica

Uma língua eletrónica (*e-tongue*) é um instrumento analítico para análise de líquidos constituído por um sistema de multi-sensores químicos não específicos, de baixa seletividade e sensibilidade cruzada para diferentes espécies em solução que fornece um perfil global de sinais representativo da amostra analisada [18]. Uma vez que os sensores apresentam baixa seletividade e podem responder a diversos analitos em uma mesma leitura, uma grande quantidade de dados complexos é gerada e, para processamento dos perfis, é necessária a utilização de técnicas matemáticas baseadas na análise multivariada de dados com a finalidade de identificar e quantificar componentes em misturas [18–20].

Estes dispositivos reproduzem artificialmente a sensação de sabor, imitando as línguas biológicas, onde os receptores respondem a uma grande variedade de substâncias [21]. Nestas, as células sensoriais das papilas gustativas são encarregadas pelo reconhecimento das substâncias responsáveis pelo sabor, que convertem essa informação num sinal elétrico transmitido pelos neurônios e interpretado pelo cérebro permitindo o reconhecimento das sensações básicas do paladar. Analogamente, nas línguas eletrônicas o sistema de sensores químicos de sensibilidade cruzada age como o receptor das substâncias, que transmite um sinal através de um transdutor para uma unidade de processamento de dados responsável pela sua interpretação [20,22]. A Figura 4 apresenta uma representação do funcionamento de uma língua eletrônica.

O conceito da utilização de línguas eletrônicas tem sido amplamente desenvolvido nos últimos anos e seu uso aliado à quimiometria, vertente da química aplicada que utiliza a análise multivariada para interpretação de sinais analíticos instrumentais, representa uma linha de pesquisa consolidada no ramo da eletroanálise [21,22]. Em função disso, o uso de línguas eletrônicas tem se mostrado uma grande alternativa para técnicas espectroscópicas e cromatográficas tradicionais na análise de alimentos, atingindo resultados satisfatórios com o preparo mínimo das amostras na avaliação do controle de qualidade, discriminação, classificação, monitoramento de processos e análise quantitativa de produtos alimentares [20,23].

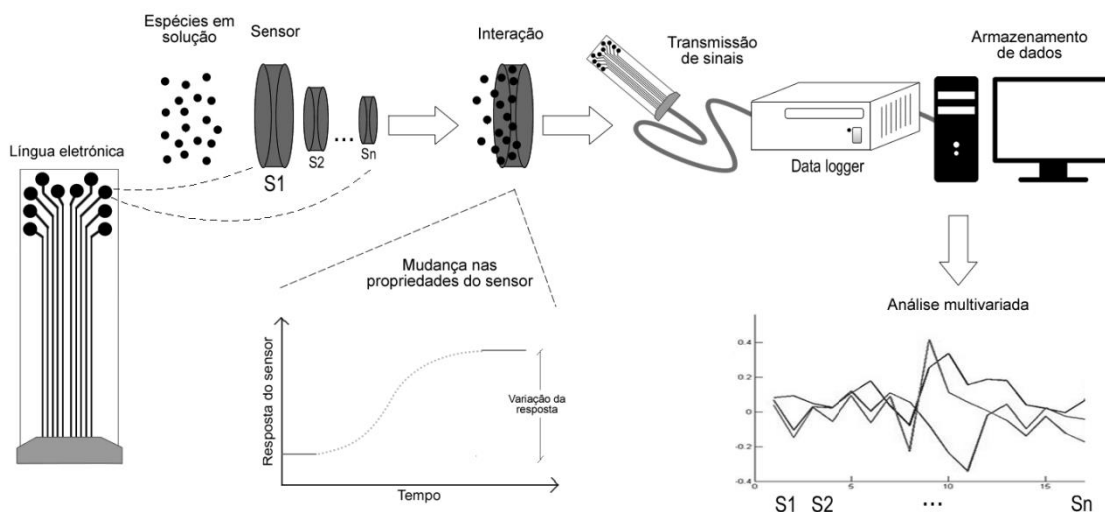


Figura 4. Representação do funcionamento de um sensor em uma língua eletrônica

Na avaliação de sabor na indústria alimentícia e farmacêutica, testes sensoriais são geralmente implementados por painéis de provadores e, embora a língua humana possa gerar padrões únicos para uma enorme variedade de substâncias, esse método pode ser problemático devido à sua baixa objetividade e reprodutibilidade uma vez que os resultados são fortemente dependentes das experiências e do estado emocional de cada indivíduo [24]. Além disso, apesar de uma série de metodologias tradicionais confiáveis utilizadas para determinação e detecção de substâncias em alimentos apresentarem boa precisão e exatidão, muitas são destrutivas, requerem elevados tempos de preparo e análise, equipamentos caros e não permitem a análise *in situ*. Nesse sentido, as línguas eletrônicas aparecem como ferramentas rápidas, de baixo custo de calibração, com desempenho satisfatório e facilmente adaptáveis para diferentes condições de utilização para avaliação da qualidade de alimentos [20,21,24].

1.3.1. Tipos de línguas eletrônicas

No que diz respeito ao tipo do sinal primário obtido através da análise com línguas eletrônicas, uma ampla variedade de sensores químicos tem sido usados no design destes sistemas, sendo os principais tipos baseados em métodos eletroquímicos, como potenciometria, voltametria, amperometria, impedimetria e condutimetria, além de métodos ópticos e enzimáticos (biossensores) [18–20,22,25].

Vidros calcogenetos e materiais cristalinos têm sido utilizados como membranas de sensores potenciométricos, e metais nobres para detecção principalmente de sinais amperométricos [22,26]. Materiais sensíveis baseados em membranas poliméricas contendo diferentes substâncias ativas em diferentes proporções têm sido utilizados tanto para sensores potenciométricos [27,28] quanto ópticos [18]. As línguas eletrônicas mais usadas estão focadas na eletroanálise, em especial na potenciometria e voltametria [10].

Embora línguas eletrônicas voltamétricas tenham sido relatadas por apresentarem algumas vantagens em relação à robustez, versatilidade do método e melhor desempenho em aplicações a baixas concentrações [19], especialmente voltametria de pulso para se obter informações de soluções multicomponentes [29], o uso dessa técnica envolve procedimentos mais complexos do que simples medidas potenciométricas. A língua eletrônica potenciométrica é também uma metodologia sensível e de detecção rápida, mais fácil de manusear, de processo de medição mais simples e com pré-

processamento da amostra mínimo. Devido a estas vantagens, este trabalho incide sobre a língua eletrônica potenciométrica para a análise de açúcares, onde o potencial elétrico é medido, na ausência de corrente, entre um eléctrodo de referência e vários eléctrodos de trabalho presentes em paralelo num sistema de multi-sensores, em geral, constituídos por membranas poliméricas que apresentam seletividade cruzada para as espécies químicas presentes na solução analisada [30].

Normalmente, o eléctrodo de referência utilizado é o de Ag/AgCl, que consiste em um fio de prata revestido com cloreto de prata colocado em uma solução de íons cloreto, baseado na reação de meia-célula [31]:



No eléctrodo de trabalho, a membrana deve ser insolúvel em água, mecanicamente estável e ter afinidade com as substâncias a analisar, o que implica na absorção destas substâncias pela membrana e desta interação resulta uma alteração no potencial de membrana que é medido. Entretanto, as principais desvantagens da utilização de sensores potenciométricos são o fato de só se poder medir espécies carregadas [24], a dependência da temperatura e a incorporação de outros componentes da solução na superfície dos sensores, o que afeta o potencial da membrana e, conseqüentemente, a intensidade do sinal medido (*drifts* no sinal medido) [10].

Os primeiros estudos onde conjuntos de sensores foram aplicados para análise multicomponente em líquidos utilizaram sensores potenciométricos [32,33] e ainda hoje este tipo permanece como sendo um dos principais aplicados em línguas eletrônicas [10]. A primeira língua eletrônica, desenvolvida por Toko *et al.* [34], era constituída por oito eléctrodos potenciométricos com membranas lipo-poliméricas.

1.3.2. Aplicações das línguas eletrônicas

Após o contato com uma amostra, a língua eletrônica fornece um conjunto multidimensional de informações que precisa ser processado e correlacionado com informações de natureza química. Devido ao grande de número de possibilidades de utilização e combinação de sensores, as línguas eletrônicas têm permitido aplicações qualitativas onde se usam impressões globais sobre as amostras para estudos de

identificação, classificação e detecção de adulterações, bem como análises quantitativas específicas em controle de processos [19].

Outra perspectiva de aplicação é a utilização do conceito de sabor artificial [35] inspirado nas papilas de uma língua biológica, onde a finalidade é realizar uma percepção virtual do sabor de amostras. Nesse sentido, um conjunto de sensores é desenvolvido para responder aos sabores básicos (doce, salgado, amargo, ácido e umami) com o objetivo de mimetizar a avaliação humana de sabor em casos como, no controle automático de processos ou de amostras em condições extremas ou tóxicas [36].

Diversos estudos aplicando línguas eletrônicas potenciométricas têm sido reportados nos últimos anos. Martínez-Máñez *et al.* [30] e Nuñez *et al.* [37] desenvolveram línguas eletrônicas para detecção de compostos e para análise quantitativa em águas. Woertz *et al.* [38] apresenta seu uso no desenvolvimento de formulações químicas. Existem também aplicações na área de meio ambiente no monitoramento de sistemas ambientais [26,37,39].

No ramo alimentício, Escuder-Gilbert & Peris [10] e Ha *et al.* [20] fizeram revisões gerais sobre a aplicação de línguas eletrônicas na análise de alimentos. No caso de línguas eletrônicas potenciométricas, cabe destacar os trabalhos nas áreas de bebidas não-alcólicas, na avaliação semi-quantitativa de refrigerantes e sucos [28,40]; de bebidas alcoólicas, na discriminação e controle de qualidade [40]; na avaliação e identificação de chás [41,42] e aplicação em conjunto com narizes eletrônicos [43,44]. Outras aplicações se referem à classificação de diferentes cultivares [45] e avaliação de propriedades sensoriais durante o armazenamento de vegetais [46], pesquisa de adulterações em leite caprino [27] e classificação de azeites [47].

No âmbito do tema deste trabalho, a análise dos açúcares glucose, frutose e sacarose, encontram-se poucos trabalhos com a aplicação de línguas eletrônicas. Sakata *et al.* [11] classifica açúcares de diferentes qualidades utilizando uma língua eletrônica baseada em espectroscopia de impedância enquanto que, Sá *et al.* [48] analisa carboidratos em cana-de-açúcar utilizando uma língua eletrônica voltamétrica. Dias *et al.* [49] analisa a qualidade de méis utilizando uma língua eletrônica potenciométrica, e embora esta esteja diretamente relacionada com sua composição em açúcares, não foram encontrados trabalhos específicos voltados para a calibração direta de uma língua eletrônica potenciométrica para a análise quantitativa de açúcares em diversas matrizes alimentares.

Para um trabalho destes, a língua eletrônica potenciométrica deverá estar acoplada a técnicas multivariadas de estatística como, por exemplo, análise de componentes principais (PCA, do inglês *principal component analysis*), para verificar a variabilidade dos dados e a presença de outliers, e para obter modelos de previsão (calibração), o método de regressão linear múltipla (MLR, do inglês *multiple linear regression*) e o método dos mínimos quadrados parciais (PLS, do inglês *partial least squares*). Com o objetivo de selecionar os melhores sensores para os modelos de previsão que podem conter ruídos ou serem fontes de informações redundantes, algoritmos de seleção de variáveis devem ser usados como, por exemplo, o algoritmo genético (GA, do inglês *genetic algorithm*).

1.4. Principais técnicas de tratamento e interpretação de sinais

O conceito de língua eletrônica envolve a utilização de um conjunto de sensores não específicos que precisam estar aliados a técnicas de processamento de dados a fim de se interpretar suas respostas complexas e relacioná-las com seu significado analítico. Além disso, problemas comuns como os *drifts* nos sinais, condições não ideais das análises ou interferências que frequentemente ocorrem durante as medidas podem ser corrigidas ou minimizadas nestes tratamentos dos dados [26].

Um dos parâmetros chave para se determinar o tipo de informação obtida por meio de uma língua eletrônica é a escolha da ferramenta de modelagem, que normalmente é realizada em duas etapas. Primeiramente, tratamentos como centralização e escalonamento dos dados são realizados a fim de se descartar o efeito das diferenças nas dimensões das unidades, remover informações redundantes e elevar a relação sinal/ruído [50]. Em seguida, através da utilização de algoritmos apropriados e ferramentas quimiométricas, é possível obter modelos de previsão entre os sinais da língua eletrônica (variáveis independentes) e as respostas do estudo (variáveis dependentes), permitindo efetuar análises qualitativas (reconhecimento, classificação ou identificação) e quantitativas compensando o efeito da matriz ou interferências da própria amostra [40].

Entre as ferramentas multivariadas disponíveis, no âmbito deste trabalho utilizam-se técnicas não supervisionadas, como PCA para análise exploratória ou redução da dimensionalidade dos dados; ou técnicas de regressão, como os métodos MLR ou PLS,

para aplicações quantitativas, aliadas a algoritmos de seleção de variáveis, o GA neste caso [4,21,45].

Em modelos de calibração, uma das etapas mais importantes é a seleção do número e quantidade de observações que serão utilizadas na construção e avaliação dos modelos uma vez que estes podem ser altamente influenciados pelos dados utilizados em cada etapa. Tendo isto em vista, um algoritmo muito utilizado para seleção de amostras é o algoritmo de Kennard-Stone [51], onde primeiramente o par de observações mais afastadas da média é separado para o conjunto de treino e, em seguida, a atribuição das observações restantes é feita com base na distância de Mahalanobis, que pode ser obtida através de uma PCA dos dados de entrada e cálculo da distância euclidiana na matriz de *scores* truncada [52].

1.4.1. Análise de componentes principais

A PCA é um método não supervisionado que reduz a dimensionalidade dos dados, agrupando as informações altamente correlacionadas em um novo sistema de eixos que consiste em uma combinação linear das variáveis originais formando componentes principais ortogonais. Esta técnica permite a visualização de possíveis agrupamentos de amostras e *outliers* [53]. O método descarta combinações lineares que têm pequenas variações e são responsáveis pela descrição do ruído instrumental, mantendo apenas os termos com variâncias significativas de modo que: a primeira componente principal explique a maior parte da variabilidade contida nos dados; a segunda componente principal seja a combinação linear com máxima variância em direção ortogonal à primeira componente e explique a maior parte da variabilidade não explicada pela primeira componente; e assim por diante [51,52].

Esta técnica resulta numa transformação matemática abstrata da matriz original de dados representada pela equação (5) e ilustrada na Figura 5.

$$\mathbf{X} = \mathbf{T} \cdot \mathbf{P} + \mathbf{E} \quad (5)$$

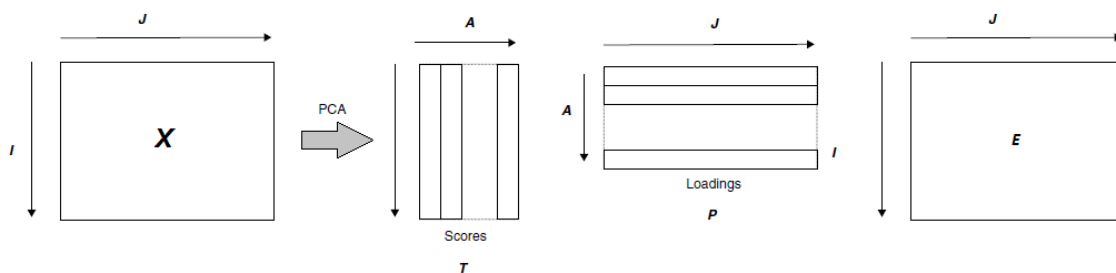


Figura 5. Decomposição da matriz original de dados (\mathbf{X}) nas matrizes de *scores* (\mathbf{T}) e *loadings* (\mathbf{P}) através de PCA

Onde:

\mathbf{X} é a matriz original de dados (I observações \times J variáveis);

\mathbf{T} é a matriz dos *scores* ($I \times A$);

\mathbf{P} é a matriz dos *loadings* (A componentes principais \times J);

\mathbf{E} é a matriz dos resíduos ($I \times J$).

O número de componentes principais possíveis de serem calculadas corresponde à dimensão comum das matrizes de *scores* e *loadings*, que não pode ser superior à menor dimensão da matriz original de dados [55]. Cada matriz de *scores* consiste numa série de vetores coluna e representam as projeções das amostras no novo conjunto de eixos. Cada matriz de *loadings* é constituída por uma série de vetores linha que representam o peso de cada variável original na decomposição dos novos eixos [56]. Com a língua eletrónica potenciométrica, a PCA foi usada, por exemplo, nos estudos de Gallardo *et al.* [21], Dias *et al* [27]., He *et al.* [20], Beullens *et al.* [45] e Cetó *et al.* [25].

1.4.2. Regressão linear múltipla

Em um modelo de regressão linear simples, uma resposta medida \mathbf{Y} está relacionada com uma única variável preditora \mathbf{X} para cada observação. Entretanto, em diversos estudos, mais de uma variável preditora precisa ser avaliada para construção de um modelo mais representativo. Dessa forma, a MLR tem como objetivo modelar a relação entre duas ou mais variáveis explanatórias e uma variável resposta através do ajuste de uma equação linear aos dados de calibração. O modelo é então aplicado a um conjunto

de amostras de teste que não estavam presentes na elaboração do modelo e sua resposta é comparada com os valores observados para este novo conjunto [57].

Assumindo a existência de p variáveis preditoras, com $i = 1,2,3,\dots, n$ amostras, o modelo MLR generalizado segue a equação 6 [58].

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon \quad (6)$$

Onde:

X_p são as variáveis independentes;

Y_i é a variável dependente;

α é o intercepto;

β_p são os coeficientes de regressão estimados para cada variável independente;

ε é a diferença entre o valor resposta observado de Y e o valor da reta da regressão.

A qualidade dos modelos de calibração pode ser avaliada através dos resultados obtidos com o grupo de teste, avaliando a relação entre os valores previstos pelo modelo obtido e os esperados (obtidos experimentalmente). Esta deverá ser linear, apresentando declive, intercepto e coeficientes de correlação ou determinação próximos de 1, 0 e 1, respectivamente, e também através da análise dos resíduos, valores de diferentes somas dos quadrados e técnicas de validação cruzada. É importante que os estimadores dos coeficientes de regressão obtidos usando o melhor modelo não sejam enviesados, embora estes contenham intervalos de confiança (normalmente ao nível de 5% de significância). Dessa forma, a linha de calibração deve ser apropriadamente descrita utilizando-se não só os valores médios mas também os erros associados, sendo que estes critérios de avaliação são aplicados a qualquer método estatístico, entre eles o MLR e PLS [59].

Quando as variáveis preditoras utilizadas em uma MLR são os *scores* das componentes principais decompostas a partir dos dados originais, temos um caso particular chamado de regressão por componentes principais (PCR, do inglês *principal component regression*) [53].

1.4.3. Mínimos quadrados parciais

A regressão pelo método PLS é considerada uma das técnicas multivariadas de regressão mais importantes para dados de elevada dimensionalidade. O modelo utiliza um conjunto de amostras de treinamento e suas respostas desejadas para decomposição da matriz de dados originais em variáveis latentes através das matrizes de *scores* e *loadings*, em analogia à PCA, estabelecendo uma relação linear entre as variáveis independentes \mathbf{X} e as variáveis dependentes \mathbf{Y} . O modelo é então aplicado a um conjunto de amostras de teste que não estavam presentes na elaboração do modelo e sua resposta é comparada com os valores observados para este novo conjunto de amostras, o que permite verificar a capacidade de previsão do modelo [57].

A abordagem PLS envolve a modelagem das variáveis preditoras contidas em \mathbf{X} e a obtenção de um modelo que busca relacionar as variáveis respostas contidas em \mathbf{Y} com as leituras em \mathbf{X} , conforme as equações (7) e (8) e ilustrado na Figura 6.

$$\mathbf{X} = \mathbf{T} \cdot \mathbf{P} + \mathbf{E} \quad (7)$$

$$\mathbf{Y} = \mathbf{T} \cdot \mathbf{q} + \mathbf{f} \quad (8)$$

Onde \mathbf{q} é uma analogia a um vetor de *loadings*, embora não seja ortogonalizado. O produto de \mathbf{T} e \mathbf{P} aproxima-se de \mathbf{X} , descontando-se os resíduos em \mathbf{E} , e o produto de \mathbf{T} e \mathbf{q} aproxima os valores resposta em \mathbf{Y} , descontando-se os resíduos em \mathbf{f} .

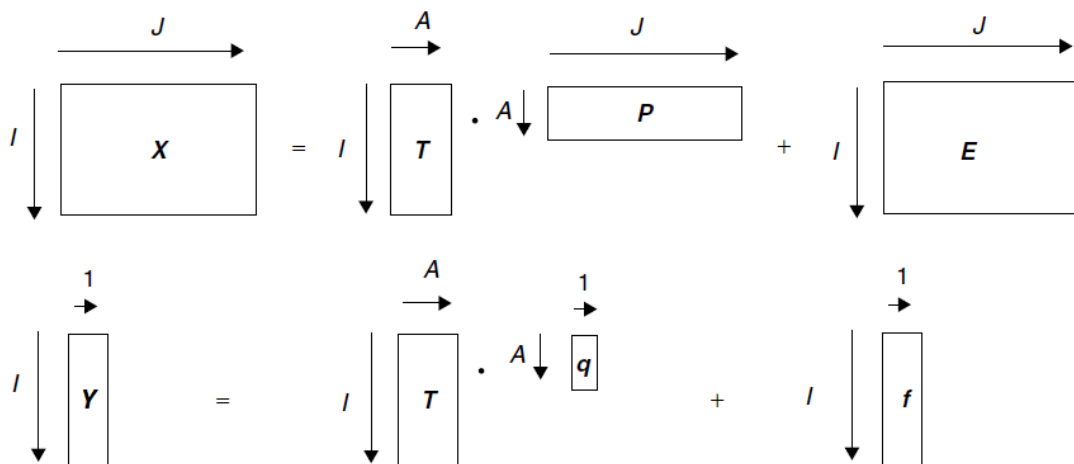


Figura 6. Decomposição das matrizes de variáveis preditoras (\mathbf{X}) e variáveis resposta (\mathbf{Y}) através de PLS

As matrizes de *scores* e *loadings* obtidas por PLS são diferentes das obtidas através de PCA, uma vez que conjuntos exclusivos dessas matrizes são obtidos para cada variável resposta do conjunto de dados, enquanto que na PCA as variáveis resposta não são levadas em conta e as matrizes dos *scores* e *loadings* são únicas.

Os estudos de Beullens *et al.* [45], Kantor *et al.* [46] e Blanco *et al.* [60] apresentam exemplos de PLS como ferramenta associada a análises com línguas eletrônicas.

1.4.4. Algoritmo genético

Na modelagem utilizando leituras de línguas eletrônicas quando se emprega uma série de sensores não-específicos, o melhor modelo é obtido testando-se vários subconjuntos de variáveis (sensores), uma vez que alguns sensores podem carregar informações redundantes ou ruído, o que prejudica o desempenho do ajuste. Dessa forma, obtêm-se modelos de previsão com um menor número de variáveis, mais simples de se interpretar e que apresenta os melhores resultados [49].

Para esta seleção de variáveis, os métodos heurísticos de seleção sequencial para frente (*forward selection*) – onde cada variável é adicionada de cada vez por ordem de maior correlação com a variável dependente até que um critério de parada seja estabelecido – ou seleção sequencial por eliminação (*backward elimination*) – onde todo o conjunto de variáveis é inicialmente considerado e as que menos contribuem para o modelo são uma a uma excluídas – não levam em consideração possíveis interações entre as variáveis que não foram selecionadas, podendo gerar adições ou exclusões redundantes de variáveis, principalmente quando o número destas é muito grande ou apresentam colinearidade, como é o caso dos sensores de sensibilidade cruzada de uma língua eletrônica [61]. A fim de contornar esse problema, algoritmos meta-heurísticos, entre eles o GA, identificam um subconjunto de variáveis dentro de um grande espaço de outros subconjuntos, capaz de gerar uma otimização local para um dado critério de desempenho previamente definido para a resolução de um problema de otimização [62].

No GA, as variáveis a serem otimizadas geralmente são codificadas sob a forma de *strings* ou vetores binários, chamados cromossomos, onde cada variável é representada por um dígito, chamado gene, e um conjunto de cromossomos é chamado de população. Inicialmente, uma população aleatória é criada e uma função de desempenho (*fitness function*) é associada a cada cromossomo representando seu grau de ajuste. Posteriormente, uma nova população é gerada através de fenômenos biologicamente

inspirados, tais como cruzamentos entre cromossomos e mutações de genes, com base na manutenção dos cromossomos que obtiveram os melhores resultados para a função de desempenho (quanto melhor o ajuste de um cromossomo, menor é a probabilidade deste sofrer cruzamento ou mutação na geração seguinte). O processo continua até que um número máximo de gerações seja atingido ou que um critério de parada seja satisfeito [63].

No contexto da seleção de variáveis, estas são escolhidas para cada modelo de regressão e estes subconjuntos tem sua capacidade preditiva testada de acordo com algum critério resultante da estimação ou previsão do modelo.

Descrições mais aprofundadas sobre o funcionamento do GA e sua aplicação no âmbito da análise combinatória para seleção de variáveis são descritas por Cadima et al. [62], Yang et al. [64], Bandyopadhyay & Pal [65] e Örkücü [63]. Na Figura 7 é apresentado um modelo generalizado do GA definido por Gendreal & Potvin [66].

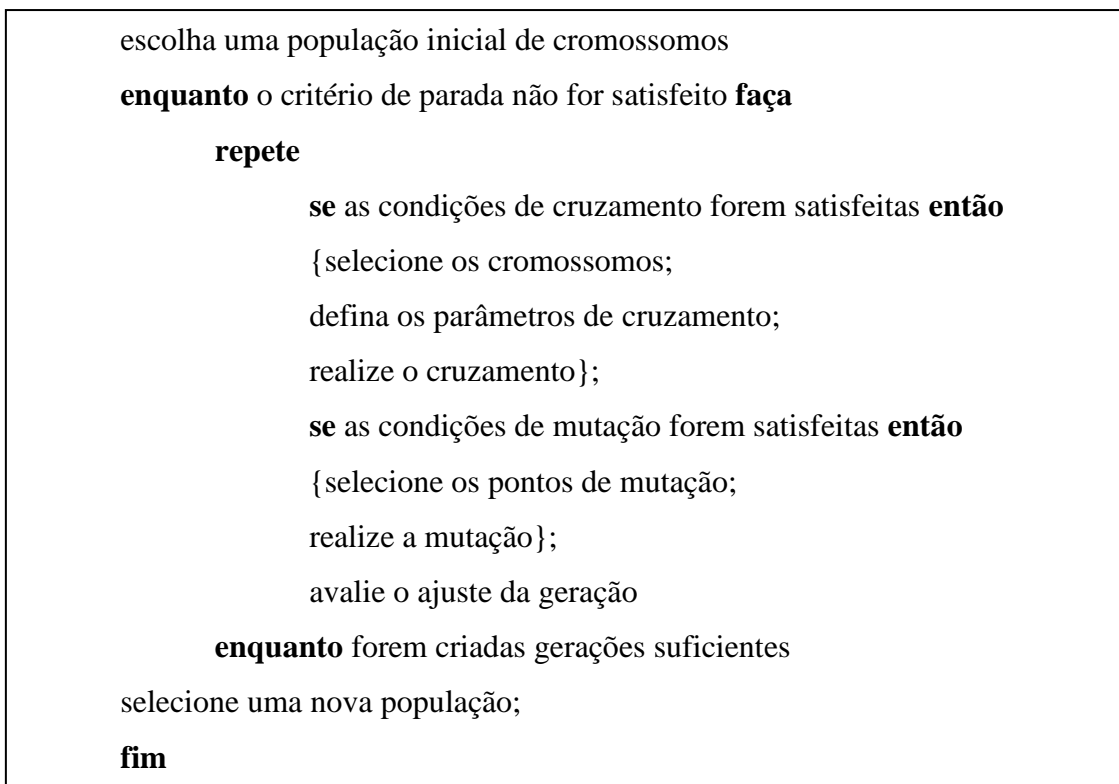
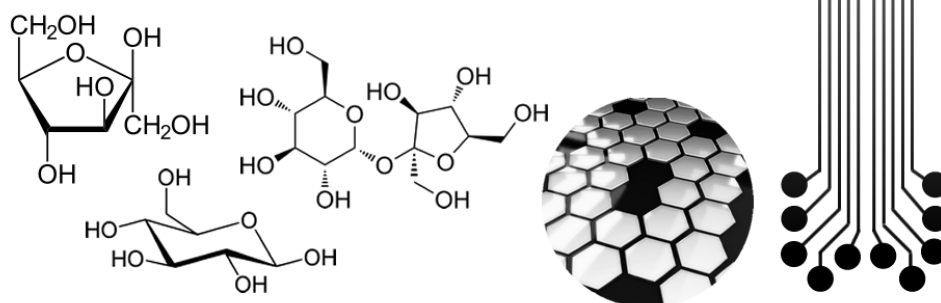


Figura 7. Modelo generalizado do algoritmo genético.



2. MATERIAIS E MÉTODOS



2.1. Reagentes e amostras

Todas as soluções usadas neste trabalho foram preparadas com água desionizada (tipo II) e, os reagentes analíticos usados no preparo das soluções padrão apresentavam pureza adequada para análise (*pro analysis*).

Na Tabela 1 estão apresentados os reagentes utilizados para preparo das soluções padrão usadas na construção das curvas de calibração do HPLC e no preparo do eluente, bem como nas análises com a língua eletrônica e das soluções com e sem fundo iônico de KCl.

Tabela 1. Reagentes utilizados nas análises em HPLC e com a língua eletrônica

Composto	Fórmula Molecular	Marca
Glucose (D)	C ₆ H ₁₂ O ₆	Fluka
Frutose (D)	C ₆ H ₁₂ O ₆	Fluka
Sacarose	C ₁₂ H ₂₂ O ₁₁	Panreac
Ácido <i>orto</i> -fosfórico 85%	H ₃ PO ₄	Panreac
Cloreto de potássio	KCl	Panreac

Neste trabalho visou-se analisar quantitativamente os açúcares em alimentos com uma língua eletrônica e para isso três amostras comerciais de méis foram adquiridas em superfícies comerciais de Bragança. Os méis foram escolhidos de forma a se ter uma gama de coloração variada, o que está relacionado com uma maior variabilidade em termos de composição química. A Tabela 2 apresenta uma breve descrição de cada uma das amostras.

Tabela 2. Descrição das amostras de méis

Código da amostra	Descrição
M1	Mel multifloral
M2	Mel com floração predominante de rosmaninho
M3	Mel com floração predominante de eucalipto

As amostras de méis foram analisadas em HPLC, considerado como método analítico de referência, para se determinar os teores de glucose, frutose e sacarose,

permitindo estabelecer os valores esperados para as amostras de mel nas análises com língua eletrônica.

2.2. Análises com HPLC

As análises de cromatografia líquida de alta eficiência foram efetuadas em equipamento da marca Varian, composto por uma bomba Prostar 220, um injetor manual Rheodyne modelo 7725i com um *loop* de 20 μL , um forno Jones Chromatography modelo 7981 e uma coluna Supelcogel C-610H com 30 cm de comprimento e 7,8 mm de diâmetro interno. Foi usado um detector de índice de refração (IR) da Varian, modelo RI-4. Cada análise teve o tempo de corrida de 30 min com fluxo isocrático de eluente de 0,5 mL/min. Os dados foram obtidos e tratados através do software Star Chromatography Workstation, version 6.4.

2.2.1. Preparo das soluções

O eluente utilizado nas análises com HPLC foi o ácido *orto*-fosfórico diluído a 0,1%, que foi filtrado em um sistema de filtração Phenomenex usando filtros Whatman de 0,2 μm de nylon acoplado a uma bomba de vácuo (Laboport) e posteriormente desgasificados em ultrassom (Elma Transsonic 460/H) durante 5 minutos.

Na construção das curvas de calibração foram utilizadas soluções padrão de misturas de sacarose, frutose e glucose em diferentes concentrações. Estas foram preparadas por medição da massa em balança analítica de cada um dos açúcares diretamente para balões volumétricos de forma a se obter misturas em sete níveis diferentes (0,10; 0,20; 0,30; 0,50; 2,0; 4,0 e 8,0 g/L). Para o estudo da precisão e exatidão das repetições e verificação da qualidade das curvas de calibração, duas soluções de controle de qualidade foram preparadas com níveis de concentração dos açúcares dentro dos limites experimentais (1,0 e 6,0 g/L) utilizando o mesmo procedimento.

As amostras de méis foram diluídas em água desionizada de forma a viabilizar a injeção no equipamento de HPLC evitando a saturação dos picos dos açúcares tendo-se pesado aproximadamente 9,0 g de mel para balões volumétricos de 50 mL e, posteriormente, transferindo-se novamente 3,5 mL destas soluções com auxílio de

micropipeta para balões volumétrico de 50 mL, de forma a se obter cromatogramas com picos de áreas dentro dos limites das curvas de calibração dos açúcares.

As soluções padrão e as diluições das amostras foram filtradas com o auxílio de microfiltros descartáveis de nylon de 0,2 μm da Whatman e seringas para dentro de *vials*, dos quais as amostras eram retiradas para serem injetadas no equipamento de HPLC.

2.2.2. Construção das curvas de calibração

As curvas de calibração isoladas para glucose, frutose e sacarose foram obtidas através do preparo de soluções padrão de misturas destes açúcares em sete níveis diferentes, descritos na secção 2.2.1. que foram injetadas em HPLC determinando-se as áreas dos picos correspondentes aos açúcares em cada nível. A identificação dos picos foi feita com base na comparação dos tempos de retenção determinados pelos trabalhos de Sequeira *et al.* [23] e Dias *et al.*[49].

As curvas foram construídas utilizando as áreas dos picos de cada açúcar em cada nível a partir de duas repetições. Duas soluções de controlo de qualidade com concentrações dos açúcares em níveis intermediários (1,0 e 6,0 g/L) foram injetadas em triplicada e suas áreas aplicadas às equações das curvas obtidas a fim de se determinar o coeficiente de variação e o erro relativo percentual destas em relação às concentrações reais.

2.3. Análises com língua eletrónica

2.3.1. Montagem do sistema de multi-sensores

Cada sistema de multi-sensores foi montado sobre as duas faces de uma placa transparente de policloreto de vinila (PVC) semi-flexível de 1,5 mm de espessura. Em cada face foram colocados 10 sensores e os circuitos impressos através de uma técnica de *print-screen* aplicando-se uma pasta epoxilica de prata condutora (EPO-TEK E4110, Epoxy Technology Inc. – Kit Part A) preparada a partir da mistura com um reagente endurecedor (Kit Part B).

Após a secagem a 40°C por 8h em estufa e limpeza do circuito, foram pulverizadas diversas camadas de resina acrílica (PLASTIK 70, Kontakt Chemie) para revestimento e

impermeabilização do sistema, mantendo vedados os pontos de aplicação dos sensores químicos. O sistema foi testado com um multivoltímetro para se verificar a integridade dos circuitos e a ausência de sobreposição dos contatos. Por fim, a placa foi ligada à uma ficha RS-232 de 25 pinos para posterior ligação em um *data logger*.

2.3.2. Sensores químicos

Os sensores químicos utilizados na construção de cada sistema correspondem a membranas poliméricas de sensibilidade cruzada preparadas a partir de diferentes combinações de plastificantes (65%) e aditivos químicos (3%), usando o PVC (32%) como polímero. Os aditivos usados foram: octadecilamina, álcool oleílico, cloreto de metiltriocetilamônio e ácido oleico. Os plastificantes utilizados foram: adipato de bis-(1-butilpentilo), sebacato de dibutilo, 2-nitrofeniloctil éter, fosfato de tris-(2-etil-hexilo) e ftalato de bis-(2-etil-hexilo).

Cada sensor foi preparado em duplicata por medição das massas de cada um dos seus três componentes (todos da marca Fluka) e dissolução em solvente tetrahidrofurano (Sigma) a fim de se obter uma solução viscosa homogênea. As membranas do sistema de multi-sensores foram formadas usando a técnica da gota, na qual adições sucessivas de gotas das soluções eram aplicadas diretamente sobre os contatos da língua eletrônica em intervalos de 3-5 minutos para evaporação completa do solvente até a obtenção de uma membrana cristalina polimérica.

A Tabela 3 apresenta a ordem das membranas no sistema de multi-sensores. A composição das membranas polimérica foi selecionada tendo em consideração os resultados obtidos por Dias *et al.* [28] na análise semi-quantitativa de glucose e frutose.

2.3.3. Equipamento

O dispositivo utilizado neste trabalho era constituído por dois sistemas de multi-sensores iguais em relação aos sensores e por um eletrodo de referência Ag/AgCl de dupla junção. Estes estavam ligados a um sistema de aquisição de dados *data logger* (Agilent 34970) para obtenção do sinal potenciométrico de cada sensor através do computador utilizando o software Agilent BenchLink Data Logger. Todas as leituras com a língua eletrônica foram feitas em soluções sob agitação magnética (VELP Scientifica ARE heating magnetic stirrer).

A Figura 8 apresenta um esquema ilustrativo da montagem do sistema de multi-sensores.

Tabela 3. Ordem das membranas no sistema de multi-sensores

Sensor		Plastificante (65%)	Aditivo (3%)
Sistema 1	Sistema 2		
1	21	2-nitrofeniloctil éter	Octadecilamina
2	22	2-nitrofeniloctil éter	Álcool oleílico
3	23	2-nitrofeniloctil éter	Cloreto de metiltrioctilamônio
4	24	2-nitrofeniloctil éter	Ácido oleico
5	25	Fosfato de tris-(2-etil-hexilo)	Octadecilamina
6	26	Fosfato de tris-(2-etil-hexilo)	Álcool oleílico
7	26	Fosfato de tris-(2-etil-hexilo)	Cloreto de metiltrioctilamônio
8	28	Fosfato de tris-(2-etil-hexilo)	Ácido oleico
9	29	Adipato de bis-(1-butilpentilo)	Octadecilamina
10	30	Adipato de bis-(1-butilpentilo)	Álcool oleílico
11	31	Adipato de bis-(1-butilpentilo)	Cloreto de metiltrioctilamônio
12	32	Adipato de bis-(1-butilpentilo)	Ácido oleico
13	33	Sebacato de dibutilo	Octadecilamina
14	34	Sebacato de dibutilo	Álcool oleílico
15	35	Sebacato de dibutilo	Cloreto de metiltrioctilamônio
16	36	Sebacato de dibutilo	Ácido oleico
17	37	Ftalato de bis-(2-etil-hexilo)	Octadecilamina
18	38	Ftalato de bis-(2-etil-hexilo)	Álcool oleílico
19	39	Ftalato de bis-(2-etil-hexilo)	Cloreto de metiltrioctilamônio
20	40	Ftalato de bis-(2-etil-hexilo)	Ácido oleico

2.3.4. Metodologia de análise

Preparo das soluções padrão. Para se determinar a concentração de glucose, frutose e sacarose das amostras utilizando a língua eletrônica, primeiro foram analisadas soluções padrão com massas conhecidas de cada um dos açúcares para construção de modelos multivariados capazes de relacionar o sinal obtido através do sistema de multi-sensores e a concentração de açúcares. O desenho experimental ortogonal utilizado para preparo das soluções padrão foi determinado por Brereton [67], o qual apresenta 5 níveis de concentração para cada um dos constituintes das soluções padrão em um conjunto de 25 misturas.

As soluções padrão foram preparadas através da pesagem em balança analítica da massa dos açúcares diretamente para balões volumétricos de 100 mL, aos quais também

foram adicionados 50 mL de solução de KCl 2M e então aferidos com água de forma a se obter um fundo iônico de 1M. Os níveis reais de concentração individual dos açúcares variaram de 0,300 g/L a 8,167 g/L, valores dentro do intervalo de concentração utilizado para construção das curvas de calibração do HPLC. Tabela 4 apresenta os níveis e as concentrações dos açúcares nas soluções padrão. O mesmo procedimento foi realizado para o preparo de soluções padrões sem fundo iônico, não sendo adicionada a estas a solução salina de KCl.

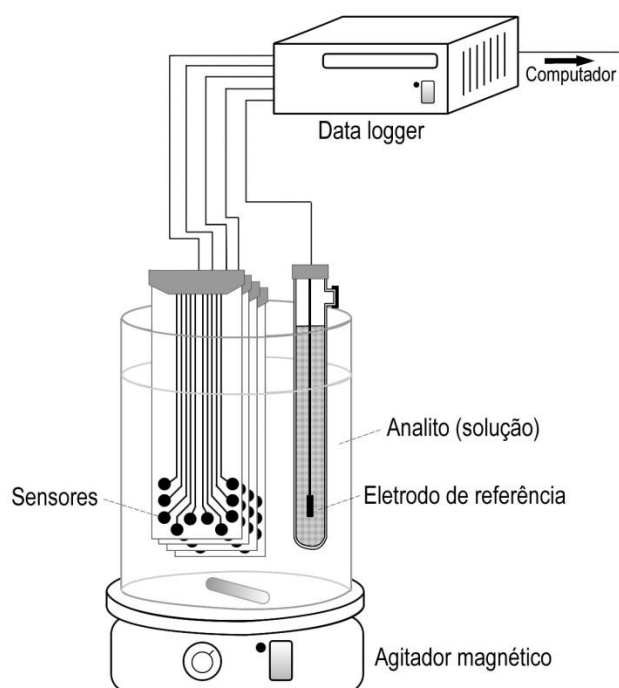


Figura 8. Esquema da montagem do sistema de multi-sensores

Na Figura 9 está representado graficamente o desenho experimental ortogonal para 25 soluções de misturas ternárias dos açúcares em 5 níveis de concentração diferentes. É possível notar que não há correlações entre as concentrações dos açúcares e que estas estão distribuídas de maneira a ocupar todo o espaço amostral. Também se verifica que o desenho permitiu obter valores de concentração de açúcares totais com aceitável variabilidade, no intervalo de 1,7 a 21,1 g/L.

Tabela 4. Níveis e concentrações dos açúcares nas soluções padrão com fundo iônico de KCl 1M

Solução	Nível			Concentração (g/L)			
	Comp. 1	Comp. 2	Comp. 3	Glucose	Frutose	Sacarose	AT
1	0	0	0	2,005	2,064	2,221	6,290
2	0	-2	-1	2,025	0,360	1,044	3,429
3	-2	-1	-2	0,327	1,047	0,304	1,678
4	-1	-2	2	1,022	0,330	8,029	9,381
5	-2	2	2	0,316	8,031	8,018	16,365
6	2	2	0	8,026	8,033	2,015	18,074
7	2	0	-1	8,003	2,043	1,006	11,052
8	0	-1	2	2,000	1,025	8,149	11,174
9	-1	2	-1	1,009	8,038	1,018	10,065
10	2	-1	1	8,018	1,077	5,020	14,115
11	-1	1	1	1,000	5,181	5,019	11,200
12	1	1	0	5,201	5,036	2,028	12,265
13	1	0	2	5,017	2,089	8,028	15,134
14	0	2	1	2,091	8,035	5,052	15,178
15	2	1	2	8,045	5,013	8,094	21,152
16	1	2	-2	5,031	8,167	0,450	13,648
17	2	-2	-2	8,012	0,300	0,340	8,652
18	-2	-2	0	0,304	0,331	2,020	2,655
19	-2	0	1	0,335	2,076	5,033	7,444
20	0	1	-2	2,069	5,085	0,324	7,478
21	1	-2	1	5,038	0,379	5,092	10,509
22	-2	1	-1	0,312	5,230	1,001	6,543
23	1	-1	-1	5,021	1,046	1,002	7,069
24	-1	-1	0	1,013	1,058	0,339	2,410
25	-1	0	-2	1,037	2,041	0,382	3,460

AT – Açúcares totais

Preparo das soluções de mel. Para serem analisadas com a língua eletrônica, foram preparadas três soluções mãe por dissolução de cada um dos méis. Primeiramente, cerca de 9,0 g de amostra foram pesadas diretamente para balões volumétricos de 50 mL, que foram aferidos com água desionizada. A partir de cada uma destas soluções foram preparadas diluições transferindo-se com auxílio de micropipeta 2,0, 3,5 e 5,0 mL destas para balões volumétricos de 50 mL, perfazendo assim 9 soluções. Foram também preparadas soluções de misturas binárias em três diferentes

níveis das soluções mãe (0,5 mL + 0,5 mL; 1,0 mL + 1,0 mL e 2 mL + 2 mL em balões volumétricos de 50 mL), somando outras 9 soluções. Estas 18 soluções foram então analisadas em paralelo com as soluções padrão descritas anteriormente, sendo ao longo das análises realizadas aleatoriamente três análises repetidas de soluções de mel. O procedimento foi realizado de forma a se obter soluções de mel com e sem fundo iônico de KCl 1M.

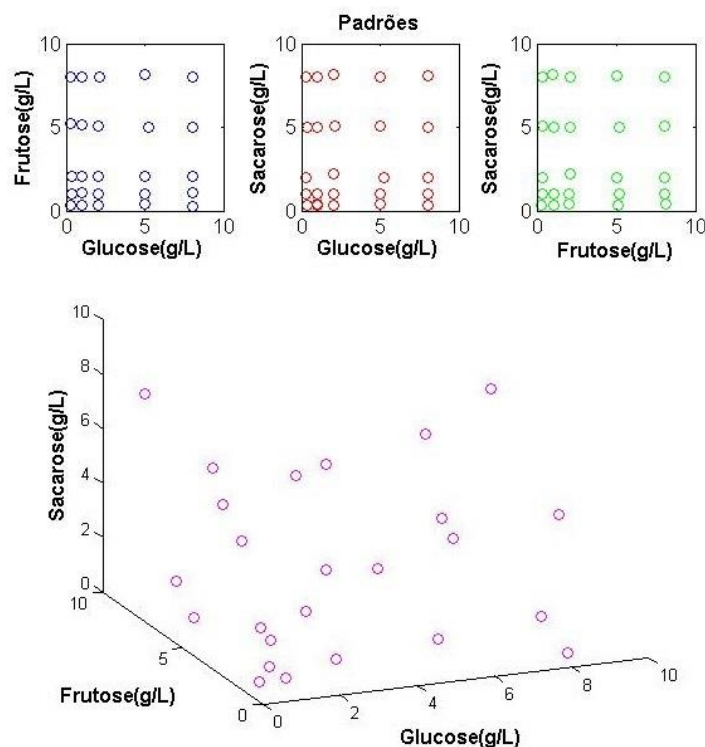


Figura 9. Desenho experimental ortogonal utilizado no preparo das soluções padrão de misturas de glucose, frutose e sacarose.

Leituras. As leituras com a língua eletrônica foram efetuadas de forma direta através da imersão dos sensores em cada uma das soluções agitadas magneticamente durante 5 minutos à temperatura ambiente. Entre cada análise o sistema analítico foi lavado cuidadosamente com água desionizada e levemente enxugado com papel absorvente.

2.4. Análise de dados

Os dados experimentais obtidos através das análises com a língua eletrónica foram tratados com o programa de estatística R versão 3.2.0 (The R Foundation for Statistical Computing, Vienna, Austria) e no MATLAB R2008b (The MathWorks Inc., Natick, USA).

O trabalho incidiu em estabelecer um procedimento de tratamento de dados com a finalidade de quantificar os açúcares em soluções padrão e de mel a partir dos sinais potenciométricos do sistema de multi-sensores através da construção de modelos de regressão. Para isto, primeiramente foi feita uma análise exploratória do conjunto de dados através da construção dos gráficos de extremos e quartis, para se verificar quais sensores apresentavam maior variabilidade nas respostas e que, por isso, poderiam apresentar grande peso na construção dos modelos de regressão, podendo prejudicá-los caso a variabilidade seja em decorrência da presença de interferentes, carregue informações redundantes ou seja consequência da má afinidade entre os sensores e os analitos. Além disso, foi também realizada uma PCA para se verificar o comportamento da nuvem de dados no que diz respeito à separação qualitativa entre soluções padrão e soluções de mel possíveis tendências qualitativas.

Ainda, antes de se iniciar a construção dos modelos de regressão, foi feito o pré-processamento dos sinais potenciométricos centralizando-os na média e escalonando-os (maiores valores assumem valor 1 e menores valores assumem valor 0) com a finalidade de se minimizar o efeito das diferenças de magnitude na intensidade dos sinais.

2.4.1. Modelos de regressão linear múltipla

Como abordagem inicial testou-se a construção de modelos de MLR com seleção de variáveis (sensores) usando algoritmo *stepwise*, mistura dos métodos de seleção sequencial para frente e seleção sequencial por eliminação. Após a eliminação dos sensores que apresentavam grande variabilidade e realização dos pré-processamentos, foram construídos modelos de estimação separadamente para cada açúcar e açúcares totais para os conjuntos de soluções padrão e soluções de mel, na ausência e presença de fundo iônico, a fim de se verificar a possibilidade do estabelecimento de uma relação linear entre a concentração dos açúcares e a intensidade dos sinais potenciométricos.

Em seguida, para se avaliar a capacidade preditiva dos modelos de MLR com algoritmo *stepwise*, os conjuntos de dados de soluções padrão e de mel foram separados em grupos de treino, para construção dos modelos, e teste, para verificação da capacidade dos modelos, através do algoritmo de Kennard & Stone [51]. O número de amostras selecionadas para os subconjuntos de treino e teste foi um parâmetro testado tendo em vista a obtenção de modelos satisfatórios com o menor número de soluções de calibração. Após a aplicação dos modelos para previsão da concentração de cada um dos açúcares e açúcares totais, os resultados previstos foram comparados com as concentrações verdadeiras destes compostos através da construção de modelos de regressão linear simples e avaliação dos coeficientes de determinação, declive e ordenada na origem.

2.4.2. Modelos com método dos mínimos quadrados parciais

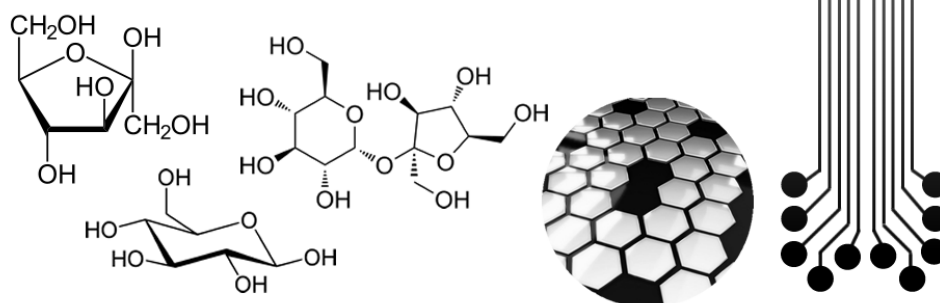
Outra ferramenta testada para construção de modelos de previsão foi a regressão pelo método PLS. Da mesma forma que para os modelos de MLR, as soluções padrão e soluções de mel foram separadas em grupos de treino e teste através do algoritmo de Kennard & Stone [51], sendo o número de observações destinadas para o grupo de treino o menor possível para que se pudessem obter modelos aceitáveis.

Com esta metodologia, a seleção das variáveis (sensores) dos modelos foi feita utilizando-se o algoritmo genético (GA) desenvolvido para o estudo presente de forma a permitir que o desempenho da previsão dos dados do grupo de teste fosse o objeto alvo de avaliação. A verificação da capacidade preditiva dos modelos foi realizada aplicando o modelo obtido com os dados de treino aos dados de teste e, comparando os resultados previstos no teste com as concentrações reais dos açúcares nas soluções. A seleção de variáveis utilizando o GA foi feita analisando-se os parâmetros da regressão linear simples realizada para os dados do grupo de teste e, ajustando-se parâmetros do algoritmo, como probabilidade de cruzamento e mutação de forma a se obter modelos de previsão com maiores valores possíveis para o critério de ajuste usado.

Em virtude do número relativamente reduzido de observações, todos os modelos referidos, seja os modelos de MLR ou de PLS com seleção de variáveis usando algoritmo *stepwise* ou com o GA, foram construídos com validação cruzada *leave-one-out* e *k-folds* (10 grupos com 10 repetições) bem como, para os dados originais.



3. RESULTADOS E DISCUSSÕES



3.1. Análises por HPLC

Visando determinar as concentrações dos açúcares majoritários presentes nas amostras, estas foram analisadas em HPLC. A Figura 10 apresenta um exemplo de cromatograma típico de cada amostra de mel obtidos com uma coluna Supelcogel C-610H (30 cm x 7,8 mm ID de dimensão) e um detetor IR. Através da comparação com padrões, verifica-se que o primeiro açúcar a eluir é a sacarose, com um tempo de retenção médio de 10,0 min, seguido dos picos da glucose e frutose, com tempos de retenção médios de 11,7 min e 12,7 min, respectivamente.

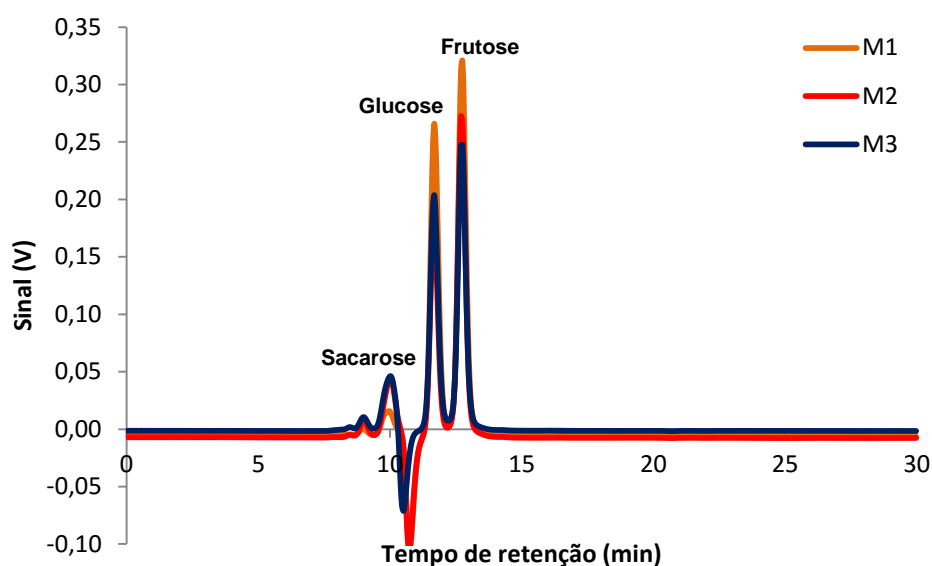


Figura 10. Cromatogramas das amostras de mel

A quantificação dos açúcares foi efetuada através da construção de curvas de calibração individuais para cada composto obtidas através da análise em duplicata de sete soluções padrão com misturas em diferentes concentrações dos açúcares. A Figura 11 apresenta as curvas de calibração obtidas através da relação linear entre a concentração dos açúcares e área dos seus respectivos picos. Na Tabela 5 indicam-se os intervalos das concentrações de cada açúcar nas soluções padrão, os parâmetros de regressão linear de cada curva (declive, ordenada na origem e coeficiente de correlação) e seus limites de detecção e quantificação calculados a partir do ajuste.

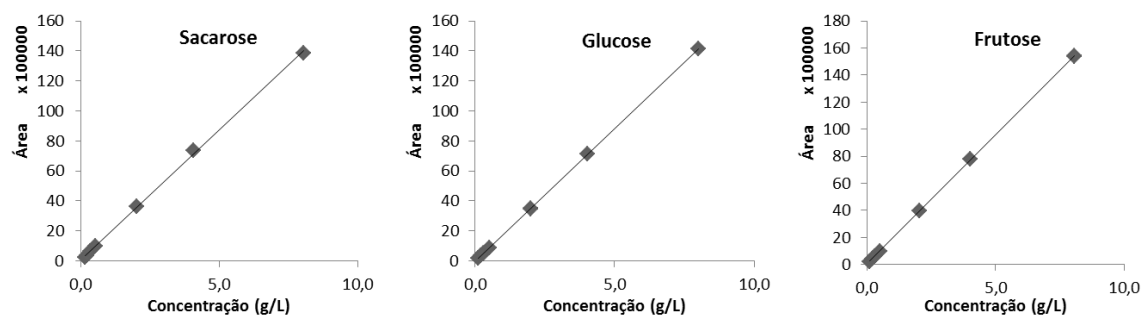


Figura 11. Curvas de calibração para sacarose, glucose e frutose obtidas por HPLC

Tabela 5. Intervalos de concentrações e parâmetros de calibração para análise dos açúcares em HPLC

Composto	Cmin (g/L)	Cmax (g/L)	a(±s)	b(±s)	R	LD (g/L)	LQ (g/L)
Sacarose	0,156	8,038	$1,7(\pm 0,1) \times 10^6$	$7,1(\pm 0,5) \times 10^4$	0,99962	0,091	0,276
Glucose	0,110	8,016	$1,77(\pm 0,01) \times 10^6$	$-3,6(\pm 0,1) \times 10^4$	0,99998	0,021	0,063
Frutose	0,121	8,064	$1,91(\pm 0,01) \times 10^6$	$1,1(\pm 0,1) \times 10^4$	0,99997	0,024	0,074

Cmin – concentração mínima; Cmax – concentração máxima; a – inclinação; b – ordenada na origem; s – desvio padrão; R – coeficiente de correlação; LD – limite de detecção; LQ – limite de quantificação.

Foram obtidos resultados semelhantes para a glucose e a frutose, com limites de detecção na ordem de 0,02 g/L e limites de quantificação de 0,06 – 0,07 g/L. Os limites de detecção e quantificação da sacarose, 0,091 g/L e 0,276 g/L, respectivamente, embora sejam superiores, são aceitáveis uma vez que a mínima concentração utilizada para preparo das soluções padrão (0,156 g/L) é superior ao limite de detecção e, a concentração de sacarose mínima esperada nas diluições das amostras de mel é superior ao limite de quantificação. Os três ajustes apresentaram bons coeficientes de correlação ($R > 0,999$).

Para avaliar a repetibilidade e exatidão das análises, duas soluções de controlo de qualidade com níveis intermediários de concentração dos açúcares (1,0 g/L e 6,0 g/L) foram analisadas em triplicata. Os desvios padrão relativos percentuais e erros relativos percentuais das concentrações foram calculados e estão apresentados na Tabela 6.

Embora, em geral, os valores mais elevados de desvio padrão relativo percentual tenham sido obtidos para a solução de controlo de qualidade de concentrações mais elevadas, os resultados de repetibilidade são satisfatórios com desvios padrão relativos inferiores a 2% para todos os açúcares. Por outro lado, os maiores valores de erro relativo percentual foram obtidos para a solução de menores concentrações dos açúcares, mas ainda assim são considerados aceitáveis ($\leq 5\%$).

Tabela 6. Avaliação da repetibilidade e exatidão das soluções de controlo de qualidade

	Composto	Ce (g/L)	Cm±s (g/L)	Sr%	Er%
Solução 1	Sacarose	1,034	1,009±0,003	0,30	2,42
	Glucose	1,002	1,062±0,006	0,54	4,99
	Frutose	1,032	1,038±0,002	0,17	0,54
Solução 2	Sacarose	6,042	6,3±0,1	1,77	3,87
	Glucose	6,338	6,61±0,08	1,14	4,23
	Frutose	6,630	6,817±0,003	0,04	2,82

Ce – concentração esperada; Cm – concentração medida; s – desvio padrão; Sr – desvio padrão relativo percentual; Er – erro relativo percentual.

Após verificação da qualidade dos modelos de calibração obtidos, foi então possível calcular a concentração de sacarose, glucose e frutose nas amostras de méis. Cada solução de amostra diluída foi analisada em duplicata para que fosse possível determinar o desvio padrão relativo percentual. Na Tabela 7 estão apresentados os resultados expressos em massa de açúcar por massa de mel, os açúcares totais (AT), as razões de intolerância à frutose livre (FI_{free}) e total (FI_{total}) e índice de doçura em sacarose equivalente (Sac.eq). Considerou-se que os açúcares totais do mel resultam da soma dos três açúcares maioritários analisados, entretanto, no caso da sacarose pode existir a influência de outros dissacarídeos presentes em baixas concentrações.

Tabela 7. Caracterização dos açúcares nas amostras de mel

Composto	Mel 1		Mel 2		Mel 3	
	Cm±s (g/100g_{mel})	Sr%	Cm±s (g/100g_{mel})	Sr%	Cm±s (g/100g_{mel})	Sr%
Sacarose	4,22±0,02	0,47	12,51±0,04	0,28	12,37±0,05	0,37
Glucose	36,9±0,2	0,31	31,5±0,5	1,48	31,90±0,01	0,03
Frutose	44,13±0,01	0,03	43,43±0,05	0,11	39,5±0,1	0,32
AT	85,3±0,1	0,13	87,4±0,5	0,55	83,8±0,2	0,21
Índice	Mel 1		Mel 2		Mel 3	
FI_{free}	0,84±0,002	0,28	0,73±0,01	1,37	0,81±0,002	0,29
FI_{total}	0,85±0,002	0,25	0,78±0,01	0,96	0,85±0,002	0,21
Sac.eq	107,89±0,09	0,08	110,9±0,4	0,35	104,3±0,3	0,26

Cm – concentração medida; s – desvio padrão; Sr – desvio padrão relativo; AT – açúcares totais; FI_{free} – razão de intolerância à frutose livre; FI_{total} – razão de intolerância à frutose total.

Os resultados de repetibilidade das análises em HPLC das amostras de mel são satisfatórios com desvios padrão relativos inferiores a 2% para todos os açúcares, para os açúcares totais e para as razões de intolerância à frutose, mostrando que a

determinação da concentração dos açúcares através deste método é adequada para amostras de mel.

Verificou-se que as três amostras de méis apresentam maior concentração em frutose, seguido de glucose e por fim sacarose, sendo que a menor concentração em sacarose foi observada no Mel 1 (o único multifloral), cerca de três vezes menor, enquanto que as concentrações dos outros açúcares são semelhantes. Os níveis de glucose e frutose obtidos estão de acordo com os referidos em outros trabalhos [49,67], nos quais constam que em geral a percentagem desses açúcares varia de 19 – 40% e 28 – 45%, respectivamente. Entretanto, os níveis de sacarose obtidos, em especial para as amostras de méis monoflorais, se encontram em níveis superiores aos referidos por tais estudos (0,1 – 4,8%). Com isso, as amostras de mel analisadas apresentaram razões de intolerância a frutose livre e total inferiores a 1 e por isso podemos afirmar que estas amostras não são adequadas para pessoas que sofrem de intolerância à esse açúcar. Ainda, em relação ao impacto sensorial, através do cálculo do índice de doçura em sacarose equivalente (Sac.eq) pode-se dizer que o Mel 2 é o que apresenta o sabor doce mais intenso, embora não se diferencie muito das outras amostras.

Através do cálculo das massas relativas dos açúcares nas amostras de méis, foi possível determinar a concentração destes nas diluições preparadas para as análises com a língua eletrônica, apresentando-se na Figura 12 a relação das concentrações entre os açúcares nas soluções de mel.

É possível observar que existe certa correlação entre a concentração dos açúcares nas soluções de mel preparadas, o que é de se esperar uma vez que os méis apresentam uma proporção relativa natural de frutose, glucose e sacarose e, quanto mais diluída for a solução menor vai ser a concentração de todos os seus açúcares, aumentando à medida que a concentração de mel na diluição aumenta. Além disso, em função da maior proporção relativa natural de frutose e glucose em relação à sacarose no mel, os intervalos de concentração para estes açúcares são mais amplos e são onde as amostras mais se distinguem.

Em função das concentrações dos açúcares nas soluções de mel apresentarem colinearidade, cabe salientar a importância de se realizar o estudo em paralelo com as soluções padrão preparadas em desenho experimental de níveis ortogonais de misturas dos três açúcares, uma vez que a colinearidade pode representar um problema nas análises com a língua eletrônica.

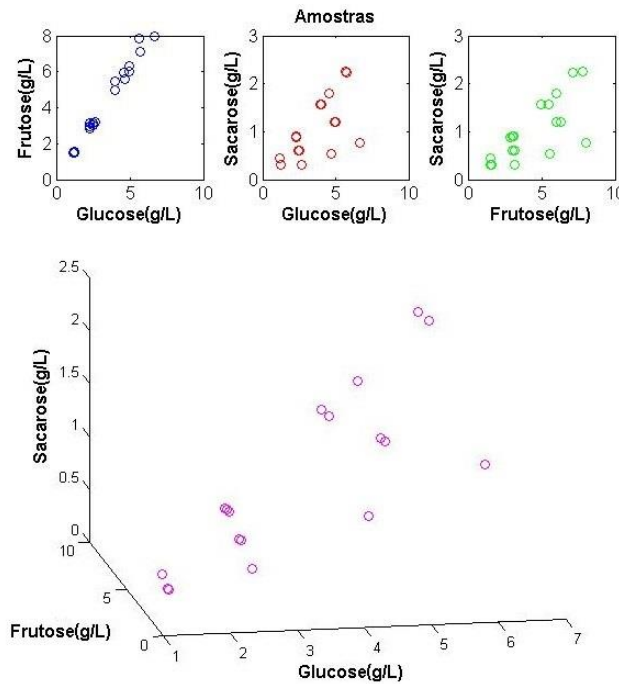


Figura 12. Concentração relativa dos açúcares nas soluções de méis para análises com língua eletrônica

Com a qualidade da quantificação verificada para a glucose, frutose e sacarose, os valores obtidos para as soluções de mel podem ser admitidos como as concentrações reais observadas (valores de referência) para que seja possível avaliar a quantificação dos açúcares através da língua eletrônica, comparando-os com os previstos pelos modelos.

3.2. Análises com língua eletrônica

Neste trabalho foi utilizada uma língua eletrônica potenciométrica com sensores de sensibilidade cruzada com o objetivo de se verificar a possibilidade da análise quantitativa simultânea de sacarose, glucose, frutose e açúcares totais em soluções de mel e em soluções padrão de misturas dos três açúcares.

A Figura 13 apresenta um exemplo da variação do potencial elétrico para os diferentes sensores da língua eletrônica em função do tempo para uma solução diluída de mel. É possível observar que existe uma estabilização da intensidade de cada sensor ao longo do tempo após a inserção do sistema de multi-sensores na solução analisada.

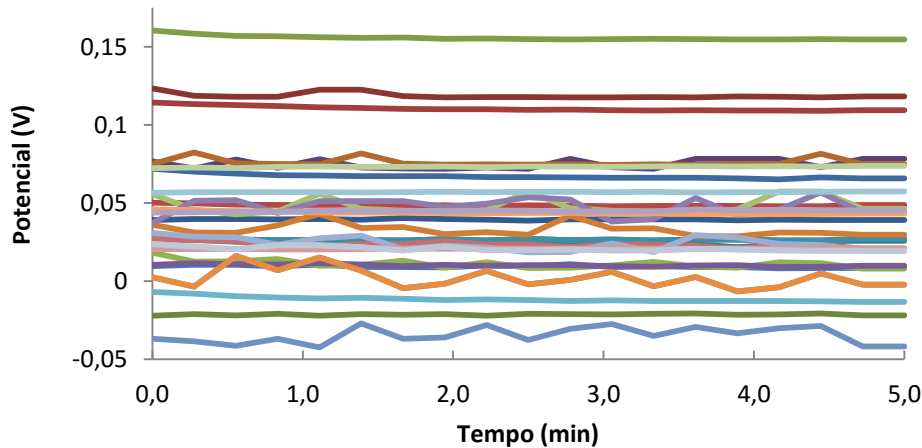


Figura 13. Variação da intensidade do sinal potenciométrico da língua electrónica em função do tempo para uma solução de mel

Todas as soluções de mel (preparadas com cada mel ou por mistura de dois méis) foram analisadas com a língua electrónica simultaneamente com as soluções padrão de misturas dos açúcares de maneira aleatória. Na Figura 14 estão representados os perfis dos sinais da língua electrónica obtidos ao final de 5 minutos de análise para todos os ensaios com e sem a presença de fundo iônico de KCl 1M e, na Figura 15 os gráficos de extremos e quartis para cada caso.

É possível observar que com a presença do fundo iônico, que tem por finalidade uniformizar a matriz da solução descontando possíveis interferências em virtude da matriz das amostras, a variabilidade da intensidade dos sinais em cada sensor é muito menor (variando numa amplitude média de 0,0075 V, não sendo considerados os sensores 11, 12, 20, 29 e 40 por apresentarem variabilidade muito elevada na intensidade dos sinais ou muitos valores extremos). Entretanto, nesta situação como a maioria das respostas de alguns sensores varia muito pouco, as intensidades dos sinais em cada análise se aproximam muito da resposta média para estes sensores, reduzindo o intervalo dos quartis e, em função disso, fazendo com que existam mais observações com valores extremos. Para as análises sem fundo iônico, a variabilidade da intensidade dos sinais é maior (com amplitude média de 0,0636 V, não sendo considerados os sensores 11, 12, 20, 29 e 40, como já referido) e, por isso, os intervalos dos quartis são mais largos, permitindo que mesmo respostas mais distantes da resposta média não sejam consideradas extremos. Para ambos os casos, as observações discordantes são pontuais, não sendo necessária a remoção de nenhuma solução para análise.

Além disso, observa-se que sensores com maior variabilidade de resposta se destacam muito mais nas análises com soluções na presença de fundo iônico, o que pode ter uma influência maior na construção dos modelos para quantificação dos açúcares, fazendo com que estes sejam altamente influenciados pela resposta destes poucos sensores que podem carregar informações redundantes e prejudicar o desempenho dos modelos.

Isto reforça a ideia já referida de que a seleção de variáveis (de sensores, no caso) é uma tarefa importante na modelagem multivariada, não só para se reduzir o número de termos necessários para explicação do problema, mas também para evitar que sejam usadas variáveis que não carregam informações relevantes.

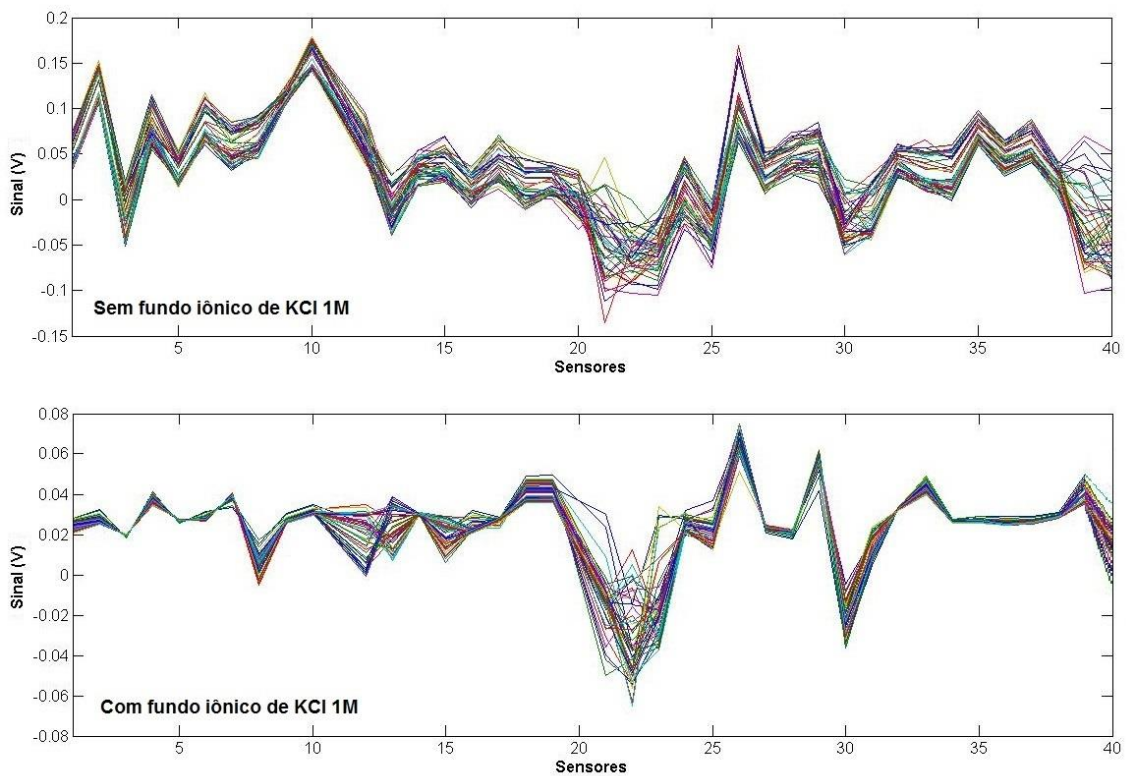


Figura 14. Perfis dos sinais potenciométricos das soluções padrão e de mel com e sem fundo iônico de KCl 1M

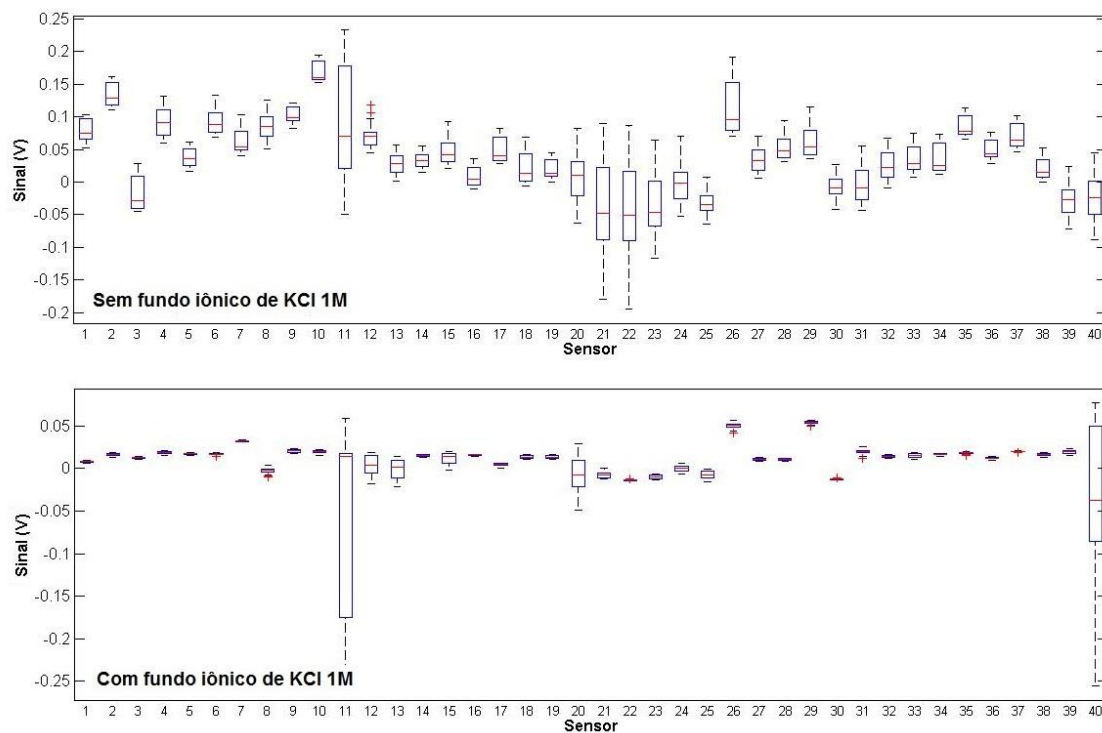


Figura 15. Extremos e quartis dos sinais potenciométricos das soluções padrão e de mel com e sem fundo iônico de KCl 1M

Outra maneira de se verificar a existência observações discordantes é através de uma análise exploratória dos dados através da PCA, que é uma técnica multivariada não supervisionada de redução da dimensionalidade dos dados que permite a visualização de agrupamentos de amostras.

A Figura 16 apresenta uma PCA realizada apenas com as respostas dos sensores após centralização e standardização dos dados para soluções padrão e de mel com fundo iônico medidas simultaneamente, com a remoção dos sensores 11, 12, 20, 29 e 40, por serem os com maior variabilidade na resposta, conforme verificado na Figura 15. Neste caso não foi possível observar agrupamentos ou tendências no espaço tridimensional que explica 79,85% da variabilidade dos dados, evidenciando sua dispersão aleatória não enviesada, permitindo a sua aplicação na construção de modelos de previsão. Não foram observados *outliers* na nuvem de dispersão e por isso não é necessário remover nenhuma das soluções. O mesmo foi observado para o conjunto de soluções medidas sem fundo iônico.

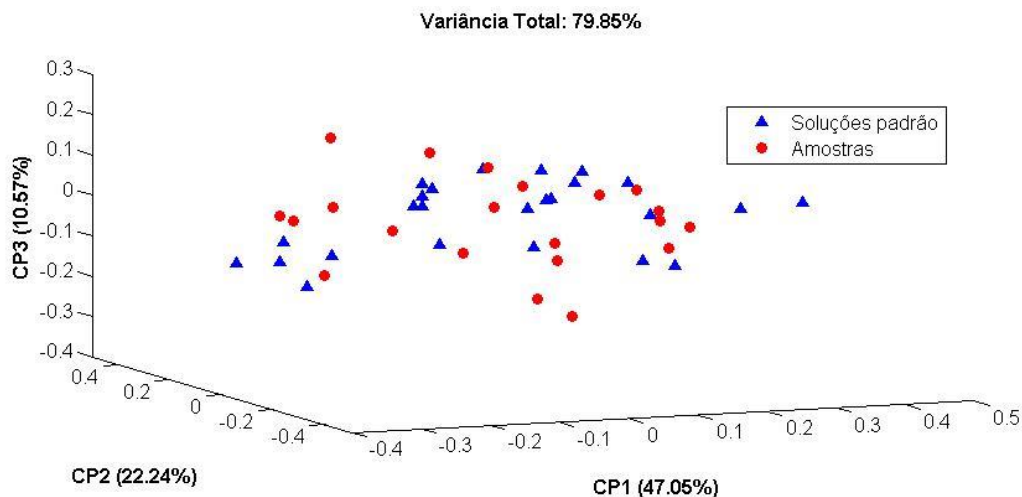


Figura 16. PCA para soluções padrão e de mel medidas simultaneamente

3.3. Modelos multivariados de estimação e previsão

Com os resultados das análises utilizando a língua electrónica, pretendeu-se estudar a construção de modelos multivariados capazes de relacionar os sinais potenciométricos do sistema de multi-sensores com a concentração de açúcares das soluções analisadas. Neste sentido, o maior problema a ser solucionado é a escolha de quais sensores devem ser considerados na construção destes modelos, uma vez que o uso de todos eles pode produzir bons modelos de ajuste para os dados de calibração, mas apresentarem mau desempenho na previsão para conjuntos de dados que não estavam presentes na calibração. Desta forma, a redução do número de variáveis predictoras pode evitar efeitos de sobre-ajuste e melhorar a previsão dos modelos.

Como primeira abordagem, para estudar a capacidade dos sensores em responder à concentração dos compostos, foi utilizado o método MLR com seleção de variáveis usando algoritmo *stepwise* para cada um dos açúcares (glucose, frutose e sacarose) e açúcares totais, obtendo-se diversos modelos de estimação e previsão para cada analito para diferentes conjuntos de dados.

3.3.1. Modelos MLR de estimação usando algoritmo *stepwise*

Para verificar se é possível obter uma estimação das concentrações dos açúcares com base nos sinais potenciométricos, modelos de MLR para cada açúcar e açúcares totais foram construídos usando todos os dados no treino do modelo para os quatro

grupos: soluções padrão e de mel, com e sem fundo iônico. O algoritmo para construção de modelos de MLR foi definido para se obter modelos de estimação contendo de 2 a 20 sensores selecionados usando o algoritmo *stepwise*.

A Tabela 8 apresenta um resumo dos modelos MLR obtidos com os sensores selecionados para construção de cada modelo de estimação. Foram escolhidos os modelos lineares que apresentaram o menor número de variáveis (sensores) nos quais fossem obtidos simultaneamente valores superiores a 0,999 para o declive e para o coeficiente de determinação ajustado na relação linear entre os valores obtidos pela previsão do modelo e os valores esperados.

Tabela 8. Modelos de estimação de MLR obtidos usando algoritmo *stepwise*

Soluções sem fundo iônico					
Soluções padrão					
Analito	N ^o sensores	Sensores	R ²	a	b
Glucose	19	1, 2, 5, 7, 9, 10, 13, 14, 16, 17, 18, 21, 22, 24, 27, 28, 32, 37, 38	0,99997	0,99999	NS
Frutose	18	1, 3, 8, 9, 10, 14, 17, 21, 23, 24, 26, 27, 28, 32, 33, 35, 37, 39	0,99923	0,99982	NS
Sacarose	17	1, 3, 4, 5, 7, 8, 10, 13, 14, 19, 21, 24, 25, 26, 30, 37, 38	0,99932	0,99980	NS
AT	18	5, 7, 8, 9, 13, 16, 17, 19, 23, 25, 26, 30, 33, 35, 36, 37, 38, 39	0,99955	0,99989	NS
Soluções de mel					
Analito	N ^o sensores	Sensores	R ²	a	b
Glucose	14	5, 8, 10, 13, 18, 23, 26, 28, 32, 33, 34, 35, 36, 37	0,99942	0,99982	NS
Frutose	14	1, 6, 13, 18, 19, 21, 22, 23, 24, 30, 33, 34, 35, 38	0,99967	0,99990	NS
Sacarose	12	1, 3, 4, 8, 10, 14, 24, 27, 32, 34, 35, 39	0,99909	0,99964	NS
AT	14	1, 2, 4, 5, 9, 17, 22, 24, 26, 27, 28, 31, 35, 36	0,99947	0,99984	NS
Soluções com fundo iônico de KCl 1M					
Soluções padrão					
Analito	N ^o sensores	Sensores	R ²	a	b
Glucose	17	1, 2, 7, 8, 10, 16, 17, 19, 22, 24, 26, 31, 33, 35, 36, 37, 38	0,99946	0,99986	NS
Frutose	17	1, 2, 3, 4, 5, 6, 7, 16, 17, 21, 22, 23, 24, 32, 33, 34, 36	0,99923	0,99981	NS
Sacarose	18	1, 2, 5, 6, 8, 14, 17, 19, 21, 23, 24, 25, 26, 28, 30, 33, 35, 38	0,99933	0,99983	NS
AT	17	2, 3, 6, 10, 16, 18, 19, 21, 22, 26, 27, 30, 31, 35, 37, 38, 39	0,99912	0,99975	NS
Soluções de mel					
Analito	N ^o sensores	Sensores	R ²	a	b
Glucose	14	7, 8, 13, 14, 15, 17, 19, 21, 23, 30, 33, 34, 37, 38	0,99915	0,99975	NS
Frutose	14	1, 3, 6, 7, 14, 16, 17, 23, 24, 26, 27, 28, 32, 34	0,99931	0,99979	NS
Sacarose	14	1, 5, 7, 10, 15, 16, 17, 18, 31, 32, 33, 34, 35	0,99971	0,99991	NS
AT	14	1, 2, 4, 5, 9, 17, 22, 24, 26, 27, 28, 31, 35, 36	0,99947	0,99984	NS

AT – açúcares totais; R² – coeficiente de determinação ajustado; a – declive; b – ordenada na origem; NS – não significativo.

Optou-se por utilizar o coeficiente de determinação ajustado uma vez que neste é feita uma correção dos graus de liberdade, impedindo que seu valor seja sempre crescente com a adição de mais variáveis explanatórias, ou seja, é uma forma de penalizar adições de variáveis que não contibuem significativamente para o modelo.

Foi observado que para todos os açúcares, seja para as soluções padrão ou para as de mel, com ou sem fundo iônico, é possível se obter modelos de estimação. O número de sensores necessários para construção destes modelos para que se obtenham simultaneamente valores de declive e coeficiente de determinação ajustado superiores a 0,999 varia de 12 a 19, verificando-se maior numero de sensores para os modelos de estimação obtidos para as soluções padrão. Em geral, não foram verificadas grandes diferenças entre o número de sensores dos modelos para as soluções com ou sem fundo iônico.

Cabe destacar que existe uma grande variação entre os modelos no que diz respeito aos sensores selecionados, não existindo modelos com os mesmos sensores capazes de estimar respostas diferentes. Além disso, o conjunto de sensores obtido para cada situação não é único para que o critério da regressão seja atingido, ou seja, é possível se obter outros modelos com o mesmo número de sensores mas usando diferentes sensores que permitem obter parâmetros semelhantes de qualidade no ajuste linear entre os valores obtidos pelos modelos e os valores esperados.

Na Figura 17 estão apresentados exemplos comparando os valores reais da concentração de açúcar (glucose) com os valores estimados através da regressão para as soluções padrão e de mel, com ou sem fundo iônico. Estes resultados mostram que a língua eletrônica fornece informação analítica quantitativa sobre os açúcares utilizando os sensores selecionados. Resultados semelhantes são obtidos para os outros açúcares.

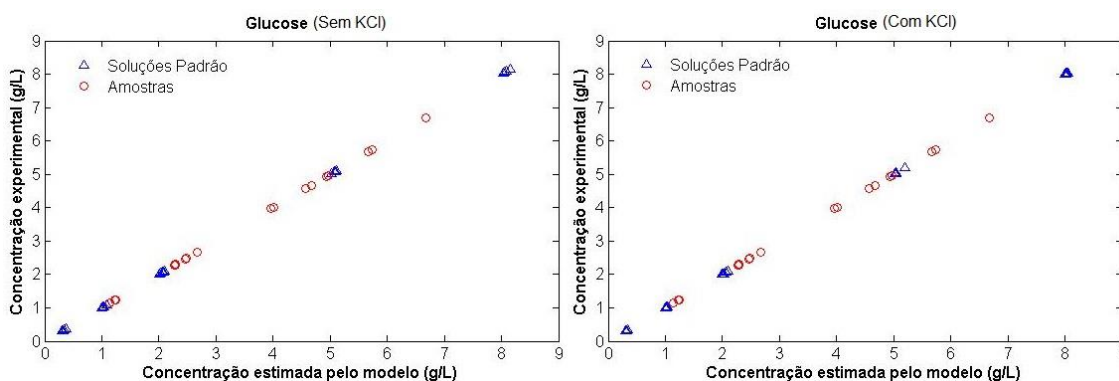


Figura 17. Comparações das concentrações experimentais e estimadas de glucose usando os modelos de estimação de MLR obtidos usando algoritmo *stepwise*

A seguir estuda-se o procedimento matemático para a obtenção de modelos de previsão que possam ser usados na quantificação dos açúcares em soluções que não pertencem aos grupos de treino.

3.3.2. Separação dos dados

A Figura 18 apresenta um exemplo da aplicação do algoritmo Kennard-Stone para as soluções de mel com fundo iônico usando os perfis de sinais obtidos da análise com a língua eletrônica, onde se mostra as 7 soluções de mel selecionadas para o grupo de teste em um espaço amostral de 21 observações.

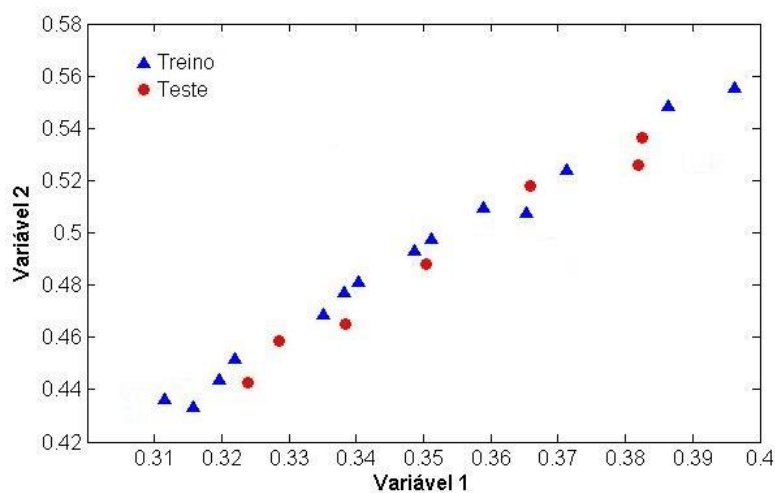


Figura 18. Exemplo de seleção de variáveis utilizando o algoritmo de Kennard & Stone

O número de observações destinadas para os grupos de treino foi definido a partir de diversas tentativas de forma a se obter os primeiros modelos de previsão satisfatórios com o menor número de soluções de treino. Isso porque a ideia é obter uma metodologia de calibração rápida, logo, um menor número de análises na calibração é um fator de interesse. A Tabela 9 apresenta a separação dos dados em grupos de treino e teste para as soluções padrão e de mel, na presença ou ausência de fundo iônico.

Tabela 9. Separação dos dados em conjuntos de treino e teste para construção dos modelos de regressão

Soluções sem fundo iônico		
Soluções Padrão	Treino (19)	1, 2, 3, 4, 5, 6, 8, 9, 13, 14, 16, 17, 18, 19, 20, 22, 23, 24, 25
	Teste (6)	7, 10, 11, 12, 15, 21
Soluções de mel	Treino (14)	1, 2, 3, 5, 8, 9, 11, 12, 14, 15, 17, 19, 20, 21
	Teste (7)	4, 6, 7, 10, 13, 16, 18
Soluções com fundo iônico de KCl 1M		
Soluções Padrão	Treino (19)	1, 2, 3, 4, 7, 8, 9, 11, 12, 13, 14, 15, 16, 17, 18, 21, 22, 23, 24
	Teste (6)	5, 6, 10, 19, 20, 25
Soluções de mel	Treino (14)	1, 3, 5, 7, 8, 10, 12, 13, 14, 17, 18, 19, 20, 21
	Teste (7)	2, 4, 6, 9, 11, 15, 16

Como já referido anteriormente, existem 25 soluções padrão com misturas ternárias ortogonais dos açúcares em diferentes concentrações e 21 soluções de diluições e misturas de mel, numeradas aleatoriamente em ambos os casos. Tanto para as análises com ou sem fundo iônico, o número de soluções destinadas para o treino e teste das soluções padrão ou de mel foi o mesmo – soluções padrão: 19 treino e 6 teste; soluções de mel: 14 treino e 7 teste. É possível observar que embora o número seja igual, as soluções escolhidas para as análises diferentes em relação ao fundo iônico não são as mesmas e, que o número de soluções necessárias para o treino das soluções com mel é inferior em relação ao das soluções padrão. Em geral, o algoritmo Kennard & Stone permitiu selecionar grupos de teste cujos níveis de concentrações se encontram inseridos no grupo de treino, incluindo todos os níveis do desenho experimental.

3.3.3. Modelos de previsão usando o algoritmo *stepwise*

Sendo então comprovada a hipótese da possibilidade de obtenção de modelos de estimação capazes de relacionar os sinais potenciométricos com a concentração de açúcares, foram estudados modelos MLR de previsão com seleção de variáveis usando o algoritmo *stepwise* através da separação dos grupos de dados em observações de treino, para construção dos novos modelos, e observações de teste que não estavam presentes no treinamento para testar a capacidade de previsão destes. Os resultados estão dispostos na Tabela 10.

Tabela 10. Modelos de MLR obtidos usando algoritmo *stepwise* e resultados da previsão para os dados de teste

Soluções sem fundo iônico								
Soluções padrão								
Analito	N ^o sensores	Sensores	Treino		Teste			
			R ²	a	R ²	a	b	
Glucose	14	4, 5, 8, 9, 10, 13, 15, 22, 27, 28, 32, 33, 36, 39	1,0000	1,0000	0,0459	0,7996	NS	
Frutose	13	1, 7, 10, 13, 14, 19, 23, 26, 32, 36, 37, 38, 39	0,9999	1,0000	0,5042	0,1735	7,6252	
Sacarose	12	3, 9, 10, 15, 16, 17, 18, 21, 22, 34, 37, 38	0,9999	0,9999	0,2377	1,0453	0,1167	
AT	12	1, 3, 4, 5, 8, 10, 18, 22, 25, 28, 35, 36	0,9999	0,9999	0,7919	1,2448	NS	
Soluções de mel								
Analito	N ^o sensores	Sensores	Treino		Teste			
			R ²	a	R ²	a	b	
Glucose	8	3, 4, 5, 9, 15, 16, 23, 32	1,0000	1,0000	0,8541	0,8403	NS	
Frutose	8	4, 6, 7, 10, 16, 22, 24, 36	1,0000	1,0000	0,4613	0,6733	NS	
Sacarose	8	5, 6, 9, 10, 18, 24, 37, 38	1,0000	1,0000	0,7562	0,8272	NS	
AT	8	4, 7, 10, 16, 22, 24, 36, 37	1,0000	1,0000	0,4349	0,6607	NS	
Soluções com fundo iônico de KCl 1M								
Soluções padrão								
Analito	N ^o sensores	Sensores	Treino		Teste			
			R ²	a	R ²	a	b	
Glucose	13	2, 3, 4, 7, 10, 18, 22, 23, 27, 30, 33, 36, 38	0,9994	0,9999	0,2981	0,9696	NS	
Frutose	12	6, 8, 10, 13, 16, 21, 22, 25, 26, 31, 33, 39	0,9998	0,9999	0,1174	-0,0371	NS	
Sacarose	13	1, 7, 8, 13, 16, 18, 26, 27, 30, 32, 33, 37, 39	0,9997	1,0000	0,2469	0,2900	NS	
AT	12	10, 13, 15, 16, 18, 19, 21, 22, 23, 28, 33, 36	0,9996	1,0000	0,7719	0,9151	NS	
Soluções de mel								
Analito	N ^o sensores	Sensores	Treino		Teste			
			R ²	a	R ²	a	b	
Glucose	9	8, 10, 17, 18, 28, 31, 36, 38, 39	1,0000	1,0000	0,2760	0,4747	NS	
Frutose	8	5, 7, 9, 21, 22, 27, 28, 31	0,9999	0,9999	0,4741	0,4667	NS	
Sacarose	8	4, 6, 7, 25, 31, 34, 35, 39	0,9999	0,9999	0,0386	0,5208	NS	
AT	8	6, 10, 21, 24, 31, 32, 37, 38	1,0000	1,0000	0,4691	0,5884	NS	

AT – açúcares totais; R² – coeficiente de determinação ajustado; a – declive; b – ordenada na origem; NS – não significativo.

Os modelos foram construídos de forma a se obter subgrupos com o número de sensores entre 2 e o número de observações para os grupos de treino. Como critério de seleção dos modelos de estimação utilizou-se o coeficiente de determinação ajustado superior a 0,999 para os dados do grupo de treino. Estes modelos foram então usados na previsão dos respectivos açúcares usando os dados presentes nos grupos de teste.

Comparando estes resultados com os apresentados na Tabela 8 (modelos MLR de estimação), verificou-se que os modelos de ajuste ao grupo de treino apresentam menor

número de sensores, variando de 8 a 14 sensores. Isso ocorre uma vez que é reduzido o número de observações utilizadas para construção dos modelos, já que parte das observações são usadas para teste e, dessa forma, é de se esperar que menos variáveis sejam necessárias para se obter modelos de estimação. Também se verificou que não se observam alterações relevantes do número de sensores necessários para estimar a concentração dos açúcares nas soluções padrão ou nas soluções de mel com e sem fundo iônico. De maneira análoga à situação obtida nos modelos de estimação, existem diversos modelos com o mesmo número de variáveis capazes de gerarem estimações satisfatórias e, em nenhum dos casos os sensores escolhidos são os mesmos.

A generalização dos mesmos modelos foi testada usando as técnicas de validação cruzada *leave-one-out* e *k-fold* (subgrupos de 10 observações e 10 repetições) não sendo notadas diferenças consideráveis ao nível da avaliação da qualidade dos modelos. Verificou-se também que, embora os modelos obtidos com o grupo de treino tenham apresentado bons resultados de ajuste, estes modelos aplicados aos conjuntos de testes dos dados não produzem boas previsões. Globalmente, os coeficientes de determinação ajustados para a avaliação da qualidade de previsão no grupo de teste (relação linear entre os valores previstos pelo melhor modelo obtido e os esperados) são inferiores a 0,85. Os bons resultados obtidos para o grupo de treino mostram a capacidade de adaptação das mesmas respostas dos sensores para diferentes problemas, ou seja, os sensores utilizados como variáveis preditoras na construção dos modelos, combinados com seus coeficientes estimados, respondem perfeitamente apenas para aquele conjunto de observações sob o qual foi construído. Dessa forma, observações que não estavam presentes na calibração, quando aplicadas ao modelo, apresentam previsões discordantes.

Sendo assim, é necessária uma nova abordagem para seleção de variáveis e ferramenta de regressão para que seja possível a obtenção de modelos que apresentem bons resultados tanto na estimação quanto na previsão sem que haja sobre-ajuste. Para isto, na seção seguinte estudou-se a aplicação do GA desenvolvido para a seleção dos sensores em modelos de regressão.

3.4. Modelos com seleção de variáveis usando algoritmo genético

Após ser verificado que embora tenha sido possível se obter ótimos modelos de estimação capazes de relacionar os sinais potenciométricos das análises com língua

eletrônica com as concentrações dos açúcares nas soluções padrão e de mel, estes modelos quando aplicados aos grupos de testes (dados que não estavam presentes na calibração) produziam previsões muito ruins.

Dessa forma, a primeira alternativa que se buscou para contornar este problema foi continuar a construir modelos de MLR, porém com o GA para seleção de variáveis (sensores). O GA desenvolvido para este estudo foi programado para incluir qualquer ferramenta de modelagem (MLR, PLS, redes neurais artificiais ou outros) e escolher algum dos parâmetros obtidos através dessas ferramentas como critério de desempenho.

Os parâmetros de configuração do GA foram otimizados com o objetivo de selecionar os melhores subconjuntos de sensores que produzissem bons modelos de previsão com o menor número possível de iterações. O número máximo de iterações estabelece o número de tentativas de obter o melhor modelo de previsão antes que a pesquisa GA seja interrompida. Neste trabalho fez-se uma pesquisa do melhor subconjunto de sensores entre 25 e 500 gerações (nº de iterações). Em cada geração implementou-se 10^9 tentativas para se obter um novo melhor modelo de previsão em relação ao estabelecido. Após estas tentativas sem qualquer melhoria, iniciava-se nova iteração. Verificou-se que com 200 iterações era possível obter modelos de previsão satisfatórios ($R > 0,995$) com um tempo de cálculo de 40 minutos para a maioria das situações. O aumento de número de iterações incrementa a qualidade do modelo de previsão, mas também o tempo de cálculo.

Outro parâmetro otimizado foi a taxa de cruzamento (probabilidade de cruzamento entre pares de cromossomos ou $p_{crossover}$), que é a função que forma descendentes através da combinação de parte da informação genética de seus pais. Tipicamente, é um grande valor e por padrão está definido para 0,8. Neste trabalho verificou-se a influência dos seus valores de probabilidade no intervalo de 0,2 a 1 e verificou-se que se estabelecendo o valor de 1, promovia um melhor modelo de previsão para as iterações definidas. Além destes, a mutação é a uma função que altera aleatoriamente os valores de alguns genes num cromossomo pai. O parâmetro $p_{mutation}$ corresponde à probabilidade de mutação num cromossoma pai. Normalmente a mutação ocorre com uma pequena probabilidade e, por padrão está definido para 0,1. Neste trabalho, após o estudo do seu efeito usando valores entre 0,1 e 1, verificou-se que eram obtidos melhores resultados de previsão para as iterações definidas quando o valor foi definido para 1. Por fim, outro parâmetro definido na construção do GA se refere ao elitismo, que corresponde ao número de melhores indivíduos com aptidão para sobreviver a cada geração. Por

padrão, os melhores 5% de indivíduos vão sobreviver a cada iteração. Nas iterações com o GA manteve-se este valor após verificação de ser a melhor probabilidade no intervalo de valores testados entre 10 e 0,5%, na obtenção do melhor modelo de previsão.

Também se testaram vários critérios de qualidade de ajuste cujas iterações do algoritmo genético deveria maximizar (por exemplo, o coeficiente de correlação, coeficiente de determinação ou coeficiente de determinação ajustado para a relação linear entre os valores previstos pelo melhor modelo obtido e os esperados do grupo de teste) ou minimizado no caso do critério de ajuste estar relacionado com os resíduos (por exemplo, a média dos resíduos ou a raiz do erro quadrático médio, RMSE; da relação linear entre os valores previstos pelo melhor modelo obtido e os esperados do grupo de teste). Neste estudo selecionou-se como critério para definição dos modelos, os valores coeficiente de determinação ajustado, pois representa o quanto o modelo consegue explicar os valores observados penalizando adições de variáveis que não contribuem significativamente para melhora dos modelos.

Como já descrito na seção 2.4.2, durante as iterações do GA, os modelos foram primeiramente construídos com o conjunto de dados de treino e então aplicados aos dados do grupo de teste. Os valores previstos obtidos para o grupo de teste foram comparados com os valores reais das concentrações dos açúcares das soluções e o critério de maximização da relação linear era utilizado para se definir nas iterações posteriores os subconjuntos de sensores selecionados. Isto fazia com que o GA procurasse uma solução que buscasse atender o parâmetro de otimização (coeficiente de determinação ajustado, no caso) às custas de um número muito grande de variáveis, em geral acima de 20 sensores, número superior ao de observações para construção dos modelos, o que representa uma evidência muito clara de sobre-ajuste em modelos de MLR, uma vez um dos pressupostos para obtenção de modelos de regressão linear é que o número de variáveis não ultrapasse o número de observações.

Dessa forma, uma alternativa buscada foi a construção de modelos de regressão pelo método PLS usando o GA como ferramenta para seleção de variáveis. Os sensores selecionados e os resultados para os dados do grupo de treino e teste estão apresentados na Tabela 11 e nas Figuras 19 e 20.

Tabela 11. Modelos de PLS obtidos usando algoritmo genético e resultados da previsão para os dados de teste

Soluções sem fundo iônico									
Soluções padrão									
Modelo	Analito	N ^o _{Iter}	N ^o _{sens}	Sensores	R ² _{treino}	R ² _{teste}	a _{teste}	b _{teste}	RMSE _{teste}
A	Glucose	200	20	1, 3, 4, 5, 6, 7, 8, 9, 14, 15, 17, 18, 23, 24, 28, 30, 32, 33, 35, 38	1,0000	0,9947	0,9903	NS	0,2311
B	Frutose	300	20	3, 6, 7, 8, 9, 13, 17, 18, 21, 23, 24, 25, 26, 27, 30, 31, 32, 33, 36, 39	1,0000	0,9916	1,0059	NS	0,4069
C	Sacarose	200	22	1, 2, 4, 5, 7, 9, 13, 14, 15, 16, 18, 19, 22, 24, 25, 26, 30, 31, 32, 35, 36, 39	1,0000	0,9956	0,9866	NS	0,1940
D	AT	300	24	3, 4, 5, 8, 9, 13, 14, 15, 16, 17, 19, 21, 23, 24, 26, 27, 32, 33, 34, 35, 36, 37, 38, 39	1,0000	0,9978	1,0226	NS	0,2072
Soluções de mel									
Modelo	Analito	N ^o _{Iter}	N ^o _{sens}	Sensores	R ² _{treino}	R ² _{teste}	a _{teste}	b _{teste}	RMSE _{teste}
E	Glucose	200	15	2, 6, 8, 9, 10, 14, 17, 18, 21, 23, 24, 25, 27, 33, 34	1,0000	0,9926	1,0069	NS	0,1668
F	Frutose	200	15	3, 9, 13, 14, 15, 19, 22, 23, 25, 31, 32, 34, 35, 37, 39	1,0000	0,9954	0,9771	NS	0,2992
G	Sacarose	200	19	2, 3, 5, 6, 7, 9, 10, 13, 15, 24, 25, 26, 27, 31, 33, 34, 35, 36, 38	1,0000	0,9996	1,0178	NS	0,0198
H	AT	200	17	1, 2, 3, 7, 9, 10, 13, 14, 15, 16, 17, 19, 22, 23, 25, 33, 35, 37	1,0000	0,9953	0,9977	NS	0,4847
Soluções com fundo iônico de KCl 1M									
Soluções padrão									
Modelo	Analito	N ^o _{Iter}	N ^o _{sens}	Sensores	R ² _{treino}	R ² _{teste}	a _{teste}	b _{teste}	RMSE _{teste}
I	Glucose	200	18	1, 2, 7, 8, 9, 10, 15, 16, 17, 18, 22, 25, 30, 33, 34, 35, 36, 37	1,0000	0,9983	1,0695	NS	0,3040
J	Frutose	200	20	1, 3, 4, 5, 6, 7, 8, 9, 10, 13, 14, 15, 16, 17, 19, 22, 23, 35, 37, 38	1,0000	0,9971	1,0151	NS	0,2045
K	Sacarose	500	19	1, 2, 5, 7, 10, 18, 19, 21, 23, 27, 28, 30, 31, 32, 34, 36, 37, 38, 39	1,0000	0,9966	1,0508	NS	0,2738
L	AT	200	17	3, 5, 6, 10, 16, 17, 18, 19, 22, 26, 27, 30, 31, 33, 35, 37, 38	1,0000	0,9961	0,9980	NS	0,3508
Soluções de mel									
Modelo	Analito	N ^o _{Iter}	N ^o _{sens}	Sensores	R ² _{treino}	R ² _{teste}	a _{teste}	b _{teste}	RMSE _{teste}
M	Glucose	200	17	1, 2, 5, 6, 8, 9, 14, 15, 21, 24, 27, 30, 32, 33, 34, 35, 36	1,0000	0,9914	0,9874	NS	0,1758
N	Frutose	200	17	1, 3, 4, 5, 6, 9, 13, 14, 15, 16, 17, 19, 26, 27, 30, 33, 37	1,0000	0,9955	0,9977	NS	0,2990
O	Sacarose	500	16	5, 7, 8, 10, 13, 14, 15, 18, 19, 22, 24, 26, 32, 35, 36, 38	1,0000	0,9857	1,0907	NS	0,1023
P	AT	200	17	1, 2, 5, 6, 10, 14, 15, 16, 21, 28, 30, 32, 33, 35, 36, 38, 39	1,0000	0,9943	0,9762	NS	0,4847

AT – açúcares totais; N^oiter – número de iterações; N^osens – número de sensores; a_{teste} – declive do teste; b_{teste} – ordenada na origem do teste; NS – Não significativo; RMSE – erro quadrático médio

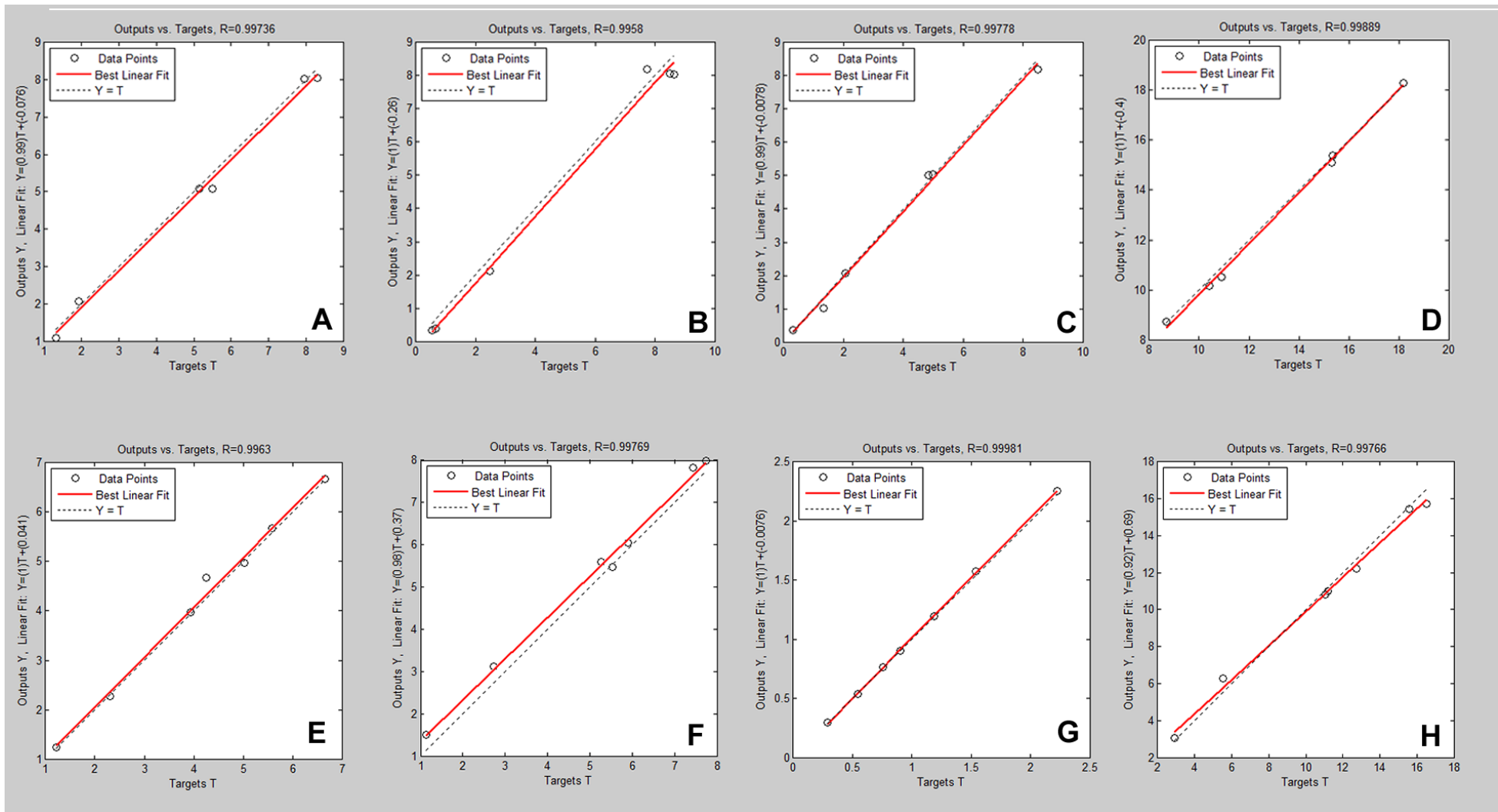


Figura 19. Resultados de previsão para os dados de teste usando os modelos PLS para soluções sem fundo iônico. Soluções padrão (A – glucose, B – frutose, C – Sacarose, D – Açúcares totais) e Soluções de mel (E – glucose, F – frutose, G – Sacarose, H – Açúcares totais).

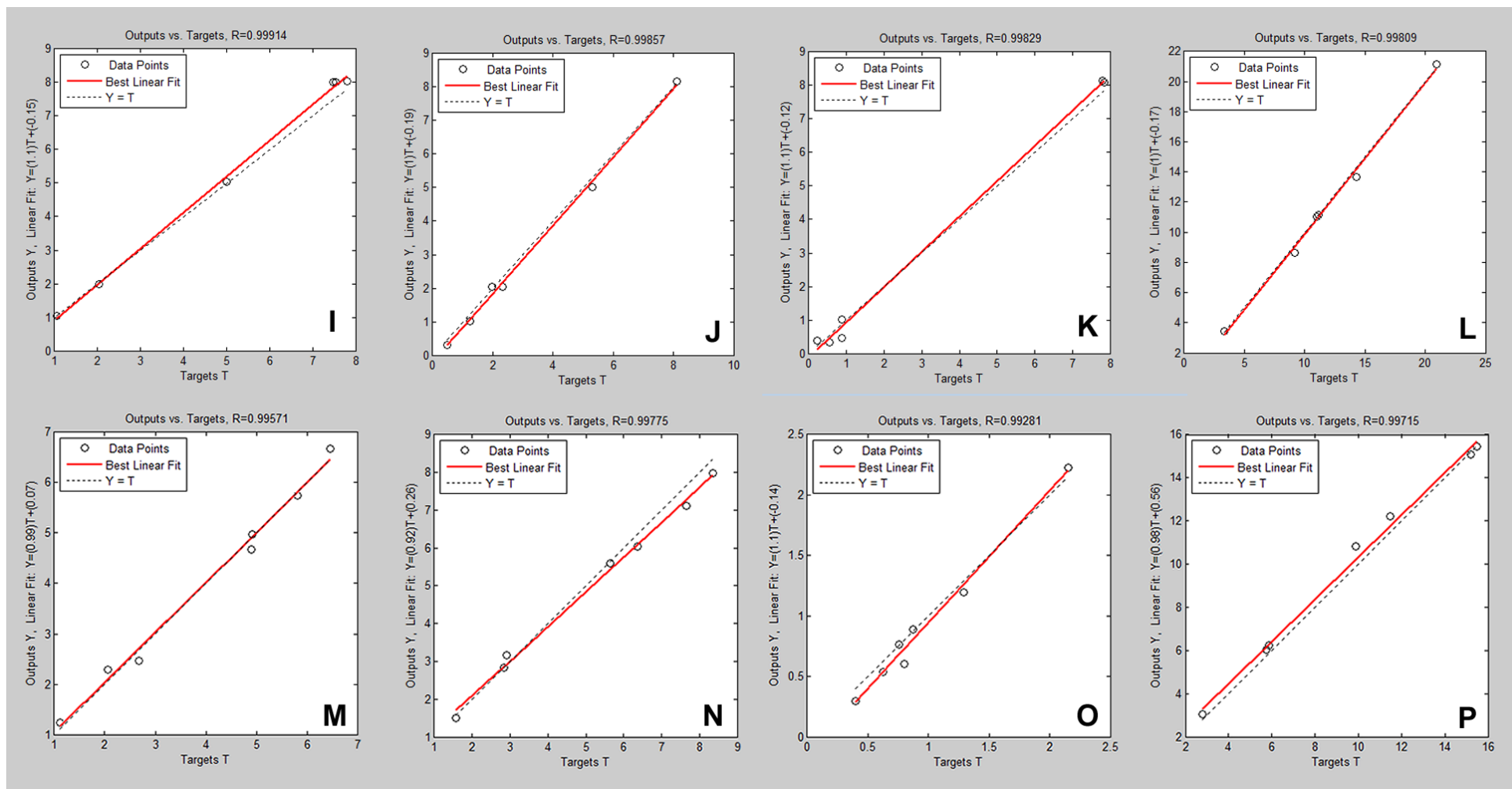


Figura 20. Resultados de previsão para os dados de teste usando os modelos PLS para soluções com fundo iônico de KCl 1M. Soluções padrão (I – glucose, J – frutose, K – Sacarose, L – Açúcares totais) e Soluções de mel (M – glucose, N – frutose, O – Sacarose, P – Açúcares totais).

Foi possível se obter modelos de previsão com desempenho satisfatório para todos os casos, com coeficientes de determinação ajustados no teste superiores a 0,99, que foi o critério utilizado pelo GA para seleção dos conjuntos de sensores, declives entre 0,9762 e 1,0907 (intervalo que contém o valor teórico ótimo 1,0), ordenadas na origem não significativas e valores de RMSE do grupo de teste aceitáveis, com valor máximo de 0,4847, demonstrando boa relação linear entre os valores previstos pelos modelos e os valores reais da concentração dos açúcares nas soluções analisadas.

Verificou-se também que não é possível apontar diferenças relevantes no que diz respeito ao número de sensores selecionados pelos modelos de PLS obtidos, para cada tipo de solução com ou sem fundo iônico, variando de maneira bem distribuída entre 15 e 24 sensores entre os modelos.

Os modelos obtidos foram alcançados com o número de iterações (ou gerações) de 200 ou 300, com exceção dos modelos K e O (Tabela 11), que precisaram de 500 iterações para encontrar um subconjunto de sensores que produzisse resultados satisfatórios para a função de desempenho. Em cada iteração um novo subconjunto de sensores era obtido e, caso apresentasse o melhor resultado para função de desempenho, era mantido como o subconjunto resposta, caso contrário o subconjunto resposta era o que obteve melhor valor desempenho nas iterações anteriores. A Figura 21 apresenta como exemplo a estabilização do valor da função de desempenho (maximização do coeficiente de determinação ajustado para os dados do grupo de teste) para o modelo E, obtido para concentração de glucose em soluções de mel sem a presença de fundo iônico usando validação cruzada *k-folds* de 10 subgrupos com 10 repetições.

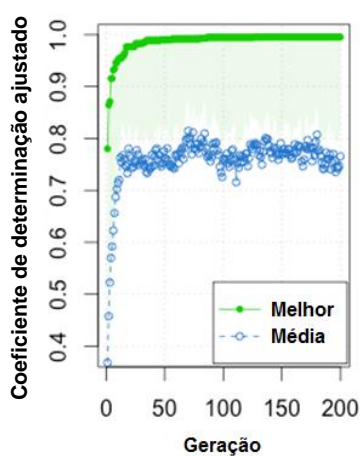


Figura 21. Estabilização da função de desempenho para o GA obtida no modelo E

No que diz respeito aos sensores, todos os modelos apresentaram subconjuntos de sensores selecionados diferentes entre si, evidenciando a capacidade de adaptação da língua eletrônica às diversas respostas através dos sensores de sensibilidade cruzada. A Figura 22 mostra quais os sensores estão presentes nos modelos construídos para as soluções padrão e de mel com presença e ausência de fundo iônico de KCl 1M. Os valores de frequência contabilizam a presença de cada sensor nos modelos, ou seja, só foi contabilizada a presença de um dos sensores mesmo estando presente em duplicado no modelo.

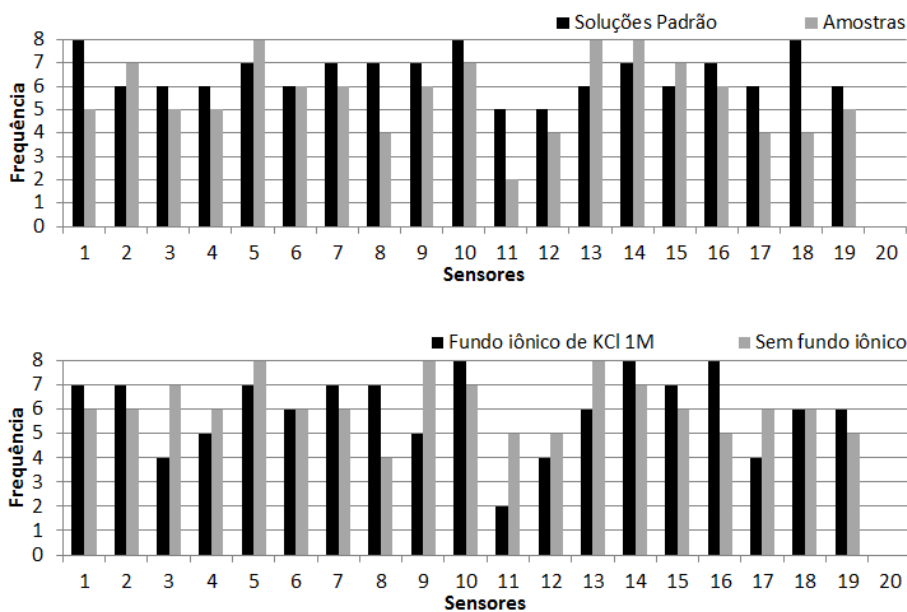


Figura 22. Frequência da presença dos sensores na construção dos modelos PLS com seleção de variáveis através do GA.

Dos 16 modelos construídos, 8 foram obtidos a partir de soluções com fundo iônico e 8 de soluções sem fundo iônico. Ainda, 8 deles se referem a modelos para soluções padrão e outros 8 para soluções de mel. Dessa forma, nos gráficos apresentados na Figura 22, a máxima frequência possível de presença de cada sensor na construção dos modelos é 8.

Com exceção do sensor 20, excluído das análises em ambas as repetições por apresentar variabilidade de resposta muito elevada, possivelmente por problemas de fixação no momento da construção da língua eletrônica, todos os outros sensores estiveram presentes na construção dos modelos, inclusive os sensores 9, 11 e 12, que

tiveram uma de suas repetições retiradas das análises depois de feita a análise exploratória dos dados. Isso demonstra que os sensores pré-selecionados para construção da língua eletrônica utilizada neste trabalho são adequados para análise dos açúcares, estando em concordância com os sensores selecionados por Dias *et al.* [28] para análise semi-quantitativa de glucose e frutose utilizando MLR e PLS.

No que diz respeito à diferenciação da escolha dos sensores nas análises das soluções de mel destacaram-se os sensores 5, 13 e 14, presentes em todos os modelos e os sensores 2, 10 e 15, presentes em 7 modelos. Nas soluções padrão, os sensores 7, 9 e 18 estavam presentes em 8 modelos e os sensores 5, 7, 8, 9, 14, e 16, presentes em 7 dos modelos. Em relação à diferenciação dos modelos ao nível dos sensores para as análises com fundo iônico cabe destacar os sensores 10, 14 e 16 presentes nos 8 modelos possíveis, e os sensores 1, 2, 5, 7, 8, e 15 presentes em 7 destes modelos. Para as análises sem fundo iônico, vale a pena citar os sensores 5, 9 e 13 presentes em todos os modelos possíveis, e os sensores 3, 10, e 14, presentes em 7 modelos.

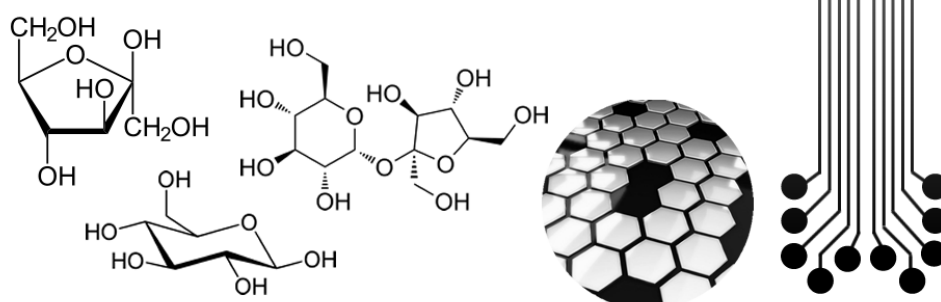
Devido essa elevada heterogeneidade de seleção dos sensores, não é possível se definir um subgrupo reduzido específico de sensores para serem utilizados na análise de açúcares com a língua eletrônica potenciométrica, ou seja, cada modelo requer subconjuntos de sensores próprios para produzir os melhores resultados ao nível de previsão da concentração dos açúcares.

Em relação à quantificação de açúcares, encontram-se alguns estudos com diferentes línguas eletrônicas e utilizando outras ferramentas de análise dos dados. Cipri *et al.* [69] utiliza uma língua bio-eletrônica voltamétrica com sensores de diferentes enzimas celobioses dehidrogenases aliada à RNAs na quantificação de açúcares em misturas, obtendo-se um modelo de previsão com R^2 de 0,726 para quantificação de glucose. Em seu estudo, Major *et al.* [70] aplica uma língua eletrônica comercial aliada às técnicas de PCA, RNA e Análise de Correlação Canônica para caracterização e classificação botânica de mel, obtendo modelos de previsão com R^2 para quantificação dos açúcares totais igual a 0,958. Tian *et al.* [71] em seu estudo sobre discriminação de gomas de damasco utilizando uma língua eletrônica comercial e aplicando RNAs e regressão por PLS determina o teor de açúcares totais das amostras analisadas, obtendo valores de R^2 na validação dos modelos de 0,9853 e 0,4999, respectivamente. Na avaliação do teor de açúcares em peras não-climatéricas utilizando uma língua eletrônica voltamétrica, Wei & Wang [72] utiliza as técnicas de PCR, PLS e Máquinas de Vetor Suporte, obtendo valores de R^2 de 0,963, 0,897 e 0,996, respectivamente.

Os resultados obtidos neste estudo são satisfatórios em relação aos melhores resultados obtidos nos trabalhos acima referidos, quer para a glucose ou açúcares totais.



4. CONCLUSÕES



Neste trabalho foi aplicada uma língua eletrônica potenciométrica com o objetivo de se estabelecer uma metodologia analítica e procedimentos matemáticos para quantificação dos açúcares que poderão ser compostos majoritários nos alimentos (glucose, frutose e sacarose) e açúcares totais. Para isto foram analisadas soluções padrão de misturas de açúcares e soluções de dissoluções de amostras comerciais de méis, cujas concentrações dos açúcares foram através de ensaios em HPLC.

No estudo da validação da quantificação dos açúcares por HPLC, verificou-se que as curvas de calibração individuais para os açúcares apresentaram bons coeficientes de correlação ($R > 0,999$) e limites de detecção e quantificação com valores inferiores à 0,1 e 0,3 g/L, respectivamente. Os resultados de repetibilidade das análises apresentaram desempenho satisfatório, com desvios padrão relativos inferiores a 2% e erros relativos percentuais inferiores a 5% para todos os açúcares. Após verificação da qualidade dos modelos de calibração obtidos, foi então possível calcular a concentração dos açúcares nas amostras de mel, obtendo-se concentrações concordantes com o já relatado pela literatura com desvios padrão relativos inferiores a 2%. Verificou-se que as três amostras de méis apresentam maior concentração em frutose, seguido de glucose e por fim sacarose, obtendo-se razões de intolerância a frutose livre e total inferiores a 1, não sendo então indicadas para pessoas que sofrem de intolerância a esse açúcar. Foi verificado que as concentrações dos açúcares nas soluções de amostras de mel que foram utilizadas na análise com língua eletrônica apresentaram colinearidade, ressaltando a importância de se realizar as análises em paralelo com soluções padrão de misturas ternárias dos açúcares em níveis ortogonais.

Realizada as análises com a língua eletrônica das soluções padrão e de mel, foi verificado após a análise exploratória dos dados que alguns sensores apresentaram variabilidades discrepantes de resposta (resultantes de uma má adesão da membrana polimérica à placa de PVC do sistema de multi-sensores) e por isso foram removidos das etapas seguintes para construção dos modelos de regressão.

Nos modelos de MLR usando o algoritmo *stepwise* foi verificado que embora fosse possível estabelecer uma boa relação entre as respostas dos sensores e as concentrações dos açúcares ($R^2_{\text{treino}} > 0,999$ para todos os modelos), a previsão dos modelos não apresentava desempenho satisfatório para observações dos grupos de teste e, que mudanças na técnica de validação dos modelos (*leave-one-out* ou *k-folds*) não geraram resultados com diferenças assinaláveis. O número de observações destinadas para os grupos de treino foi definido a partir de diversas tentativas de forma a se obter os

primeiros modelos de previsão satisfatórios com o menor número de soluções de treino, definindo-se assim 6 observações de teste para as soluções padrão e 7 para as soluções de mel.

Dessa forma, foram então construídos modelos PLS com seleção de variáveis utilizando o GA. Os parâmetros de configuração do GA foram otimizados com o objetivo de selecionar os melhores subconjuntos de sensores que produzissem bons modelos de previsão com o menor número possível de iterações. Entre os critérios de avaliação da qualidade dos modelos de previsão, foi escolhido o coeficiente de determinação ajustado para a relação entre os valores previstos pelos modelos e os experimentais, uma vez que este representa o quanto o modelo consegue explicar os valores observados penalizando adições de variáveis que não contribuem significativamente para melhora dos modelos.

Foi possível se obter modelos de previsão com desempenho satisfatório para todos os casos, com valores de R^2_{ajustado} no teste superiores a 0,99, declives entre 0,9762 e 1,0907, ordenadas na origem não significativas e valores de RMSE aceitáveis (com valor máximo de 0,4847), demonstrando boa relação linear entre os valores previstos pelos modelos e os valores reais da concentração dos açúcares nas soluções analisadas. Não é possível apontar diferenças relevantes em relação ao número de sensores selecionados pelos modelos de PLS, variando de maneira distribuída de 15 a 24 sensores entre os modelos. Não foram verificadas diferenças expressivas no número de sensores e qualidade dos modelos em relação às análises na ausência ou presença de fundo iônico, mostrando que embora o intervalo de variação da resposta dos sensores para as análises com fundo iônico fosse bastante reduzido em relação às análises sem fundo iônico, este não é um fator que interfere na construção dos modelos para as soluções analisadas.

Devido à elevada heterogeneidade na seleção dos sensores, que se mostraram adequados para análise de açúcares, não foi possível se definir um subgrupo reduzido específico de sensores para serem utilizados na análise de açúcares com a língua eletrônica potenciométrica, ou seja, cada modelo requer subconjuntos de sensores próprios para produzir os melhores resultados ao nível de previsão da concentração dos açúcares.

Os resultados obtidos neste trabalho mostram a potencialidade da análise de açúcares em alimentos com a língua eletrônica potenciométrica. Como **trabalhos futuros**, sugere-se a fusão dos dados obtidos com as soluções padrão e de mel para se

definir um modelo matemático de previsão que ultrapasse as diferenças da composição da matriz entre elas. O objetivo é completar o presente trabalho mostrando que é possível calibrar uma língua electrónica potenciométrica para a análise de açúcares no mel usando soluções padrão de mistura de açúcares puros, com níveis de concentração ortogonais através de desenho experimental, e amostras de mel (em pequeno número). Ou seja, superar a necessidade de se ter um grande número de amostras de mel padronizadas e não colineares (situação pouco provável de se conseguir) para dar robustez à calibração multivariada. As primeiras tentativas de tratamento de dados sugerem que o problema é não linear e, por isso, entre outros tratamentos de dados, o GA desenvolvido poderá ser adaptado para fazer também para modelos de regressão polinomiais quadráticos e com redes neurais artificiais, uma vez que na bibliografia consultada não existem trabalhos de regressão quadrática nos estudos de análise multivariada quantitativa usando GA e, por isso, esta metodologia seria aplicada pela primeira vez no tratamento de dados de línguas electrónicas.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] H.-D. Belitz, W. Grosch, and P. Schieberle, *Food Chemistry*, vol. 107. Springer Berlin Heidelberg, 2009, p. 1070.
- [2] M. Cox and D. Nelson, *Lehninger Principles of Biochemistry*, 4th ed. 2004, p. 1100.
- [3] P. Singh, K. Valentas, and E. Rotstein, *Handbook of Food Engineering Practice*. New York: CRC Press, 1997, p. 698.
- [4] L. G. Dias, C. Sequeira, A. C. a Veloso, M. E. B. C. Sousa, and A. M. Peres, “Evaluation of healthy and sensory indexes of sweetened beverages using an electronic tongue”, *Anal. Chim. Acta*, vol. 848, pp. 32–42, 2014.
- [5] F. B. Hu and V. S. Malik, “Sugar-sweetened beverages and risk of obesity and type 2 diabetes: epidemiologic evidence.”, *Physiol. Behav.*, vol. 100, no. 1, pp. 47–54, 2010.
- [6] J. E. Riby, T. Fujisawa, and N. Kretchmer, “Fructose absorption”, *Am. J. Clin. Nutr.*, vol. 58, no. 5, pp. 748–753, 1993.
- [7] M. E. Latulippe and S. M. Skoog, “Fructose malabsorption and intolerance: effects of fructose with and without simultaneous glucose ingestion”, *Crit. Rev. Food Sci. Nutr.*, vol. 51, pp. 583–592, 2011.
- [8] L. Dias, C. Sequeira, A. Veloso, J. Morais, M. Sousa, and A. Peres, “A Size Exclusion HPLC Method for Evaluating the Individual Impacts of Sugars and Organic Acids on Beverage Global Taste by Means of Calculated Dose-Over-Threshold Values”, *Chromatography*, vol. 1, pp. 141–158, 2014.
- [9] L. Dias, C. Sequeira, A. Veloso, J. Sá Morais, M. Sousa, and A. Peres, “Classification of commercial juices using sensory and healthy ratios calculated from organic acids and sugars HPLC profiles”, *J. Food Qual.*, pp. 0–29, 2013.
- [10] L. Escuder-Gilabert and M. Peris, “Review : Highlights in recent applications of electronic tongues in food analysis”, *Anal. Chim. Acta*, vol. 665, no. 1, pp. 15–25, 2010.
- [11] T. Sakata, K. Faceli, T. Almeida, A. Júnior, and W. Steluti, “The Assessment of the Quality of Sugar using Electronic Tongue and Machine Learning Algorithms”, in *11th International Conference on Machine Learning and Applications*, 2012, pp. 538–541.
- [12] J. P. Marais, J. L. Wit, and G. V. Quicke, “A critical examination of the Nelson-Somogyi method for the determination of reducing sugars”, *Anal. Biochem.*, vol. 15, no. 3, pp. 373–381, 1966.

- [13] R. D. N. Silva, V. N. Monteiro, J. D. X. Alcanfor, E. M. Assis, and E. R. Asquiere, “Comparação de métodos para a determinação de açúcares redutores e totais em mel”, *Ciência e Tecnol. Aliment.*, vol. 23, no. 3, pp. 337–341, 2003.
- [14] F. L. C. Machado, G. Campos, and M. T. G. Souza, “Comparação entre os métodos de Lane-Eynon e polarimétrico para determinação de amido em farinha de mandioca”, *Rev. Inst. Adolfo Lutz*, vol. 68, no. 1, pp. 155–159, 2009.
- [15] E. Giannoccaro, Y. J. Wang, and P. Chen, “Comparison of two HPLC systems and an enzymatic method for quantification of soybean sugars”, *Food Chem.*, vol. 106, pp. 324–330, 2008.
- [16] F. J. Rambla, S. Garrigues, and M. de la Guardia, “PLS-NIR determination of total sugar, glucose, fructose and sucrose in aqueous solutions of fruit juices”, *Anal. Chim. Acta*, vol. 344, pp. 41–53, 1997.
- [17] L. Xie, X. Ye, D. Liu, and Y. Ying, “Quantification of glucose, fructose and sucrose in bayberry juice by NIR and PLS”, *Food Chem.*, vol. 114, no. 3, pp. 1135–1140, 2009.
- [18] Y. Vlasov, A. Legin, A. Rudnitskaya, C. Di Natale, and A. D’Amico, “Nonspecific sensor arrays (‘electronic tongue’) for chemical analysis of liquids (IUPAC Technical Report)”, *Pure Appl. Chem.*, vol. 77, no. 11, pp. 1965–1983, 2005.
- [19] M. Del Valle, “Electronic tongues employing electrochemical sensors”, *Electroanalysis*, vol. 22, pp. 1539–1555, 2010.
- [20] D. Ha, Q. Sun, K. Su, H. Wan, H. Li, N. Xu, F. Sun, L. Zhuang, N. Hu, and P. Wang, “Recent achievements in electronic tongue and bioelectronic tongue as taste sensors”, *Sensors Actuators B Chem.*, vol. 207, pp. 1136–1146, 2015.
- [21] J. Gallardo, S. Alegret, and M. Del Valle, “Application of a potentiometric electronic tongue as a classification tool in food analysis”, *Talanta*, vol. 66, pp. 1303–1309, 2005.
- [22] H. Gupta, A. Sharma, S. Kumar, and S. K. Roy, “E-tongue: a tool for taste evaluation.”, *Recent Pat. Drug Deliv. Formul.*, vol. 4, pp. 82–89, 2010.
- [23] C. B. Sequeira, “Análise de açúcares e ácidos orgânicos em sumos comerciais: aplicação de HPLC-SEC-UV-IR e língua eletrônica.” Bragança, p. 79, 2012.
- [24] F. Winqvist, C. Krantz-Rülcker, and I. Lundström, “Electronic Tongues and Combinations of Artificial Senses”, *Sensors Updat.*, vol. 11, pp. 279–306, 2002.
- [25] X. Cetó, M. Gutiérrez-Capitán, D. Calvo, and M. del Valle, “Beer classification by means of a potentiometric electronic tongue.”, *Food Chem.*, vol. 141, pp. 2533–40, 2013.

- [26] A. Mimendia, J. M. Gutiérrez, L. Leija, P. R. Hernández, L. Favari, R. Muñoz, and M. del Valle, “A review of the use of the potentiometric electronic tongue in the monitoring of environmental systems”, *Environ. Model. Softw.*, vol. 25, pp. 1023–1030, 2010.
- [27] L. A. Dias, A. M. Peres, A. C. Veloso, F. S. Reis, M. Vilas-Boas, and A. A. S. C. Machado, “An electronic tongue taste evaluation: Identification of goat milk adulteration with bovine milk”, *Sensors Actuators, B Chem.*, vol. 136, pp. 209–217, 2009.
- [28] L. G. Dias, A. M. Peres, T. P. Barcelos, J. S. Morais, and A. A. S. C. Machado, “Semi-quantitative and quantitative analysis of soft drinks using an electronic tongue”, *Sensors Actuators B. Chem.*, vol. 154, no. 2, pp. 111–118, 2011.
- [29] P. Ivarsson, Y. Kikkawa, F. Winqvist, C. Krantz-Rülcker, N. E. Höjer, K. Hayashi, K. Toko, and I. Lundström, “Comparison of a voltammetric electronic tongue and a lipid membrane taste sensor”, *Anal. Chim. Acta*, vol. 449, pp. 59–68, 2001.
- [30] R. Martínez-Mañez, J. Soto, E. Garcia-Breijo, L. Gil, J. Ibáñez, and E. Llobet, “An ‘electronic tongue’ design for the qualitative analysis of natural waters”, *Sensors Actuators, B Chem.*, vol. 104, pp. 302–307, 2005.
- [31] M. Liu, M. Wang, J. Wang, and D. Li, “Comparison of random forest, support vector machine and back propagation neural network for electronic tongue data classification: Application to the recognition of orange beverage and Chinese vinegar”, *Sensors Actuators, B Chem.*, vol. 177, pp. 970–980, 2013.
- [32] Y. Vlasov and A. Legin, “Non-selective chemical sensors in analytical chemistry: from ‘electronic nose’ to ‘electronic tongue,’” *Fresenius. J. Anal. Chem.*, vol. 361, pp. 255–260, 1998.
- [33] M. Otto and J. D. R. Thomas, “Model studies on multiple channel analysis of free magnesium, calcium, sodium, and potassium at physiological concentration levels with ion-selective electrodes”, *Anal. Chem.*, vol. 57, no. 13, pp. 2647–2651, 1985.
- [34] K. Toko, K. Hayashi, M. Yamanaka, and K. Yamafuji, “Multichannel taste sensor using lipid membranes”, *Sensors Actuators B Chem.*, vol. 2, no. 3, pp. 205–213, 1990.
- [35] K. Toko, “Taste sensor”, *Sensors Actuators B Chem.*, vol. 64, pp. 205–215, 2000.
- [36] K. Toko, “Taste sensor with global selectivity”, *Mater. Sci. Eng. C*, vol. 4, pp. 69–82, 1996.
- [37] L. Nunez, X. Ceto, M. I. Pividori, M. V. B. Zanoni, and M. Del Valle, “Development and application of an electronic tongue for detection and monitoring of nitrate, nitrite and ammonium levels in waters”, *Microchem. J.*, vol. 110, pp. 273–279, 2013.

- [38] K. Woertz, C. Tissen, P. Kleinebudde, and J. Breitzkreutz, “A comparative study on two electronic tongues for pharmaceutical formulation development”, *J. Pharm. Biomed. Anal.*, vol. 55, no. 2, pp. 272–281, 2011.
- [39] C. Krantz-Rülcker, M. Stenberg, F. Winquist, and I. Lundström, “Electronic tongues for environmental monitoring based on sensor arrays and pattern recognition: A review”, *Anal. Chim. Acta*, vol. 426, pp. 217–226, 2001.
- [40] A. M. Peres, L. G. Dias, T. P. Barcelos, J. Sá Morais, and A. A. S. C. Machado, “An electronic tongue for juice level evaluation in non-alcoholic beverages”, *Procedia Chem.*, vol. 1, no. 1, pp. 1023–1026, 2009.
- [41] W. He, X. Hu, L. Zhao, X. Liao, Y. Zhang, M. Zhang, and J. Wu, “Evaluation of Chinese tea by the electronic tongue: Correlation with sensory properties and classification according to geographical origin and grade level”, *Food Res. Int.*, vol. 42, no. 10, pp. 1462–1467, 2009.
- [42] Q. Chen, J. Zhao, and S. Vittayapadung, “Identification of the green tea grade level using electronic tongue and pattern recognition”, *Food Res. Int.*, vol. 41, pp. 500–504, 2008.
- [43] R. Banerjee, B. Tudu, L. Shaw, A. Jana, and N. Bhattacharyya, “Instrumental testing of tea by combining the responses of electronic nose and tongue”, *J. Food Eng.*, vol. 110, no. 3, pp. 356–363, 2012.
- [44] M. Cole, J. a. Covington, and J. W. Gardner, “Combined electronic nose and tongue for a flavour sensing system”, *Sensors Actuators B Chem.*, vol. 156, no. 2, pp. 832–839, 2011.
- [45] K. Beullens, P. Mészáros, S. Vermeir, D. Kirsanov, A. Legin, S. Buysens, N. Cap, B. M. Nicolai, and J. Lammertyn, “Analysis of tomato taste using two types of electronic tongues”, *Sensors Actuators, B Chem.*, vol. 131, pp. 10–17, 2008.
- [46] D. B. Kantor, G. Hitka, A. Fekete, and C. Balla, “Electronic tongue for sensing taste changes with apricots during storage”, *Sensors Actuators B Chem.*, vol. 131, pp. 43–47, 2008.
- [47] L. G. Dias, A. Fernandes, A. C. A. Veloso, A. A. S. C. Machado, J. A. Pereira, and A. M. Peres, “Single-cultivar extra virgin olive oil classification using a potentiometric electronic tongue”, *Food Chem.*, vol. 160, pp. 321–329, 2014.
- [48] A. C. de Sá, A. Cipri, A. González-Calabuig, N. R. Stradiotto, and M. del Valle, “Resolution of galactose, glucose, xylose and mannose in sugarcane bagasse employing a voltammetric electronic tongue formed by metals oxyhydroxide/MWCNT modified electrodes”, *Sensors Actuators B Chem.*, vol. 222, pp. 645–653, 2016.
- [49] M. Dias, “Língua eletrónica potenciométrica: uma ferramenta para análise da qualidade do mel”, Universidade do Porto, 2013.

- [50] X. Cetó, F. Céspedes, and M. Del Valle, “Comparison of methods for the processing of voltammetric electronic tongues data”, *Microchim. Acta*, vol. 180, pp. 319–330, 2013.
- [51] R. W. Kennard and L. A. Stone, “Computer Aided Design of Experiments”, *Technometrics*, vol. 11, no. 1, pp. 137–148, 1969.
- [52] R. De Maesschalck, D. Jouan-Rimbaud, and D. L. L. Massart, “The Mahalanobis distance”, *Chemom. Intell. Lab. Syst.*, vol. 50, pp. 1–18, 2000.
- [53] P. Wold, K.; Geladi, “Principal component analysis”, *Chemom. Intell. Lab. Syst.*, vol. 2, no. 37, 1987.
- [54] S. Haykin, *Redes neurais: princípios e prática*, 2nd ed. Porto Alegre: Bookman, 2001, p. 900.
- [55] R. G. Brereton, *Chemometrics: Data Analysis for the Laboratory and Chemical Plant*, vol. 8. 2003, p. 471.
- [56] C. M. Bishop, *Pattern recognition and machine learning*. New York: Springer, 2006, p. 738.
- [57] C. K. Bayne and R. Kramer, “Chemometric Techniques for Quantitative Analysis”, *Technometrics*, vol. 41. p. 173, 1999.
- [58] J. M. Sá, *Applied statistics using SPSS, Statistica, MATLAB and R*, 2nd ed., vol. 53. Springer Berlin Heidelberg, 2007, p. 505.
- [59] F. Raposo, “Evaluation of analytical calibration based on least-squares linear regression for instrumental techniques: A tutorial review”, *Trends Anal. Chem.*, vol. 77, pp. 167–185, 2016.
- [60] C. A. Blanco, R. De La Fuente, I. Caballero, and M. L. Rodríguez-Méndez, “Beer discrimination using a portable electronic tongue based on screen-printed electrodes”, *J. Food Eng.*, vol. 157, pp. 57–62, 2015.
- [61] J. M. Graham, “Problems with stepwise procedures in discriminat analysis.”, *Annu. Meet. Southwest Educ. Res. Assoc.*, 2001.
- [62] J. Cadima, J. O. Cerdeira, and M. Minhoto, “Computational aspects of algorithms for variable selection in the context of principal components”, *Comput. Stat. Data Anal.*, vol. 47, pp. 225–236, 2004.
- [63] H. Örkücü, “Subset selection in multiple linear regression models: A hybrid of genetic and simulated annealing algorithms”, *Appl. Math. Comput.*, vol. 219, pp. 11018–11028, 2013.
- [64] X.-S. Yang, *Nature-Inspired Metaheuristic Algorithms*, 2nd ed. Luniver Press, 2010, p. 147.

- [65] S. Bandyopadhyay and S. Pal, *Classification and learning using genetic algorithms - Applications in Bioinformatics and Web Intelligence*. Springer, 2010, p. 327.
- [66] M. Gendreau and J.-Y. Potvin, *Handbook of metaheuristics*, 2nd ed., vol. 157. Springer, 2010, p. 648.
- [67] R. G. Brereton, *Applied Chemometrics for Scientists*. Wiley, 2007, p. 379.
- [68] S. Bogdanov, J. Tomislav, R. Sieber, and P. Gallmann, “Honey for Nutrition and Health: a Review”, *Am. J. Coll. Nutr.*, vol. 27, pp. 677–689, 2008.
- [69] A. Cipri, C. Schulz, R. Ludwig, L. Gorton, and M. del Valle, “A novel bio-electronic tongue using different cellobiose dehydrogenases to resolve mixtures of various sugars and interfering analytes”, *Biosens. Bioelectron.*, vol. 79, pp. 515–521, 2016.
- [70] N. Major, K. Marković, M. Krpan, G. Šarić, M. Hruškar, and N. Vahčić, “Rapid honey characterization and botanical classification by an electronic tongue”, *Talanta*, vol. 85, pp. 569–574, 2011.
- [71] X. Tian, J. Wang, and X. Zhang, “Discrimination of preserved licorice apricot using electronic tongue”, *Math. Comput. Model.*, vol. 58, no. 3–4, pp. 737–745, 2013.
- [72] Z. Wei and J. Wang, “The evaluation of sugar content and firmness of non-climacteric pears based on voltammetric electronic tongue”, *J. Food Eng.*, vol. 117, no. 1, pp. 158–164, 2013.