

## TÉCNICA NÃO DESTRUTIVA DE ANÁLISE DE TANINO EM CAFÉ EMPREGANDO ESPECTROSCOPIA NO INFRAVERMELHO E ALGORITMO GENÉTICO

*Marco Flôres Ferrão<sup>1</sup>*

*João Carlos Furtado<sup>2</sup>*

*Luiz Gustavo Neumann<sup>3</sup>*

*Pedro Henrique de Almeida Konzen<sup>4</sup>*

*Marcelo Antonio Morgano<sup>5</sup>*

*Neura Bragagnolo<sup>6</sup>*

*Márcia Miguel Castro Ferreira<sup>7</sup>*

### RESUMO

Neste trabalho é proposta uma nova metodologia de análise para a quantificação de tanino em café cru, fazendo-se uso da espectroscopia por reflexão difusa no infravermelho com transformada de Fourier (DRIFTS) associada a regressão por mínimos quadrados parciais (PLS). Para tanto, espectros DRIFTS foram obtidos de amostras de café e os melhores modelos de regressão PLS foram alcançados através da seleção dos mais adequados conjuntos de frequências espectrais, empregando-se algoritmos genéticos. Para avaliar o processo de otimização (seleção das melhores frequências espectrais), foram empregados os erros padrões de calibração (SEC) e de validação (SEV). Todos os modelos otimizados se mostraram superiores àqueles onde não foi empregada a seleção de variáveis (frequências espectrais), o que reflete a importância da otimização para a construção de modelos mais eficientes. Como resultado deste trabalho podemos destacar que as rotinas de análise, baseadas em DRIFTS/PLS e otimizadas por AG, para a quantificação de tanino em café cru, são técnicas não destrutivas, rápidas e com menor

<sup>1</sup> Professor do Departamento de Química e Física da Universidade de Santa Cruz do Sul – UNISC - Santa Cruz do Sul – RS, ferrao@unisc.br

<sup>2</sup> Professor do Departamento de Informática - UNISC - RS

<sup>3</sup> Acadêmico do Curso de Engenharia da Produção – UNISC – RS

<sup>4</sup> Acadêmico do Curso de Matemática Computacional - UNISC - RS

<sup>5</sup> Assistente de Pesquisa do Instituto de Tecnologia de Alimentos – ITAL – Campinas - SP

<sup>6</sup> Professora da Faculdade de Engenharia de Alimentos - UNICAMP – Campinas - SP

<sup>7</sup> Professora do Instituto de Química – UNICAMP – Campinas - SP

custo, sendo adequadas para o emprego em análises de rotina visando o controle de qualidade de insumos para as indústrias alimentícias.

**Palavras-chave:** reflexão difusa, otimização, regressão multivariada, café cru, algoritmo genético, tanino

### ABSTRACT

In this work a new methodology of analysis for the quantification of tannin in green coffee is proposed, using Genetic Algorithm (GA) optimization with Diffuse Reflectance Infrared Fourier Transform Spectroscopy (DRIFTS) and Partial-Least Squares regression method (PLS). After the best wavelength selection, which results in models with better coefficient regression ( $R^2$ ), these were improved using routines based on genetic algorithm heuristic method. For the evaluation of the optimization Standard Error Validation (SEV) and Standard Error Calibration (SEC) were used. The results reflect the importance to choose the set variable selection, as well as the kind of pre-processing applied. These results let us conclude that good models aiming the prediction of tannin levels can be obtained, and that reflection techniques are adequate to allow a fast obtain of spectra of the green coffee milling and not to create wastes which are harmful to the environment.

**Keywords:** genetic algorithm, multivariate regression, green coffee, diffuse reflection, optimization, tannin

### INTRODUÇÃO

Com a constante evolução dos processos produtivos, cada vez mais se fazem necessárias ferramentas analíticas que não causem impacto ambiental, bem como não necessitem de tratamentos demorados na amostra e ainda que não alterem as características inerentes dos insumos industriais.

Neste sentido o presente grupo vem pesquisando ferramentas analíticas para a caracterização das matérias-primas empregadas na industrialização de alimentos. Em especial vem sendo utilizadas técnicas espectroscópicas baseadas na radiação infravermelha, como por exem-

plo, a reflexão no infravermelho próximo (NIRR) e a reflexão difusa no infravermelho médio com transformada de Fourier (DRIFT). Aliados a estas técnicas de análise, ferramentas quimiométricas de regressão, como o método dos mínimos quadrados parciais (PLS), tem apresentado excelentes resultados.<sup>1</sup>

Apesar do bom desempenho apresentado pelas combinações NIRR/PLS e DRIFT/PLS, uma limitação ao seu emprego como uma técnica de rotina nas empresas do ramo alimentício reside no fato de conhecer qual porção do espectro infravermelho deva ser empregada para a modelagem de cada parâmetro de qualidade a ser avaliado.

Como solução desta limitação propomos o emprego de algoritmos genéticos, os quais tem logrado êxito em problemas de natureza semelhante, sendo de fácil implementação e com os quais estamos obtendo resultados representativos.

Em particular, este trabalho propõe a implementação de um método de controle de qualidade do café arábica cru, a partir de dados de infravermelho, buscando determinar o teor de tanino presente em diversas amostras de café nacional de diferentes procedências. Como estratégia empregamos a técnica DRIFT/PLS para construção dos modelos de calibração, sendo os mesmos otimizados por uma rotina de algoritmo genético (AG).

### **Composição Química do Café e sua Influência sobre a Qualidade**

O componente mais conhecido e valorizado do café é a cafeína. Mas o café possui uma série de outros componentes também importantes. O café possui desde o grão até a bebida final componentes tão importantes quanto a cafeína para o ser humano - os quinídeos (ou lactonas) formados a partir dos ácidos clorogênicos existentes nos grãos verdes de café após uma torrefação adequada. Apesar de importante, pouco se sabe sobre qual o teor real destes quinídeos no café e qual a sua variação nas diferentes marcas provenientes de diferentes processos de torrefação.<sup>2</sup>

Para se ter uma ligeira idéia de multiplicidade dos componentes de um grão de café, a seguir estão citados alguns dos mais importantes:

- Cafeína: é o principal princípio ativo do café. O grão de café cru contém em média, cerca de 1% de cafeína. É estimulante do sistema nervoso central;
- Colina: responsável pelo sabor específico da infusão de café;

- Trigonelina;
- Ácidos clorogênicos: atuam de forma benéfica no cérebro;
- Hidratos de carbono;
- Sais orgânicos;
- Aromáticos;
- Proteínas;
- Vitaminas;
- Taninos: interferem no sabor;
- Lipídios: responsáveis pelo aroma;
- Açúcares;
- Celulose: estimula os intestinos;
- Minerais: aproveitados no organismo;
- Água.

Muitos fatores determinantes da qualidade de bebida do café são conseqüências da composição dos grãos do café cru. Muitos estudos têm sido conduzidos para identificar a participação destes componentes na formação das características desejáveis da bebida. No entanto, a composição do café cru é dependente de condições ambientais, práticas culturais e principalmente de sua origem genética. A altitude e o sombreamento afetam significativamente os componentes de aroma do café.<sup>3</sup>

### **Tanino**

Os taninos são compostos não cristalizáveis e na presença de água formam soluções coloidais que apresentam reação ácida e forte sabor adstringente.

Os taninos compreendem um grande grupo de substâncias complexas muito disseminadas no reino vegetal; em quase todas as famílias botânicas há espécies que contêm taninos. Estes, quando ocorrem em grande quantidade, geralmente se localizam em determinados órgãos da planta como as folhas, os frutos, o córtex ou o caule.

Quimicamente, os taninos são substâncias complexas que em geral ocorrem como misturas de polifenóis difíceis de separar por não se cristalizarem. Alguns autores preferem usar a designação “extratos de tanino” em vez de “taninos”. Recentemente, com o uso de métodos cromatográficos, foi possível não só confirmar a complexidade dos extratos de tanino como também identificar os polifenóis simples, presentes em pequenas quantidades em tais misturas. A determinação

destes últimos é importante porque em geral se acredita que os taninos complexos se originem de polifenóis simples através de polimerização. Como muitos taninos condensados nunca foram isolados nem caracterizados, sua origem biogenética não é totalmente conhecida.<sup>4</sup>

Tradicionalmente os taninos são classificados segundo sua estrutura química em dois grupos: taninos hidrolisáveis e taninos condensados.

Taninos hidrolisáveis são caracterizados por um poliol central, geralmente beta-*D*-glucose, cujas funções hidroxilas são esterificadas com o ácido gálico. O composto beta-1,2,3,4,6-pentagalactil-*D*-glucose representa o padrão máximo de substituição alcançado, sendo considerado o precursor imediato para ambas as classes de taninos hidrolisáveis (galotaninos e elagitaninos).<sup>5</sup>

Os taninos condensados são oligômeros e polímeros formados pela policondensação de duas ou mais unidades flavan-3-ol e flavan-3,4-diol. Essa classe de taninos também é denominada como proantocianidina devido ao fato de os taninos condensados produzirem pigmentos avermelhados da classe das antocianidinas, tais como cianidinas e delphinidina, após degradação com ácido mineral diluído à quente.

Nas últimas décadas, vários grupos têm investigado as atividades farmacológica e biológica dos taninos. Testes *in vitro* realizados com extratos ricos em taninos ou com taninos puros têm identificado diversas atividades biológicas dessa classe de substâncias. Dentre essas atividades podem-se citar: ação bactericida e fungicida apresentada por Scalbert<sup>6</sup>, antiviral reportada por Okuda *et al.*<sup>7</sup>, moluscicida estudada por Marston e Hostettman,<sup>8</sup> como inibidores de enzimas na forma de glucosiltransferases de *Streptococcus mutans* e *S. sobrinus*<sup>9</sup> e ainda ação antitumoral.<sup>10</sup>

Acredita-se que as atividades farmacológicas dos taninos são devidas, pelo menos em parte, a três características gerais que são comuns em maior ou menor grau aos dois grupos de taninos, condensados e hidrolisáveis. Essas características são:

- 1) complexação com íons metálicos (ferro, manganês, vanádio, cobre, alumínio, cálcio, entre outros);
- 2) atividade antioxidante e seqüestradora de radicais livres;
- 3) habilidade de complexar com outras moléculas incluindo macromoléculas tais como, proteínas e polissacarídeos.

## Reflexão Difusa

A técnica de reflexão difusa está amplamente fundamentada na literatura,<sup>11-16</sup> sendo mais aplicada quando associada aos equipamentos que operam no infravermelho próximo, com os quais geralmente é designada de espectroscopia de reflexão no infravermelho próximo (NIRS ou NIRR); ou associada àqueles que operam na região do infravermelho médio, sendo conhecida por espectroscopia por reflexão difusa no infravermelho médio com transformada de Fourier (DRIFTS).

A reflexão difusa ocorre em superfícies não totalmente planas, podendo o substrato ser contínuo ou fragmentado (na forma de pó). Neste processo de reflexão o feixe incidente penetra a superfície da amostra interagindo com a matriz, retornando à superfície da mesma, após absorção parcial e múltiplos espalhamentos.

Na reflexão difusa, a radiação incidente entra em contato diversas vezes com as partículas da amostra sendo conseqüentemente atenuada. Com isto a radiação que sofre a reflexão difusa fornece informações qualitativas e quantitativas, sobre a natureza química da amostra.

É importante salientar que a radiação que retorna de forma difusa de um substrato é geralmente muito inferior em magnitude que a radiação incidente, em função disso, a maior parte dos acessórios de reflexão difusa apresentam esquemas óticos que visam concentrar a radiação, para posteriormente a mesma ser focada sobre o sistema de detecção dos instrumentos.

Do ponto de vista da análise quantitativa, além das anomalias que podem ser atribuídas à presença do fenômeno da reflexão especular, merece ainda destaque o efeito que os diferentes tamanhos de partículas podem ter sobre o espectro. Estas anomalias estão amplamente descritas nos trabalhos publicados por Olinger e Griffiths.<sup>17-18</sup>

Para reduzir o efeito do tamanho de partícula, a exemplo de Devaux *et al.*,<sup>19</sup> muitos autores têm proposto transformações no espectro, tais como a primeira e a segunda derivadas ou correção do espalhamento multiplicativo (MSC).

## Técnicas Heurísticas de Otimização

Problemas de otimização combinatória aparecem em praticamente todas as áreas do conhecimento: negócios, ciências físicas, químicas e biológicas, engenharia, arquitetura, economia e administração. Tais

problemas são, em geral, difíceis de serem tratados, pois apresentam um grande número de variáveis e de possíveis soluções.

Problemas lineares contínuos possuem no simplex um algoritmo muito eficiente para a solução exata, já os problemas lineares discretos, salvos alguns casos, normalmente não são adequados a este algoritmo. Desta forma, quando não se é possível utilizar um método exato para resolver o problema de otimização, seja por questão de praticidade ou de tempo de execução, normalmente, se recorrem aos métodos heurísticos. As técnicas heurísticas não garantem obter a melhor solução, mas boas aproximações em tempo hábil podem ser obtidas.

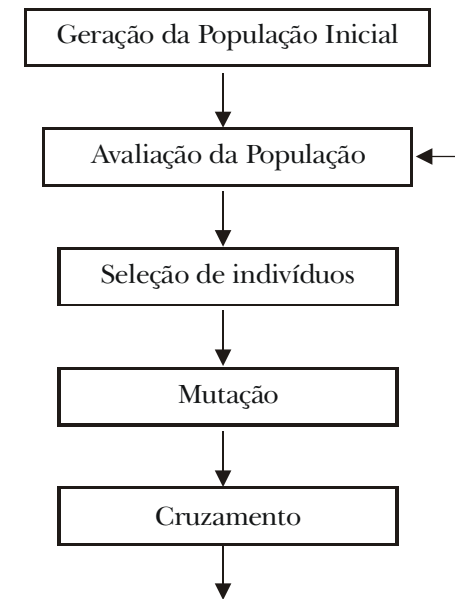
### Algoritmos Genéticos

Em meados do século XIX, o trabalho do naturalista Carolus Linnaeus sobre a classificação biológica de organismos despertou o interesse pela similaridade entre certas espécies, levando a acreditar na existência de uma certa relação entre elas. Mas foi somente em 1858 que Charles Darwin apresentou sua teoria de evolução através de seleção natural. Seus trabalhos influenciaram não só o futuro da Biologia, Botânica e Zoologia, mas também tiveram grande influência sobre o pensamento religioso, filosófico, político e econômico da época.

Pelos anos de 1950 e 1960, muitos biólogos começaram a desenvolver simulações computacionais de sistemas genéticos. Entretanto, foi John Holland quem começou, seriamente, a desenvolver as primeiras pesquisas neste tema. A idéia inicial de Holland foi de tentar imitar algumas etapas do processo de evolução natural das espécies incorporando-as a um algoritmo computacional. Gradativamente, Holland foi refinando suas idéias e em 1975 publicou o seu livro “Adaptation in Natural and Artificial Systems”, onde expõe todo o seu estudo sobre Algoritmos Genéticos (AGs).<sup>20</sup> Desde então, estes algoritmos vêm sendo aplicados com sucesso nos mais diversos problemas de otimização e no aprendizado de máquinas.

Os Algoritmos Genéticos (AGs), desenvolvidos por Holland, são baseados na evolução natural das espécies. Basicamente, os AGs, de forma similar à teoria biológica dos sistemas naturais, consistem que, os indivíduos mais adaptados sobrevivem e geram descendentes com suas características hereditárias. Estes elementos, que comporão as novas gerações, tendem a ter a mesma aparência, ou fenótipo, que os antecessores.

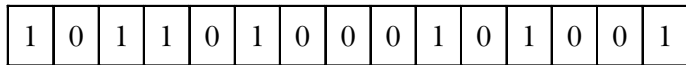
Resumidamente, os AGs constituem de uma população inicial (configurações possíveis) e, através de operadores, chamados operadores genéticos, promovem alterações na representação genética de cada indivíduo (configuração) presente na população. Através de avaliações na população, são identificados indivíduos mais adaptados que outros e, assim, é estabelecido regras de sobrevivência entre os indivíduos ao longo das gerações. Pode-se dividir um AG em 5 etapas: geração da população inicial, avaliação da população, seleção de indivíduos, mutação e cruzamento,<sup>21-25</sup> estes encontram-se apresentadas na Figura 1.



**Figura 1:** Principais etapas de um algoritmo genético.

O procedimento de geração da população inicial, normalmente, é um processo aleatório, podendo ser incorporado algum conhecimento prévio sobre o problema. A codificação binária das possíveis soluções em cromossomos (indivíduos) foi inicialmente proposta por Holland. Desta forma, cada indivíduo é representado por uma *string* binária, conforme ilustra a Figura 2. Nem sempre, essa é a melhor escolha. Codificações utilizando números reais ou apenas números inteiros podem apresentar melhor desempenho dependendo do problema. O número de indivíduos (tamanho da população) é um parâmetro de entrada dos AGs. O

tamanho da população pode variar de problema para problema. É importante ressaltar que quanto maior o número de indivíduos melhor é o desempenho dos AGs, mas maior é o tempo computacional exigido pelo algoritmo. Por outro lado, um número pequeno de indivíduos pode comprometer a busca, pois os indivíduos podem ficar presos em mínimos locais.<sup>24-25</sup>

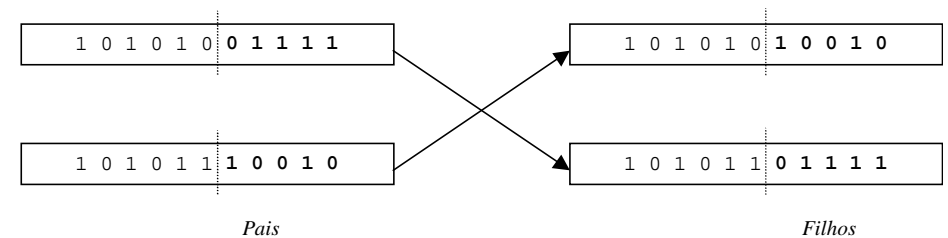


**Figura 2:** Exemplo de um cromossomo utilizando a representação binária.

Na avaliação dos indivíduos os cromossomos são decodificados e, então, avaliados. Uma boa escolha da função objetivo é fundamental nos AGs, pois os indivíduos serão selecionados conforme seu valor. O valor da função objetivo quando aplicada a um indivíduo é chamado de *fitness*. A *fitness* é o parâmetro de adaptabilidade de cada indivíduo.<sup>25</sup>

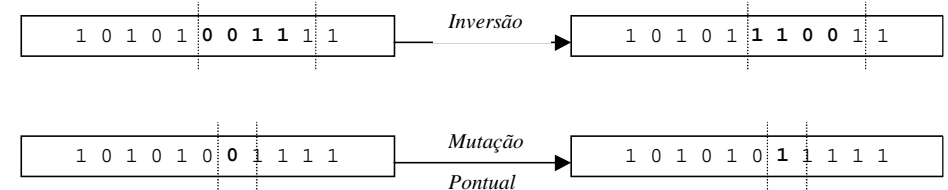
O processo de seleção tem por objetivo selecionar os indivíduos da população mais adaptados. Existem vários meios de fazer esta seleção, mas aqui destacamos o procedimento de seleção por roleta viciada e por torneio. Na seleção por roleta viciada cada indivíduo recebe uma fatia da roleta de tamanho proporcional a sua *fitness*. Desta forma, os indivíduos mais adaptados recebem uma fatia maior que os indivíduos menos adaptados. A roleta é girada e o indivíduo escolhido permanece na população. O número de indivíduos selecionados também é um parâmetro de entrada dos AGs. No procedimento de seleção por torneio, escolhe-se dois indivíduos aleatoriamente na população e, então, o mais adaptado é escolhido. Conceitos de elitismo podem ser incorporados ao processo de seleção.<sup>24-26</sup>

Os indivíduos selecionados são, então, levados ao operador genético de cruzamento. A idéia deste operador é cruzar dois indivíduos e obter um ou mais filhos com as características dos pais. Existem várias maneiras de implementar o operador de cruzamento. Uma destas formas é o operador de cruzamento simples, onde um ponto de corte é utilizado para separar os cromossomos em duas partes. A concatenação destas partes formam novos indivíduos (filhos), conforme ilustrado pelo esquema na Figura 3.<sup>27-28</sup>



**Figura 3:** Esquema de cruzamento utilizando um ponto de corte com geração de dois filhos.

O operador genético de mutação é responsável, basicamente, por assegurar a diversificação da população. Neste procedimento são escolhidos genes aleatoriamente e estes são, então, alterados. O número de indivíduos que deve sofrer mutação é um parâmetro de entrada dos AGs. Destacamos os processos de mutação por inversão e o pontual. No processo de mutação por inversão uma faixa do cromossomo é escolhida e é feita a inversão das posições dos genes desta faixa escolhida. No procedimento pontual, apenas são sorteados alguns genes e estes têm seu valor alterado.<sup>24-25,27</sup> Estes processos são apresentados na Figura 4.



**Figura 4:** Exemplos de mutação por inversão e pontual em cromossomos.

Conforme Goldberg e Luna,<sup>28</sup> podemos definir formalmente um AG da seguinte forma:

$$AG = (N, P, F, \Theta, \Omega, \Psi) \quad (1)$$

Onde P é uma população com N indivíduos,  $P = \{S_1, S_2, \dots, S_N\}$ . Cada indivíduo  $S_i$ , com  $i = 1, 2, \dots, N$ , é uma cadeia, ou conjunto, de valores inteiros de comprimento  $n$ , representando uma solução do problema. Já, F representa a função *fitness* (expressão 2) que retorna um valor positivo e real na avaliação de cada indivíduo e,  $\Theta$  é o operador de

seleção de pais que escolhe  $r$  indivíduos de  $P$  (expressão 3). O conjunto de operadores genéticos é representado por  $\Omega$ , incluindo o cruzamento ( $\Omega_C$ ), o operador de mutação ( $\Omega_M$ ) e qualquer outro operador específico que produza  $s$  filhos a partir de  $r$  pais (expressão 4).

$$F : S_i \rightarrow R^+, i = 1, 2, \dots, N \quad (2)$$

$$\Theta : P \rightarrow \{p_1, p_2, \dots, p_r\} \quad (3)$$

$$\Omega = \{\Omega_C, \Omega_M, \dots\} : \{p_1, p_2, \dots, p_r\} \rightarrow \{f_1, f_2, \dots, f_s\} \quad (4)$$

Por último, é o operador de remoção que retira  $s$  indivíduos selecionados na população  $P$ , permitindo, por exemplo, que  $s$  filhos (novas configurações) sejam adicionados à nova população  $P_{t+1}$  (expressão 5):

$$P_{t+1} = P_t \setminus \Psi(P_t) + \{f_1, f_2, \dots, f_s\} \quad (5)$$

## PROCEDIMENTO EXPERIMENTAL

O método de referência utilizado para a determinação do teor de tanino foi o de Folin-Denis.<sup>29</sup> Foram pesados  $600 \pm 0,1$  mg de amostra de café, transferiu-se para balão volumétrico de 500 mL com água destilada, agitou-se e completou-se o volume. Pipetou-se 5 mL da solução e transferiu-se para balão volumétrico de 1000 mL, contendo 75 mL de água destilada. Adicionou-se 5 mL do reagente de Folin-Denis<sup>8</sup> e 10 mL de solução saturada de carbonato de sódio. Diluiu-se para 100 mL, agitou-se, filtrou-se após 30 minutos e determinou-se a absorvância em espectrofotômetro a 760 nm. O teor de tanino foi determinado construindo-se uma curva analítica com padrão de ácido tânico e reação com o reagente de Folin-Denis.

<sup>8</sup> 100 g de tungstato de sódio dihidratado p.a., 20 g de ácido fosfomolibdico e 50 mL de ácido fosfórico p.a. concentrado em 750 mL de água destilada. Refluxa-se por 2 horas, esfria-se e dilui-se para um litro

Os espectros no infravermelho médio foram coletados em um espectrofotômetro Nicolet 520, sendo realizadas 2 réplicas para cada amostra, com resolução de  $4 \text{ cm}^{-1}$ , 16 varreduras para cada espectro e a região espectral utilizada foi de  $650\text{-}4000 \text{ cm}^{-1}$ . O instrumento foi equipado com um acessório de reflectância difusa (JASCO), sendo os sinais expressos em  $\log(1/R)$ . Foram utilizadas 40 amostras no total, selecionando-se 22 para a construção dos modelos (44 espectros) e 18 para a validação dos mesmos (36 espectros).

## Modelagem Quimiométrica

Os modelos foram desenvolvidos em ambiente MATLAB, empregando os espectros em forma de  $\log(1/R)$ . A calibração foi realizada com o método de regressão por mínimos quadrados parciais (PLS) utilizando os espectros normalizados.

Para a avaliação dos modelos foram empregados os erros padrões de calibração (SEC) e de validação (SEV), conforme equação 6, onde  $y_i$  é o valor de referência para a  $i$ ésima amostra, e  $\hat{y}_i$  é a previsão para o valor desta amostra.

$$SEC = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (6)$$

Os modelos PLS foram otimizados empregando-se rotina baseada em AG com base na *fitness* ( $(SEV + SEC) + |(SEV - SEC)|$ ).

Como parâmetros de entrada da rotina AG foram empregados:

- a. População: 50 cromossomos
- b. Percentual de genes selecionados: 5%
- c. Percentual de mutação na população: 10%
- d. Percentual de mutação nas variáveis latentes: 30%
- e. Percentual de mutação de genes: 5%
- f. Número de iterações: 500
- g. Número máximo de variáveis latentes: 20

## RESULTADOS

Nas tabelas 1 e 2 são apresentados os resultados da otimização utilizando o algoritmo genético, sendo realizadas 5 repetições do experimento, empregando 250 e 500 iterações.

**Tabela 1** - Valores dos parâmetros da otimização na modelagem DRIFT/PLS-AG empregando 250 iterações.

Repetição	Fitness	SEV (%)	Var. Sel.	VL's	R <sup>2</sup>
A	0,926010	0,4188	108	4	0,7490
B	0,813500	0,3660	125	4	0,8052
C	0,895778	0,4051	87	4	0,7638
D	0,563484	0,2548	121	6	0,9074
E	0,853990	0,3850	103	4	0,7853

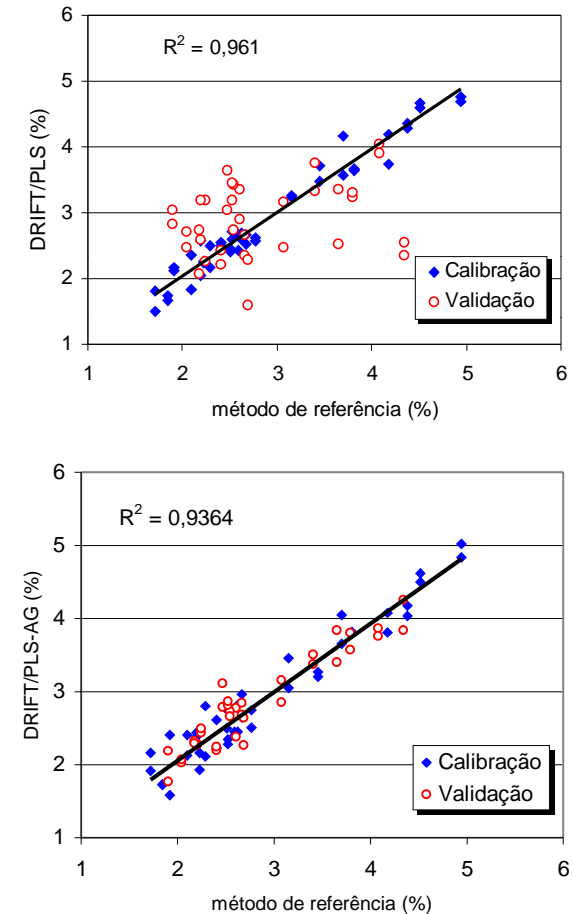
SEV: erro padrão de validação; Var. Sel.: número total de variáveis selecionadas; VL's: número de variáveis latentes selecionado; R<sup>2</sup>: coeficiente de correlação

**Tabela 2** - Valores dos parâmetros da otimização na modelagem DRIFT/PLS-AG empregando 500 iterações.

Repetição	Fitness	SEV (%)	Var. Sel.	VL's	R <sup>2</sup>
A	0,726062	0,3262	116	4	0,8447
B	0,533501	0,2406	142	5	0,9162
C	0,746170	0,3337	88	5	0,8360
D	0,470390	0,2323	130	6	0,9364
E	0,751752	0,3369	108	4	0,8335

SEV: erro padrão de validação; Var. Sel.: número total de variáveis selecionadas; VL's: número de variáveis latentes selecionado; R<sup>2</sup>: coeficiente de correlação

Com base nas Tabelas 1 e 2 foi selecionado o modelo mais representativo desta série de iterações, podendo-se visualizar na Figura 6 as curvas de calibração com e sem otimização empregando 6 variáveis latentes. Também são apresentadas nas tabelas 3 e 4 os resultados para o método de referência ( $y_i$ ) e para o método espectroscópico (DRIFT) proposto ( $\hat{y}_i$ ), para as amostras utilizadas na validação, bem como os respectivos desvios padrões.



**Figura 5:** Curvas de calibração para o modelo sem otimização (à esquerda) e com otimização empregando algoritmo genético (à direita) para o infravermelho médio (DRIFTS).

**Tabela 3** - Resultados para as amostras de validação empregando os dados DRIFT/PLS sem otimização.

Nº Amostra	$y_i$ (%)	$\sigma y_i$	$\hat{y}_i$ (%)	$\sigma \hat{y}_i$
1	3,654	0,060	2,943	0,600
4	4,079	0,043	3,970	0,105
8	4,339	0,035	2,453	0,127
11	3,790	0,099	3,265	0,049
16	3,407	0,098	3,549	0,305
21	2,607	0,104	3,136	0,311
24	2,404	0,161	2,320	0,159
25	2,243	0,044	2,722	0,657
27	2,470	0,041	3,351	0,417
29	2,199	0,136	2,894	0,422
32	2,540	0,040	3,088	0,497
33	2,173	0,198	2,402	0,477
36	2,663	0,222	2,516	0,224
37	2,040	0,166	2,604	0,169
41	1,901	0,001	2,935	0,143
46	2,684	0,090	1,941	0,490
48	3,074	0,015	2,821	0,491
55	2,520	0,032	3,327	0,190

$y_i$ : valor médio da *i*ésima amostra obtida pelo método de referência

$\sigma y_i$ : desvio-padrão do valor médio da *i*ésima amostra obtida pelo método de referência

$\hat{y}_i$ : valor médio da *i*ésima amostra estimada pelo método DRIFT/PLS

$\sigma \hat{y}_i$ : desvio-padrão do valor médio da *i*ésima amostra estimada pelo método DRIFT/PLS

**Tabela 4** - Resultados para as amostras de validação empregando os dados DRIFT/PLS-AG com otimização.

Nº Amostra	$y_i$ (%)	$\sigma y_i$	$\hat{y}_i$ (%)	$\sigma \hat{y}_i$
1	3,654	0,060	3,618	0,307
4	4,079	0,043	3,814	0,072
8	4,339	0,035	4,046	0,290
11	3,790	0,099	3,689	0,164
16	3,407	0,098	3,441	0,093
21	2,607	0,104	2,576	0,280
24	2,404	0,161	2,223	0,034
25	2,243	0,044	2,463	0,039
27	2,470	0,041	2,950	0,227
29	2,199	0,136	2,276	0,013
32	2,540	0,040	2,697	0,049
33	2,173	0,198	2,310	0,022
36	2,663	0,222	2,762	0,117
37	2,040	0,166	2,048	0,030
41	1,901	0,001	1,979	0,296
46	2,684	0,090	2,456	0,261
48	3,074	0,015	3,003	0,217
55	2,520	0,032	2,841	0,033

$y_i$ : valor médio da *i*ésima amostra obtida pelo método de referência

$\sigma y_i$ : desvio-padrão do valor médio da *i*ésima amostra obtida pelo método de referência

$\hat{y}_i$ : valor médio da *i*ésima amostra estimada pelo método DRIFT/PLS-AG

$\sigma \hat{y}_i$ : desvio-padrão do valor médio da *i*ésima amostra estimada pelo método DRIFT/PLS-AG

## CONCLUSÕES

Os resultados apresentados demonstram a vantagem da otimização dos modelos de regressão empregando algoritmos genéticos. Tanto os modelos com 250 iterações, quanto os com 500, possibilitam a obtenção de modelos de calibração com  $R^2$  superiores a 0,90, porém a principal diferença reside no fato dos modelos com 500 iterações apresentarem  $R^2$  superiores e valores de SEV menores.

Podemos verificar que as concentrações de tanino em amostras de café cru podem ser facilmente determinadas empregando regressão por mínimos quadrados parciais em conjunto com esta técnica de otimização



combinatória (AG), obtendo-se assim rotinas de análise que podem ser facilmente empregadas para o controle de matérias-primas na indústria alimentícia, tendo como vantagens a não destruição das amostras, a não necessidade de descarte de produtos nocivos ao ambiente, além do baixo tempo de análise.

### AGRADECIMENTOS

Ao Consórcio Brasileiro de Pesquisa e Desenvolvimento do Café e à FUNAPE, pelo apoio financeiro para a realização deste trabalho. À Fundação de Amparo a Pesquisa do Rio Grande do Sul pela bolsa concedida (processo número 01503479).

### REFERÊNCIAS

- CARVALHO, C. W.; MÜLLER, E. I.; FERRÃO, M. F.; DAVANZO, C. U. *Redes*, 7 (número especial), p. 141-167, 2002.
- SCHOLZ, M. B. S.; PRETE, C. E. C.; CRUDI, E. & MAGRI, T. B. Composição química de variedades (coffee arábica). In: Simpósio de Pesquisa dos Cafés do Brasil. *Resumos expandidos*, EMBRAPA Café, Belo Horizonte: Minosplan, p.673-676, 2000.
- MEDIKKA. <http://www.medikka.com/revistamed/tema.html>; Acesso em: 25 de fev. de 2003.
- ROBERTS, J. E.; SPEEDIE, M. K.; TYLER, V. E. *Farmacognosia e Farmacobioteecnologia*. São Paulo: Premier, 1997.
- SIMÕES, C. M. O. *Farmacognosia: da planta ao medicamento*. 4. ed. PortoAlegre/ Florianópolis: Ed. da universidade/UFRGS/Ed. da UFSC, 2002.
- SCALBERT, A. *Phytochemistry*, n. 30, p. 3875-3883, 1991.
- OKUDA, T.; YOSHIDA, T.; HATANO, T. *Phytochemistry*, n. 32, p. 507-521, 1993.
- MARSTON, A.; HOSTETTMANN, K. *Phytochemistry*, n. 24, p. 639-652, 1985.
- HATTORI, M.; KUSUMOTO, I.; NAMBA, T.; ISHIGAMI, T. & HARA, Y. *Chemical and Pharmaceutical. Bulletin*. n. 38, p. 717-720, 1990.
- OKUDA, T.; YOSHIDA, H.; HATANO, T. *Planta Medica*, n. 55, p. 117-122, 1989.
- FERRÃO, M. F. *Tecno-Lógica*, n. 5, p. 63-85, 2001.
- ABDULLAH, A.H.; SHERMAN, W.F. *Vibrational Spectroscopy*. n. 13, p. 133-142, 1997.

- BIRTH, G.S.; HECHT, H.G. The physics of near-infrared reflectance – in *Near-infrared technology in the agricultural and food industries* – by Phil Williams & Karl Norris, St. Paul: American Association of Cereal Chemists, 1987.
- CULLER, S.R. *Diffuse reflectance infrared spectroscopy: Sampling techniques for qualitative/quantitative analysis of solids* – in *Practical sampling techniques for infrared analysis* – by Patricia B. Coleman, Boca Raton: CRC Press, 1993.
- OSBORNE, B. G.; FEARN, T.; HINDLE, P. H. *Practical NIR spectroscopy with applications in food and beverage analysis* – Singapore: Longman Scientific & Technical, 1993.
- WETZEL, D. L. *Analytical. Chemistry*, n. 55, p. 1165A-1176A, 1983.
- OLINGER, J. M.; GRIFFITHS, P. R. *Applied Spectroscopy*, n. 47, p. 687-694, 1993.
- OLINGER, J. M.; GRIFFITHS, P. R. *Applied Spectroscopy*, n. 47, p. 695-701, 1993.
- DEVAUX, M.F.; DUFOUR, N. N.; ROBERT, P.; BERTRAND, D. *Applied Spectroscopy*, n. 49, p. 84-91, 1995.
- HOLLAND, J. H. *Adaptation in natural and artificial systems*. MIT Press, p. 11-147, 1975.
- BÄCK, TH.; SCHWEFEL, H. P. *Evolutionary Computation*. n. 1, p. 1-23, 1993.
- GOLDBERG, D. E. *Genetic algorithms in search, optimization and machine learning*. Addison-Wesley, Reading, MA, p. 11-172, 1989.
- LORENA, L.A.N.; FURTADO, J.C.; *Evolutionary Computation*, n. 9, p. 309-327, 2001.
- MITCHELL, M. *An Introduction to Genetic Algorithms*. MIT Press, Cambridge, England, 1996.
- VIANA, G.V.R.. *Meta-heurísticas e programação paralela em otimização combinatória.*, Fortaleza: EUFC, 1998, 250 p.
- BAKER, J. E. *Adaptive selection methods for genetic algorithms*. In: J. J. Grefenstette, ed., *Proceedings of the First International Conference on Genetic Algorithms and Their Applications*. Erbaum. 1985.
- FOGARTY, T.C. *Varying the Probability of Mutation in the Genetic Algorithm*. Procs. of the 3<sup>rd</sup>. Int. Conf. on Genetic Algorithms (ICGA), J. D. Schaffer (ed.), Morgan Kaufmann Publ., p. 104-109, 1989.
- GOLDBARG, M.C. *Otimização combinatória e programação linear: modelos e algoritmos.*, Rio de Janeiro: Campus, 2000.
- ASSOCIATION OF OFFICIAL ANALYTICAL CHEMISTS. *Official methods of analysis of the A.O.A.C.* Washington D.C., U.S.A. 10<sup>th</sup> 1965 (n. 9.061).