



**UNIVERSIDADE TECNOLÓGICA  
FEDERAL DO PARANÁ**  
Programa de Pós-Graduação em Tecnologia de  
Alimentos

**Estudo dos Genótipos de Café Arábica Utilizando FTIR  
e Redes Neurais Artificiais**

**Jade Varaschim Link**

Campo Mourão  
2013

**Jade Varaschim Link**

**Estudo dos Genótipos de Café Arábica Utilizando FTIR  
e Redes Neurais Artificiais**

Dissertação apresentada ao programa de Pós Graduação em Tecnologia de Alimentos da Universidade Tecnológica Federal do Paraná, como parte dos requisitos para obtenção do título de mestre em Tecnologia de Alimentos.

Campo Mourão

2013



## **TERMO DE APROVAÇÃO**

### **ESTUDO DOS GENÓTIPOS DE CAFÉ ARÁBICA UTILIZANDO FTIR E REDES NEURAIAS ARTIFICIAIS**

Por

**JADE VARASCHIM LINK**

Essa dissertação foi apresentada às 14 horas e 30 minutos do dia 25 de fevereiro de 2013, como requisito parcial para a obtenção do Título de Mestre em Tecnologia de Alimentos, Linha de Pesquisa Processos Tecnológicos na Indústria de Alimentos, Programa de Pós-Graduação em Tecnologia de Alimentos, da Universidade Tecnológica Federal do Paraná. O candidato foi arguido pela Banca Examinadora composta pelos professores abaixo assinados. Após deliberação, a Banca Examinadora considerou o trabalho aprovado.

---

Prof. Dr. Evandro Bona (Orientador – PPGTA)

---

Prof. Dr. Wyrllen Everson de Souza (Membro Externo)

---

Prof. Dr. Dionísio Borsato (Membro Externo)

A folha de aprovação assinada encontra-se na Coordenação do Programa

**Orientador**  
**Prof Dr Evandro Bona**

## **AGRADECIMENTOS**

Ao professor Dr. Evandro Bona, pela orientação e dedicação durante esses 2 anos de trabalho.

Aos professores do Programa de Pós-Graduação em Tecnologia de Alimentos - PPGTA e para a secretária Marlene Alamini Bau, pela contribuição, apoio e cooperação.

Ao professor Dr. Dionísio Borsato, a técnica de laboratório Msc. Ivanira Moreira e a todos do Departamento de Química da Universidade Estadual de Londrina (UEL) pelo apoio nas análises no equipamento FTIR.

A Dra. Maria Brígida dos Santos Scholz, pela colaboração, e ao pessoal do Instituto Agrônomo do Paraná - Londrina (IAPAR) pelo fornecimento das amostras de café e realização das análises químicas.

Ao aluno de Iniciação Científica André Luis Guimarães Lemes por todo o auxílio durante a pesquisa.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pelo suporte financeiro ao projeto.

À Coordenação de Aperfeiçoamento de Pessoal em Nível Superior (CAPES), pela concessão da bolsa de mestrado.

## **BIOGRAFIA**

Possui graduação em Engenharia de Alimentos pela Universidade Estadual do Centro-Oeste (UNICENTRO) (2009). Foi monitor da disciplina de Engenharia de Processos I (Fenômenos de transporte I e Operações unitárias I) do curso de Engenharia de Alimentos.

## GENERAL ABSTRACT

The climatic conditions of the coffee crop give special attributes to the beverage and could improve its value. However, it is essential to prove the geographical and genotype origin of the cultivar using reliable methods. Several mathematical methods have been developed in an attempt to reproduce the human capability of pattern recognition. The artificial neural networks (ANNs) are a set of techniques based on statistical principles, which is currently growing in the food science to perform tasks of regression and pattern recognition. ANNs are techniques that can represent complex and nonlinear input-output relationships. In this study three types of artificial neural network were used, namely the self-organizing maps (unsupervised learning), the multilayer perceptron (supervised learning) and the radial-basis function network (hybrid learning process), for the recognition and geographic and genotypic classification of arabic coffee. For this purpose, spectra obtained in the Fourier transform infrared (FTIR) were analyzed using different ANNs. Other techniques were also applied during neural network development: ensemble averaging to improve the network generalization and sequential simplex optimization to select parameters. It was concluded that the SOM were able to recognize the coffee genotypes and geographical origin using the chemical profile data. After the results it was concluded that the optimized multilayer perceptrons were able to classify the samples of arabica coffee geographically. However, for genotypic classification the performance was not satisfactory. The optimized radial-basis function networks showed more satisfactory results since these networks were able to classify the samples of arabica coffee geographically and genotypically. The radial-basis function networks presents the best results to genotypic classification and in addition a smaller number of free parameters compared with multilayer perceptrons, which has a high number of synaptic weights, thus a large degrees of freedom database is necessary to produce a network with generalization capability.

## RESUMO GERAL

As condições climáticas específicas de cada região de cultivo do café fornecem atributos especiais para a bebida e podem assim aumentar seu valor agregado. No entanto, é essencial provar a origem geográfica e genotípica do cultivar por meio de métodos confiáveis. Vários métodos matemáticos e estatísticos foram desenvolvidos na tentativa de reproduzir a capacidade humana de reconhecimento de padrões. As redes neurais artificiais (RNAs) são um conjunto de técnicas baseadas em princípios matemáticos e estatísticos, que vem atualmente ganhando espaço para realizar tarefas de regressão e reconhecimento de padrões. As RNAs são técnicas capazes de realizar o mapeamento de relações complexas e não lineares entre múltiplas variáveis de entrada e saída. Neste trabalho foram utilizados três tipos de rede neural artificial, o mapa auto-organizável (aprendizagem não-supervisionada), o perceptron de múltiplas camadas (aprendizagem supervisionada) e a rede de base radial (processo de aprendizagem híbrido), para o reconhecimento e classificação geográfica e genotípica de café arábica. Para esse fim a composição química e os espectros obtidos no equipamento de espectroscopia no infravermelho com transformada de Fourier (FTIR) foram analisados através do emprego de diferentes RNAs. No desenvolvimento das redes, foram aplicadas metodologias para: melhor generalização das redes (média de ensemble) e escolha dos parâmetros de rede (otimização simplex sequencial). O mapa auto-organizável apresentou foi capaz de reconhecer os genótipos de café e a sua origem geográfica, utilizando os dados do perfil químico das amostras. Após os resultados obtidos foi possível concluir que os perceptrons de múltiplas camadas otimizados foram capazes de classificar as amostras de café arábica geograficamente. Entretanto, para a classificação genotípica o desempenho não foi totalmente satisfatório. As redes de função de base radial otimizadas apresentaram resultados mais satisfatórios pois foram capazes de classificar as amostras de café arábica geograficamente e genotipicamente. Além dos melhores resultados obtidos para a classificação genotípica, as redes de função de base radial apresentam um número menor de parâmetros livres quando comparada com os perceptrons de múltiplas camadas, que apresentam um número muito elevado de pesos sinápticos, não proporcionando número suficiente de graus de liberdade para que a rede neural possa aprender de maneira confiável.



# LISTA DE ILUSTRAÇÕES

## CAPÍTULO 1

Figura 1 - Mapa auto-organizável segundo o modelo de Kohonen .....	22
Figura 2 - Convergência do erro de quantização médio durante o treinamento do mapa com 7 neurônios em cada dimensão.....	25
Figura 3 - Mapas de pesos para a variável açúcares redutores e os mapas topológicos da região e genótipos de café arábica.....	27
Figura 4 - Mapas de pesos para a variável cafeína e os mapas topológicos da região e genótipos de café arábica .....	28
Figura 5 - Mapas de pesos para a variável proteínas e os mapas topológicos da região e genótipos de café arábica .....	30
Figura 6 - Mapas de pesos para a variável ácidos clorogênicos e os mapas topológicos da região e genótipos de café arábica.....	31

## CAPÍTULO 2

Figura 1 - Mapa da distância entre as cidades dos cafés estudados.....	40
Figura 2 - Representação de um perceptron de múltiplas camadas (MLP) .....	43
Figura 3 - Espectro obtido para o genótipo IPR 105 cultivado em Cornélio Procópio .....	48

## CAPÍTULO 3

Figura 1 - Mapa da distância entre as cidades dos cafés estudados.....	64
Figura 2 - Representação de uma rede de função de base radial (RBF) .....	67
Figura 3 - Espectros normalizados obtidos para os genótipos de café arábica.....	71

# LISTA DE TABELAS

## CAPÍTULO 1

Tabela 1 - Genótipos de café arábica estudados.....	21
Tabela 2 - Médias e desvios padrão das variáveis em relação à região de cultivo .....	26
Tabela 3 - Médias e desvios padrão das variáveis em relação aos genótipos de café .....	29

## CAPÍTULO 2

Tabela 1 - Genótipos de café arábica utilizados .....	39
Tabela 2 - Parâmetros otimizados através do simplex sequencial.....	47
Tabela 3 - Resultados obtidos para os perceptron de múltiplas camadas propostos para a classificação geográfica de café arábica .....	49
Tabela 4 - Resultados obtidos os perceptron de múltiplas camadas propostos para a classificação genotípica de café arábica .....	50
Tabela 5 - Tabela de confusão com os resultados de reconhecimento e classificação genotípica da Rede 11 para as amostras de teste.....	53

## CAPÍTULO 3

Tabela 1 - Genótipos de café arábica utilizados .....	63
Tabela 2 - Parâmetros otimizados através do simplex sequencial.....	70
Tabela 3 - Resultados obtidos para as redes de base radial propostas para a classificação geográfica de café arábica .....	73
Tabela 4 - Resultados obtidos para as redes de base radial propostas para a classificação genotípica de café arábica .....	74
Tabela 5 - Comparação entre as RBFs desenvolvidas e outros métodos que foram propostos para a classificação geográfica e genotípica de café arábica.....	75
Tabela 6 - Tabela de confusão com os resultados de reconhecimento e classificação genotípica da Rede L para as amostras de teste.....	77

# LISTA DE ABREVIATURAS

**ACG** - Ácidos Clorogênicos

**ACP** - Análise de Componentes Principais

**AE** - Autoescalamento

**AR** - Açúcares Redutores

**AT** - Açúcares Totais

**Caf** - Cafeína

**CB** - *City Block*

**COR** - Correlação

**COS** - Cosseno

**CP** - Cornélio Procópio

**DE** - Distância Euclidiana

**Dens** - Densidade

**DP** - Dados Puros

**EQM** - Erro Quadrado Médio

**FL** - Função Logística

**FTIR** - Espectroscopia de infravermelho com transformada de Fourier

**G** - Gaussiana

**IAPAR** - Instituto Agronômico do Paraná

**L** - Linear

**Lip** - Lipídios

**MLP** - *Multi Layer Perceptron* (Perceptron de múltiplas camadas)

**MM** - Máximo e Mínimo (minimax)

**MQ** - Multiquadrática

**MQI** - Multiquadrática Inversa

**Pro** - Proteínas

**PV** - Paranavaí

**RBF** - *Radial Basis Function Networks* (Redes de base radial)

**RNA** - Redes Neurais Artificial

**SIMCA** - *Soft Independent Modelling of Class Analogies*

**SOM** - *Self-organizing maps* (Mapas auto-organizáveis)

**Tan** - Taninos

**THS** - Tangente Hiperbólica Sigmoidal

**VU** - Vetor Unitário

# SUMÁRIO

## **CAPÍTULO 1:.....16**

<b>1 SEGMENTAÇÃO GEOGRÁFICA E GENOTÍPICA DE CAFÉ ARÁBICA UTILIZANDO OS MAPAS AUTO-ORGANIZÁVEIS .....</b>	<b>17</b>
<b>1.1 INTRODUÇÃO .....</b>	<b>18</b>
<b>1.2 MATERIAIS E MÉTODOS .....</b>	<b>20</b>
1.2.1 AMOSTRAS DE CAFÉ ARÁBICA .....	20
1.2.2 DENSIDADE APARENTE E COMPOSIÇÃO DO GRÃO VERDE .....	21
1.2.3 MAPAS AUTO-ORGANIZÁVEIS (SOM) .....	22
<b>1.3 RESULTADOS E DISCUSSÕES .....</b>	<b>25</b>
<b>1.4 CONCLUSÃO .....</b>	<b>31</b>
<b>1.5 REFERÊNCIAS .....</b>	<b>32</b>

## **CAPÍTULO 2:.....34**

<b>2 PERCEPTRON DE MÚLTIPLAS CAMADAS OTIMIZADO PARA A CLASSIFICAÇÃO GEOGRÁFICA E GENOTÍPICA DE QUATRO GENÓTIPOS DE CAFÉ ARÁBICA .....</b>	<b>35</b>
<b>2.1. INTRODUÇÃO .....</b>	<b>36</b>
<b>2.2. MATERIAIS E MÉTODOS .....</b>	<b>38</b>
2.2.1. AMOSTRAS DE CAFÉ ARÁBICA .....	39
2.2.2. ESPECTROSCOPIA DE INFRAVERMELHO COM TRANSFORMADA DE FOURIER (FTIR).....	40
2.2.3. TRANSFORMAÇÃO DOS ESPECTROS.....	41
2.2.4. NORMALIZAÇÃO .....	42
2.2.5. PERCEPTRON DE MÚLTIPLAS CAMADAS (MLP) .....	42
2.2.6. MÉDIA DE ENSEMBLE .....	45
2.2.7. OTIMIZAÇÃO DOS PARÂMETROS DA REDE .....	46
<b>2.3. RESULTADOS E DISCUSSÕES .....</b>	<b>47</b>
2.3.1. PERCEPTRON DE MÚLTIPLAS CAMADAS (MLP) .....	48
<b>2.4. CONCLUSÃO .....</b>	<b>54</b>
<b>2.5. REFERÊNCIAS .....</b>	<b>55</b>

## **CAPÍTULO 3:.....58**

<b>3 CLASSIFICAÇÃO GEOGRÁFICA E GENOTÍPICA DE CAFÉ ARÁBICA UTILIZANDO FTIR E REDES DE BASE RADIAL .....</b>	<b>59</b>
<b>3.1 INTRODUÇÃO .....</b>	<b>60</b>
<b>3.2 MATERIAIS E MÉTODOS .....</b>	<b>62</b>
3.2.1 AMOSTRAS DE CAFÉ ARÁBICA .....	62
3.2.2 ESPECTROSCOPIA DE INFRAVERMELHO COM TRANSFORMADA DE FOURIER (FTIR).....	64
3.2.3 TRANSFORMAÇÃO DOS ESPECTROS.....	65
3.2.4 NORMALIZAÇÃO .....	66
3.2.5 REDES DE FUNÇÃO DE BASE RADIAL (RBF) .....	66
3.2.6 MÉDIA DE ENSEMBLE .....	69
3.2.7 OTIMIZAÇÃO DOS PARÂMETROS DA REDE .....	70

<b>3.3 RESULTADOS E DISCUSSÕES</b> .....	71
3.3.1 REDES DE FUNÇÃO DE BASE RADIAL (RBF) .....	72
<b>3.4 CONCLUSÃO</b> .....	78
<b>3.5 REFERÊNCIAS</b> .....	79

**CONCLUSÃO GERAL.....81**

**APÊNDICE.....83**

<b>4.1 APÊNDICE A</b> .....	84
4.1.1 MAPAS DE PESO PARA AS DEMAIS VARIÁVEIS ESTUDADAS E MAPAS TOPOLÓGICOS DA REGIÃO E GENÓTIPOS DE CAFÉ ARÁBICA .....	84
<b>4.2 APÊNDICE B</b> .....	86
4.2.1. OTIMIZAÇÃO SIMPLEX .....	86
4.2.2. REFERÊNCIAS .....	90
<b>4.3 APÊNDICE C</b> .....	92
4.3.1. ALGORITMO K-MÉDIAS (K-MEANS) .....	92
4.3.2. REFERÊNCIAS .....	93

**ANEXOS.....94**

<b>5.1 ANEXO A</b> .....	95
5.1.1 NORMAS PARA PUBLICAÇÃO NO PERIÓDICO JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY .....	95
<b>5.2 ANEXO B</b> .....	109
5.2.1 NORMAS PARA PUBLICAÇÃO NA REVISTA BRASILEIRA DE PESQUISA EM ALIMENTOS (REBRAPA) .....	109
<b>5.3 ANEXO C</b> .....	112
5.3.1 NORMAS PARA PUBLICAÇÃO NO PERIÓDICO FOOD CHEMISTRY .....	112

# APRESENTAÇÃO

Esta dissertação é composta por 3 capítulos que foram elaborados na forma de artigos para publicação e estão colocados nas normas das revistas para onde serão encaminhados.

**Capítulo 1** - Autores: Jade Varaschim Link; André Luis Guimarães Lemes; Izabele Marquetti; Maria Brígida dos Santos Scholz; Evandro Bona. Título: Segmentação geográfica e genotípica de café arábica utilizando os mapas auto-organizáveis. Periódico: Journal of Agricultural and Food Chemistry.

**Capítulo 2** - Autores: Jade Varaschim Link; André Luis Guimarães Lemes; Herily Pereira Sato; Maria Brígida dos Santos Scholz; Evandro Bona. Título: Perceptron de múltiplas camadas otimizado para a classificação geográfica e genotípica de quatro genótipos de café arábica: Revista Brasileira de Pesquisa em Alimentos (REBRAPA).

**Capítulo 3** - Autores: Jade Varaschim Link; André Luis Guimarães Lemes; Izabele Marquetti; Maria Brígida dos Santos Scholz; Evandro Bona. Título: Classificação geográfica e genotípica de café arábica utilizando FTIR e redes de base radial. Periódico: Food Chemistry.

## **CAPÍTULO 1:**

Segmentação geográfica e genotípica de café arábica utilizando os mapas auto-organizáveis



# 1 SEGMENTAÇÃO GEOGRÁFICA E GENOTÍPICA DE CAFÉ ARÁBICA UTILIZANDO OS MAPAS AUTO-ORGANIZÁVEIS

## Resumo

Vários métodos estatísticos e matemáticos foram desenvolvidos na tentativa de reproduzir a capacidade humana de reconhecimento de padrões. Os mapas auto-organizáveis (SOM) são um tipo de rede neural artificial (RNA) com aprendizagem não supervisionada desenvolvido para examinar a estrutura de dados multidimensionais. Este estudo teve como objetivo realizar uma segmentação geográfica e genotípica do café arábica cultivado no Brasil utilizando o algoritmo SOM para análise de agrupamentos. Foram coletados 14 genótipos de café arábica de duas diferentes cidades (Paranavaí e Cornélio Procópio). A densidade, os teores de cafeína, ácidos clorogênicos, taninos, açúcares totais e redutores, proteínas e lipídios dos grãos de café verde foram analisados. Utilizando esses dados o SOM foi capaz de discriminar os 14 genótipos e também foi observada uma segmentação da origem geográfica. Os açúcares redutores, cafeína e ácidos clorogênicos foram as variáveis mais importantes para a separação da região de cultivo dos genótipos de café arábica. Os grupos formados e as relações entre eles foram muito semelhantes às obtidas pela ACP (análise de componentes principais), porém o SOM apresentou uma melhor representação gráfica dos agrupamentos. Concluiu-se que o SOM foi capaz de reconhecer os genótipos de café e a sua origem geográfica, utilizando o perfil químico das amostras.

*Palavras-chave:* café verde, SOM, análise de componentes principais, redes neurais artificiais

## 1.1 INTRODUÇÃO

O reconhecimento de padrões é um processo pelo qual um sinal recebido é atribuído a uma classe dentre um número predeterminado de categorias. Os seres humanos possuem excelência nata para a aprendizagem e reconhecimento de padrões. Nesse sentido, os métodos estatístico-matemáticos têm sido desenvolvidos na tentativa de mimetizar, mesmo de maneira rudimentar, tal aptidão humana <sup>1,2</sup>. Tradicionalmente, para um estudo exploratório e compactação de dados, a Análise de Componentes Principais (ACP) é uma das primeiras metodologias multivariadas escolhidas para a análise de agrupamentos. No entanto, a técnica de transformação da ACP tem algumas desvantagens. Primeiramente, é explicitamente assumido que a estrutura dimensional (topologia) dos dados de entrada pode ser reduzida a uma forma linear. O segundo ponto crítico na ACP é a qualidade do resultado que pode ser influenciado por amostras discrepantes. Terceiro, o poder de visualização dessas técnicas de transformação deterioram consideravelmente se o número de dimensões relevantes no espaço multivariado permanecer elevado <sup>3,4</sup>.

Um método extremamente versátil para realizar o mapeamento de relações complexas e não lineares entre múltiplas variáveis de entrada e saída são as redes neurais artificiais (RNAs). As RNAs são um conjunto de técnicas baseadas em princípios matemáticos, que vem crescentemente ganhando espaço para realizar tarefas de regressão e reconhecimento de padrões. É um processo maciçamente paralelamente distribuído constituído de unidades de processamentos simples, que tem a propensão natural para armazenar conhecimento experimental e torná-lo disponível para uso <sup>2,5,6</sup>. Um tipo de RNA de aprendizagem não supervisionada desenvolvida para examinar a estrutura de dados multidimensionais são os mapas auto-organizáveis (SOM), introduzidos por Kohonen <sup>7</sup>. Em um SOM os neurônios estão colocados em nós de uma grade que é normalmente bidimensional. Esses neurônios

passam por um processo de aprendizagem competitiva, ou seja, competem entre si para serem ativados. O neurônio vencedor se torna seletivamente mais adaptado ao padrão de entrada incidente. Assim, é formado um mapa topográfico dos padrões de entrada no qual as localizações espaciais dos neurônios na grade são indicativas das características estatísticas intrínsecas contidas nos padrões de entrada. Devido ao fato de um mapa auto-organizável ser inerentemente não linear, ele pode ser considerado como uma generalização não linear da análise de componentes principais <sup>2</sup>. Os mapas auto-organizáveis têm sido amplamente aplicados na área de ciência e tecnologia de alimentos na caracterização de variedades de morangos <sup>8</sup>; identificação de misturas binárias de azeite de oliva italianos <sup>9</sup>; caracterização de amostras de alecrim de acordo com suas origens geográficas <sup>10</sup>; reconhecimento de padrões aromáticos de café solúvel <sup>11</sup>.

O café é um dos mais valiosos produtos primários comercializados no mundo. Seu cultivo, processamento, comercialização, transporte e mercado proporcionam milhões de empregos em todo o mundo <sup>12</sup>. O consumo brasileiro de café aumenta a cada ano, no período entre maio de 2011 e abril de 2012 foi registrado o consumo de 19,975 milhões de sacas, representando um acréscimo de 3,05% em relação ao período anterior correspondente <sup>13</sup>. Além do mercado interno, as exportações de café verde do Brasil somaram 116,63 mil toneladas em março de 2012, baixa de 19,2% em relação ao mesmo mês de 2011 <sup>14</sup>.

Duas são as espécies principais de café, o *Coffea arabica* variedade Arábica e o *Coffea canephora* variedade Robusta <sup>15,16,17</sup>. Estas espécie apresentam uma composição química muito distinta e o café arábica fornece um bebida com qualidade e aroma superior ao café robusta <sup>16</sup>. As condições climáticas específicas de cada região de cultivo do café fornecem atributos especiais para a bebida e podem assim aumentar seu valor agregado. No entanto, é essencial provar a origem geográfica e genotípica do cultivar por meio de métodos confiáveis

<sup>4</sup>. Devido a este fato, este trabalho teve como objetivo realizar uma avaliação geográfica e genotípica do café arábica cultivado no Brasil. Para este fim foi realizada uma análise de agrupamentos por meio dos mapas auto-organizáveis (SOM) utilizando os dados de densidade do grão e de composição química dos genótipos.

## **1.2 MATERIAIS E MÉTODOS**

### **1.2.1 AMOSTRAS DE CAFÉ ARÁBICA**

Aproximadamente 3kg de 14 genótipos (27 amostras analisadas em duplicata) de café arábica catalogados pelo Instituto Agrônomo do Paraná - Londrina (IAPAR) foram coletados (safra de 2010) e transportados para Londrina. No IAPAR, as amostras foram imediatamente colocadas em caixas de madeira com uma malha de fundo e movidas oito vezes por dia até que a umidade dos grãos chegou a 11-12% e logo após as amostras foram beneficiadas (remoção da casca e pergaminho). Para as análises, os grãos de café foram congelados com nitrogênio líquido e moídos em moinho de disco (modelo Perten 3600) com 0,6 milímetros de tamanho de partícula final <sup>18</sup>.

Os grãos verdes, moídos, peneirados e embalados dos genótipos de café arábica foram utilizados posteriormente para análise. Para alguns genótipos foram fornecidas amostras cultivadas em diferentes cidades. As análises para se determinar a composição química das amostras de café verde foram realizadas no IAPAR.

Tabela 1 - Genótipos de café arábica estudados.

<b>Genótipo</b>	<b>Cidade de Cultivo</b>
IPR 97	Paranavaí
IPR 99	Cornélio Procópio Paranavaí
IPR 100	Paranavaí
IPR 101	Cornélio Procópio
IPR 102	Paranavaí
IPR 105	Cornélio Procópio Paranavaí
IPR 106	Cornélio Procópio Paranavaí
IA 59	Cornélio Procópio Paranavaí
IA 59 enxertado	Paranavaí
IPR 105 enxertado	Paranavaí
Bourbon	Cornélio Procópio
Catuai	Paranavaí
Tupi	Paranavaí
Mundo Novo	Cornélio Procópio

### 1.2.2 DENSIDADE APARENTE E COMPOSIÇÃO DO GRÃO VERDE

A densidade aparente do café verde foi determinada pelo método de queda livre, segundo método de Buenaventura-Serrano e Castaño-Castrillón <sup>19</sup>. Os resultados foram expressos em g.mL<sup>-1</sup>. A cafeína foi extraída com óxido de magnésio e determinada pelo método espectrofotométrico <sup>20</sup>. Os ácidos clorogênicos totais foram verificados de acordo com a metodologia proposta por Clifford e Wight <sup>21</sup>. Taninos totais foram determinados com o reagente Folin-Ciocalteu usando ácido gálico como padrão <sup>22</sup>. Açúcares totais e redutores foram determinados de acordo com a metodologia especificada por Somogyi-Nelson <sup>23</sup>. Proteínas e lipídios foram determinados pelo respectivo método proposto pela AOAC (Association of Official Agricultural Chemists) <sup>22</sup>.

### 1.2.3 MAPAS AUTO-ORGANIZÁVEIS (SOM)

Neste trabalho foi introduzido o algoritmo SOM (em inglês *Self-organizing Maps*) bidimensional <sup>2</sup>. O principal objetivo do mapa auto-organizável é transformar um padrão de sinal incidente de dimensão arbitrária em um mapa discreto bidimensional e realizar esta transformação adaptativamente de uma maneira topologicamente ordenada <sup>2,8</sup>. No modelo de Kohonen (Figura 1) cada neurônio da grade está totalmente conectado com todos os nós de fonte da camada de entrada. O algoritmo responsável pela formação do mapa auto-organizável começa primeiramente inicializando os pesos sinápticos da grade. Isto é feito atribuindo-lhes valores pequenos (entre 0 e 1) tomados de um gerador de números aleatórios. Fazendo desta forma, nenhuma organização prévia é imposta ao mapa de características. Uma vez que a grade tenha sido apropriadamente inicializada, há três processos essenciais envolvidos na formação do mapa auto-organizável: competição, cooperação e adaptação sináptica.

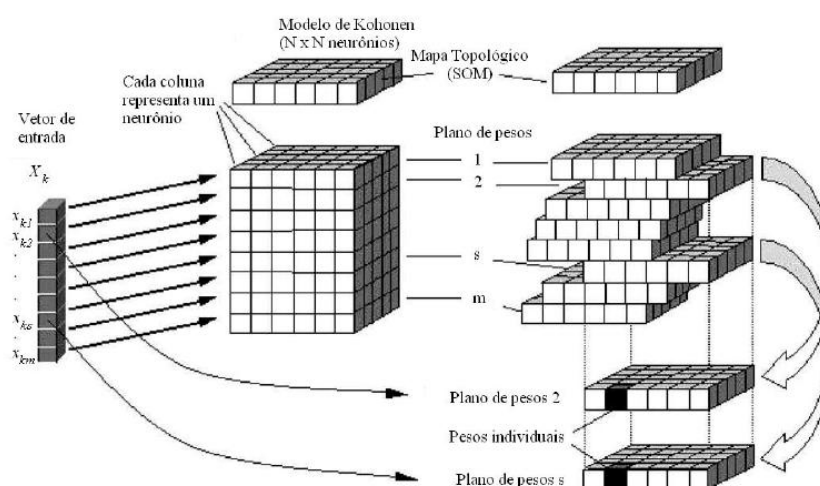


Figura 1 - Mapa auto-organizável segundo o modelo de Kohonen. Na figura  $k$  representa o número de padrões de entrada,  $m$  é a quantidade de variáveis de entrada e  $N$  é o número de neurônios em cada dimensão <sup>24</sup>.

No processo de competição, para cada padrão de entrada, os neurônios da grade calculam seus respectivos valores de distância euclidiana. Estes valores fornecem a base para a competição entre os neurônios<sup>2,9,10</sup>. O neurônio com menor distância euclidiana é declarado vencedor da competição. A função escolhida para representar a vizinhança topológica deve satisfazer as seguintes propriedades: (i) ser simétrica em relação ao ponto de máximo definido pela posição do neurônio vencedor; (ii) sua amplitude deve ser monotonamente decrescente com o aumento da distância em relação ao neurônio vencedor. Uma escolha típica que satisfaz estas exigências é a função gaussiana apresentada na equação (1)<sup>2,10</sup>.

$$h_{j,i} = \exp\left(\frac{d_{j,i}^2}{2\sigma^2}\right) \quad (1)$$

Na equação (1) o parâmetro  $\sigma$  é o raio efetivo da vizinhança topológica.

Uma outra característica única do algoritmo SOM é que o tamanho da vizinhança topológica diminui com o tempo de treinamento. Esta exigência é satisfeita fazendo-se com que a largura da função de vizinhança topológica diminua com o tempo garantindo a formação de regiões topológicas especializadas. Uma escolha para implementar essa redução é a partir de um decaimento exponencial (equação 2).

$$\sigma(n) = \sigma_0 \exp\left(-\frac{n}{\tau_1}\right) \quad n = 0, 1, 2, \dots, \quad (2)$$

onde  $\sigma_0$  é o valor de  $\sigma$  na inicialização do algoritmo SOM,  $\tau_1$  é uma constante de tempo, sendo recomendado o uso de  $\tau_1 = 1000 / \log \sigma_0$  e  $n$  representa uma época de treinamento, ou seja, uma apresentação completa de todos os padrões de entrada<sup>2</sup>.

No processo adaptativo sináptico, para que a grade seja auto-organizável, é necessário que o vetor de peso sináptico  $\mathbf{w}_j$  do neurônio  $j$  da grade se modifique em relação ao vetor de entrada  $\mathbf{x}$ . Para realizar o processo adaptativo sináptico, é utilizada uma modificação do postulado de aprendizagem de Hebb, usando um formalismo de tempo discreto,

matematicamente, o processo adaptativo é dado pela equação (3), que é aplicada a todos os neurônios da grade que se encontram dentro da vizinhança topológica do neurônio vencedor<sup>11</sup>.

$$\mathbf{w}_j(n+1) = \mathbf{w}_j(n) + \eta(n)h_{j,i(x)}(n) (\mathbf{x} - \mathbf{w}_j(n)) \quad n = 0, 1, 2, \dots \quad (3)$$

onde  $\eta(n)$  é a taxa de aprendizagem que, como representado, deve ser variável e decrescente ao longo as  $n$  épocas de treinamento.

É possível decompor a adaptação dos pesos sinápticos da grade em duas fases: uma fase de ordenação ou de auto-organização seguida por uma fase de convergência. Na etapa de ordenação a taxa de aprendizagem deve ser maior para garantir que o mapa topológico sofra maiores ajustes em seus pesos sinápticos, tornando-se ordenado. Na fase de convergência é realizada uma sintonia fina do mapa usando uma taxa de aprendizagem menor<sup>2,11</sup>. O decréscimo da taxa de aprendizagem também pode ser realizado através de um decaimento exponencial, como apresentado na equação (4).

$$\eta(n) = \eta_0 \exp\left(-\frac{n}{\tau_2}\right), \quad n = 0, 1, 2, \dots \quad (4)$$

onde  $\eta_0$  é a taxa de aprendizagem inicial e  $\tau_2$  é uma outra constante de tempo, sendo 1000 o valor recomendado<sup>2</sup>.

Após a fase de treinamento concluída, todo o conjunto de amostras é apresentada mais uma vez à rede e cada uma das posições no mapa bidimensional são rotuladas de acordo com a forma que as amostras foram mapeadas para ele, a rotulagem do mapa pode ser feita de diferentes formas, de acordo com qual propriedade da amostra se quer enfatizar (membro de classes, distribuição dos valores de um índice definido experimentalmente, ou apenas a identidade da amostra)<sup>5,8</sup>.

Foram fornecidas como variável de entrada para os mapas auto-organizáveis os dados de densidade e composição química das amostras. Para os 14 genótipos de café arábica (Tabela 1) as análises foram realizadas em duplicata, totalizando assim, 54 amostras. Foram



testados mapas bidimensionais que apresentavam 7 e 8 neurônios em cada dimensão. O mapa foi treinado ao longo de 7000 épocas para garantir uma convergência do erro de quantização médio.

Todas as análises matemáticas e/ou estatísticas, assim como os mapas auto-organizáveis e a ACP (análise de componentes principais) foram realizadas no software MATLAB R2007b (The MathWorks Inc., Natick, USA).

### 1.3 RESULTADOS E DISCUSSÕES

Os mapas bidimensionais com 7 e 8 neurônios em cada dimensão apresentaram distribuição e visualização semelhantes dos agrupamentos. Portanto, foi utilizado o mapa com 7 neurônios por ser um modelo neural mais simples e demandar menor tempo de treinamento. De acordo com o gráfico de treinamento que mostra a variação do erro de quantização médio (Figura 2), as 7000 épocas de treinamento foram suficientes para garantir a convergência.

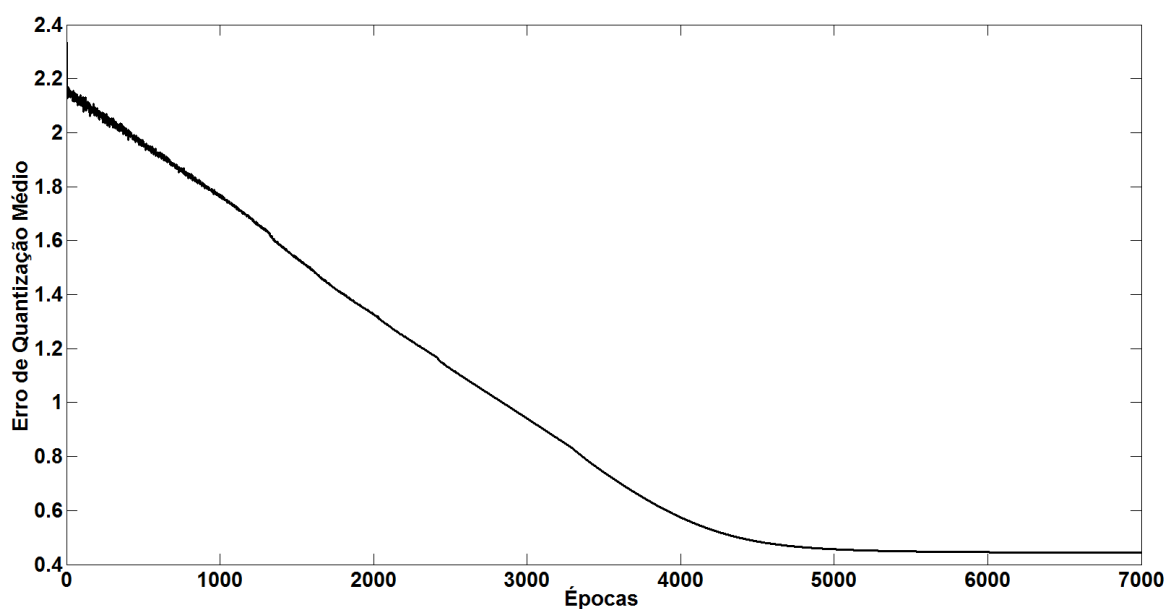


Figura 2 - Convergência do erro de quantização médio durante o treinamento do mapa com 7 neurônios em cada dimensão.

Para visualizar a diferença entre os cafés em relação à região de cultivo e ao genótipo, foram elaborados oito mapas de pesos referentes às variáveis estudadas (densidade, açúcares totais, açúcares redutores, cafeína, proteínas, ácidos clorogênicos, lipídios e taninos) e os mapas topológicos das cidades de cultivo e genótipos de café arábica.

Tabela 2 - Médias, desvios padrão e p valores das variáveis em relação à cidade de cultivo.

Variável	Cidade		p (ANOVA)
	Paranavaí	Cornélio Procópio	
Densidade	0,690 ± 0,021 <sup>a</sup>	0,680 ± 0,021 <sup>a</sup>	0,096
Açúcares Totais	7,813 ± 0,447 <sup>a</sup>	7,981 ± 0,472 <sup>a</sup>	0,197
Açúcares Redutores	0,396 ± 0,081 <sup>a</sup>	0,320 ± 0,054 <sup>b</sup>	0,000
Cafeína	1,183 ± 0,125 <sup>b</sup>	1,332 ± 0,091 <sup>a</sup>	0,000
Proteínas	16,386 ± 0,593 <sup>a</sup>	16,176 ± 0,507 <sup>a</sup>	0,191
Ácidos Clorogênicos	9,676 ± 1,363 <sup>a</sup>	8,648 ± 0,991 <sup>b</sup>	0,005
Lipídios	14,818 ± 1,531 <sup>a</sup>	15,313 ± 1,025 <sup>a</sup>	0,205
Taninos	6,054 ± 0,957 <sup>a</sup>	6,280 ± 0,960 <sup>a</sup>	0,405

\* Médias da mesma linha com a mesma letra não diferem significativamente ( $p > 0,05$ ).

As Figuras 3 a 6 apresentam os mapas de pesos e topológicos para as variáveis açúcar redutor, cafeína, proteína e ácidos clorogênicos. Essas variáveis foram escolhidas por apresentarem uma segmentação mais clara das amostras, os demais mapas se encontram no Apêndice A. A Tabela 2 apresenta as médias e desvios padrão das variáveis em relação à cidade de cultivo.

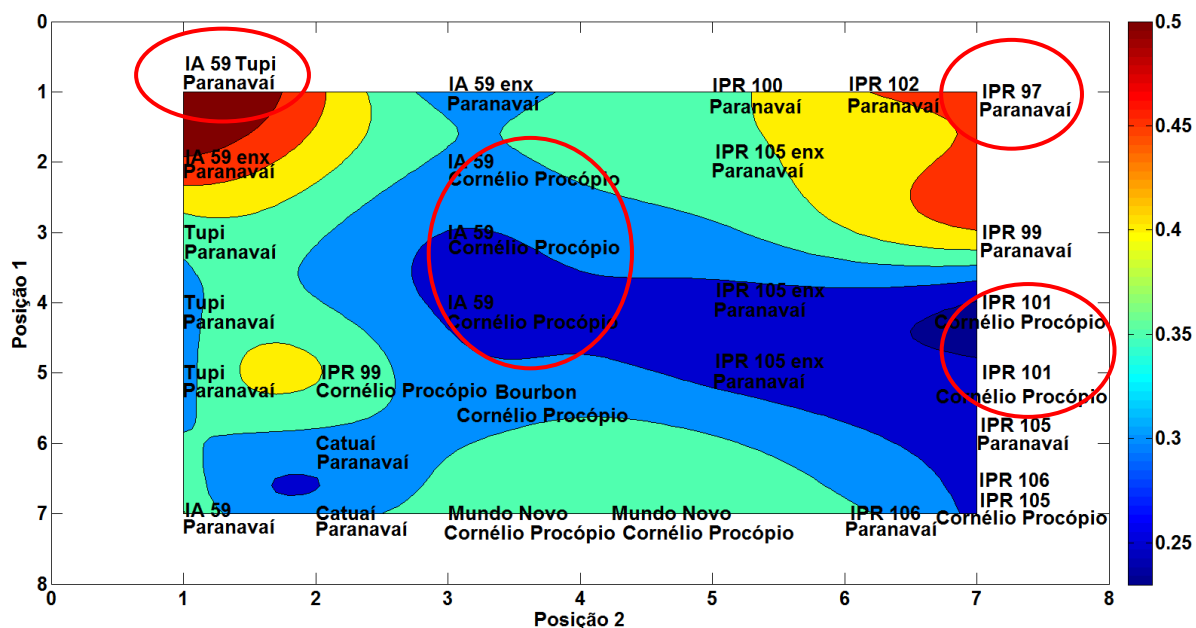


Figura 3 - Mapas de pesos para a variável açúcares redutores e os mapas topológicos das cidades e genótipos de café arábica. Na escala de cores as unidades são g/100g.

A região no mapa de pesos da Figura 3 de maior concentração de açúcares redutores se formou onde haviam alguns cafés de Paranaíba (IA 59, IPR 97 e IA 59 enx). De acordo com o teste de Tukey (Tabela 2) é observada diferença estatisticamente significativa entre os cafés de Paranaíba e Cornélio Procópio. Essa influência regional no teor de açúcar redutor fica mais evidente quando comparados os genótipos IA 59, IPR 106 e IPR 99 que foram cultivados em ambas as regiões sendo que as amostras de Paranaíba apresentaram sempre maiores índices de açúcar redutor, com exceção do genótipo IPR 105 que apresentou índices baixos para as duas regiões. Em relação aos genótipos as amostras IPR 97 e IPR 101 apresentaram, respectivamente, as maiores e menores quantidades de açúcares redutores. Os resultados estão de acordo com os dados da Tabela 3.

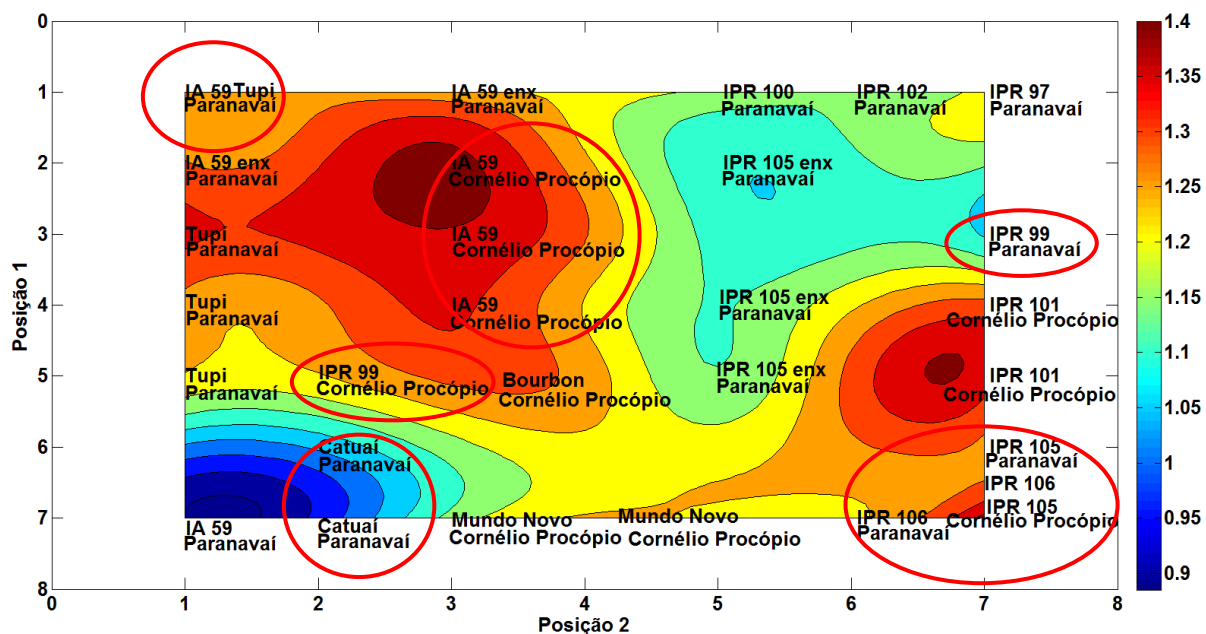


Figura 4 - Mapas de pesos para a variável cafeína e os mapas topológicos das cidades e genótipos de café arábica. Na escala de cores as unidades são g/100g.

A região de maior concentração de cafeína que se pode observar na Figura 4 se formou onde haviam alguns cafés de Cornélio Procópio (IA 59, IPR 105, IPR 106 e IPR 101) resultado que está de acordo com a Tabela 2. A região de menor concentração de cafeína se formou onde havia uma mescla de cafés de Cornélio Procópio e Paranaíba. Os genótipos presentes nesta região são Catuaí e IA 59. Os genótipos IA 59, IPR 106, IPR 99 e IPR 105 que foram cultivados nas duas regiões apresentaram diferenças de cafeína, as amostras de Paranaíba apresentaram menores concentrações de cafeína em relação as amostras de Cornélio Procópio.

Tabela 3 - Médias e desvios padrão das variáveis em relação aos genótipos de café.

Genótipo	Variável							
	Densidade	Açúcares Totais	Açúcares Redutores	Cafeína	Proteínas	Ácidos Clorogênicos	Lipídios	Taninos
IPR 100	0,667 <sup>cd</sup>	7,685 <sup>abcd</sup>	0,388 <sup>ab</sup>	1,187 <sup>ab</sup>	16,744 <sup>abcd</sup>	8,198 <sup>defg</sup>	13,643 <sup>abcd</sup>	4,954 <sup>cde</sup>
	± 0,000	± 0,092	± 0,001	± 0,001	± 0,005	± 0,101	± 0,053	± 0,125
IPR 102	0,654 <sup>d</sup>	7,538 <sup>cd</sup>	0,439 <sup>ab</sup>	1,153 <sup>ab</sup>	17,038 <sup>ab</sup>	8,034 <sup>defg</sup>	13,173 <sup>bcd</sup>	4,199 <sup>e</sup>
	± 0,006	± 0,012	± 0,005	± 0,006	± 0,088	± 0,155	± 0,031	± 0,182
IPR 105	0,667 <sup>cd</sup>	7,984 <sup>abc</sup>	0,280 <sup>ab</sup>	1,319 <sup>a</sup>	16,731 <sup>abcd</sup>	8,917 <sup>def</sup>	15,903 <sup>abc</sup>	4,689 <sup>de</sup>
	± 0,010	± 0,131	± 0,012	± 0,039	± 0,228	± 0,473	± 0,402	± 0,169
IPR 106	0,678 <sup>bcd</sup>	8,585 <sup>a</sup>	0,328 <sup>ab</sup>	1,363 <sup>a</sup>	16,435 <sup>abcd</sup>	8,929 <sup>def</sup>	15,139 <sup>abc</sup>	4,799 <sup>de</sup>
	± 0,005	± 0,100	± 0,036	± 0,146	± 0,247	± 0,159	± 0,340	± 0,186
IPR 97	0,658 <sup>d</sup>	7,311 <sup>cd</sup>	0,499 <sup>a</sup>	1,222 <sup>ab</sup>	17,340 <sup>a</sup>	9,608 <sup>bcd</sup>	11,542 <sup>d</sup>	5,738 <sup>bcd</sup>
	± 0,000	± 0,005	± 0,002	± 0,016	± 0,027	± 0,018	± 0,037	± 0,279
IPR 99	0,697 <sup>abc</sup>	7,091 <sup>cd</sup>	0,426 <sup>ab</sup>	1,169 <sup>ab</sup>	16,199 <sup>abcd</sup>	8,536 <sup>def</sup>	15,300 <sup>abc</sup>	6,114 <sup>bc</sup>
	± 0,020	± 0,707	± 0,025	± 0,089	± 0,300	± 0,054	± 0,205	± 0,632
Bourbon	0,704 <sup>ab</sup>	7,399 <sup>cd</sup>	0,316 <sup>ab</sup>	1,315 <sup>ab</sup>	15,794 <sup>bcd</sup>	7,784 <sup>fg</sup>	15,408 <sup>abc</sup>	6,501 <sup>ab</sup>
	± 0,000	± 0,056	± 0,016	± 0,009	± 0,193	± 0,066	± 0,056	± 0,309
Catuaí	0,702 <sup>ab</sup>	8,015 <sup>abc</sup>	0,326 <sup>ab</sup>	1,011 <sup>b</sup>	15,718 <sup>cd</sup>	8,151 <sup>defg</sup>	16,631 <sup>a</sup>	7,072 <sup>ab</sup>
	± 0,005	± 0,234	± 0,008	± 0,057	± 0,343	± 0,367	± 0,455	± 0,539
IA 59	0,704 <sup>a</sup>	7,984 <sup>abc</sup>	0,397 <sup>ab</sup>	1,238 <sup>ab</sup>	16,235 <sup>abcd</sup>	10,274 <sup>abc</sup>	14,482 <sup>abcd</sup>	6,759 <sup>ab</sup>
	± 0,009	± 0,216	± 0,102	± 0,220	± 0,435	± 0,846	± 1,474	± 0,352
Mundo Novo	0,671 <sup>cd</sup>	8,502 <sup>ab</sup>	0,382 <sup>ab</sup>	1,212 <sup>ab</sup>	15,467 <sup>d</sup>	7,203 <sup>g</sup>	15,682 <sup>abc</sup>	6,575 <sup>ab</sup>
	± 0,005	± 0,078	± 0,004	± 0,052	± 0,171	± 0,502	± 0,603	± 0,203
Tupi	0,708 <sup>a</sup>	7,901 <sup>abc</sup>	0,407 <sup>ab</sup>	1,310 <sup>ab</sup>	16,003 <sup>bcd</sup>	11,142 <sup>ab</sup>	15,509 <sup>abc</sup>	6,282 <sup>bc</sup>
	± 0,012	± 0,093	± 0,093	± 0,047	± 0,702	± 0,340	± 0,588	± 0,761
IPR 101	0,658 <sup>d</sup>	7,494 <sup>cd</sup>	0,255 <sup>b</sup>	1,347 <sup>a</sup>	16,343 <sup>abcd</sup>	9,637 <sup>bcd</sup>	15,425 <sup>abc</sup>	7,474 <sup>a</sup>
	± 0,000	± 0,258	± 0,014	± 0,042	± 0,261	± 0,586	± 2,032	± 0,312
IPR 105 enx	0,678 <sup>bcd</sup>	7,864 <sup>abc</sup>	0,336 <sup>ab</sup>	1,130 <sup>ab</sup>	16,328 <sup>abcd</sup>	8,645 <sup>def</sup>	13,507 <sup>bd</sup>	6,251 <sup>bc</sup>
	± 0,005	± 0,097	± 0,057	± 0,241	± 0,195	± 0,153	± 0,454	± 0,339
IA 59 enx	0,707 <sup>a</sup>	7,990 <sup>abc</sup>	0,411 <sup>ab</sup>	1,284 <sup>ab</sup>	17,006 <sup>ab</sup>	11,397 <sup>a</sup>	15,823 <sup>ab</sup>	6,244 <sup>bc</sup>
	± 0,010	± 0,410	± 0,094	± 0,026	± 0,210	± 0,193	± 0,571	± 0,252

\* Médias da mesma coluna com a mesma letra não diferem significativamente (p&gt;0,05).

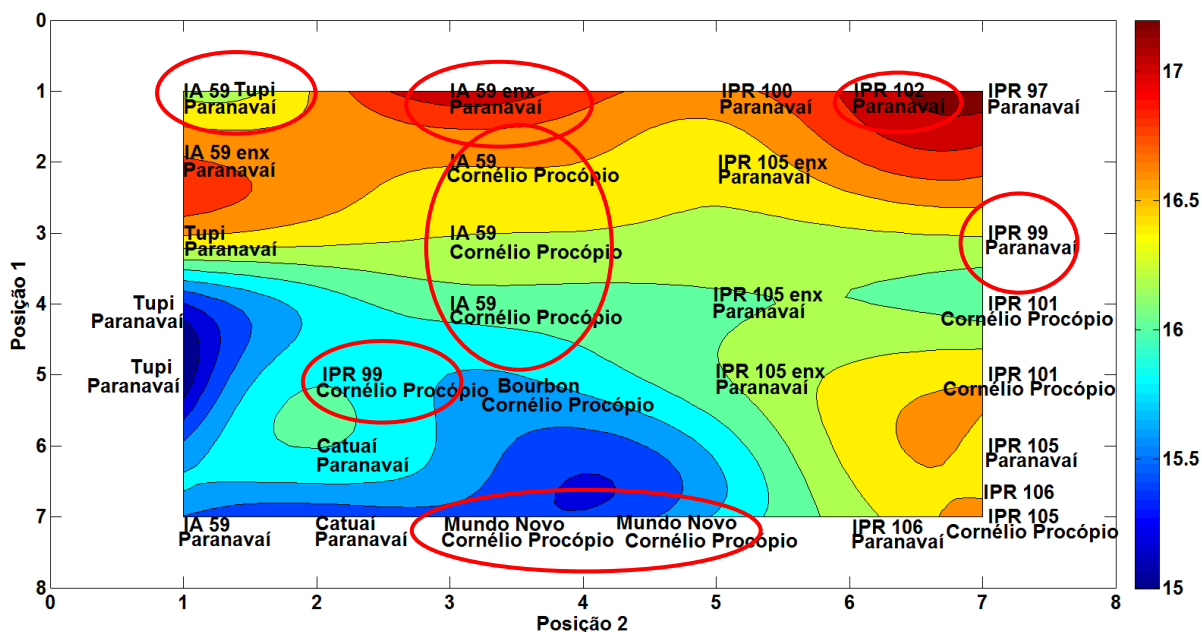


Figura 5 - Mapas de pesos para a variável proteína e os mapas topológicos das cidades e genótipos de café arábica. Na escala de cores as unidades são g/100g.

Observando a Figura 5, nota-se que a região do mapa de maior concentração de proteínas se formou onde haviam cafés de Paranavaí (IA 59 enx, IPR 97 e IPR102). As observações do mapa topológico e mapa de peso estão de acordo com os dados da Tabela 3. A região de menor concentração de proteínas se formou onde havia uma mescla de cafés de Cornélio Procópio e Paranavaí. Os genótipos presentes nesta região são Mundo Novo, Catuaí, IA 59 e Tupi. Os genótipos IA 59, IPR 106, IPR 99 e IPR 105 que foram cultivados em diferentes regiões apresentaram pouca diferença na concentração de proteínas, essa informação também está de acordo com a Tabela 2 e 3.

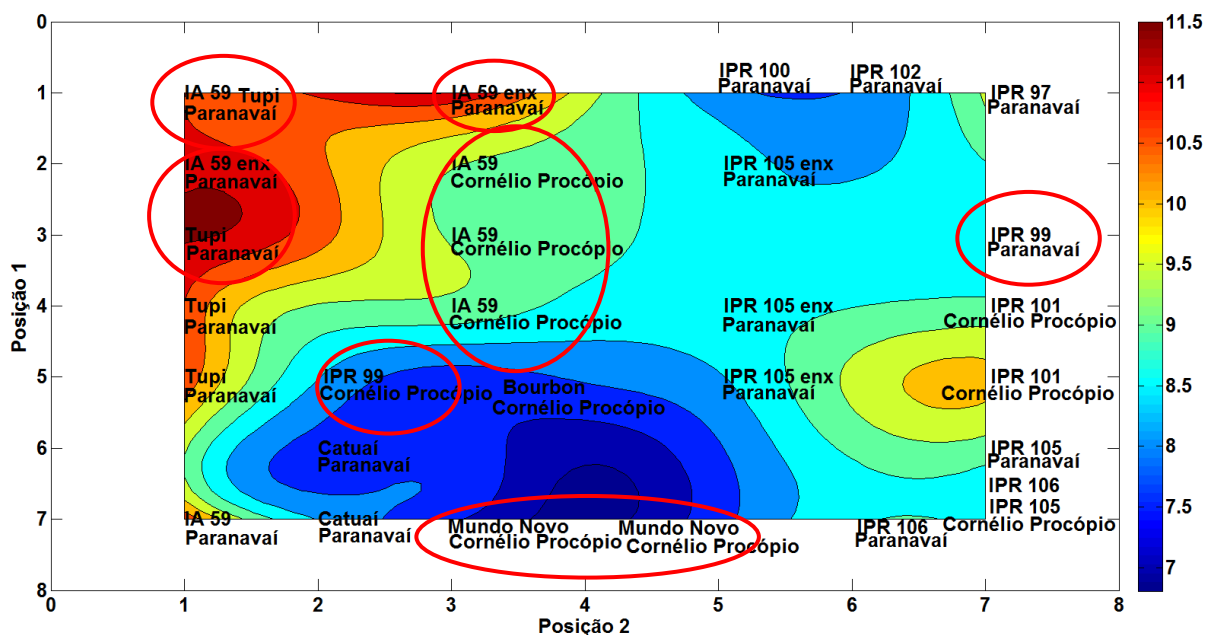


Figura 6 - Mapas de pesos para a variável ácidos clorogênicos e os mapas topológicos das cidades e genótipos de café arábica. Na escala de cores as unidades são g/100g.

A região de maior concentração de ácidos clorogênicos (Figura 6) se formou onde haviam cafés de Paranavaí ( IA 59 e IA 59 enx). A região de menor concentração de ácidos clorogênicos se formou onde havia uma mescla de cafés de Cornélio Procópio e Paranavaí. Os genótipos presentes nesta região são Catuaí, Mundo Novo, Bourbon, IPR 100, IPR 102 e IPR99. Os genótipos IA 59, IPR 106 e IPR 99 que foram cultivados nas duas regiões apresentaram diferenças na concentração de ácidos clorogênicos, as amostra de Paranavaí apresentaram maiores concentrações em relação às amostras de Cornélio Procópio. Todas as observações acima descritas estão de acordo com as Tabelas 2 e 3.

## 1.4 CONCLUSÃO

O SOM foi capaz de reconhecer os genótipos de café e sua origem geográfica utilizando os dados de perfil químico das amostras com destaque para os teores de cafeína,

ácidos clorogênicos, proteínas e açúcares redutores. Além disso, a possibilidade de gerar mapas de peso para cada variável permitiu compreender com mais detalhes como cada fator estudado interfere na segmentação observada. Portanto, a combinação desses métodos mostrou ser uma ferramenta eficiente para a segmentação de amostras de café arábica, podendo ser aplicada também em outras tarefas como a comparação entre produtos, identificação de adulterações, entre outros.

## 1.5 REFERÊNCIAS

1. Bishop, C. M., *Neural Networks for Pattern Recognition*. Oxford: Oxford University, 1995; pp.482.
2. Haykin, S., *Redes Neurais: Princípios e Prática*. 2ª edição ed.; Bookman: Porto Alegre, 2001; pp. 900.
3. Melssen, W.; Wehrens, R.; Buydens, L., Supervised Kohonen networks for classification problems. *Chemometrics and Intelligent Laboratory Systems*. **2006**, *83* (2), 99-113.
4. Borsato, D.; Pina, M. V. R.; Spacino, K. R.; Scholz, M. B. d. S.; Filho, A. A., Application of artificial neural networks in the geographical identification of coffee samples. *Eur Food Res Technol*. **2011**, *233*, 533-543.
5. Marini, F., Artificial neural networks in foodstuff analyses: Trends and perspectives A review. *Analytica Chimica Acta*. **2009**, *635*, 121–131.
6. Lucia, S. M. D.; Minim, L. A., *Redes Neurais Artificiais: Fundamentos e Aplicações*. 2006; pp. 173-194.
7. Kohonen, T., *Self-organizing maps*. 3. Ed. ed.; Berlin; Heidelberg; New York; Barcelona; Hong Kong; London; Milan; Paris; Singapore; Tokyo, 2001.
8. De Boishebert, V.; Giraudel, J.-L.; Montury, M., Characterization of strawberry varieties by SPME–GC–MS and Kohonen self-organizing map. *Chemometrics and Intelligent Laboratory Systems*. **2006**, *80* (1), 13-23.
9. Marini, F.; Magrì, A. L.; Bucci, R.; Magrì, A. D., Use of different artificial neural networks to resolve binary blends of monocultivar Italian olive oils. *Analytica Chimica Acta*. **2007**, *599*, 232-240.
10. Tigrine-Kordjani, N.; Chemat, F.; Meklati, B. Y.; Tuduri, L.; Giraudel, L.; Montury, M., Relative characterization of rosemary samples according to their geographical origins using microwave-accelerated distillation, solid-phase microextraction and Kohonen self-organizing maps. *Analytical and Bioanalytical Chemistry*. **2007**, *389*, 631-641.
11. Bona, E.; Silva, R. S. d.; Borsato, D.; Bassoli, D. G., Self-organizing maps as a chemometric tool for aromatic pattern recognition of soluble coffee. *Acta Scientiarum Technology*. **2012**, *34*, 111-119.
12. SINDICAFÉ. Dados Estatísticos. (acessado em 24 de Setembro de 2012).
13. ABIC. Indicadores da indústria de café no Brasil – 2012. (acessado em 26 de Agosto de 2012).
14. CECAFÉ. Dados estatísticos. (acessado em 19 de Outubro de 2012).
15. Rubayiza, A. B.; Meurens, M., Chemical discrimination of arabica and robusta coffees by Fourier transform Raman spectroscopy. *Journal of Agriculture and Food Chemistry*. 2005, *53*, p.4654-4659.
16. Farah, A., *Coffee as a speciality and functional beverage*. Cambridge: Woodhead Publishing Limited, 2009; pp. 370-395.
17. Higdon, J. V.; Frei, B., Coffee and Health: A review of recent human research. *Critical Reviews in Food Science and Nutrition*. **2006**, *46*, 101-123.
18. BRASIL. Regulamento técnico de identidade e de qualidade para a classificação do café beneficiado e de grão verde. Instrução Normativa nº 8 de 11 jun. 2003. (acessado em 09 de Novembro de 2011).
19. Buenaventura-Serrano, C. E.; Castaño-Castrillón, J. J., Influencia de la altitud en la calidad de la bebida de muestras de café procedente del ecotopo 206B en Colombia. *Cenicafé*. **2002**, *53*, 119-131.
20. Instituto Adolf Lutz, *Métodos físico-químicos para análise de alimentos*. Edição IV, 1ª edição digital ed.; SÃO PAULO, 2008.



21. Clifford, M. N.; Wight, J. C., Chorogenic acids and other cinnamates nature, occurrence and dietary burden. *Journal of the Science of Food and Agriculture*. **1976**, *27*, 73–84.
22. AOAC, *Official methods of analysis*. Washington, 1990.
23. Southgate, D., *Determination of food carbohydrates*. Applied Science Publishers. 1976; Londres.
24. Marini, F.; Zupan, J.; Magri, A. L., Class-modeling using Kohonen artificial neural networks. *Analytica Chimica Acta*. **2005**, *544*, 306-314.

## **CAPÍTULO 2:**

Perceptron de múltiplas camadas otimizado para a classificação geográfica e genotípica de quatro genótipos de café arábica

## 2 Perceptron de múltiplas camadas otimizado para a classificação geográfica e genotípica de quatro genótipos de café arábica

*Resumo.* As condições climáticas específicas de cada região de cultivo do café fornecem atributos especiais para a bebida e podem assim aumentar seu valor agregado. No entanto, é essencial provar a origem geográfica e genotípica do cultivar por meio de métodos confiáveis. Vários métodos estatísticos foram desenvolvidos na tentativa de reproduzir a capacidade humana de reconhecimento de padrões. Os perceptrons de múltiplas camadas (MLP) são um tipo de rede neural artificial (RNA) com aprendizagem supervisionada que são muito utilizados para classificação de padrões. Este estudo objetivou a utilização de uma RNA para a classificação geográfica e genotípica de café arábica. Para esse fim os espectros obtidos no equipamento de espectroscopia no infravermelho com transformada de Fourier (FTIR) foram analisados através do emprego de MLPs otimizados através do simplex sequencial. As redes que utilizaram a faixa de  $1900-800\text{ cm}^{-1}$  do espectro FTIR obtiveram menor erro quadrado médio e maior porcentagem de classificação correta para a segmentação das amostras de café arábica tanto por região de cultivo (100%) quanto por genótipo (77,78%). Os MLPs otimizados foram capazes de classificar geograficamente as amostras de café arábica. Entretanto, para a classificação genotípica o desempenho não foi totalmente satisfatório. Além disso, o MLP desenvolvido para classificação genotípica apresenta um número muito elevado de pesos sinápticos, não proporcionando número suficiente de graus de liberdade para que a rede neural possa aprender de maneira confiável.

**Palavras-chave:** *café verde, espectro infravermelho, MLP, redes neurais artificiais*

## 2.1. Introdução

O reconhecimento de padrões é um processo pelo qual um sinal recebido é atribuído a uma classe dentre um número predeterminado de categorias. Os seres humanos possuem excelência nata para a aprendizagem e reconhecimento de padrões. Nesse sentido, os métodos matemáticos têm sido desenvolvidos na tentativa de mimetizar, mesmo de maneira rudimentar, tal aptidão humana (BISHOP, 1995; HAYKIN, 2001). As redes neurais artificiais (RNAs) são um conjunto de técnicas baseadas em princípios estatísticos, que vem crescentemente ganhando espaço para realizar tarefas de regressão e reconhecimento de padrões. As RNAs são extremamente versáteis para realizar o mapeamento de relações complexas e não-lineares entre múltiplas variáveis de entrada e saída (BISHOP, 1995). Originalmente as RNAs foram projetadas para ser um esquema, tão preciso quanto possível, do modelo da atividade do cérebro humano. O córtex cerebral é capaz de armazenar padrões de comportamento, mesmo em presença de dados ruidosos, tornando-o mais poderoso do que qualquer computador existente (MARINI, 2009). Para alcançarem bom desempenho, as redes neurais empregam uma interligação maciça de células computacionais simples denominadas neurônios. Ela se assemelha ao cérebro em dois aspectos principais; o primeiro é que o conhecimento é adquirido pela rede a partir de seu ambiente através de um processo de aprendizagem. O segundo aspecto que assemelha as RNAs com o cérebro são as forças de conexão entre os neurônios, conhecidas como pesos sinápticos, as quais armazenam o conhecimento e servem para ponderar a entrada recebida de cada neurônio. Tal estrutura conduz a uma capacidade de generalização, ou seja, uma RNA pode produzir saídas adequadas para entradas que não estavam presentes durante o treinamento. Esses atributos tornam possível a aplicação das RNAs em problemas complexos. Os perceptrons de múltiplas camadas (MLP) são um tipo de rede neural artificial (RNA) com aprendizagem

supervisionada que são muito utilizados para classificação de padrões (HAYKIN, 2001; LUCIA e MINIM, 2006).

A importância do café na economia mundial é indiscutível, é um dos mais valiosos produtos primários comercializados no mundo. Existem cerca de 72 países produtores do grão no mundo, essa multiplicidade de produtores, deve-se à extensa faixa apta à produção do cafeeiro, graças à sua versatilidade. Os principais países produtores concentram-se na América do Sul (Brasil e Colômbia), América Central e Ásia (Vietnã) (SINDICAFÉ, 2011). O consumo brasileiro de café aumenta a cada ano, no período entre maio de 2011 e abril de 2012 foi registrado o consumo de 19,975 milhões de sacas, representando um acréscimo de 3,05% em relação ao período anterior correspondente (ABIC, 2012). Além do mercado interno, as exportações de café verde do Brasil somaram 116,63 mil toneladas em março de 2012, baixa de 19,2% em relação ao mesmo mês de 2011 (CECAFÉ, 2012).

O café tem procedência de uma árvore do gênero *Coffea* e pertence à família das rubiáceas englobando mais de 500 gêneros e 6.000 espécies. Dentro do gênero *Coffea* existem duas espécies principais, o *Coffea arabica* variedade Arábica e o *Coffea canephora* variedade Robusta (RUBAYIZA e MEURENS, 2005). As espécies *Coffea arabica* e *Coffea canephora* (robusta) apresentam uma composição química muito distinta e o café arábica fornece um bebida com qualidade e aroma superior ao café robusta (FARAH, 2009). A mistura de grãos de alto valor com outros de menor valor comercial e adição de substitutos são formas de adulteração em cafés (BRIANDET *et al.*, 1996). Além disso, o Brasil apresenta diversidade de clima e solo, assim os mais variados tipos de cafés podem ser produzidos. Essas condições climáticas conferem atributos especiais para a bebida do café (corpo, acidez e sabor). Quando a qualidade do produto pode ser atribuída ao seu local de produção, ocorre um aumento do valor agregado do café. No entanto, neste processo, devem haver mecanismos confiáveis que

comprovem a origem geográfica e genotípica do café (BORSATO *et al.*, 2011). Assim, torna-se imperativo a utilização de técnicas analíticas rápidas e de baixo custo que permitam a diferenciação e certificação da composição e qualidade dos cafés comercializados. Uma metodologia que reúne tais atributos é a espectroscopia com transformada de Fourier na região do infravermelho médio (FTIR). O FTIR permite a detecção dos componentes de maior importância na caracterização de alimentos, sendo uma metodologia que origina rapidamente uma informação sumária sobre o produto. Outra vantagem reside no fato do preparo da amostra, geralmente, não requerer a extração com solventes diminuindo assim o tempo de análise e o impacto ambiental (KAROUI *et al.*, 2010). Ressalta-se ainda que o FTIR já foi utilizado com sucesso para autenticar misturas de cafés arábica e robusta ou adição de adulterantes (KEMSLEY *et al.*, 1995; BRIANDET *et al.*, 1996; DOWNEY *et al.*, 1997; WANG *et al.*, 2009; WANG *et al.*, 2011).

Este estudo objetivou a utilização de uma rede neural artificial para a classificação geográfica e genotípica de café arábica. Para esse fim os espectros obtidos no FTIR foram analisados através do emprego de perceptrons de múltiplas camadas otimizados através do simplex sequencial.

## **2.2. Materiais e Métodos**

Todos os pré-processamentos dos espectros foram realizados no software IRsolution 1.5 (Shimadzu Corporation, Kyoto, Japão) que acompanha o equipamento. As análises matemáticas e/ou estatísticas, assim como a ACP, as redes neurais artificiais e o simplex sequencial foram realizadas no software MATLAB R2007b (The MathWorks Inc., Natick, USA).

### 2.2.1. Amostras de café arábica

Foram utilizados 4 genótipos de café arábica catalogados pelo Instituto Agronômico do Paraná - Londrina (IAPAR) da safra de 2010 (Tabela 1).

Tabela 1 - Genótipos de café arábica utilizados.

<b>Genótipo</b>	<b>Local</b>	<b>Total Amostras</b>
IPR 99	Cornélio Procópio	1
	Paranavaí	1
	Mandaguari	1
	Londrina	1
IPR 105	Cornélio Procópio	1
	Paranavaí	1
	Mandaguari	1
	Londrina	1
IPR 106	Cornélio Procópio	1
	Paranavaí	1
	Mandaguari	1
	Londrina	1
IA 59	Cornélio Procópio	2
	Paranavaí	2
	Mandaguari	1
	Londrina	1

No IAPAR, as amostras foram imediatamente colocadas em caixas de madeira com uma malha de fundo e movidas oito vezes por dia até que a umidade dos grãos chegasse a 11-12% e logo após as amostras foram beneficiadas (remoção da casca e pergaminho) (BRASIL, 2011). Os grãos verdes, moídos, peneirados e embalados dos genótipos de café arábica fornecidos pelo IAPAR, foram recebidos e armazenados em um freezer a  $-18^{\circ}\text{C}$  e utilizados posteriormente para análise. Foram utilizados cafés cultivados em 4 diferentes regiões (Cornélio Procópio, Paranavaí, Mandaguari e Londrina) totalizando 18 amostras. A Figura 1 apresenta a localização das cidades aonde os cafés foram cultivados e a distância máxima entre as mesmas.



Figura 1 - Mapa da distância entre as cidades dos cafés estudados.

### 2.2.2. Espectroscopia de infravermelho com transformada de Fourier (FTIR)

Para preparar as pastilhas foram adicionados em torno de 100 mg de KBr seco (padrão cromatográfico) e aproximadamente 1 mg de amostra finamente moída. A mistura foi, então, prensada em uma prensa hidráulica (Bovenau, P15 ST) usando um molde (ICL, ICL's Macro/Micro KBr die) empregando aproximadamente  $360 \text{ kgf/cm}^2$ . Produziu-se, assim, uma pastilha transparente. Antes da análise de cada amostra, o FTIR (Shimadzu, FTIR - 8300) foi programado para realizar um espectro de *background* do ar, sendo o mesmo utilizado para descontar a influência dos componentes do ar no espectro. Na sequência, a pastilha foi posicionada no feixe do instrumento e os espectros foram obtidos na faixa de  $3750$  a  $750 \text{ cm}^{-1}$ . Foram usadas 32 varreduras acumuladas para formar o espectro final e realizadas 5 repetições (pastilhas) para cada amostra, totalizando assim 90 espectros. Foram utilizados 72 desses espectros (80%) como amostras de treinamento das redes neurais artificiais empregadas e 18 espectros (20%) como amostras de teste (uma repetição, escolhida aleatoriamente, de cada um dos cafés estudados).



### 2.2.3. Transformação dos espectros

Após obtenção dos espectros foram aplicadas algumas transformações nos mesmos. Primeiramente, foi realizada a normalização do espectro (a maior banda obteve absorbância 1 e a menor 0), correção da linha de base, a suavização do espectro (17 *smoothing points*) e uma remoção da banda referente a concentração de CO<sub>2</sub>, desprezando assim as variações de CO<sub>2</sub> entre as amostras. Para a ACP e a RNA foi considerado o uso da região entre 3750 e 750 cm<sup>-1</sup> do espectro desconsiderando ruídos que se apresentavam além desta região. Também foi avaliada a região entre 1900 e 800 cm<sup>-1</sup> do espectro pois a mesma contém as bandas de absorção devidas à deformação axial simétrica da carbonila (ésteres, aldeídos e cetonas), deformação angular simétrica do metileno, deformações axiais e angulares simétricas do C-O (ésteres e alcoóis). Portanto, essa região contém a informação de *fingerprint* para discriminação de diferentes amostras de cafés (BRIANDET *et al.*, 1996; LYMAN *et al.*, 2003; WANG *et al.*, 2009; WANG *et al.*, 2011).

Após estes pré-processamentos foi utilizada a análise de componentes principais (ACP) para a redução da dimensionalidade dos dados (SÁ, 2007). A ACP foi realizada nos dados normalizados assim como na 1ª e 2ª derivada dos espectros. A aplicação da 1ª e da 2ª derivada sobre os dados espectrais brutos é um procedimento que pode destacar ombros espectrais, bem como minimizar o efeito de inclinações provocadas na linha de base dos espectros. Ao aplicarmos as operações de derivação as informações contidas ao longo dos diferentes números de onda são geralmente acentuadas (os ruídos também tornam-se acentuados), portanto deve-se ter cuidado com a qualidade dos espectros com os quais se deseja aplicar o cálculo das derivadas (COSTA FILHO e POPPI, 2002). A quantidade de componentes principais empregadas foi um dos parâmetros otimizados através do simplex sequencial.

#### 2.2.4. Normalização

Os vetores de entrada (as componentes principais escolhidas) foram pré-processados antes de serem alimentados na rede neural para que as funções de ativação dos neurônios artificiais não fossem facilmente saturadas ou ocorresse um erro de *overflow* (quando um número é grande demais para a estrutura de dados reservada para ele) (HAYKIN, 2001). Os métodos de pré-processamento utilizados foram: máximo e mínimo (minimax), transformação para uma escala entre -1 e 1 e autoescalamento, vetor de entrada com média zero e variância unitária (PÉREZ-MAGARIÑO *et al.*, 2004).

#### 2.2.5. Perceptron de múltiplas camadas (MLP)

Foi empregada uma rede neural artificial do tipo Perceptron de Múltiplas Camadas (MLP - *Multi Layer Perceptron*) que é muito utilizada para classificação de padrões (BONA *et al.*, 2011; BORSATO *et al.*, 2011; GALÃO *et al.*, 2011). A arquitetura desta rede é constituída por uma camada de entrada com  $n$  neurônios (um para cada componente principal utilizada), uma camada oculta (responsável pela separação dos padrões através da formação de fronteiras de decisão) contendo uma quantidade de neurônios a ser definida e uma camada de saída com 4 neurônios, um para cada região ou genótipo de café, que constrói combinações das fronteiras de decisão formadas pelos neurônios ocultos. Assim, o vetor de resposta desejada apresenta dimensão igual a 4, para uma amostra pertencente a classe  $k$  o  $k$ -ésimo valor é igual a 1 e todos os outros são zerados. A Figura 2 apresenta um esquema representativo de um perceptron de múltiplas camadas (HAYKIN, 2001).

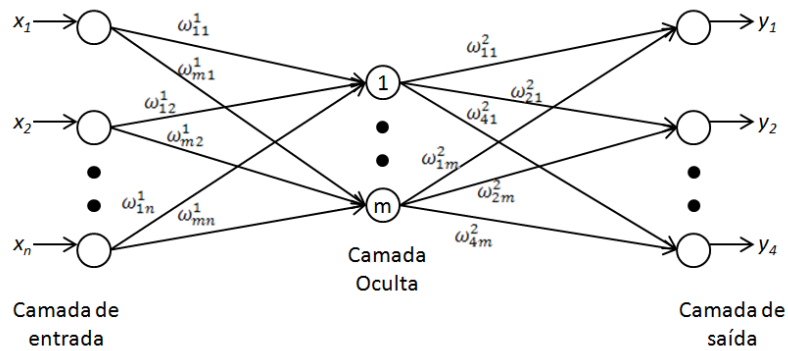


Figura 2 - Representação de um perceptron de múltiplas camadas (MLP).

Para cada conexão entre os neurônios (Figura 2) existe um peso  $w_{ij}^l$  associado, sendo  $i$  o índice do neurônio de entrada do sinal,  $j$  o neurônio de saída do sinal e  $l$  a camada onde está localizado o neurônio de entrada. Em cada um dos  $N^l$  neurônios das  $L$  camadas da rede MLP é realizado um somatório ponderado pelos pesos sinápticos dos sinais provenientes dos neurônios da camada anterior. Esta soma, chamada de campo local induzido (1), é aplicada a uma função de ativação não linear (2) que irá produzir a saída do neurônio (HAYKIN, 2001).

$$v_i^l = \sum_{j=0}^{m_i} w_{ij}^l y_j^{l-1} \quad i = 1, 2, \dots, N^l; l = 1, 2, \dots, L. \quad (1)$$

$$y_i^l = \varphi(v_i^l) \quad i = 1, 2, \dots, N^l; l = 1, 2, \dots, L. \quad (2)$$

Nas equações (1) e (2)  $v_i^l$  é o campo local induzido do neurônio  $i$  na camada  $l$ ,  $w_{ij}^l$  é o peso sináptico associado ao sinal de saída  $y_j^{l-1}$  do neurônio  $j$  da camada anterior ( $l - 1$ ),  $m_i$  é o número de entradas associadas ao neurônio  $i$ . Para  $j = 0$ , temos  $y_0^{l-1} = 1$  e  $w_{i0}^l$  é chamado de termo de polarização, ou *bias*, que tem o efeito de aumentar ou diminuir a entrada líquida da função de ativação, dependendo se ele é positivo ou negativo. Para a rede MLP desenvolvida foram testadas para a camada oculta de neurônios artificiais as funções do tipo logística e tangente hiperbólica sigmoide que são definidas, respectivamente, por,

$$\varphi = \frac{1}{1 + \exp(-av)} \quad (3)$$

$$\varphi = \frac{2}{1 + \exp(-2v)} - 1 \quad (4)$$

Para a função logística (3) o símbolo  $\alpha$  representa a inclinação e pode ser considerado como um parâmetro a ser otimizado pelo projetista da rede, embora em muitos casos é utilizado um valor unitário para este parâmetro (HUANG *et al.*, 2007). Para a camada de saída além destas também a função de ativação linear foi testada (HAYKIN, 2001).

Uma rede neural extrai seu poder computacional através de sua habilidade de aprender e, portanto de generalizar. A generalização se refere ao fato de a rede neural produzir saídas adequadas para entradas que não estavam presentes durante o treinamento (aprendizagem). O processo utilizado para realizar o processo de aprendizagem é chamado de algoritmo de aprendizagem, cuja função é modificar os pesos sinápticos da rede de uma forma ordenada para alcançar um objetivo de projeto desejado (HAYKIN, 2001; LUCIA e MINIM, 2006). Para a rede neural artificial construída foi utilizado o algoritmo de Levenberg-Marquardt. De maneira geral, o algoritmo propõe uma solução de compromisso entre o algoritmo do gradiente descendente (retropropagação) e o método de Gauss-Newton, que utiliza uma expansão da série de Taylor para aproximar o modelo de regressão não linear com termos lineares e, então, aplica mínimos quadrados para estimar os parâmetros. No algoritmo de Levenberg-Marquardt existe um parâmetro que regula o tamanho do passo das correções de peso. Para valores muito pequenos desse parâmetro a regra da atualização de pesos leva em consideração o método Gauss-Newton, mas para valores grandes a regra da atualização de pesos leva em consideração o gradiente descendente (BISHOP, 1995).

O processo de aprendizagem é mantido até os pesos sinápticos e os níveis de *bias* se estabilizarem e o erro quadrado médio (equação 5) convergir para um valor mínimo (HAYKIN, 2001).

$$EQM = \frac{1}{2N} \sum_{n=1}^N \sum_{i=1}^C e_i^2(n) \quad (5)$$

Na equação (5), N é igual ao tamanho do conjunto de treinamento e C igual ao número de neurônios na camada de saída. Tendo como objetivo uma boa generalização, foi utilizado o método de treinamento com parada antecipada para evitar que a rede acabe sendo excessivamente ajustada aos dados de treinamento. É possível identificar o início do excesso de treinamento através do uso de amostras de validação (11 amostras ou 15% das amostras de treinamento), onde além dos dados de treinamento é usado um segundo conjunto de dados que após cada época (uma apresentação completa do conjunto de treinamento) é submetido à rede para verificar a capacidade de generalização desta com dados não utilizados para o ajuste de pesos. Normalmente, o modelo não apresenta resultados tão bons sobre o conjunto de validação quanto sobre o de treinamento, em relação ao qual foi baseado. A curva de aprendizagem, ou curva do erro quadrado médio, decresce monotonamente para um número crescente de épocas. Diferentemente, para as amostras de validação, a curva do erro quadrado médio decresce para um mínimo e então começa a crescer conforme o treinamento continua, ou seja, o que a rede aprende após este ponto é essencialmente o ruído contido nos dados de treinamento. Esta heurística sugere que o ponto mínimo na curva de aprendizagem de validação seja usado como critério de parada da sessão de treinamento da rede (BISHOP, 1995; HAYKIN, 2001).

### 2.2.6. Média de ensemble

Um modo de melhorar a capacidade de generalização de uma rede neural é a combinação das respostas de várias RNAs, também chamada de máquinas de comitê, as quais foram treinadas de maneira semelhante de modo que o resultado final é a fusão das respostas

de várias redes e é superior àquele alcançável apenas por uma única RNA. Um dos métodos de implementar o comitê de máquinas é a média de ensemble, que consiste em combinar linearmente as respostas de um certo número de redes treinadas através de uma média. O uso de um ensemble diminui o erro final devido a *bias*, pois há um ajuste excessivo proposital de cada rede e também devido a variância pelo uso de diferentes condições iniciais de treinamento de cada rede (HAYKIN, 2001). Neste trabalho foi utilizada uma média de *ensemble* com 50 redes, sendo que para cada rede um novo conjunto de pesos iniciais e de exemplos de treinamento era gerado aleatoriamente.

### **2.2.7. Otimização dos parâmetros da rede**

Antes da escolha da melhor rede neural para o problema proposto, alguns parâmetros foram otimizados através do simplex sequencial (Tabela 2). A otimização foi realizada visando minimizar o erro quadrado médio para o conjunto de amostras utilizadas para teste (BONA *et al.*, 2011). O princípio do simplex básico foi deslocar uma figura regular, por exemplo um triângulo equilátero sobre uma superfície, quando duas variáveis estavam sendo consideradas (SPLENDLEY *et al.*, 1962). Os parâmetros escolhidos para a otimização estão representados na Tabela 2. O algoritmo utilizado nesse trabalho está descrito em maiores detalhes em Gao e Han (2012) e no APÊNDICE B. A otimização segue até que o valor do erro quadrado médio varie apenas dentro da tolerância estabelecida que foi de 0,001 ou pela avaliação gráfica que tem por finalidade auxiliar na visualização da otimização, que está representada como uma suavização na variação das respostas e variáveis independentes.

Tabela 2 - Parâmetros otimizados através do simplex sequencial.

<b>Rede</b>	<b>Parâmetros</b>	<b>Variação</b>
<b>Perceptron de Múltiplas Camadas</b>	Quantidade de neurônios na camada oculta	5 a 15
	Função de pré-processamento das entradas	Minimax ou Autoescalonamento
	Função de ativação utilizada na camada oculta	Logística ou Tangente hiperbólica sigmoide
	Função de ativação utilizada na camada de saída	Logística, Tangente hiperbólica sigmoide e Linear
	Quantidade de variáveis independentes	5 a 20

### 2.3. Resultados e discussões

Um espectro FTIR típico obtido e tratado para o genótipo IPR 105 cultivado em Cornélio Procópio está representado na Figura 3. Através da análise do espectro destaca-se a banda em  $1744\text{ cm}^{-1}$  resultante da deformação axial simétrica do C=O de ésteres de ácidos graxos (WANG *et al.*, 2011). As bandas na região de  $1654\text{ cm}^{-1}$  e  $1603\text{ cm}^{-1}$  estão relacionadas com a carbonila da cafeína (WANG *et al.*, 2011). A região em torno de  $1285\text{ cm}^{-1}$  está relacionada com os ácidos clorogênicos (BRIANDET *et al.*, 1996). As regiões de  $1200\text{ cm}^{-1}$  a  $900\text{ cm}^{-1}$  e de  $1500\text{ cm}^{-1}$  a  $1400\text{ cm}^{-1}$  estão relacionadas ao conteúdo de carboidratos (BRIANDET *et al.*, 1996; RIBEIRO *et al.*, 2010).

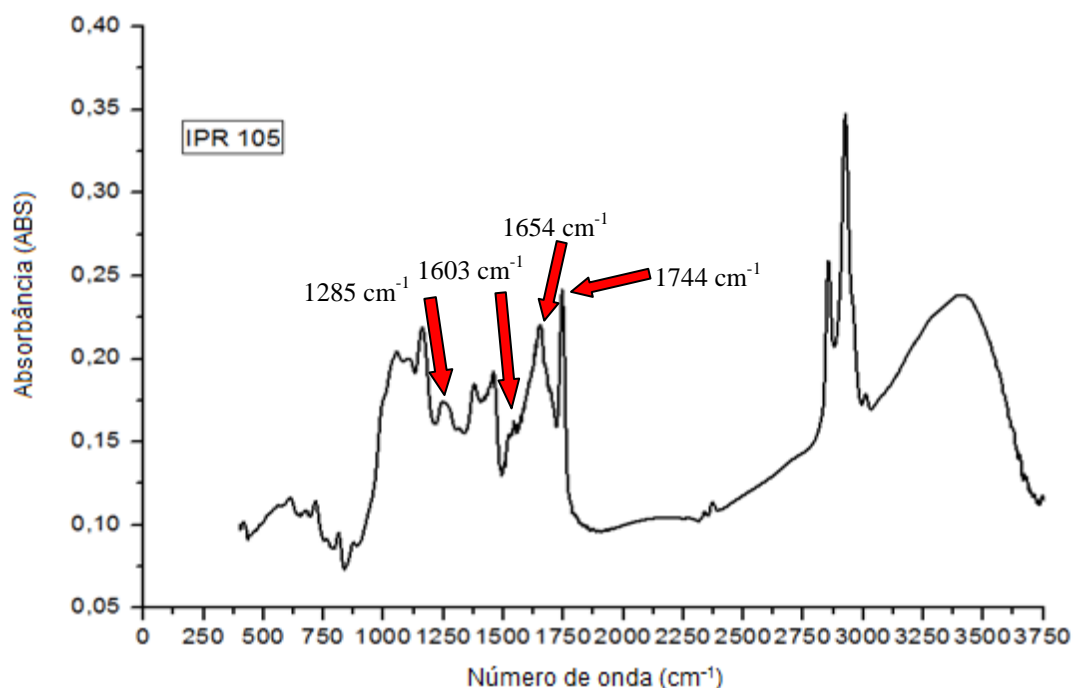


Figura 3 – Espectro obtido para o genótipo IPR 105 cultivado em Cornélio Procópio.

### 2.3.1. Perceptron de múltiplas camadas (MLP)

Todas as redes utilizadas possuem uma camada oculta. Após a otimização dos parâmetros selecionados (Tabela 2), foram criadas mil redes para cada faixa de trabalho no FTIR e tipo de tratamento do espectro (Tabelas 3 e 4). A melhor rede de cada tipo foi selecionada pelo menor erro quadrado médio e maior porcentagem de classificação para as amostras de teste. Foi utilizado também a média de *ensemble* com 50 redes (HAYKIN, 2001). Os resultados obtidos para os perceptron de múltiplas camadas propostos para a classificação geográfica e genotípica de café arábica estão representados nas Tabelas 3 e 4. É importante salientar que as condições de cada problema influenciam na escolha dos parâmetros ótimos obtidos para cada rede (SUKTHOMYA e TANNOCK, 2005).



Tabela 3 - Resultados obtidos para os perceptron de múltiplas camadas propostos para a classificação geográfica de café arábica.

	Número de onda					
	3750-750 cm <sup>-1</sup>			1900-800 cm <sup>-1</sup>		
	Rede 1	Rede 2	Rede 3	Rede 4	Rede 5	Rede 6
Espectros	Original	1 <sup>a</sup> Derivada	2 <sup>a</sup> Derivada	Original	1 <sup>a</sup> Derivada	2 <sup>a</sup> Derivada
Neurônios (camada oculta)	4	8	10	2	2	10
Função de pré-processamento (entradas) <sup>a</sup>	MM	MM	AE	MM	MM	MM
Função de Ativação (Camada oculta) <sup>b</sup>	FL	THS	THS	THS	FL	THS
Função de Ativação (Saída) <sup>b</sup>	THS	THS	L	THS	THS	THS
Componentes utilizadas	20	17	20	20	20	18
<b>Desempenho das melhores redes obtidas para os dados de treinamento</b>						
Erro quadrado médio (Ensemble)	0,0014	0,0016	0,0330	0,0015	0,0031	0,0161
Erro quadrado médio (Melhor Rede)	0,0002	0,0066	0,0463	0,0377	0,0000	0,0196
%Classificação correta (Ensemble)	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%
%Classificação correta (Melhor Rede)	100,00%	98,61%	94,44%	95,83%	100,00%	93,06%
<b>Desempenho das melhores redes obtidas para os dados de teste</b>						
Erro quadrado médio (Ensemble)	0,0250	0,0209	0,0792	0,0064	0,0240	0,0650
Erro quadrado médio (Melhor Rede)	0,2147	0,0406	0,1042	0,0353	0,0000	0,1337
%Classificação correta (Ensemble)	94,44%	100,00%	88,89%	100,00%	94,44%	88,89%
%Classificação correta (Melhor Rede)	55,56%	88,89%	72,22%	100,00%	83,33%	93,06%

<sup>a</sup> Função de pré-processamento: MM (Minimax) e AE (Autoescalonamento).

<sup>b</sup> Função de ativação: FL (Função Logística), THS (Tangente Hiperbólica Sigmoidal) e L (Linear)

A partir de uma análise da Tabela 3, é possível concluir que, o perceptron de múltiplas camadas que obteve menor erro quadrado médio e maior porcentagem de classificação correta para a segmentação das amostras de café arábica por região de cultivo foi a Rede 4, que foi treinada com os espectros originais (sem derivá-los) e com o número de onda entre 1900-800 cm<sup>-1</sup>, uma faixa espectral mais específica que contem as impressões bioquímicas (*fingerprints*) para discriminação de diferentes amostras de cafés (BRIANDET *et al.*, 1996; LYMAN *et al.*, 2003; WANG *et al.*, 2009).

Tabela 4 - Resultados obtidos os perceptron de múltiplas camadas propostos para a classificação genotípica de café arábica.

	Número de onda					
	3750-750 cm <sup>-1</sup>			1900-800 cm <sup>-1</sup>		
	Rede 7	Rede 8	Rede 9	Rede 10	Rede 11	Rede 12
Espectros	Original	1 <sup>a</sup> Derivada	2 <sup>a</sup> Derivada	Original	1 <sup>a</sup> Derivada	2 <sup>a</sup> Derivada
Neurônios (camada oculta)	9	6	5	3	6	8
Função de pré-processamento (entradas) <sup>a</sup>	AE	MM	MM	MM	MM	AE
Função de Ativação (Camada oculta) <sup>b</sup>	THS	FL	THS	FL	THS	THS
Função de Ativação (Saída) <sup>b</sup>	L	L	L	L	THS	THS
Componentes utilizadas	20	19	20	19	16	13
<b>Desempenho das melhores redes obtidas para os dados de treinamento</b>						
Erro quadrado médio (Ensemble)	0,0563	0,0305	0,0698	0,0280	0,0251	0,0235
Erro quadrado médio (Melhor Rede)	0,0811	0,0667	0,0171	0,0944	0,0213	0,0229
%Classificação correta (Ensemble)	100,00%	98,61%	97,22%	98,61%	100,00%	100,00%
%Classificação correta (Melhor Rede)	86,11%	91,67%	95,83%	79,17%	95,83%	94,44%
<b>Desempenho das melhores redes obtidas para os dados de teste</b>						
Erro quadrado médio (Ensemble)	0,1381	0,0961	0,1365	0,0862	0,0835	0,1316
Erro quadrado médio (Melhor Rede)	0,1766	0,1557	0,9482	0,1092	0,0986	0,2476
%Classificação correta (Ensemble)	61,11%	77,78%	55,56%	72,22%	83,33%	61,11%
%Classificação correta (Melhor Rede)	55,56%	72,22%	50,00%	72,22%	77,78%	50,00%

<sup>a</sup> Função de pré-processamento: MM (Minimax) e AE (Autoescalamento).

<sup>b</sup> Função de ativação: FL (Função Logística), TH (Tangente Hiperbólica Sigmoidal) e L (Linear).

Para a segmentação genotípica (Tabela 4) o MLP que obteve menor erro quadrado médio e maior porcentagem de classificação correta (Rede 11) foi o que utilizou a primeira derivada do espectro e, assim como para a classificação geográfica, com a faixa espectral entre 1900-800 cm<sup>-1</sup>. Os espectros variam muito pouco por se tratarem de uma mesma variedade de café, por isso, a primeira derivada do espectro obteve melhores resultados, pois ela acentua as diferenças espectrais removendo ruídos em uma mesma região espectral e/ou definindo bandas que se encontram sobrepostas (COSTA FILHO e POPPI, 2002). Através da análise da performance dos MLPs nas Tabelas 3 e 4 fica evidente que a classificação

genotípica é um problema mais difícil de ser modelado. Os MLPs desenvolvidos apresentam melhor desempenho quando comparados ao SIMCA (*Soft Independent Modelling of Class Analogies*) que obteve 83,33% de acerto para a classificação geográfica e 55,56% para a classificação genotípica (SATO *et al.*, 2012). Devido ao SIMCA, ser um método linear de classificação que se baseia na classificação de uma nova amostra a partir de sua projeção nas componentes principais de cada classe (BRANDEN e HUBERT, 2005), era esperado um desempenho superior dos MLPs desenvolvidos, já que o MLP realiza o mapeamento de relações não lineares entre múltiplas variáveis de entrada e saída.

Destaca-se também que para a classificação das amostras por região as redes com função de ativação da camada oculta do tipo tangente hiperbólica sigmoide obtiveram um erro quadrado médio menor e uma porcentagem de classificação superior aquelas com função de ativação do tipo função logística, provando que para a segmentação geográfica a tangente hiperbólica sigmoide é a função de ativação adequada. Para a classificação dos genótipos a combinação da função de ativação do tipo tangente hiperbólica sigmoide tanto na camada oculta quanto na camada de saída apresentou melhores resultados, sendo esta combinação adequada para este outro problema.

Com relação à normalização, ficou evidente que o método minimax foi superior ao autoescalamento, devido as redes que apresentaram melhores resultados (Rede 4 e Rede 11) apresentarem o mesmo método de normalização.

O número de variáveis independentes utilizadas na camada de entrada de neurônios também foi otimizado através do simplex sequencial. A média das componentes principais utilizadas foi de 19 para a classificação geográfica e 18 para a classificação genotípica. Tais valores encontram-se próximos do que era esperado, pois qualquer componente principal deve explicar a variância de pelo menos uma variável se a mesma há de ser mantida para

interpretação. Cada variável contribui com um valor 1 do autovalor total. Logo, apenas os fatores que apresentam autovalor maior que 1 são considerados significantes (HAIR *et al.*, 2005). Para este estudo obteve-se 19 componentes principais que apresentaram autovalores maiores que 1, comprovando assim, a capacidade que a ACP possui para compactar os dados e a capacidade de otimização do simplex sequencial.

A partir de uma análise das Tabelas 3 e 4, observa-se uma redução no erro quadrado médio e um aumento na porcentagem de classificação correta quando são comparadas as respostas da melhor rede e da média de ensemble. No trabalho realizado por Bona *et al.* (2011) foram observados resultados semelhantes, comprovando assim, que o uso da resposta do ensemble melhora o desempenho e aumenta a precisão da rede que se torna estável e convergente, duas qualidades importantes em um modelo neural (HAYKIN, 2001).

A Rede 4 obteve 100% de classificação correta (Tabela 3), ou seja, o classificador neural obtido classificou todas as amostras corretamente apesar das cidades estudadas estarem geograficamente próximas. A maioria dos trabalhos realizados para classificação de cafés utilizam amostras geograficamente muito distantes, de espécies diferentes e com diferentes graus de torra (KEMSLEY *et al.*, 1995; BRIANDET *et al.*, 1996; LYMAN *et al.*, 2003; WANG *et al.*, 2009). Neste trabalho, além das amostras serem de uma mesma espécie de café verde (arábica), os cafés são de regiões geograficamente muito próximas, sendo aproximadamente 195 km a distância máxima entre as cidades (Figura 1), comprovando assim, a eficácia do modelo neural obtido.

Tabela 5 - Tabela de confusão com os resultados de reconhecimento e classificação genotípica da Rede 11 para as amostras de teste.

Classe Prevista	Classe Observada			
	IPR 105	IPR 106	IPR 99	IA 59
IPR 105	3	0	1	0
IPR 106	0	3	0	1
IPR 99	1	1	3	0
IA 59	0	0	0	5

A classificação por genótipos obteve porcentagens de classificação correta inferiores (Tabela 4) por se tratar de um problema mais complexo como já comentado anteriormente. A Tabela 5 apresenta a tabela de confusão com os resultados de reconhecimento e classificação genotípica da Rede 11 para as amostras de teste. Esta rede, que alcançou melhores resultados para a classificação por genótipo, obteve 77,78% de classificação correta (classificação incorreta de 4 amostras), duas amostras (IPR 105 e IPR 106) foram confundidas com IPR99, uma amostra IA 59 foi confundida com IPR 106 e uma amostra IPR 99 foi confundida com IPR 105. Novamente ressalta-se que estes resultados demonstram que a classificação genotípica é muito mais complexa do que a segmentação geográfica.

A Rede 4 apresenta 20 neurônios na camada de entrada (componentes utilizadas), 2 neurônios na camada oculta e 4 na camada de saída. A Rede 11 apresenta 16 neurônios na camada de entrada, 6 neurônios na camada oculta e 4 na camada de saída. Devido ao elevado número de neurônios da Rede 11, ela apresenta um número muito alto de pesos sinápticos (130). Este número de pesos sinápticos deve ser menor que o número de exemplos utilizados no treinamento da rede neural (72 exemplos) para que ela possa aprender de maneira confiável, ou seja, devem haver graus de liberdade suficiente para que a aprendizagem da rede

seja segura, o que não se observa na Rede 11. Portanto, para melhorar o desempenho da classificação genotípica, os autores sugerem a utilização de outro tipo de RNA com menos parâmetros livres.

## **2.4. Conclusão**

Os perceptrons de múltiplas camadas otimizados foram capazes de classificar geograficamente as amostras de café arábica. Entretanto, para a classificação genotípica o desempenho não foi totalmente satisfatório. Além disso, o perceptron de múltiplas camadas desenvolvido para classificação genotípica apresenta um número muito elevado de pesos sinápticos, demandando assim, um número alto de graus de liberdade (amostras) para que a rede neural possa aprender de maneira confiável. Confirmando a sugestão da literatura consultada, as redes que utilizaram a faixa de  $1900-800\text{ cm}^{-1}$  do espectro FTIR obtiveram menor erro quadrado médio e maior porcentagem de classificação correta para a segmentação das amostras de café arábica tanto por região de cultivo quanto por genótipo. Também ficou evidente que o uso da média de ensemble aumentou o desempenho e reduziu a variabilidade das respostas das redes. O uso do simplex sequencial contornou a dificuldade para se determinar o melhor conjunto de parâmetros das redes além de realizar a maximização do desempenho e a minimização do tamanho das redes de uma maneira heurística. Assim, a combinação das redes MLP e a espectroscopia de infravermelho com transformada de Fourier (FTIR) mostrou ser uma ferramenta eficiente e rápida para o reconhecimento e classificação geográfico das amostras de café arábica.

## 2.5. Referências

ABIC. **Indicadores da indústria de café no Brasil – 2012.** <http://www.abic.com.br/publique/cgi/cgilua.exe/sys/start.htm?sid=61#1389>, 2012. Acesso em: 26 de Agosto de 2012.

BISHOP, C. M. **Neural Networks for Pattern Recognition.** Oxford: Oxford University: 1995. 482p.

BONA, E. et al. Optimized Neural Network for Instant Coffee Classification through an Electronic Nose. **International Journal of Food Engineering**, v. 7, 2011.

BORSATO, D. et al. Application of artificial neural networks in the geographical identification of coffee samples. **Eur Food Res Technol**, v. 233, p. 533-543, 2011.

BRANDEN, K. V.; HUBERT, M. Robust classification in high dimensions based on the SIMCA Method. **Chemometrics and Intelligent Laboratory Systems**, v. 79, p. 10-21, 2005.

BRASIL. **Regulamento técnico de identidade e de qualidade para a classificação do café beneficiado e de grão verde.** Instrução Normativa nº 8 de 11 jun. 2003. [http://www.claspar.pr.gov.br/arquivos/File/pdf/cafebenef008\\_03.pdf](http://www.claspar.pr.gov.br/arquivos/File/pdf/cafebenef008_03.pdf), 2011. Acesso em: 09 de novembro de 2011.

BRIANDET, R.; KEMSLEY, E. K.; WILSON, R. H. Approaches to Adulteration Detection in Instant Coffees using Infrared Spectroscopy and Chemometrics. **J Sci Food Agric**, v. 71, p. 359-366, 1996.

\_\_\_\_\_. Discrimination of Arabica and Robusta in Instant Coffee by Fourier Transform Infrared Spectroscopy and Chemometrics. **J. Agric. Food Chem.**, v. 44, p. 170-174, 1996.

CECAFÉ. **Dados estatísticos.** <http://www.cncafe.com.br/ba/file/CECAFE%20-%20Resumo%20das%20Exportacoes%20de%20Cafe%20MARCO%202012.pdf>, 2012. Acesso em: 19 de Outubro de 2012.

COSTA FILHO, P. A. D.; POPPI, R. J. Aplicação de algoritmos genéticos na seleção de variáveis em espectroscopia no infravermelho médio: determinação simultânea de glicose, maltose e frutose. **Química Nova**, v. 25, p. 46-52, 2002. ISSN 0100-4042.

DOWNEY, G. et al. Near- and Mid-Infrared Spectroscopies in Food Authentication: Coffee Varietal Identification. **J. Agric. Food Chem.**, v. 45, p. 4357-4361, 1997.

FARAH, A. **Coffee as a speciality and functional beverage.** Cambridge: Woodhead Publishing Limited: 2009. p.370-395.

GALÃO, O. F. et al. Artificial Neural Networks in the Classification and Identification of Soybean Cultivars by Planting Region. **J. Braz. Chem. Soc.**, v. 22, p. 142-147, 2011.

HAIR, J. F. et al. **Análise Fatorial em: Análise Multivariada de Dados**. 5ªed. Porto Alegre: Bookman, 2005. 38p.

HAYKIN, S. **Redes Neurais: Princípios e Prática**. 2ª edição. Porto Alegre: Bookman, 2001. 900p.

HUANG, Y.; KANGAS, L. J.; RASCO, B. A. Applications of artificial neural networks (ANNs) in food science. **Critical Reviews in Food Science and Nutrition**, v. 47, p. 13, 2007.

KAROUI, R.; DOWNEY, G.; BLECKER, C. Mid-Infrared Spectroscopy Coupled with Chemometrics: A Tool for the Analysis of Intact Food Systems and the Exploration of Their Molecular Structure-Quality Relationships - A Review. **Chemical Reviews**, v. 110, p. 6144-6167, 2010.

KEMSLEY, E. K.; RUAULT, S.; WILSON, R. H. Discrimination between *Coffea arabica* and *Coffea canephora* variant robusta beans using infrared spectroscopy. **Food Chemistry**, v. 54, p. 321-327, 1995.

LUCIA, S. M. D.; MINIM, L. A. **Redes Neurais Artificiais: Fundamentos e Aplicações**. 2006. p.173-194.

LYMAN, D. J. et al. FTIR-ATR Analysis of Brewed Coffee: Effect of Roasting Conditions. **J. Agric. Food Chem.**, v. 51, p. 3268-3273, 2003.

MARINI, F. Artificial neural networks in foodstuff analyses: Trends and perspectives A review. **Analytica Chimica Acta**, v. 635, p. 121-131, 2009.

PÉREZ-MAGARIÑO, S. et al. Comparative study of artificial neural network and multivariate methods to classify Spanish DO rose wines. **Talanta**, v. 62, n. 5, p. 983-990, 2004. ISSN 0039-9140. Disponível em: <  
<http://www.sciencedirect.com/science/article/pii/S0039914003006155> >.

RIBEIRO, J. S.; SALVA, T. J.; FERREIRA, M. M. C. Chemometric studies for quality control of processed brazilian coffees using drifts. **Journal of Food Quality**, v. 33, p. 212-227, 2010.

RUBAYIZA, A. B.; MEURENS, M. Chemical discrimination of arabica and robusta coffees by Fourier transform Raman spectroscopy. **Journal of Agriculture and Food Chemistry**, v. 53, p. p.4654-4659, 2005.

SATO, H. P. et al. Viabilidade de uso do FTIR e análise multivariada para a segmentação de amostras de café verde. **XVII Seminário de Iniciação Científica e Tecnológica da UTFPR - SICITE**. Curitiba: 8 p. 2012.

SINDICAFÉ. Dados Estatísticos. <http://sindicafe.com.br/plus/modulos/conteudo/?tac=cafe-no-mundo>, 2011. Acesso em: 24 de Novembro de 2012.



SUKTHOMYA, W.; TANNOCK, J. The optimization of neural network parameters using Taguchi's design of experiments approach: an application in manufacturing process modeling. **Neural Computation & Application**, v. 14, p. 7, 2005.

SÁ, J. P. M. D. **Applied Statistics Using SPSS, STATISTICA, MATLAB and R**. 2<sup>nd</sup> ed. New York: 2007.

WANG, J. et al. Fourier Transform Infrared Spectroscopy for Kona Coffee Authentication. **Journal of food science**, v. 74, p. 385-389, 2009.

WANG, N.; FU, Y.; LIM, L.-T. Feasibility Study on Chemometric Discrimination of Roasted Arabica Coffees by Solvent Extraction and Fourier Transform Infrared Spectroscopy. **J. Agric. Food Chem.**, v. 59, p. 3220-3226, 2011.

## **CAPÍTULO 3:**

Classificação geográfica e genotípica de café arábica utilizando FTIR e redes de base radial

### 3 CLASSIFICAÇÃO GEOGRÁFICA E GENOTÍPICA DE CAFÉ ARÁBICA UTILIZANDO FTIR E REDES DE BASE RADIAL

#### Resumo

As condições climáticas específicas de cada região de cultivo do café fornecem atributos especiais para a bebida e podem assim aumentar seu valor agregado. No entanto, é essencial provar a origem geográfica e genotípica do cultivar por meio de métodos confiáveis. Um exemplo de rede neural artificial (RNA) que tem sido muito utilizada para classificação de padrões é a rede de base radial (RBF). No contexto de uma RBF, as unidades ocultas fornecem um conjunto de funções que constituem uma base arbitrária para os padrões de entrada, quando eles são expandidos para um espaço oculto multidimensional. Estas funções são chamadas de funções de base radial. Este estudo objetivou o desenvolvimento de uma RNA para a classificação geográfica e genotípica de café arábica. Para esse fim os espectros obtidos por espectroscopia no infravermelho com transformada de Fourier (FTIR) foram analisados através do emprego de RBFs otimizadas através do simplex sequencial. As RBFs otimizadas foram capazes de classificar as amostras de café arábica, tanto geograficamente (100% de classificação correta), como genotipicamente (94,44%). As RBFs obtidas apresentam melhor desempenho que o SIMCA (*Soft Independent Modelling of Class Analogies*) e o perceptron de múltiplas camadas (MLP) desenvolvidos para classificação dos genótipos de café arábica, além de apresentar um número menor de parâmetros livres e serem mais rápidas durante o treinamento.

**Palavras-chave:** *café verde, espectro infravermelho, RBF, redes neurais artificiais*

### 3.1 INTRODUÇÃO

O café tem procedência de uma árvore do gênero *Coffea* e pertence à família das rubiáceas englobando mais de 500 gêneros e 6.000 espécies. Dentro do gênero *Coffea* existem duas espécies principais, o *Coffea arabica* variedade Arábica e o *Coffea canephora* variedade Robusta (Rubayiza & Meurens, 2005). O consumo brasileiro de café aumenta a cada ano, no período entre maio de 2011 e abril de 2012 foi registrado o consumo de 19,975 milhões de sacas, representando um acréscimo de 3,05% em relação ao período anterior correspondente (Abic, 2012). Além do mercado interno, as exportações de café verde do Brasil somaram 116,63 mil toneladas em março de 2012, baixa de 19,2% em relação ao mesmo mês de 2011. Já o consumo per capita foi de 6,18 kg de café em grão cru ou 4,94 kg de café torrado, quase 83 litros para cada brasileiro por ano, registrando uma evolução de 1,23% em relação ao período anterior. O consumo per capita brasileiro continua sendo um dos mais elevados mesmo quando comparado com o de países europeus. Os campeões de consumo, entretanto, ainda são os países nórdicos – Finlândia, Noruega, Dinamarca – com uma quantidade próxima dos 13 kg/por habitante/ano (Cecafé, 2012).

O Brasil apresenta diversidade de clima e solo, assim os mais variados tipos de cafés podem ser produzidos. Essas condições climáticas conferem atributos especiais para a bebida do café (corpo, acidez e sabor), e quando a qualidade do produto pode ser atribuída ao seu local de produção, ocorre um aumento do valor agregado do café. No entanto, neste processo, devem haver mecanismos confiáveis que comprovem a origem geográfica e genotípica do café (Borsato, Pina, Spacino, Scholz & Filho, 2011). Assim, torna-se imprescindível a utilização de técnicas analíticas rápidas e de baixo custo que permitam a diferenciação e certificação da composição e qualidade dos cafés comercializados. Uma metodologia que reúne tais atributos é a espectroscopia com transformada de Fourier na região do

infravermelho médio (FTIR). O FTIR permite a detecção dos componentes de maior importância na caracterização de alimentos, sendo uma metodologia que origina rapidamente uma informação sumária sobre o produto. Outra vantagem reside no fato do preparo da amostra, geralmente, não requerer a extração com solventes diminuindo assim o tempo de análise e o impacto ambiental (Karoui, Downey & Blecker, 2010). Ressalta-se ainda que o FTIR já foi utilizado com sucesso para autenticar misturas de cafés arábica e robusta ou adição de adulterantes (Briandet, Kemsley & Wilson, 1996; Downey, Briandet, Wilson & Kemsley, 1997; Kemsley, Ruault & Wilson, 1995; Wang, Jun, Bittenbender, Gautz & Li, 2009; Wang, Fu & Lim, 2011).

Os seres humanos possuem excelência nata para a aprendizagem e reconhecimento de padrões. Nesse sentido, os métodos estatístico-matemáticos têm sido desenvolvidos na tentativa de mimetizar, mesmo de maneira rudimentar, tal aptidão humana (Bishop, 1995; Haykin, 2001). As redes neurais artificiais (RNAs) são um conjunto de técnicas baseadas em princípios estatísticos, que vem crescentemente ganhando espaço para realizar tarefas de regressão e reconhecimento de padrões. As RNAs são extremamente versáteis para realizar o mapeamento de relações complexas e não lineares entre múltiplas variáveis de entrada e saída (Bishop, 1995). Um exemplo de rede neural artificial (RNA) que tem sido muito utilizada para classificação de padrões é a rede de base radial (RBF) (Tudu, Jana, Metla, Ghosh, Bhattacharyya & Bandyopadhyay, 2009). No contexto de uma RBF, as unidades ocultas fornecem um conjunto de funções radiais que constituem uma base arbitrária para os padrões de entrada, quando eles são expandidos para um espaço oculto multidimensional (Haykin, 2001). Para realizar tarefas de classificação as redes RBF, geralmente, oferecem maior flexibilidade na arquitetura de rede do que o perceptron de múltiplas camadas (MLP) e o

treinamento das redes RBF são muito mais rápidas do que outras arquiteturas de redes neurais (Tudu *et al.*, 2009).

Este estudo objetivou o desenvolvimento de redes de função de base radial para a classificação geográfica e genotípica de café arábica. Para esse fim os espectros obtidos no FTIR foram analisados através do emprego de redes de base radial otimizadas através do simplex sequencial.

## **3.2 MATERIAIS E MÉTODOS**

Todos os pré-processamentos dos espectros foram realizados no software IRsolution 1.5 (Shimadzu Corporation, Kyoto, Japão) que acompanha o equipamento. As análises matemáticas e/ou estatísticas, assim como a ACP, as redes de base radial e o simplex sequencial foram realizadas no software MATLAB R2007b (The MathWorks Inc., Natick, USA).

### **3.2.1 AMOSTRAS DE CAFÉ ARÁBICA**

Foram utilizados 4 genótipos de café arábica catalogados pelo Instituto Agrônomo do Paraná - Londrina (IAPAR) da safra de 2010 (Tabela 1).

Tabela 1 - Genótipos de café arábica utilizados.

<b>Genótipo</b>	<b>Local</b>	<b>Total Amostras</b>
IPR 99	Cornélio Procópio	1
	Paranavaí	1
	Mandaguari	1
	Londrina	1
IPR 105	Cornélio Procópio	1
	Paranavaí	1
	Mandaguari	1
	Londrina	1
IPR 106	Cornélio Procópio	1
	Paranavaí	1
	Mandaguari	1
	Londrina	1
IA 59	Cornélio Procópio	2
	Paranavaí	2
	Mandaguari	1
	Londrina	1

No IAPAR, as amostras foram imediatamente colocadas em caixas de madeira com uma malha de fundo e movidas oito vezes por dia até que a umidade dos grãos chegou a 11-12% e logo após as amostras foram beneficiadas (remoção da casca e pergaminho) (Brasil, 2011). Os grãos verdes, moídos, peneirados e embalados dos genótipos de café arábica fornecidos pelo IAPAR, foram recebidos e armazenados em um freezer a -18°C e utilizados posteriormente para análise. Foram utilizados cafés cultivados em 4 diferentes regiões (Cornélio Procópio, Paranavaí, Mandaguari e Londrina) totalizando 18 amostras. A Figura 1 apresenta a localização das cidades aonde os cafés foram cultivados e a distância máxima entre as mesmas.



Figura 1 - Mapa da distância entre as cidades dos cafés estudados.

### 3.2.2 ESPECTROSCOPIA DE INFRAVERMELHO COM TRANSFORMADA DE FOURIER (FTIR)

Para preparar as pastilhas foram adicionados em torno de 100 mg de KBr seco (padrão cromatográfico) e aproximadamente 1 mg de amostra finamente moída. A mistura foi, então, prensada em uma prensa hidráulica (Bovenau, P15 ST) usando um molde (ICL, ICL's Macro/Micro KBr die) empregando aproximadamente  $360 \text{ kgf/cm}^2$ . Produziu-se, assim, uma pastilha transparente. Antes da análise de cada amostra, o FTIR (Shimadzu, FTIR - 8300) foi programado para realizar um espectro de *background* do ar, sendo o mesmo utilizado para descontar a influência dos componentes do ar no espectro. Na sequência, a pastilha foi posicionada no feixe do instrumento e os espectros foram obtidos na faixa de  $4000$  a  $400 \text{ cm}^{-1}$ . Foram usadas 32 varreduras acumuladas para formar o espectro final e realizadas 5 repetições (pastilhas) para cada amostra, totalizando assim 90 espectros. Foram utilizados 72 desses espectros (80%) como amostras de treinamento das redes neurais artificiais empregadas e 18 espectros (20%) como amostras de teste (uma repetição de cada um dos cafés estudados).



### 3.2.3 TRANSFORMAÇÃO DOS ESPECTROS

Após obtenção dos espectros foram aplicadas algumas transformações nos mesmos. Primeiramente, foi realizada a normalização do espectro (a maior banda obteve absorbância 1 e a menor 0), correção da linha de base, a suavização do espectro (17 *smoothing points*) e uma remoção da banda referente a concentração de CO<sub>2</sub>, desprezando assim as variações de CO<sub>2</sub> entre as amostras. Para a ACP e a RNA foi considerado o uso da região entre 3750 e 750 cm<sup>-1</sup> do espectro desconsiderando ruídos que se apresentavam além desta região. Também foi avaliada a região entre 1900 e 800 cm<sup>-1</sup> do espectro pois a mesma contém as bandas de absorção devidas a deformação axial simétrica da carbonila (ésteres, aldeídos e cetonas), deformação angular simétrica do metileno, deformações axiais e angulares simétricas do C-O (ésteres e alcoóis). Portanto, essa região contém a informação de *fingerprint* para discriminação de diferentes amostras de cafés (Briandet *et al.*, 1996; Lyman, Benck, Dell, MErle, & Murray-Wijelath, 2003; Wang *et al.*, 2009; Wang *et al.*, 2011).

Após estes pré-processamentos foi utilizada a análise de componentes principais (ACP) para a redução da dimensionalidade dos dados (SÁ, 2007). A ACP foi realizada nos dados normalizados assim como na 1ª e 2ª derivada dos espectros. A aplicação da 1ª e da 2ª derivada sobre os dados espectrais brutos é um procedimento que pode destacar ombros espectrais, bem como minimizar o efeito de inclinações provocadas na linha de base dos espectros. Ao aplicarmos as operações de derivação as informações contidas ao longo dos diferentes números de onda são geralmente acentuadas (os ruídos também tornam-se acentuados), portanto deve-se ter cuidado com a qualidade dos espectros com os quais se deseja aplicar o cálculo das derivadas (Costa Filho & Poppi, 2002). A quantidade de componentes principais empregadas foi um dos parâmetros otimizados através do simplex sequencial.

### 3.2.4 NORMALIZAÇÃO

Os vetores de entrada (as componentes principais escolhidas) foram pré-processados antes de serem alimentados na rede neural para que as funções de ativação dos neurônios artificiais não sejam facilmente saturadas ou ocorra um erro de *overflow* (Haykin, 2001). Os métodos de pré-processamento utilizados foram: máximo e mínimo (minimax), transformação para uma escala entre -1 e 1; autoescalamento, vetor de entrada com média zero e variância unitária; vetor unitário, norma euclidiana igual a 1 (Pérez-Magariño, Ortega-Heras, González-San José & Boger, 2004).

### 3.2.5 REDES DE FUNÇÃO DE BASE RADIAL (RBF)

A arquitetura de uma rede neural RBF, mostrada na Figura 2, envolve três camadas com papéis totalmente diferentes entre si. Os neurônios de entrada conectam a rede ao seu ambiente. A segunda camada, a única camada oculta da rede, aplica uma transformação não linear do espaço de entrada para um espaço oculto, também conhecido como espaço de características. Essas unidades ocultas fornecem um conjunto de funções radiais que constituem uma base arbitrária para os padrões de entrada. O espaço oculto é de alta dimensionalidade, permitindo assim, uma separação linear dos grupos (teorema de Cover sobre a separabilidade de padrões). A camada de saída faz uma combinação linear das bases radiais, fornecendo a resposta da rede ao padrão de ativação aplicado à camada de entrada (Haykin, 2001).

O número de neurônios da camada de entrada é determinado pela quantidade de componentes principais utilizadas que é um parâmetro que foi determinado pelo simplex sequencial. Na camada oculta existem  $\varphi_{jk}$  funções de base radiais, onde  $j$  representa o

número da classe e  $k$  representa a  $k$ -ésima função de base radial. A quantidade de bases radiais utilizadas também foi um parâmetro determinado pelo simplex sequencial. O número de nós ocultos associados a cada classe é diretamente dependente da complexidade dos padrões a serem separados (Tudu *et al.*, 2009).

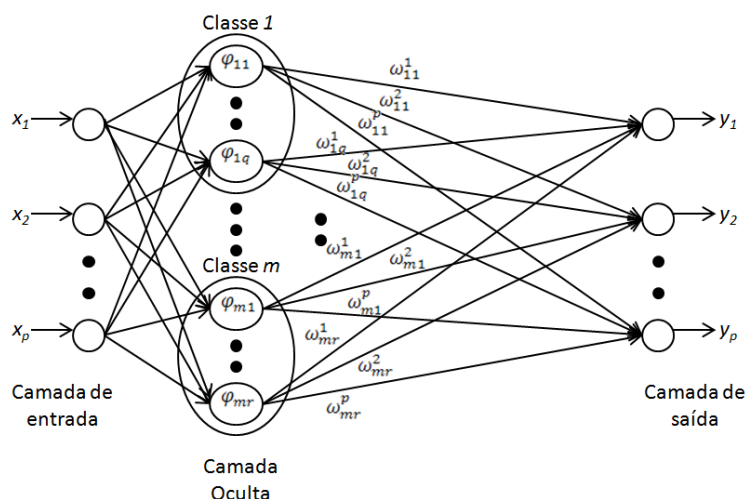


Figura 2 - Representação de uma rede de função de base radial (RBF).

Há uma grande classe de funções de base radial, esta classe inclui as seguintes funções que são de particular interesse no estudo de redes RBF:

1. Multiquadráticas:

$$\varphi(r) = (r^2 + \sigma^2)^{1/2} \quad (1)$$

2. Multiquadráticas inversas:

$$\varphi(r) = \frac{1}{(r^2 + \sigma^2)^{1/2}} \quad (2)$$

3. Funções gaussianas:

$$\varphi(r) = \exp\left(-\frac{r^2}{2\sigma^2}\right) \quad (3)$$

onde  $r$  é a distância entre os centros das bases e  $\sigma$  é um parâmetro cujo valor controla a suavidade da função de interpolação.

O método de aprendizagem aplicado foi o processo de aprendizagem híbrido (Bishop, 1995; Haykin, 2001), que é constituído de dois diferentes estágios:

- Estágio de aprendizagem auto-organizada: cujo propósito é estimar localizações adequadas para os centros das funções de base radial na camada oculta.
- Estágio de aprendizagem supervisionada: que completa o projeto da rede estimando os pesos lineares da camada de saída.

Para o processo de aprendizagem auto-organizada foi utilizado um método de agrupamento que separa o conjunto fornecido de dados em subgrupos, cada um dos quais sendo tão homogêneo quanto possível. Para tanto, foi empregado o algoritmo de K-médias (*K-means*), que coloca os centros das funções de base radial apenas em regiões densamente povoadas do espaço multidimensional de entrada.

Tendo identificado os centros individuais das funções de base radial utilizando o algoritmo de agrupamento K-médias, o próximo estágio do processo de aprendizagem foi avaliar a distância máxima entre as bases (dispersão) e estimar os pesos da camada de saída.

O cálculo da distância máxima entre as bases (dispersão) deve ser realizado empregando uma função cujo desvio padrão é fixado de acordo com o espalhamento dos centros conforme a equação (4) (Bishop, 1995):

$$\sigma = \frac{d_{max}}{\sqrt{2m_1}} \quad (4)$$

onde  $m_1$  é o número de centros e  $d_{max}$  é a distância máxima entre os centros escolhidos. A equação (6) assegura que as funções de base radial não sejam pontiagudas demais ou planas demais; estas duas condições extremas devem ser evitadas.

Para a determinação dos pesos lineares da camada de saída da rede foi utilizado um procedimento direto que utiliza a pseudo-inversa da matriz  $\Phi$  (equação 5) (Haykin, 2001).

$$W = \Phi^\dagger T \quad (5)$$

onde  $T$  é o vetor resposta desejada do conjunto de treinamento e  $\Phi^\dagger$  é a pseudo-inversa da matriz  $\Phi$  que é calculada por:

$$\Phi^\dagger = (\Phi^T \Phi)^{-1} \Phi^T \quad (6)$$

onde  $\Phi^T$  é a matriz transposta de  $\Phi$ , que é definida por:

$$\Phi = \varphi_k \left( \|x_j - t_i\|^2, \sigma \right), j = 1, 2, \dots, N; i = 1, 2, \dots, m_1 \quad (7)$$

onde  $x_j$  representa o  $j$ -ésimo vetor de entrada da amostra de treinamento,  $k$  representa a  $k$ -ésima função de base radial e  $t_i$  é o número de centros (Tudu *et al.*, 2009).

As redes de base radial (RBF) e os perceptrons de múltiplas camadas (MLP) são exemplos de redes em camadas alimentadas adiante, não lineares e ambos são aproximadores universais, entretanto, estas duas redes diferem entre si em vários aspectos importantes (número de camadas ocultas, argumento da função de ativação, algoritmo de treinamento e velocidade de treinamento) (Bishop, 1995; Haykin, 2001; Marini, 2009; Tudu *et al.*, 2009).

### 3.2.6 MÉDIA DE ENSEMBLE

Um modo de melhorar a capacidade de generalização de uma rede neural é a combinação das respostas de várias RNAs, também chamada de máquinas de comitê, as quais foram treinadas de maneira semelhante de modo que o resultado final é a fusão das respostas de várias redes e é superior àquele alcançável apenas por uma única RNA. Um dos métodos de implementar o comitê de máquinas é a média de ensemble, que consiste em combinar linearmente as respostas de um certo número de redes treinadas através de uma média. O uso de um ensemble diminui o erro final devido a *bias*, pois há um ajuste excessivo proposital de

cada rede e também devido a variância pelo uso de diferentes condições iniciais de treinamento de cada rede (Haykin, 2001). Neste trabalho foi realizada uma média de *ensemble* com as redes com performance de classificação maior que 70% sendo cada rede inicializada com condições diferentes.

### 3.2.7 OTIMIZAÇÃO DOS PARÂMETROS DA REDE

Para escolha da melhor rede neural para o problema proposto, alguns parâmetros foram otimizados através do simplex sequencial (Tabela 2). A otimização foi realizada visando minimizar o erro quadrado médio para o conjunto de amostras utilizadas para teste (Bona, Silva, Borsato & Bassoli, 2011).

Tabela 2 - Parâmetros otimizados através do simplex sequencial.

<b>Rede</b>	<b>Parâmetro</b>	<b>Variação</b>
<b>Redes de Função de Base Radial</b>	Quantidade de bases radiais	5 a 15
	Função de pré-processamento das entradas	Dados Puros, Minimax, Autoescalamento ou Vetor Unitário
	Função de distância para o algoritmo K-médias	Distância euclidiana, <i>City Block</i> , Cosseno ou Correlação
	Método de inicialização do algoritmo K-médias	Aleatório ou Uniforme
	Tipos de base radial	Gaussiana, Multiquadrática ou Multiquadrática inversa
	Quantidade de variáveis independentes	7 a 30

O princípio do simplex básico foi deslocar uma figura regular, por exemplo um triângulo equilátero sobre uma superfície, quando duas variáveis estavam sendo consideradas (Splendley, Himsforth & Hext, 1962).

O algoritmo utilizado nesse trabalho está descrito em maiores detalhes em Gao e Han (2010) e no APÊNDICE B. A otimização segue até que o valor do erro quadrado médio varie apenas dentro da tolerância estabelecida que foi de 0,001 ou pela avaliação gráfica que tem por finalidade auxiliar na visualização da otimização, que está representada como uma suavização na variação das respostas e variáveis independentes.

### 3.3 RESULTADOS E DISCUSSÕES

Os espectro FTIR obtidos e tratado para os genótipos de café arábica estão representados na Figura 3.

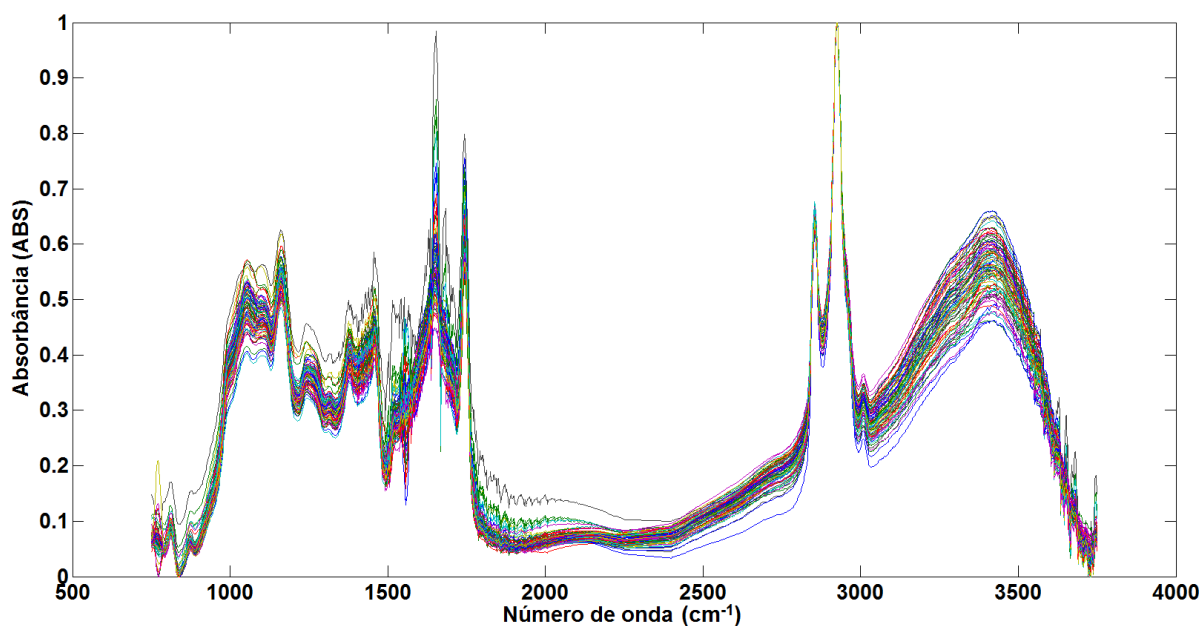


Figura 3 – Espectros normalizados obtidos para os genótipos de café arábica.

Através da análise dos espectros destaca-se a banda em  $1744\text{ cm}^{-1}$  resultante da deformação axial simétrica do C=O de ésteres de ácidos graxos (Wang *et al.*, 2011). As bandas na região de  $1654\text{ cm}^{-1}$  e  $1603\text{ cm}^{-1}$  estão relacionadas com a carbonila da cafeína (Wang *et al.*, 2011). A região de  $1285\text{ cm}^{-1}$  está relacionada com os ácidos clorogênicos

(Briandet *et al.*, 1996). As regiões de  $1200\text{ cm}^{-1}$  a  $900\text{ cm}^{-1}$  e de  $1500\text{ cm}^{-1}$  a  $1400\text{ cm}^{-1}$  estão relacionadas ao conteúdo de carboidratos (Briandet *et al.*, 1996; Ribeiro, Salva & Ferreira, 2010).

### 3.3.1 REDES DE FUNÇÃO DE BASE RADIAL (RBF)

Após a otimização dos parâmetros selecionados (Tabela 2), foram criadas dez mil redes para cada faixa de trabalho no FTIR e tipo de tratamento do espectro (Tabelas 3 e 4). A melhor rede de cada tipo foi selecionada pelo menor erro quadrado médio e maior porcentagem de classificação para as amostras de teste e também foi realizada uma média de ensemble com as redes com performance de classificação maior que 70%. Os resultados obtidos para as redes de função de base radial propostas para a classificação geográfica e genotípica de café arábica estão representados nas Tabelas 3 e 4. É importante salientar que as condições de cada problema influenciam na escolha dos parâmetros ótimos obtidos para cada rede (Sukthomya & Tannock, 2005).

A partir de uma análise da Tabela 3, é possível concluir que, as redes RBF que obtiveram o menor erro quadrado médio e maior porcentagem de classificação correta para a segmentação geográfica de café arábica foram a Rede A e a Rede D. Ambas apresentaram 100% de classificação correta das amostras de teste tanto para a média de ensemble quanto para suas melhores redes. Porém a Rede D apresenta menores valores do erro quadrado médio para as amostras de teste e, também obteve melhores resultados para as amostras de treinamento. Além disso, por possuir um número menor de neurônios na camada de entrada, um indicativo de modelo mais simples, a mesma foi escolhida como a melhor opção para a classificação geográfica dos cafés. Esta rede (Rede D) foi treinada com os espectros originais (sem derivá-los) e com o faixa espectral entre  $1900\text{-}800\text{ cm}^{-1}$ , uma faixa espectral mais



específica que contem as impressões bioquímicas (*fingerprints*) para discriminação de diferentes amostras de cafés (Briandet *et al.*, 1996; Lyman *et al.*, 2003; Wang *et al.*, 2009).

Tabela 3 - Resultados obtidos para as redes de base radial propostas para a classificação geográfica de café arábica.

	Número de onda					
	3750-750 cm <sup>-1</sup>			1900-800 cm <sup>-1</sup>		
	Rede A	Rede B	Rede C	Rede D	Rede E	Rede F
Espectro	Original	1 <sup>a</sup> Derivada	2 <sup>a</sup> Derivada	Original	1 <sup>a</sup> Derivada	2 <sup>a</sup> Derivada
Número de bases radiais	15	15	11	15	15	13
Função de distância para o K-médias <sup>a</sup>	CB	CB	CB	CB	CB	CB
Método de inicialização do K-médias	Aleatório	Aleatório	Aleatório	Aleatório	Aleatório	Aleatório
Tipo de base radial <sup>b</sup>	MQ	MQ	G	MQ	MQ	G
Componentes utilizadas	20	23	26	18	25	16
Função de pré-processamento (entradas) <sup>c</sup>	AE	VU	MM	AE	AE	AE
<b>Desempenho das melhores redes obtidas para os dados de treinamento</b>						
Erro quadrado médio (Ensemble)	0,0496	0,0424	0,0998	0,0364	0,0456	0,0976
Erro quadrado médio (Melhor Rede)	0,0487	0,0512	0,1082	0,0407	0,0491	0,1078
% Classificação correta (Ensemble)	95,83%	95,83%	86,11%	95,83%	95,83%	77,78%
% Classificação correta (Melhor Rede)	98,61%	94,44%	72,22%	100,00%	95,83%	72,22%
<b>Desempenho das melhores redes obtidas para os dados de teste</b>						
Erro quadrado médio (Ensemble)	0,0745	0,0652	0,1265	0,0412	0,0568	0,1070
Erro quadrado médio (Melhor Rede)	0,0658	0,0681	0,1059	0,0433	0,0430	0,0962
% Classificação correta (Ensemble)	100,00%	94,44%	77,78%	100,00%	94,44%	83,33%
% Classificação correta (Melhor Rede)	100,00%	100,00%	88,89%	100,00%	100,00%	94,44%

<sup>a</sup> Função de distância para o K-médias: DE (Distância euclidiana), CB (*City Block*), COS (Cosseno) e COR (Correlação).

<sup>b</sup> Tipo de base radial: G (Gaussiana), MQ (Multiquadrática), MQI (Multiquadrática inversa).

<sup>c</sup> Função de pré-processamento: DP (Dados Puros), MM (Minimax), AE (Autoescalonamento) e VU (Vetor Unitário).

Para a segmentação genotípica (Tabela 4) a RBF que obteve menor erro quadrado médio e maior porcentagem de classificação correta (Rede L) utilizou a primeira derivada do espectro e, assim como para a classificação geográfica, com a faixa espectral entre 1900-800

cm<sup>-1</sup>. Os espectros variam muito pouco por se tratarem de uma mesma variedade de café, por isso, a primeira derivada do espectro obteve melhores resultados, pois ela acentua as diferenças espectrais removendo ruídos em uma mesma região espectral e/ou definindo bandas que se encontram sobrepostas (Costa Filho & Poppi, 2002). Através da análise da performance das RBF nas Tabelas 3 e 4 fica evidente que a classificação genotípica é um problema mais difícil de ser modelado.

Tabela 4 - Resultados obtidos para as redes de base radial propostas para a classificação genotípica de café arábica.

	Número de onda					
	3750-750 cm <sup>-1</sup>			1900-800 cm <sup>-1</sup>		
	Rede G	Rede H	Rede I	Rede J	Rede L	Rede M
Espectros	Original	1 <sup>a</sup> Derivada	2 <sup>a</sup> Derivada	Original	1 <sup>a</sup> Derivada	2 <sup>a</sup> Derivada
Número de bases radiais	15	14	12	14	14	11
Função de distância para o K-médias <sup>a</sup>	DE	CB	CB	DE	CB	CB
Método de inicialização do K-médias	Aleatório	Aleatório	Aleatório	Aleatório	Aleatório	Aleatório
Tipo de base radial <sup>b</sup>	MQ	MQ	MQ	MQ	MQ	MQ
Componentes utilizadas	24	25	24	26	21	26
Função de pré-processamento (entradas) <sup>c</sup>	MM	AE	MM	AE	MM	AE
<b>Desempenho das melhores redes obtidas para os dados de treinamento</b>						
Erro quadrado médio (Ensemble)	0,0857	0,0807	0,1056	0,0691	0,0800	0,1089
Erro quadrado médio (Melhor Rede)	0,0982	0,1016	0,1076	0,0858	0,0819	0,1197
% Classificação correta (Ensemble)	80,56%	88,89%	75,00%	90,28%	88,89%	73,61%
% Classificação correta (Melhor Rede)	80,56%	76,39%	75,00%	83,33%	88,89%	70,83%
<b>Desempenho das melhores redes obtidas para os dados de teste</b>						
Erro quadrado médio (Ensemble)	0,1174	0,0989	0,1240	0,0959	0,0941	0,1167
Erro quadrado médio (Melhor Rede)	0,1142	0,0934	0,1258	0,0925	0,0840	0,1193
% Classificação correta (Ensemble)	77,78%	83,33%	83,33%	83,33%	83,33%	66,67%
% Classificação correta (Melhor Rede)	88,89%	94,44%	83,33%	94,44%	94,44%	83,33%

<sup>a</sup> Função de distância para o K-médias: DE (Distância euclidiana), CB (*City Block*), COS (Cosseno) e COR (Correlação).

<sup>b</sup> Tipo de base radial: G (Gaussiana), MQ (Multiquadrática), MQI (Multiquadrática inversa).

<sup>c</sup> Função de pré-processamento: DP (Dados Puros), MM (Minimax), AE (Autoescalamento) e VU (Vetor Unitário).

As RBFs desenvolvidas apresentam melhores resultados quando comparados com outros métodos que foram aplicados ao mesmo banco de dados (Tabela 5) (Sato, Bona, Link & Lemes, 2012). Devido a RBF ser um método não linear de classificação, o desempenho foi superior em comparação ao SIMCA (*Soft Independent Modelling of Class Analogies*), que é um método linear, e se baseia na classificação de uma nova amostra a partir de sua projeção nas componentes principais de cada classe (Branden & Hubert, 2005). A RBF desenvolvida apresenta performance semelhante ao perceptron de múltiplas camadas (MLP) proposto para a classificação geográfica, porém, apresenta melhor desempenho para a classificação genotípica, sendo um melhor modelo neural para este problema, além da RBF apresentar um menor número de parâmetros livres, geralmente, oferecer maior flexibilidade na arquitetura de rede do que o MLP e também o treinamento das redes RBF ser muito mais rápido do que outras arquiteturas de redes neurais (Bishop, 1995; Haykin, 2001; Tudu *et al.*, 2009).

Tabela 5 - Comparação entre as RBFs desenvolvidas e outros métodos que foram propostos para a classificação geográfica e genotípica de café arábica.

<b>Porcentagem de classificação correta para os dados de teste</b>		
	Classificação Geográfica	Classificação Genotípica
<b>SIMCA</b>	83,33%	55,56%
<b>MLP</b>	100,00%	77,78%
<b>RBF</b>	100,00%	94,44%

O número de variáveis independentes utilizadas na camada de entrada de neurônios também foi otimizado através do simplex sequencial. A média de componentes principais utilizadas foi de 21 para a classificação geográfica e 24 para a classificação genotípica, esses dados também confirmam que a separação dos genótipos é mais complexa. Tais valores encontram-se dentro do que era esperado, pois de acordo o critério da raiz latente qualquer componente principal deve explicar a variância de pelo menos uma variável se a mesma há de

ser mantida para interpretação. Cada variável contribui com um valor 1 do autovalor total. Logo, apenas os fatores que têm raízes latentes maiores que 1 são considerados significantes (Hair, Anderson, Tatham, & Black, 2005). Para este estudo obteve-se 19 componentes principais que apresentaram valores de raízes latentes maiores que 1, comprovando assim, a capacidade que a ACP possui para compactar os dados e a capacidade de otimização do simplex sequencial.

Resalta-se ainda que tanto para a classificação geográfica quanto para a classificação dos genótipos as redes com base radial do tipo multiquadrática obtiveram um erro quadrado médio menor e uma porcentagem de classificação superior, provando que para a segmentação geográfica e genotípica as bases radiais do tipo multiquadrática são as mais adequadas. Dentre as funções de distância para o algoritmo K-médias, ficou comprovado que a *City Block* apresenta uma superioridade em relação as demais funções testadas. Com relação a normalização, ficou evidente que o método autoescalonamento é superior aos demais para a classificação geográfica, devido a maioria das redes que apresentaram melhores resultados utilizarem este mesmo método de normalização. Para a classificação genotípica, dois métodos de normalização obtiveram resultados satisfatórios (autoescalonamento e minimax), mas a rede que obteve os melhores resultados (Rede L), apresentou o uso da normalização minimax, sendo esta então a melhor opção para a classificação genotípica dos cafés. A observação dos resultados deste trabalho permite reforçar que, os melhores parâmetros de uma rede são dependentes do problema. Portanto, a utilização de uma metodologia de otimização é fundamental para encontrar o melhor conjunto de parâmetros que irá permitir a construção de uma rede neural eficiente e confiável (Ludermir, Yamazaki, & Zanchetin, 2006).

Em relação à média de ensemble os resultados obtidos para as amostras de teste da Rede L (classificação genotípica) não foram superiores aos obtidos pela melhor rede, devido

ao elevado número de redes com performance de classificação superior a 70% e essas redes não apresentarem resultados próximos aos obtidos pela melhor. Essa observação é contrária ao observado por Link *et al.* (2013) e por Bona *et al.* (2011) que relatam uma melhoria na capacidade de generalização de redes do tipo MLP.

Tabela 6 - Tabela de confusão para o reconhecimento e classificação genotípica da Rede L para as amostras de teste.

Classe Prevista	Classe Observada			
	IPR 105	IPR 106	IPR 99	IA 59
IPR 105	3	0	0	0
IPR 106	0	4	0	0
IPR 99	1	0	4	0
IA 59	0	0	0	6

A Rede D obteve 100% de classificação correta (Tabela 3), ou seja, o classificador neural obtido classificou todas as amostras corretamente apesar das cidades estudadas estarem geograficamente próximas. A maioria dos trabalhos realizados para classificação de cafés utilizam amostras geograficamente muito distantes, de espécies diferentes e com diferentes graus de torra (Briandet *et al.*, 1996; Briandet *et al.*, 1996; Kemsley *et al.*, 1995; Lyman *et al.*, 2003; Wang *et al.*, 2009). Neste trabalho, além das amostras serem de uma mesma espécie de café verde (arábica), os cafés são de regiões geograficamente muito próximas, sendo aproximadamente 195 km a distância máxima entre as cidades (Figura 1), comprovando assim, a eficácia do modelo neural obtido.

A Tabela 6 apresenta a tabela de confusão para o reconhecimento e classificação genotípica das amostras de teste pela Rede L que alcançou os melhores resultados (94,44% de

classificação correta). A Rede L classificou incorretamente apenas 1 amostra de teste (IPR 105 que foi confundida com IPR 99). Novamente ressalta-se que estes resultados demonstram que a classificação genotípica é muito mais complexa do que a segmentação geográfica.

### **3.4 CONCLUSÃO**

As redes de base radial otimizadas foram capazes de classificar as amostras de café arábica tanto geograficamente como genotipicamente. As redes que utilizaram a faixa de 1900-800  $\text{cm}^{-1}$  do espectro FTIR obtiveram menor erro quadrado médio e maior porcentagem de classificação correta para a segmentação das amostras de café arábica tanto por região de cultivo quanto por genótipo. As redes de base radial desenvolvidas apresentam melhor desempenho que perceptrons de múltiplas camadas para classificação dos genótipos de café arábica, além de apresentar um número menor de parâmetros livres e serem mais rápidas durante o treinamento. O uso da média de ensemble aumentou o desempenho das respostas da maioria das redes, mas devido a se tratar de uma média simples, alguns resultados não foram satisfatórios. O simplex sequencial demonstrou ser uma metodologia eficiente para se determinar o melhor conjunto de parâmetros, maximizando o desempenho e minimizando o tamanho das redes de uma maneira heurística. Assim, a combinação das redes de funções de base radial e a espectroscopia de infravermelho com transformada de Fourier (FTIR) mostrou ser uma ferramenta rápida e eficaz para o reconhecimento e classificação de amostras de café arábica.

### 3.5 REFERÊNCIAS

- ABIC. (2012). Indicadores da indústria de café no Brasil – 2012. <http://www.abic.com.br/publique/cgi/cgilua.exe/sys/start.htm?sid=61#1389>.
- Bishop, C. M. (1995). *Neural Networks for Pattern Recognition*. Oxford: Oxford University.
- Bona, E., Silva, R. S. S. D., Borsato, D., & Bassoli, D. G. (2011). Optimized Neural Network for Instant Coffee Classification through an Electronic Nose. *International Journal of Food Engineering*, 7.
- Borsato, D., Pina, M. V. R., Spacino, K. R., Scholz, M. B. d. S., & Filho, A. A. (2011). Application of artificial neural networks in the geographical identification of coffee samples. *Eur Food Res Technol*, 233, 533-543.
- Branden, K. V., & Hubert, M. (2005). Robust classification in high dimensions based on the SIMCA Method. *Chemometrics and Intelligent Laboratory Systems*, 79, 10-21.
- Brasil. (2011). Regulamento técnico de identidade e de qualidade para a classificação do café beneficiado e de grão verde. Instrução Normativa nº 8 de 11 jun. 2003. [http://www.claspar.pr.gov.br/arquivos/File/pdf/cafefenef008\\_03.pdf](http://www.claspar.pr.gov.br/arquivos/File/pdf/cafefenef008_03.pdf): Ministério da Agricultura Pecuária e Abastecimento.
- Briandet, R., Kemsley, E. K., & Wilson, R. H. (1996). Approaches to Adulteration Detection in Instant Coffees using Infrared Spectroscopy and Chemometrics. *J Sci Food Agric*, 71, 359-366.
- Briandet, R., Kemsley, E. K., & Wilson, R. H. (1996). Discrimination of Arabica and Robusta in Instant Coffee by Fourier Transform Infrared Spectroscopy and Chemometrics. *J. Agric. Food Chem.*, 44, 170-174.
- CECAFÉ. (2012). Dados estatísticos. In, vol. 2012). <http://www.cncafe.com.br/ba/file/CECAFÉ%20-%20Resumo%20das%20Exportacoes%20de%20Cafe%20MARCO%202012.pdf>.
- Costa Filho, P. A. d., & Poppi, R. J. (2002). Aplicação de algoritmos genéticos na seleção de variáveis em espectroscopia no infravermelho médio: determinação simultânea de glicose, maltose e frutose. *Química Nova*, 25, 46-52.
- Downey, G., Briandet, R., Wilson, R. H., & Kemsley, E. K. (1997). Near- and Mid-Infrared Spectroscopies in Food Authentication: Coffee Varietal Identification. *J. Agric. Food Chem.*, 45, 4357-4361.
- Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (2005). *Análise Fatorial*. In: \_\_\_\_\_. *Análise Multivariada de Dados* (5ª ed. ed.): Porto Alegre: Bookman.
- Haykin, S. (2001). *Redes Neurais: Princípios e Prática* (2ª edição ed.). Porto Alegre: Bookman.
- Karoui, R., Downey, G., & Blecker, C. (2010). Mid-Infrared Spectroscopy Coupled with Chemometrics: A Tool for the Analysis of Intact Food Systems and the Exploration of Their Molecular Structure-Quality Relationships - A Review. *Chemical Reviews*, 110, 6144-6167.
- Kemsley, E. K., Ruault, S., & Wilson, R. H. (1995). Discrimination between Coffea arabica and Coffea canephora variant robusta beans using infrared spectroscopy. *Food Chemistry*, 54, 321-327.
- Link, J. V., Lemes, A. L. G., Sato, H. P., Scholz, M. B. S., Bona, E. Optimized Multilayer perceptron for the geographical and genotypic classification of four genotypes of arabica coffee. *Revista Brasileira de Pesquisa em Alimentos (REBRAPA)*, 2013.
- Ludermir, T. B., Yamazaki, A., & Zanchetin, C. (2006). An optimization methodology for neural network weights and architectures. *IEEE Transactions on Neural Networks*, 17, 7.
- Lyman, D. J., Benck, R., Dell, S., Merle, S., & Murray-Wijelath, J. (2003). FTIR-ATR Analysis of Brewed Coffee: Effect of Roasting Conditions. *J. Agric. Food Chem.*, 51, 3268-3273.
- Marini, F. (2009). Artificial neural networks in foodstuff analyses: Trends and perspectives A review. *Analytica Chimica Acta*, 635, 121-131.
- Pérez-Magariño, S., Ortega-Heras, M., González-San José, M. L., & Boger, Z. (2004). Comparative study of artificial neural network and multivariate methods to classify Spanish DO rose wines. *Talanta*, 62(5), 983-990.
- Ribeiro, J. S., Salva, T. J., & Ferreira, M. M. C. (2010). Chemometric studies for quality control of processed brazilian coffees using drifts. *Journal of Food Quality*, 33, 212-227.
- Rubayiza, A. B., & Meurens, M. (2005). Chemical discrimination of arabica and robusta coffees by Fourier transform Raman spectroscopy. *Journal of Agriculture and Food Chemistry*, 53, p.4654-4659.
- Sato, H. P., Bona, E., Link, J. V., & Lemes, A. L. G. (2012). Viabilidade de uso do FTIR e análise multivariada para a segmentação de amostras de café verde. In *XVII Seminário de Iniciação Científica e Tecnológica da UTFPR - SICITE*, (pp. 8). Curitiba.
- Sukthomya, W., & Tannock, J. (2005). The optimization of neural network parameters using Taguchi's design of experiments approach: an application in manufacturing process modeling. *Neural Computation & Application*, 14, 7.
- Sá, J. P. M. D. (2007). *Applied Statistics Using SPSS, STATISTICA, MATLAB and R* (2<sup>nd</sup> ed. ed.). New York.

- Tudu, B., Jana, A., Metla, A., Ghosh, D., Bhattacharyya, N., & Bandyopadhyay, R. (2009). Electronic nose for black tea quality evaluation by an incremental RBF network. *Sensors and Actuators B: Chemical*, 138, 90-96.
- Wang, J., Jun, S., Bittenbender, H. C., Gautz, L., & Li, Q. X. (2009). Fourier Transform Infrared Spectroscopy for Kona Coffee Authentication. *Journal of food science*, 74, 385-389.
- Wang, N., Fu, Y., & Lim, L.-T. (2011). Feasibility Study on Chemometric Discrimination of Roasted Arabica Coffees by Solvent Extraction and Fourier Transform Infrared Spectroscopy. *J. Agric. Food Chem.*, 59, 3220-3226.



## CONCLUSÃO GERAL

Neste trabalho foi observado que a utilização de diferentes tipos de redes neurais artificiais, tanto para a análise de agrupamentos (Mapas Auto-Organizáveis) quanto para o reconhecimento e classificação de padrões (Perceptron de Múltiplas Camadas e Redes de Funções de Base Radial), é uma estratégia eficaz para a segmentação geográfica e genotípica do café arábica verde.

Os Mapas Auto-Organizáveis foram capazes de distinguir os 14 genótipos de café arábica verde utilizando os dados de perfil químico das amostras com destaque para os teores de cafeína, ácidos clorogênicos, proteínas e açúcares redutores. A possibilidade de gerar mapas de peso para cada variável permitiu uma análise mais fundamentada a respeito da influência de cada variável analisada na segmentação observada para as amostras.

Os Perceptrons de Múltiplas Camadas (MLP) e as Redes de Funções de Base Radial (RBF) foram eficientes na classificação geográfica das amostras, porém a classificação genotípica demonstrou ser mais complexa. O MLP desenvolvido para classificação genotípica apresentou um número muito elevado de pesos sinápticos, demandando assim, um número alto de graus de liberdade (amostras) para que a rede neural possa aprender de maneira confiável. Por sua vez, o modelo RBF, por apresentar menos parâmetros livres, obteve um melhor desempenho além de ser mais rápido durante o treinamento. Ficou evidente que o uso da média de ensemble aumentou o desempenho e reduziu a variabilidade das respostas da maioria das redes, mas devido a se tratar de uma média simples, em alguns casos não houve melhora.

Confirmando a sugestão da literatura consultada, as redes que utilizaram a faixa de 1900-800  $\text{cm}^{-1}$  do espectro FTIR obtiveram menor erro quadrado médio e maior porcentagem de classificação correta para a segmentação das amostras de café arábica tanto por região de cultivo quanto por genótipo.

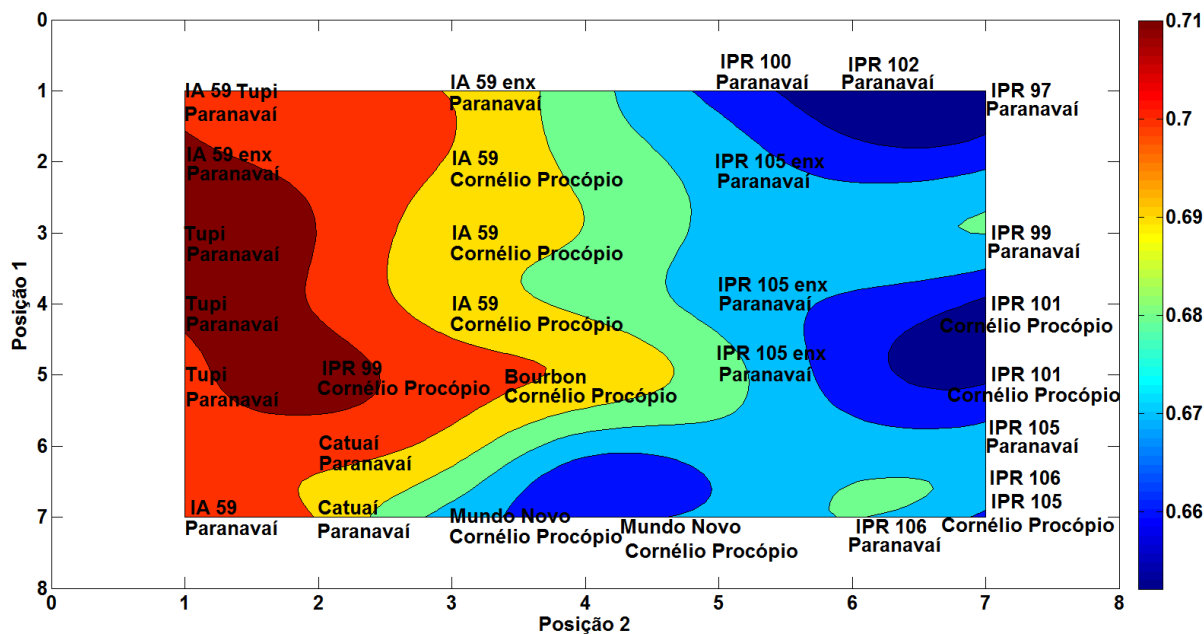
O simplex sequencial demonstrou ser uma metodologia eficiente para se determinar o melhor conjunto de parâmetros de rede, maximizando o desempenho e minimizando o tamanho das redes de uma maneira heurística.

Portanto, a combinação de redes neurais artificiais e a espectroscopia de infravermelho com transformada de Fourier (FTIR) mostrou ser uma ferramenta eficiente e rápida para o reconhecimento e classificação de amostras de café arábica verde, podendo ser aplicada também em outras tarefas como a comparação entre produtos, identificação de adulterações, entre outros.

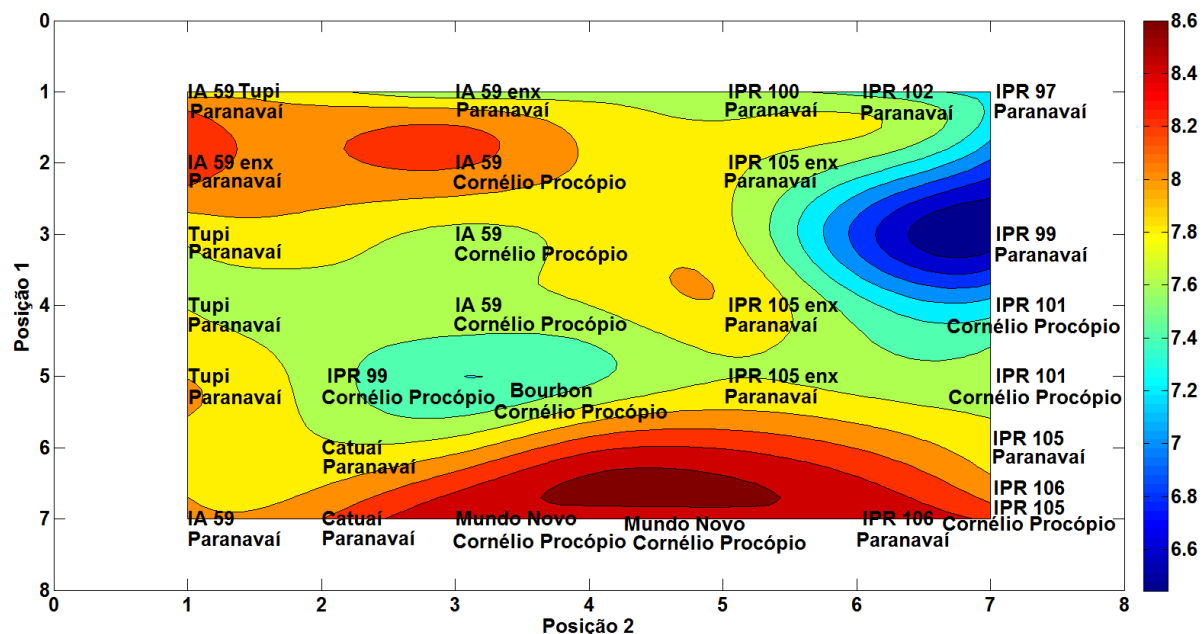
## **APÊNDICE**

## 4.1 APÊNDICE A

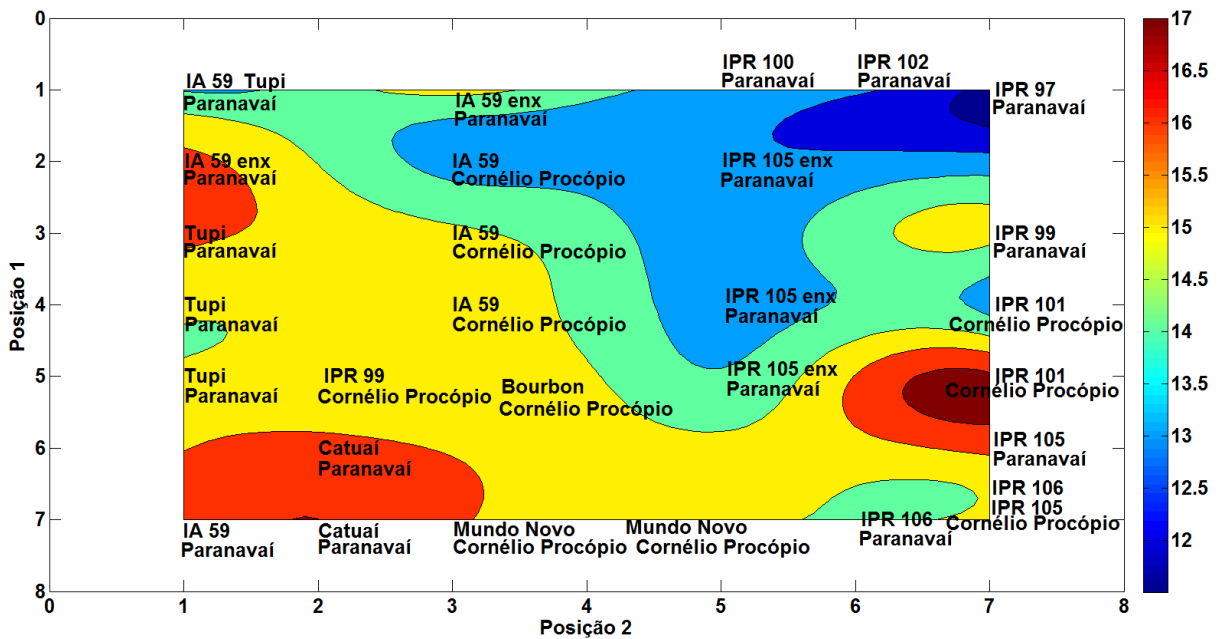
### 4.1.1 MAPAS DE PESO PARA AS DEMAIS VARIÁVEIS ESTUDADAS E MAPAS TOPOLÓGICOS DA REGIÃO E GENÓTIPOS DE CAFÉ ARÁBICA



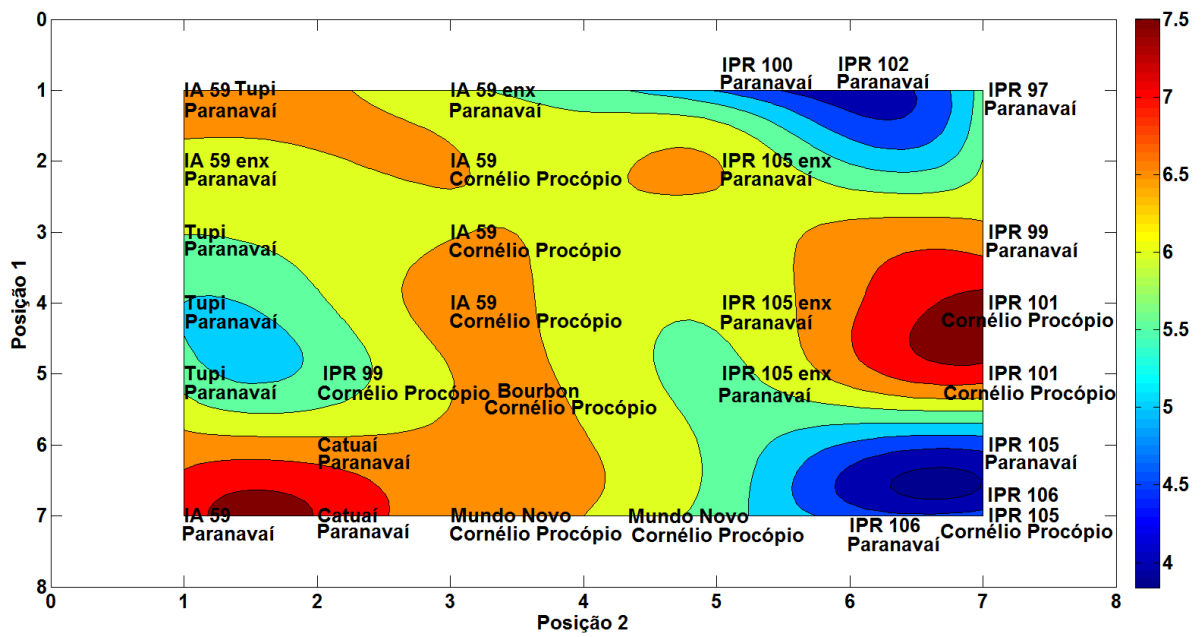
Mapas de pesos para a variável densidade e os mapas topológicos da região e genótipos de café arábica. Na escala de cores as unidades são g.mL<sup>-1</sup>.



Mapas de pesos para a variável açúcares totais e os mapas topológicos da região e genótipos de café arábica. Na escala de cores as unidades são g/100g.



Mapas de pesos para a variável lipídios e os mapas topológicos da região e genótipos de café arábica. Na escala de cores as unidades são g/100g.



Mapas de pesos para a variável taninos e os mapas topológicos da região e genótipos de café arábica. Na escala de cores as unidades são g/100g.

## 4.2 APÊNDICE B

### 4.2.1. OTIMIZAÇÃO SIMPLEX

O princípio do simplex básico foi deslocar uma figura regular, por exemplo um triângulo equilátero sobre uma superfície, quando duas variáveis estavam sendo consideradas (SPLENDLEY *et al.*, 1962).

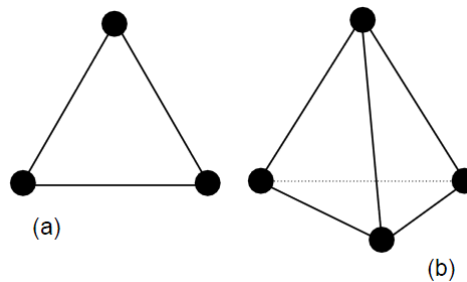


Figura 1 - Interpretação geométrica do simplex para (a) duas variáveis (b) três variáveis.

O método do simplex modificado é o método mais utilizado para resolver o problema de otimização sem restrições, devido a modificação do simplex básico, alterado em seu tamanho e forma, para adaptar-se melhor à superfície de resposta (NELDER e MEAD, 1965). Posteriormente foi desenvolvido um método onde inicia-se a otimização (maximização ou minimização) atribuindo-se limites inferiores ( $L_i$ ) e superiores ( $U_i$ ) para cada fator que será controlado (PIRES *et al.*, 1998). Foi apresentada, uma técnica que permite calcular as demais coordenadas do simplex inicial segundo as expressões (6) e (7), onde  $n$  é o número de variáveis (contínuas ou qualitativas) e  $t$  a distância entre dois vértices (geralmente tomada como 1) (SPLENDLEY *et al.*, 1962).

$$p = \frac{t}{n\sqrt{2}} (\sqrt{n+1} + n - 1) \quad (6)$$

$$q = \frac{t}{n\sqrt{2}}(\sqrt{n+1} - 1) \quad (7)$$

Foram adicionados ao simplex, denominado supermodificado, os limites inferiores e superiores dos fatores utilizados nas expressões (8) e (9) (NAKAI *et al.*, 1984).

$$m_1 = L_i + p(U_i - L_i) \quad (8)$$

$$m_2 = L_i + q(U_i - L_i) \quad (9)$$

A partir da geometria analítica pode-se determinar que as coordenadas dos vértices de um simplex regular são representadas pela matriz  $\mathbf{M}$ , na qual as colunas representam os componentes dos vértices, numerados de  $\mathbf{1}$  até  $\mathbf{n+1}$  e as linhas representam as coordenadas,  $\mathbf{i} = \mathbf{1}$  até  $\mathbf{n}$  (HIMMELBLAU, 1972).

$$\mathbf{M} = \begin{bmatrix} L_i & m_1 & m_2 & m_2 \\ L_i & m_2 & m_1 & m_2 \\ \dots & \dots & \dots & \dots \\ L_i & m_2 & m_2 & m_1 \end{bmatrix} \text{ Matriz } n \times n+1$$

O algoritmo simplex é o método mais utilizado para resolver o problema de otimização sem restrições, como por exemplo, encontrar o mínimo de uma função. Através das respostas obtidas em cada interação, os vértices do simplex são ordenados de acordo com seus valores em  $\mathbf{B}$  (melhor),  $\mathbf{N}$  (intermediários) e  $\mathbf{W}$  (pior) (GAO e HAN, 2012).

$$\mathbf{B} \leq \mathbf{N} \leq \dots \leq \mathbf{W}$$

O novo simplex será determinado rejeitando-se o vértice correspondente à pior resposta e substituindo-se esse vértice por uma operação. O algoritmo utiliza cinco operações possíveis: a reflexão, a expansão, a contração externa e interna e o encolhimento sendo cada uma delas associadas a um parâmetro escalar:  $\alpha$  (reflexão),  $\beta$  (expansão),  $\gamma$  (contração externa e interna) e  $\delta$  (encolhimento). Os valores destes parâmetros devem satisfazer  $\alpha > 0$ ,  $\beta > 1$ ,  $0 < \gamma < 1$ , e  $0 < \delta < 1$  (GAO e HAN, 2012). No simplex estes parâmetros foram calculados adaptativamente as  $n$  dimensões do problema de acordo com as expressões (10, 11, 12 e 13),

$$\alpha = 1 \quad (10)$$

$$\beta = 1 + \frac{2}{n} \quad (11)$$

$$\gamma = 0.75 - \frac{1}{2n} \quad (12)$$

$$\delta = 1 - \frac{1}{n} \quad (13)$$

A direção do movimento de reflexão (Figura 3) é determinada pelo centroide ( $\bar{P}$ ) formado pelos pontos remanescentes (NETO *et al.*, 2010).



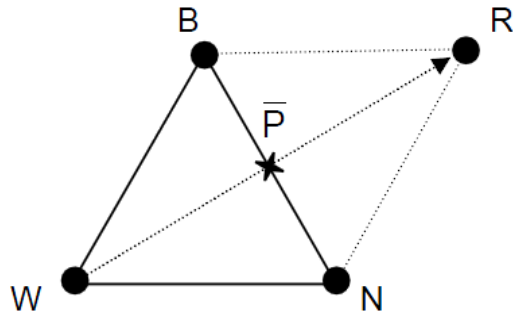


Figura 3 - Movimento de reflexão em um simplex para duas variáveis.

Os passos realizados pelo simplex para a realização da otimização estão citados abaixo (NETO *et al.*, 2010; GAO e HAN, 2012).

1- Ordenar: Avaliar a função (rede neural artificial) nos  $n+1$  vértices calculados pelo simplex inicial e classificar os vértices em **B** (melhor), **N** (intermediários) e **W** (pior).

2- Reflexão: Calcular o ponto de reflexão ( $R$ ) a partir de,

$$R = \bar{P} + \alpha(\bar{P} - W) \quad (14)$$

Avaliar  $R$  Se  $B \leq R \leq N$ , substitua  $W$  por  $R$ .

3- Expansão: Se  $R < B$ , em seguida, calcular o ponto de expansão ( $S$ ) a partir de,

$$S = \bar{P} + \beta(R - W) \quad (15)$$

e avaliar  $S$  . Se  $S < R$ , substitua  $W$  por  $S$ , caso contrário substitua  $W$  por  $R$ .

4- Contração externa: Se  $N \leq R < W$ , calcular o ponto de contração externa ( $U$ ) a partir de,

$$U = \bar{P} + \gamma(R - \bar{P}) \quad (16)$$

e avaliar  $U$ . Se  $U \leq R$ , substitua  $W$  por  $U$ , caso contrário, prosseguir para o passo 6.

5- Contração interna: Se  $R \geq W$ , calcular o ponto de contração interna ( $T$ ) a partir de,

$$T = \bar{P} - \gamma(R - \bar{P}) \quad (17)$$

e avaliar  $T$ . Se  $T < W$ , substituir  $W$  com  $T$ , caso contrário, prosseguir para o passo 6.

6- Encolhimento: Para  $2 \leq i \leq W$ , definir,

$$E_i = B + \delta(R_i - B) \quad (18)$$

A otimização segue até que o valor do erro quadrado médio varie apenas dentro da tolerância estabelecida que foi de 0,001 ou pela unidade gráfica que tem por finalidade auxiliar na visualização da otimização que está representada como um abrandamento na variação das respostas e variáveis independentes.

#### 4.2.2. REFERÊNCIAS

GAO, F.; HAN, L. Implementing the Nelder-Mead simplex algorithm with adaptive parameters. **Comput Optim Appl**, v. 51, p. 259–277, 2012.

HIMMELBLAU, D. M. *Applied Nonlinear Programming*. McGraw-Hill Book Company, New York, 1972. 498p.

NAKAI, S.; KOIDE, R.; EUGESTER, K. A. A new mapping super-simplex optimization for food products and process development. *J. Food Sci.*, v. 49, p. 1143-1148, 1984.

NELDER, J. A.; MEAD, R. A. simplex method for function minimization. *Computer J*, v. 7, p. 308-312, 1965.

NETO, B. D. B.; SCARMINIO, I. S.; BRUNS, R. E. **Como fazer experimentos. Pesquisa e desenvolvimento na ciência e na indústria**. 4ªed. Porto Alegre: Bookman, 2010. 413p.

PIRES, M. V. P.; BORSATO, D.; SILVA, R. S. F. Desenvolvimento de aplicativo para microinformática visando a otimização de sistemas alimentares, 16, 1998, Rio de Janeiro. Anais ... SBCTA. In: **Congresso Brasileiro de Ciência e Tecnologia de Alimentos**, v. 3, p. 1565-1568, 1998.

SPLENDLEY, W.; HIMSWORTH, F. R.; HEXT, G. R. **Sequential application of simplex designs in optimization and evolutionary operation**. *Technometrics*, v. 4, p. 441-461, 1962.

## 4.3 APÊNDICE C

### 4.3.1. ALGORITMO K-MÉDIAS (K-MEANS)

Para o processo de aprendizagem auto-organizada das redes de funções de base radial foi utilizado um método de agrupamento que separa o conjunto fornecido de dados em subgrupos, cada um dos quais sendo tão homogêneo quanto possível. Para tanto, foi empregado o algoritmo de K-médias (*K-means*), que coloca os centros das funções de base radial apenas em regiões densamente povoadas do espaço multidimensional de entrada. Considere que  $m_1$  represente o número de funções de base radial e que  $\{t_k(n)\}_{k=1}^{m_1}$  represente os centros das funções de base radial na interação  $n$  do algoritmo. Então, o algoritmo de agrupamento K-médias atua como segue (Haykin, 2001):

1. Inicialização: Escolha de valores aleatórios ou uniformemente distribuídos para os centros iniciais  $t_k(0)$ ; a única restrição é que estes valores iniciais sejam diferentes.
2. Amostragem: Retira-se um vetor  $\mathbf{x}$  do espaço de entrada com certa probabilidade. O vetor  $\mathbf{x}$  é apresentado à entrada do algoritmo na interação  $n$ .
3. Casamento de similaridade: considere que  $k(\mathbf{x})$  represente o índice do centro com o melhor casamento (vencedor) com o vetor de entrada  $\mathbf{x}$ . Encontre  $k(\mathbf{x})$  na interação  $n$  usando a distância euclidiana mínima (equação 4), ou outra métrica de distância como critério:

$$k(\mathbf{x}) = \arg \min_k \|\mathbf{x}(n) - t_k(n)\|, k = 1, 2, \dots, m_1 \quad (4)$$

onde  $t_k(n)$  é o centro da  $k$ -ésima função de base radial na interação  $n$ .

4. Atualização: Ajuste os centros das funções de base radial, usando a regra da atualização (equação 5):

$$t_k(n+1) = \begin{cases} t_k(n) + \eta[x(n) - t_k(n)], & k = k(x) \\ t_k(n), & \text{caso contrário} \end{cases} \quad (5)$$

onde  $\eta$  é um parâmetro da taxa de aprendizagem no intervalo  $0 < \eta < 1$ .

5. Continuação: Incrementa-se  $n$  de 1 e é retornado ao passo 2. O procedimento é continuado até que não sejam mais observadas modificações nos centros  $t_k$ .

Tendo identificado os centros individuais das funções de base radial utilizando o algoritmo de agrupamento K-médias, o próximo estágio do processo de aprendizagem das redes de funções de base radial foi avaliar a distância máxima entre as bases (dispersão) e estimar os pesos da camada de saída.

#### 4.3.2. REFERÊNCIAS

HAYKIN, S. **Redes Neurais: Princípios e Prática**. 2ª edição. Porto Alegre: Bookman, 2001.

## **ANEXOS**

## 5.1 ANEXO A

### 5.1.1 NORMAS PARA PUBLICAÇÃO NO PERIÓDICO JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

#### IMPORTANT MANUSCRIPT SUBMISSION REQUIREMENTS

Manuscripts and revised manuscripts must be submitted via the ACS Paragon Plus Web site (<http://paragonplus.acs.org/login>). E-mailed submissions and hardcopy submissions will not be processed. An overview of and complete instructions for the Web submission process are available at the ACS Paragon Plus Web site.

The Paragon Plus Web site employs state-of-the-art security mechanisms to ensure that all electronically submitted papers are secure. These same security mechanisms are also utilized throughout the peer-review process, permitting access only to editors and reviewers who are assigned to a particular paper.

When submitting, please be aware of the following requirements.

- All manuscripts must be accompanied by a **cover letter** that includes an **explanation of the manuscript's significance**, including its originality, its contribution to new knowledge in the field, and its relevance to research in agricultural and food chemistry.
- The system requires authors to supply the names, e-mail addresses and affiliations of at least four recommended reviewers. The recommended reviewers should not be anyone who is or, in the previous two years, has been a former adviser/advisee, colleague in the same institution, research collaborator, and/or coauthor of papers and patents or in any other way has a conflict of interest.
- The author's preference for manuscript category is indicated during the submission process. However, the final decision on the category under which the manuscript will be listed lies with the Editor.
- The manuscript abstract and text must appear in a single, double-spaced column; lines in the abstract and text must be numbered consecutively from beginning to end in a separate column at the left.
- All coauthors listed on the title page of the manuscript must be entered into the Paragon Plus System at step 2 in the manuscript submission process. Only one corresponding author is allowed for each manuscript in Paragon Plus. Additional corresponding authors may be designated on the manuscript title page.
- Authors selecting the *Just Accepted* manuscript option when submitting should be sure that the form of author and coauthor names as entered into the Paragon Plus System matches the form on the manuscript title page.
- References must be numbered in the order in which they appear in the text.
- All of the text (including the title page, abstract, all sections of the body of the paper, figure captions, scheme or chart titles and footnotes, and references) and tabular material should be in one file, with the complete text first followed by the tabular material.
- A separate conclusion section is not to be used. Conclusions should be incorporated into the results and discussion section.

Complete instructions for manuscript preparation and a Journal Publishing Agreement form are available at the *Journal's* Web site. Please conform to these instructions when submitting manuscripts.

Authors whose manuscripts are published in the *Journal* will be expected to review manuscripts submitted by other researchers from time to time.

#### JOURNAL SCOPE

The *Journal of Agricultural and Food Chemistry* publishes high-quality, cutting edge original research representing complete studies and research advances dealing with the chemistry and biochemistry of agriculture and food. The *Journal* also encourages papers with chemistry and/or biochemistry as a major component combined with biological/sensory/nutritional/toxicological evaluation related to

agriculture and/or food. As a general rule, manuscripts dealing with herbal remedies or those testing specific compounds in cell-based assays related to disease states (e.g., “anticancer” activity) will no longer be considered within the scope of the *Journal* and should be submitted elsewhere. Manuscripts describing properties of extracts, without detailing the chemical composition of the extracts responsible for the described properties, will generally not be accepted for review.

The *Journal* is organized into the following sections:

Agricultural and Environmental Chemistry  
Analysis and Chemosensory Perception of Flavor  
Analytical Methods  
Bioactive Constituents and Functions  
Biofuels and Biobased Products  
Chemical Aspects of Biotechnology and Molecular Biology  
Food and Beverage Chemistry/Biochemistry  
Food Safety and Toxicology

## MANUSCRIPT TYPES

**Research articles** must report **original research that is expected to have a definable impact on the advancement of science and technology, incorporating a significant component of innovative chemistry**. Originality will be documented by novel experimental results, theoretical treatments, interpretations of data, and absence of prior publications on the same/similar topics. Fragmentation of work into an incremental series of manuscripts is not acceptable.

**Expedited Handling.** There is no separate Rapid Communications, Notes, or Letters section. However, manuscripts describing results deemed to be highly important and urgent in a field of research will be considered for expedited processing and review. **Only manuscripts reporting complete research, as opposed to preliminary results, will be considered.** A request for expedited handling, along with justification for the request, must be included in the cover letter accompanying the manuscript.

**Review articles** will be considered that summarize information in a field in which the literature is scattered and/or treat published data or other information so as to provide a new approach or stimulate further research. Authors considering the preparation of a review **should submit a synopsis to the Editor** before submission to establish whether the manuscript will meet these guidelines.

**Perspectives**, which explore needs and opportunities in agricultural and food chemistry in a less technical format than a review article, will be considered. Authors should **contact the Editor** to outline the area to be covered before submitting a Perspectives manuscript. For an example, see *J. Agric. Food Chem.* **2008**, *56*, 7587–7592.

**Comments** related to published papers will be considered from readers if the correspondence is **received within six months of the date of publication of the original paper**; the authors of the original paper will be given the **opportunity to reply** to such comments within two months, if they so desire. Both comments and replies should not exceed 1000 words each, including citations, and will be published consecutively in the same issue of the *Journal* after peer review. For examples, see *J. Agric. Food Chem.* **2007**, *55*, 7213–7214 and *J. Agric. Food Chem.* **2007**, *55*, 7215–7216.

**Symposia or Topical Collections.** The Editor will consider publication of a series of manuscripts reporting or synthesizing original research that are presented in a symposium or otherwise clustered around a single topic. Prospective organizers should **contact the Editor well in advance** to determine whether the subject matter conforms to the *Journal*'s goals, criteria, and available space and to obtain specific instructions for submission of the manuscripts. For an example, see *J. Agric. Food Chem.* **2008**, *56*, 5983–6184. Each manuscript will be subject to the normal peer-review process.

**Additions/Corrections.** Corresponding authors wishing to submit a correction to a paper already published in print should submit the item via the Paragon Plus Web site. In your cover letter, include the manuscript number of the paper to be corrected. In the correction document, include the full title of the original publication, all author names, the volume and page numbers of the print publication, the original manuscript number, and a brief description of the correction(s) needed. If a figure is to be



corrected, please include the figure in the correction document. Please note that the Editor has final approval as to whether an addition/correction will be published.

### **ETHICS, CONFLICT OF INTEREST**

Authors and coauthors are responsible for the integrity of their manuscripts. The Editor may impose a two year submission moratorium on authors and coauthors that are found to be in violation of the ethical guidelines.

Authors and coauthors should familiarize themselves by reading the entire *Ethical Guidelines to Publication of Chemical Research*, which is available at the *Journal's* Web site.

A statement describing any financial conflicts of interest or lack thereof is published with each manuscript. During the submission process, the corresponding author must provide this statement on behalf of all authors of the manuscript. The statement should describe all potential sources of bias, including affiliations, funding sources, and financial or management relationships, that may constitute conflicts of interest (please see <http://pubs.acs.org/ethics>, ACS Ethical Guidelines). The statement will be published in the final paper. If no conflict of interest is declared, the following statement will be published in the paper: "The authors declare no competing financial interest."

### **AUTHOR LIST**

During manuscript submission, the submitting author must provide contact information (full name, email address, institutional affiliation and mailing address) for all of the co-authors. Because all of the author names are automatically imported into the electronic Journal Publishing Agreement, the names must be entered into ACS Paragon Plus in the same sequence as they appear on the first page of the manuscript. (Note that co-authors are not required to register in ACS Paragon Plus.) The author who submits the manuscript for publication accepts the responsibility of notifying all co-authors that the manuscript is being submitted. Deletion of an author after the manuscript has been submitted requires a confirming letter to the Editor-in-Chief from the author whose name is being deleted. For more information on ethical responsibilities of authors, see the Ethical Guidelines to Publication of Chemical Research.

### **EDITORIAL PEER REVIEW PROCESS**

Peer review is used to help ensure the **highest possible quality** in published manuscripts. For a discussion of this, see "The Importance of Peer Review" by H. L. Wheeler and W. B. Wheeler, *J. Agric. Food Chem.* (Editorial) **2006**, *54*, 8983–8983. Scientists with expertise in the subject matter being treated will evaluate the manuscript for validity of the experimental design and results, originality, significance, and appropriateness to the *Journal*. **The Editors may exercise their prerogative to decline a manuscript without peer review if that paper is judged to be outside the scope of the *Journal* (lacks significant chemistry/biochemistry), poorly written or formatted, fragmentary and marginally incremental, or lacking in significance.** Manuscripts describing properties of extracts, without detailing the chemical composition of the extracts responsible for the described properties, will generally not be accepted for review.

All manuscripts submitted are reviewed and handled by the Editor-in-Chief or assigned to one of the Associate Editors. The Associate Editor and Editorial Assistant are then responsible for the assigned manuscripts, including evaluating the content and format of the paper, selecting reviewers, monitoring the progress of the review process, evaluating the comments of reviewers and forwarding them to the authors for their response, communicating ultimate acceptance or rejection to the corresponding author, and carrying out a final check of accepted manuscripts for appropriate format and style. **Typically, three reviewers are selected** per paper on the basis of the subject matter, available expertise, and the Editor's knowledge of the field. Potential reviewers for each paper are identified by various means, including a computerized search of the subject area. Authors must submit the names and addresses (including e-mail addresses) of at least four potential reviewers who do not have conflicts of interest with the authors or manuscript content; however, the Editors are under no obligation to use specific individuals. Reviewers are normally asked to provide their assessments

within two to three weeks. Anonymous copies of the reviews and the Editor's decision regarding the acceptability of the manuscript are sent to the corresponding author. If the reviewers' evaluations of the manuscript disagree, or if reviewer's and Editor's comments are not satisfactorily addressed by the authors, the Editor may reject the manuscript or select additional reviewers. These additional reviews are used by the Editor to assist in reaching the final decision regarding disposition of the manuscript. The obligations of the Editors and Reviewers are outlined in the *Ethical Guidelines*. Aids for reviewers titled "A Guide to a Review" and "Components of a Manuscript to be Considered in a Review" are available at the Reviewer Information Web site (<http://pubs.acs.org/4authors>).

**Just Accepted Manuscripts.** *Just Accepted* manuscripts are peer-reviewed, accepted manuscripts that are published on the ACS Publications Web site prior to technical editing, formatting for publication, and author proofing—usually within 30 minutes to 24 hours of acceptance by the editorial office. During the manuscript submission process, authors can choose to have their manuscript published online as a *Just Accepted* manuscript. Authors choosing this option must ensure that all intellectual property/patent issues are resolved. To ensure rapid delivery of the accepted manuscript to the Web, authors must adhere carefully to all requirements in the journal's Scope, Policy, and Instructions for authors. For further information, please refer to the *Just Accepted* FAQ, at <http://services.acs.org/pubshelp/passthru.cgi?action=kb&item=244>. Note that publishing a manuscript as *Just Accepted* is not a means by which to comply with the NIH Public Access Mandate.

**ASAP Publication.** Accepted manuscripts will be published on the "Articles ASAP" page on the *Journal's* Web site as soon as page proofs are corrected and all author concerns are resolved. Publication on the Web usually occurs within 4 working days of receipt of page proof corrections, and this can be anywhere from 2 to 6 weeks in advance of the cover date of the issue. Manuscripts assigned to a special issue often remain published ASAP for several months. Authors should take this schedule into account when planning intellectual and patent activities related to a manuscript. The date on which an accepted paper is published on the Web is recorded on the Web version of the manuscript and on the first page of the PDF version.

## MANUSCRIPT PREPARATION

**Manuscript Format.** Manuscripts must be prepared using accepted word-processing software, and all parts must be double-spaced. All pages must be numbered consecutively starting with the title page and including tables and figures. **Lines in the abstract and text should be numbered consecutively from beginning to end in a separate column at the left. Do not put line numbers on pages with tables or figures.** A standard font, in a size of 12 points or greater, must be used. The *Journal* requires authors to stay within a **20 typed page limit**, not including references, tables, and figures.

Standard American English usage is required. Authors who are not familiar with standard American English are urged to seek assistance; deficiencies in grammar may be a serious hindrance during the review process.

*The ACS Style Guide* (3rd ed., 2006; ISBN 0-8412-3999-1), available from Oxford University Press, Order Department, 201 Evans Road, Cary, NC 27513, provides a detailed treatment of the fundamentals of manuscript preparation. Refer to a current issue of the *Journal* for general style.

The various sections of the manuscript should be assembled in the following sequence:

Title and authorship (single page)  
 Abstract and keywords (single page)  
 Introduction  
 Materials and Methods  
 Results /Discussion  
 Abbreviations Used  
 Safety  
 Acknowledgment  
 Supporting Information description  
 Literature Cited

Figure captions  
Tables  
Figures  
Graphic for table of contents

### TITLE AND AUTHORSHIP

The title, authorship, and institutional affiliations should be included on a single page.

**Title.** The title should be specific, informative, and concise. Keywords in the title assist in effective literature retrieval. If a plant is referred to in the title or elsewhere in the text by its common or trivial name, it should be identified by its scientific name in parentheses immediately following its first occurrence. This term should also be provided as one of the keywords. If trade names are mentioned, give generic names in parentheses.

**Authorship.** Be consistent in authorship designation on the manuscript and on all correspondence.

**First name, middle initial, and last name** are generally adequate for correct identification, but omit titles. Give the complete mailing address of all institutions where work was conducted and identify the affiliation of each author. If the current address of an author is different, include it in a footnote on the title page. The name of the author to whom inquiries about the paper should be addressed must be marked with an asterisk; provide the telephone and fax numbers and e-mail address of this correspondent.

### ABSTRACT AND KEYWORDS

**Abstract.** Authors' abstracts are used directly for *Chemical Abstracts*. The abstract should be a clear, concise (100–150 words), one-paragraph summary, informative rather than descriptive, giving scope and purpose, experimental approach, significant results, and major conclusions. Write for literature searchers as well as journal readers.

**Keywords.** Provide significant keywords to aid the reader in literature retrieval. The keywords are published immediately before the text, following the abstract.

### INTRODUCTION

Discuss relationships of the study to previously published work, but do not reiterate or attempt to provide a complete literature survey. Use of *Chemical Abstracts/Scifinder* and other appropriate databases is encouraged to ensure that important prior publications or patents are cited and that the manuscript does not duplicate previously published work. **The purpose or reason for the research being reported, and its significance, originality, or contribution to new knowledge in the field, should be clearly and concisely stated.**

Do not include or summarize current findings in this section.

### MATERIALS AND METHODS

Apparatus, reagents, and biological materials used in the study should be incorporated into a general section. List devices of a specialized nature or instruments that may vary in performance, such that the model used may affect the quality of the data obtained (e.g., spectroscopic resolution).

List and describe preparation of special reagents only. Reagents normally found in the laboratory and preparations described in standard handbooks or texts should not be listed.

Specify the source, vendor [city and state (or city and country if non-U.S.)], and availability of special equipment, reagents, kits, etc. Do not include catalog numbers.

Biological materials should be identified by scientific name (genus, species, authority, and family) and cultivar, if appropriate, together with the site from which the samples were obtained. Specimens obtained from a natural habitat should be preserved by deposit of samples in an herbarium for plants or in a culture collection for microorganisms, with a corresponding collection or strain number listed.

Manuscripts describing studies in which live animals or human subjects are used must include a statement that such experiments were performed in compliance with the appropriate laws and institutional guidelines and also name the institutional committee that approved the experiments. Authors are encouraged to note the approval code or number or give the name of the approving office or official. (See Reporting Specific Data: Animal or Human Studies.) Manuscripts reporting data from inhumane treatment of experimental animals will be rejected.

Specific experimental methods should be sufficiently detailed for others to repeat the experiments unequivocally. Omit details of procedures that are common knowledge to those in the field. Brief highlights of published procedures may be included, but details must be left to the Literature Cited, and verbatim repeat of previously published methods, even if done by the authors, will not be permitted unless a quotation from a published work is included, and placed in quotation marks, with the reference to the source included at the end of the quotation. Describe pertinent and critical factors involved in reactions so the method can be reproduced, but avoid excessive description. For information on the reporting of certain types of data see Reporting Specific Data. Describe statistical design and methods in this section.

## RESULTS/DISCUSSION

Results and discussion may be presented in separate sections or combined into a single section, whichever format conveys the results in the most lucid fashion without redundancy. Be complete but concise in discussing findings, comparing results with previous work and proposing explanations for the results observed.

**All data must be accompanied by appropriate statistical analyses, including complete information on sampling, replication, and how the statistical method employed was chosen.**

Avoid comparisons or contrasts that are not pertinent, and avoid speculation unsupported by the data obtained.

A separate summary or conclusion section is not to be used; any concluding statements are to be incorporated under Results and Discussion.

## ABBREVIATIONS AND NOMENCLATURE

Standard abbreviations, without periods, should be used throughout the manuscript.

Refer to *The ACS Style Guide* for the preferred forms of commonly used abbreviations. Specialized abbreviations may be used provided they are placed in parentheses after the word(s) for which they are to substitute at first point of use and are again defined in this section. Avoid trivial names and “code” abbreviations (e.g., NAR for naringenin) unless such codes are in common usage (e.g., MTBE for methyl *tert*-butyl ether).

If trade names are used, define at point of first use. If nomenclature is specialized, include a “Nomenclature” section at the end of the paper, giving definitions and dimensions for all terms. Use SI units insofar as possible. Refer to *The ACS Style Guide* for lists of SI units and a discussion of their use.

Write all equations and formulas clearly and number equations consecutively. Place superscripts and subscripts accurately; avoid superscripts that may be confused with exponents. Identify typed letters and numbers that might be misinterpreted, such as “oh” for zero or “ell” for one. Chemistry numbering requiring primes should be identified as such (i.e., 3,3′-dihydroxy-), not by an apostrophe (e.g., 3,3′-dihydroxy-).

It is the authors' responsibility to provide correct nomenclature. Structures should be included for uncommon chemicals, particularly when the systematic or common name is too complex or unclear to readily denote the structure. Such structures should be included as a figure or table. All nomenclature must be consistent and unambiguous and should conform with current American usage. Insofar as possible, authors should use systematic names similar to those used by Chemical Abstracts Service, the International Union of Pure and Applied Chemistry, and the International Union of Biochemistry and Molecular Biology. *Chemical Abstracts* (CA) nomenclature rules are described in Appendix IV of the *Chemical Abstracts Index Guide*. For CA nomenclature advice, consult the Manager of

Nomenclature Services, Chemical Abstracts Service, P.O. Box 3012, Columbus, OH 43210-0012. A name generation service is available for a fee through CAS Client Services, 2540 Olentangy River Road, P.O. Box 3343, Columbus, OH 43210-0334 [telephone (614) 447-3870; fax (614) 447-3747; e-mail [answers@cas.org](mailto:answers@cas.org)]. In addition, the ACS Web site has links to nomenclature recommendations at <http://chemistry.org>.

## SAFETY

Authors are required to call special attention in their manuscripts to safety considerations such as explosive tendencies, special precautionary handling procedures, and toxicity.

## ACKNOWLEDGMENT

Include essential credits but hold to an absolute minimum. Omit academic and social titles. Meeting presentation data and acknowledgment of financial support of the work should not be included here; give these instead in a note following the Literature Cited. It is the responsibility of the corresponding author to notify individuals named in the Acknowledgment.

## LITERATURE CITED

Consult *The ACS Style Guide* and current issues of the *Journal* for examples of reference format.

Authors should cite all prior published work directly pertinent to the manuscript. However, extensive bibliographies that go beyond a direct connection with the manuscript are discouraged. Prior work can often be covered by citation of a few leading references or of review articles. As a general guideline, authors should attempt to limit the literature cited to approximately 50 or fewer citations.

Authors are responsible for the accuracy of their references. References taken from a review or other secondary source should be checked for accuracy with the primary source.

References should be listed on a separate page and numbered in the order in which they are cited in the text.

References should be cited in the text by superscript numbers, for example, 1,2–5, etc.

Give complete information, using the last name and initials of the author, patentee, or equivalent; do not use “Anonymous”.

Follow *Chemical Abstracts Service Source Index* for abbreviations of journal titles. Because subscribers to the Web edition of the *Journal* are now able to click on the “Chemport” or other tag following each reference to retrieve the corresponding abstract from various Web resources, reference accuracy is critical.

Typical references follow the styles given below.

For journals:

1. Brown, J.; Jones, M.; Green, D. Article title. *J. Agric. Food Chem.* **1980**, 28, 1–4. (Use issue number only if each issue of the periodical begins with page 1.)

For books:

2. Smith, L; Caldwell, A. Chapter title. In *Book Title*, edition no.; Keys, F., Park, G., Eds.; Publisher: City, State (or Country if non-U.S.), Year; Vol. no., pp.

For Web pages:

3. Black, A.; White, B. Page title. URL (<http://etc.>) (most recent access date).

Papers should not depend for their usefulness on unpublished material, and excessive reference to material “in press” is discouraged. Reference to the authors’ own unpublished work is permitted if the subject is of secondary importance to the manuscript in question, but any unpublished results of central importance must be described in sufficient detail within the manuscript. **If pertinent references are “in press” or unpublished for any reason, furnish copies to enable reviewers to evaluate the manuscript. An electronic copy of these materials should be uploaded according to the directions for review-only Supporting Information.**

“In press” references should include the Digital Object Identifier (DOI) assigned by the potential publisher.

## TABLES AND ARTWORK

The tables and graphics (illustrations) should be inserted in the manuscript file after the Literature Cited section. Do not upload tables and graphics which are to be published in the manuscript as Supporting Information files.

**Tables and figures should be carefully designed to maximize presentation and comprehension of the experimental data with superfluous information excluded.** Useful information not directly relevant to the discussion may be included under Supporting Information.

**Tables.** Tables may be created using a word processor's text mode or table format feature. The table format feature is preferred. Ensure each data entry is in its own table cell. If the text mode is used, separate columns with a single tab and use a line feed (enter) at the end of each row.

Tables should be numbered consecutively with Arabic numerals and should be grouped after the Literature Cited section. Footnotes in tables should be given letter designations and be cited in the table by italic superscript letters. The sequence of letters should proceed by row rather than by column. Each table should be provided with a descriptive heading, which, together with the individual column headings, should make the table, as nearly as possible, self-explanatory. In setting up tabulations, authors are requested to keep in mind the type area of the journal page (17.8 × 25.4 cm), and the column width (8.5 cm), and to make tables conform to the limitations of these dimensions.

**Arrangements that leave many columns partially filled or that contain much blank space should be avoided.** Conversely, arrangements that include >20 columns should be broken into two tables if possible. If *significance of values* is to be indicated, use a lower case letter, on line, one space after the value.

**Figures and Artwork.** Insert the illustrations into the word-processing file following the Literature Cited. Artwork should be sequentially numbered using Arabic numbers. Schemes and charts may have titles and footnotes; figures should have captions.

For bar charts, bars with hatching patterns generally reproduce well. Bars that range in shading from light to dark gray to black can usually be reproduced successfully, although we do not recommend any more than two shades of gray. A legend needs to be included within the figure itself rather than the patterns or shades included in the caption.

For manuscripts containing gel patterns, use of a high-resolution digital scanner is recommended. Only high-quality digital reproductions will allow reviewers to correctly verify the experimental results. For an example of gel patterns see *J. Agric. Food Chem.* **2004**, *52*, 5717–5723, Figures 2 and 3.

Only readable and accurately represented images are acceptable; the **Editors reserve the option to reject images that do not satisfactorily support points made in the manuscript or that are not of satisfactory quality for publication.**

The quality of the illustrations published in the *Journal* largely depends on the quality of the originals provided. Figures cannot be modified or enhanced by the journal production staff. Contrast is important. Each figure or photograph should be properly labeled.

Illustrations must fit a one- or two-column format on the journal page. **For efficient use of journal space, single-column illustrations are preferred.**

For best results, submit illustrations in the actual size at which they should appear in the journal. Illustrations that do not need to be reduced to fit a single or double column will yield the best quality. Lettering should be no smaller than 4.5 points. (Helvetica or Arial type works well for lettering.) Lines should be no thinner than 0.5 point. Lettering and lines should be of uniform density. Avoid the use of very large and very small lettering within the same figure.

If artwork that must be reduced will be submitted, use larger lettering and thicker lines so that, when reduced, the artwork meets the above-mentioned parameters.

Avoid using complex textures and shading to achieve a three-dimensional effect. To show a pattern, choose a simple crosshatch design.

Color illustrations should be submitted only if they are essential for clarity of communication.

Reproduction of color illustrations will be provided at no cost to the author. Do not submit color prints to be printed in black and white.

**Structural Formulas. Structural formulas should be included for all new chemicals and for existing chemicals for which chemical nomenclature and/or trivial names do not convey the structure adequately.** Structural formulas are valuable in expressing concisely the precise nature of the compounds under discussion and revealing the essence of the subject to readers unfamiliar with the topic, without their necessary recourse to reference materials. The use of chemical names without accompanying structures may cause readers to overlook the significance of the paper.

Structures should be produced with the use of a drawing program such as ChemDraw. Structure drawing preferences (preset in the ACS Stylesheet in ChemDraw) are as follows: as drawing settings select...

chain angle	120°
bond spacing	18% of width
fixed length	14.4 points (0.508 cm, 0.2 in.)
bold width	2.0 points (0.071 cm, 0.0278 in.)
line width	0.6 point (0.021 cm, 0.0084 in.)
margin width	1.6 points (0.056 cm, 0.0222 in.)
hash spacing	2.5 points (0.088 cm, 0.0347 in.)
as text settings select...	
font	Arial or Helvetica
size	10 points
under preferences choose...	
units	points
tolerances	3 pixels
under page setup choose...	
paper	US Letter
scale	100%

Using the ChemDraw ruler or appropriate margin settings, create structure blocks, schemes, and equations having maximum widths of 11.3 cm (one-column format) or 23.6 cm (two-column format). Note: if the foregoing preferences are selected as cm values, the ChemDraw ruler is calibrated in cm. Also note that a standard sheet of paper is only 21.6 cm wide, so all graphics submitted in two-column format must be prepared and printed in landscape mode.

Use boldface type for compound numbers but not for atom labels or captions.

Authors using other drawing packages should, as far as possible, modify their program's parameters to reflect the above guidelines.

## TABLE OF CONTENTS GRAPHICS

Authors of research articles, perspectives, and reviews are required to include a suitable graphic for publication in the table of contents (TOC) in the Web edition of the *Journal*. Submission of this graphic is mandatory. This graphic should capture the reader's attention and, in conjunction with the manuscript's title, give the reader a quick visual impression of the type of chemistry described. Structures should be constructed as specified under Structural Formulas above. The TOC graphic may be up to 4.7 in. (12.0 cm) wide and 1.8 in. (4.6 cm) tall. (See detailed instructions at the Paragon Plus Web site.) Text should be limited to labels for compounds, reaction arrows, and figures. The use of color to enhance the scientific value is encouraged. The TOC graphic should be inserted on a separate page at the end of the manuscript file.

## SUPPORTING INFORMATION

Extensive tables, graphs, spectra, calculations, and other material beyond a modest content in the published paper may be included in the Web edition of the *Journal*. These will **not** be part of the published article but can be accessed separately on the Web by readers.

Supporting Information must be submitted at the same time as the manuscript and uploaded separately to the ACS Paragon Plus environment. A list of acceptable file types is available on the Web. All Supporting Information files of the same type should be prepared as a single file (rather than

submitting a series of files containing individual images or structures). For example, all Supporting Information available as PDF files should be contained in one PDF file.

The material should be described in a paragraph inserted between the Acknowledgment and the Literature Cited sections, using the following format: "Supporting Information Available: Description. This material is available free of charge via the Internet at <http://pubs.acs.org>."

Components of the Supporting Information should be clearly labeled.

**DO NOT UPLOAD FIGURES AND TABLES THAT ARE TO BE PUBLISHED IN THE ARTICLE INTO THE SUPPORTING INFORMATION FILE.** Figures and tables that will appear in the published article are to be inserted in the manuscript directly after the Literature Cited section.

### **CURRENTLY ACCEPTABLE WORD-PROCESSING PACKAGES**

Refer to the Paragon Plus environment Web site for acceptable software packages.

LaTeX users should follow the guidelines given on the Web.

### **REVISIONS AND RESUBMISSIONS**

#### **For all revisions:**

- Clearly identify the manuscript as a revision; reference the manuscript number.
- Include an itemized list of changes, with a response to each comment made by the Editor and by each reviewer.
- Be aware that the manuscript may be sent for additional review, to the same or additional reviewers, at the discretion of the Editor.
- Please upload the signed Journal Publishing Agreement or fax it to the assigned Editor.

#### **For all resubmissions:**

- Clearly identify all resubmissions; reference the previous manuscript number.
- Include an itemized list of changes, including a response to each comment made by the Editor and by each reviewer.
- Please upload the signed Journal Publishing Agreement or fax it to the assigned Editor.

### **JOURNAL PUBLISHING AGREEMENT**

A properly completed and signed Journal Publishing Agreement (JPA) must be submitted for each manuscript. ACS Paragon Plus provides an electronic version of the JPA that will be available on the **My Authoring Activity** tab of the Corresponding Author's Home page once the manuscript has been assigned to an Editor. A PDF version of the Agreement is also available, but **authors are strongly encouraged to use the electronic JPA**. If the PDF version is used, **all pages of the signed PDF JPA must be submitted**. If the corresponding author cannot or should not complete either the electronic or PDF version for any reason, another author should complete and sign the PDF version of the form. Forms and complete instructions are available at

<http://pubs.acs.org/page/copyright/journals/index.html>. **For questions about the form or about signing the form, contact the ACS Copyright Office at (202) 872-4368 or -4367.**

**Note:** Authors who are not U.S. Government employees or bona fide agents should sign Part A of the form only. If ALL of the authors were employees or bona fide agents of the U.S. Government when the paper was prepared, the work is a work of the U.S. Government and only Part B, "U.S. Government Employees", should be signed if BOTH of the following circumstances apply:

- ALL authors are or were bona fide officers or employees of the U.S. Government when the paper was prepared.
- The work is a work of the U.S. Government, prepared by an officer/employee of the U.S. Government as part of official duties.

**If the work was prepared under a U.S. Government contract or is coauthored by a non-U.S. Government employee, the work is not a work of the U.S. Government; DO NOT SIGN PART B.**



Sign only Part A of the form. Call the ACS Copyright Office at the above telephone number for assistance.

## PROOFS AND REPRINTS

**Proofs.** The corresponding author of an accepted manuscript will receive e-mail notification and complete instructions when page proofs are available for review via a secure Web site. It is the responsibility of the corresponding author to ascertain that all coauthors agree with the corrections before the corrections are returned. Corrections should be designated by galley proof line number. Galley proof corrections should be returned within 48 h of receipt to ensure timely publication of the manuscript. Routine rephrasing of sentences or additions are not permitted at the page proof stage. Alterations should be restricted to serious changes in interpretation or corrections of data. Extensive or important changes on page proofs, including changes to the title or list of authors, are subject to Editorial review.

**ACS Policies for E-prints and Reprints.** Under the ACS Articles on Request policy, the Society will provide (free of charge) to all contributing authors a unique URL within the ACS Web site that they may e-mail to colleagues or post on external Web sites. These author-directed links are designed to facilitate distribution of an author's published work to interested colleagues in lieu of direct distribution of the PDF file by the author. The ACS Articles on Request policy allows 50 downloads within the first year after Web publication and unlimited access via the same author-directed links 12 months after Web publication.

The ACS AuthorChoice option establishes a fee-based mechanism for authors or their research funding agencies to sponsor the open availability of their articles on the Web at the time of online publication. Under this policy, the ACS as copyright holder will enable unrestricted Web access to a contributing author's publication from the Society's Web site in exchange for a fixed payment from the sponsoring author. ACS AuthorChoice will also enable participating authors to post electronic copies of published articles on their own personal Web sites and institutional repositories for noncommercial scholarly purposes and allow immediate open access to an article as soon as it is published on the ACS Web site.

When authors are sent the proof of their paper, they will receive a link to a Web site where they may order author reprints. They may also call Cierant Corporation, (866) 305-0111, from 9 a.m. to 5 p.m. EST. Reprints will be shipped within two weeks after the issue publication date. Neither the Editors nor the Washington ACS Office keeps a supply of reprints; requests for single copies of papers should be addressed to the corresponding author of the paper concerned.

## REPORTING SPECIFIC DATA

**Bioactivity.** Manuscripts reporting on bioactivity of plant-derived or other extracts must also include identification and characterization of individual chemicals responsible for the observed bioactivity.

**Gas Chromatographic Methods.** For manuscripts in which gas chromatographic methods are used, see "Reporting of Gas Chromatographic Methods", by Morton Beroza and Irwin Hornstein [*J. Agric. Food Chem.* **1973**, *21*, 7A (located at the back of the January 1973 issue or as a link from the *Journal's* Author Information page)]. Consult recent issues for examples of GC, LC, and other instrument parameter descriptions.

**Spectroscopic Data.** This is a guide only; in certain cases different methods of data presentation may be more suitable. Authors are encouraged to consult examples of data presentation published in recent issues of the *Journal* for appropriate style and format. **Complete infrared, NMR, mass, or other spectra will be published only if novel or necessary to substantiate points made under the Results or Discussion sections.** Such presentations take up valuable space, and essentially the same information can frequently be put into a much more compact form by simply listing the position and intensity of the maxima. It is usually not necessary to list all of the maxima in the spectra to provide an adequate description. Report the type of instrument used (e.g., in mass spectrometry, whether magnetic, quadrupole, etc.) and also the type of cell, the solvent (if any), and the state of the sample (whether liquid, gas, solution, etc.).

**Mass Spectra.** List the molecular ion and about 10 of the major ions with their intensities in parentheses, or more preferably use the method outlined by H. S. Hertz, R. A. Hites, and K. Biemann (*Anal. Chem.* **1971**, *43*, 681–691). This method involves dividing the spectrum into consecutive regions of 14 mass units starting at  $m/z$  6 (i.e., 6–19, 20–33, 34–47, 48–61, etc.). The two most intense ions in each region are then listed. Intensities, relative to the most intense ion, the intensity of which is taken as 100, are shown in parentheses immediately following the  $m/z$  value; for example: hexanal, mass spectrum found (70 eV, two most intense ions each 14 mass units above  $m/z$  34): 43 (86), 44 (100), 56 (86), 57 (65), 71 (28), 72 (33), 82 (18), 85 (5), 97 (2), 100 (2). If the molecular ion does not appear in this presentation, the author should indicate it separately.

**Nuclear Magnetic Resonance (1H NMR or 13C NMR) Spectra.** The frequency used, the solvent, and also temperature (if other than ambient) are first specified. The type of unit used ( $\delta$  or  $\tau$ ) is then stated, followed by the position of the center of gravity of the sharp line, broad line, or spin–spin multiplet in these units. This is then followed by information in parentheses which (1) describes the type of splitting, that is, singlet as s, doublet as d, triplet as t, quadruplet as qd, multiplet as m; (2) gives the value of the number of protons the area represents; (3) gives the coupling constant  $J$ ; and (4) gives the part of the molecule connected with the particular absorption with the protons involved underlined.

An example would be 1H NMR for ethanol (60 MHz, CCl<sub>4</sub>):  $\delta$  1.22 (t, 3,  $J$  = 7 Hz, CH<sub>2</sub>CH<sub>3</sub>), 2.58 (s, 1, OH), 3.70 (qd, 2,  $J$  = 7 Hz, OCH<sub>2</sub>CH<sub>3</sub>).

**Other Spectra.** In general, list position and intensity of the maxima. In some cases it may be desirable to list points of inflection.

A brief explanation should be given for any abbreviations not in common use.

Examples:

- Reporting liquid chromatography (HPLC) and HPLC/MS: “Analysis of Polyphenolic Antioxidants from the Fruits of Three *Pouteria* Species by Selected Ion Monitoring Liquid Chromatography–Mass Spectrometry”, by Jun Ma et al. *J. Agric. Food Chem.* **2004**, *52*, 5873–5878.
- Reporting data in detail, including UV shifts and IR spectra: “Characterization of Vegetable Oils: Detailed Compositional Fingerprints Derived from Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry”, by Zhigang Wu et al. *J. Agric. Food Chem.* **2004**, *52*, 5322–5328.

**Novel Compound Characterization.** For a discussion of the *Journal's* expectations for compound characterization, please read “Compound Identification: A *Journal of Agricultural and Food Chemistry* Perspective” by R. J. Molyneux and P. Schieberle. *J. Agric. Food Chem.* **2007**, *55*, 4625–4629 (DOI: 10.1021/jf070242j). It is essential that novel compounds, either synthetic or isolated from natural sources, be characterized rigorously and unequivocally. Supporting data normally include physical form, melting point (if solid), UV/IR spectra if appropriate, 1H and 13C NMR, mass spectrometric data, and optical rotation (when compounds have chiral centers).

Examples:

- Reporting X-ray data: “Racemic and Enantiopure Synthesis and Physicochemical Characterization of the Novel Taste Enhancer *N*-(1-Carboxyethyl)-6-(hydroxymethyl)pyridinium-3-ol Inner Salt”, by Renaud Villard et al. *J. Agric. Food Chem.* **2004**, *51*, 4040–4045.
- Reporting data in detail, including UV shifts: “Novel Flavonol Glycoside, 7-*O*-Methyl Mearnsitrin, from *Sageretia theezans* and Its Antioxidant Effect”, by Shin-Kyo Chung et al. *J. Agric. Food Chem.* **2004**, *52*, 4664–4668.
- Reporting data for previously known compounds: “Phenolic Constituents and Antioxidant Activity of *Wendita calysina* Leaves (Burrito), a Folk Paraguayan Tea”, by Anna Lisa Piccinelli et al. *J. Agric. Food Chem.* **2004**, *52*, 5863–5868.

**Flavor Constituents.** Manuscripts reporting on flavor constituents should conform to the recommendations made by the International Organization of the Flavor Industry [for details, see the Editorial in the October 1996 issue of *J. Agric. Food Chem.* (*44*, 2941–2941)]. In brief, any identification of a flavoring substance must pass scrutiny of the latest forms of available analytical

techniques. **In practice, this means that any particular substance must have its identity confirmed by at least two methods, for example, comparison of chromatographic and spectrometric data (which may include GC, MS, IR, and NMR) with those of an authentic sample.** If only one method has been applied (MS data alone or retention index or Kovats index alone), the identification shall be labeled “tentative”. In addition, authors are encouraged to include at least semiquantitative data on the concentration of an identified component in the original source, for example, foodstuff or plant part. Ranges such as <1 µg/kg, 1–10 µg/kg, and 10–100 µg/kg are acceptable.

Flavor is evoked by smell (aroma) and taste. A good example showing the correct characterization of taste compounds is the study by Czepa and Hofmann (*J. Agric. Food Chem.* **2003**, *51*, 3865–3873). A good example for aroma compound identification is the study by Milo and Grosch (*J. Agric. Food Chem.* **1996**, *48*, 2366–2371).

The use of reference compounds is a must, if data on sensory properties of single compounds are reported. Odor, which is perceived during sniffing of a food extract at a certain retention index, may be indicative of the presence of a given compound, but not conclusive unless substantiated by chromatographic and/or spectrometric data and comparison with an authentic reference compound. **Soil Classification.** Soils used in research should be described down to the family level according to the soil classification scheme given in *Soil Taxonomy, A Basic System of Soil Classification for Making and Interpreting Soil Surveys*, 2nd ed. (Agricultural Handbook 436; U.S. Government Printing Office: Washington, DC, 1999) (available on-line at <http://soils.usda.gov/technical/classification/taxonomy/>). Also give series name if known.

This requirement is to allow comparison and extrapolation to other work giving similar soil classifications, as published in journals such as the *Journal of Soil Science*, *Soil Science Society of America Journal*, *Journal of Environmental Quality*, and *Geoderma*. If information is unavailable to classify the soils at the desired family level, classification should be described or estimated at least to the great group level in the same classification system.

**Statistics.** Manuscripts reporting analytical, biological activity, composition, and related data must include relevant statistical information to support discussion of differences or similarities in data sets. Refer to a standard statistics reference such as *Statistical Methods*, 8th ed.; Snedecor, G. W., Cochran, W. G., Eds.; University Press: Ames, IA, 1989.

**Animal or Human Studies.** Manuscripts describing studies in which the use of live animals or human subjects is involved must include under Materials and Methods a statement that such experiments were performed in compliance with the appropriate laws and institutional guidelines, and also name the institutional committee that approved the experiments. For experiments with human subjects, a statement that informed consent was obtained from each individual must be included and the consent forms made available to the *Journal* on request. Reviewers of manuscripts involving animal or human experiments will be asked to comment specifically on the appropriateness and conformity to regulations of such experiments. **Authors are encouraged to note the approval code or number or give the name of the approving office of official.**

*Animal Subjects.* The use of animals in a study should be employed only when there are no alternative methods for investigating the fundamental questions of the study. In such cases, it is the ethical responsibility of all authors to ensure that the care of animals is of the highest possible order, that pain and/or distress is minimized, and that the numbers involved are strictly limited to those essential to fulfill the experimental design. In the United States the care and use of laboratory animals is regulated by the U.S. Department of Agriculture (USDA) under the Animal Welfare Act. Links to the regulations and other information are available at [http://www.aphis.usda.gov/animal\\_welfare/links.shtml](http://www.aphis.usda.gov/animal_welfare/links.shtml). It is recognized that researchers in other countries may be governed by different laws and regulations. In such cases, experiments should be designed to conform either to the above USDA regulations or to the International Guiding Principles for Biomedical Research Involving Animals (1985), available at [http://www.cioms.ch/publications/guidelines/1985\\_texts\\_of\\_guidelines.htm](http://www.cioms.ch/publications/guidelines/1985_texts_of_guidelines.htm).

*Human Subjects.* The use of human subjects in experimental studies requires informed consent. Such consent requires that the subjects be informed completely not only about the procedures involved but

also about the aims, design, and expected outcomes of the study. Consent must be obtained not only when subjects are involved directly in the study but also when samples (tissue, blood, plasma, etc.) are required for in vitro experiments. In the United States the protection of human research subjects is regulated by the U.S. Department of Health and Human Services (HHS). Regulations are available at <http://www.hhs.gov/ohrp/>. Laws and regulations governing researchers in other countries must be observed, but experiments should be designed to conform to the intent of the HHS regulations as far as possible.

In relation to the subject matter of the *Journal*, experiments involving taste and food quality evaluation and consumer acceptance are exempt from the above regulations [CFR 46.101 (b) (6)]. However, it should be noted that this would not exempt studies in which extracts, isolates, pure compounds, etc., obtained from conventional food sources are subjected to such evaluation.

**The *Journal* will reject any manuscript for which there is reason to believe that animals have been subjected to unnecessary pain or distress or when informed consent of human subjects is absent or incomplete. Editor**

**Contact Information:**

James N. Seiber, Editor

*Journal of Agricultural and Food Chemistry*

Department of Environmental Toxicology

University of California

One Shields Avenue

Davis, California 95616

U.S.A.

Telephone (530) 754-7005

E-mail [jafc@jafc.acs.org](mailto:jafc@jafc.acs.org)

## 5.2 ANEXO B

### 5.2.1 NORMAS PARA PUBLICAÇÃO NA REVISTA BRASILEIRA DE PESQUISA EM ALIMENTOS (REBRAPA)

Diretrizes para Autores

#### **Informações aos Autores e Formatação dos Manuscritos**

A Revista Brasileira de Pesquisa em Alimentos publica artigos e comunicações científicas na área de Ciência, Tecnologia e Engenharia de Alimentos. Os trabalhos podem ser apresentados em português, inglês ou espanhol, devendo observar as disposições normativas da revista não podendo exceder 6000 palavras (excluindo resumo, abstract, tabelas, figuras, legendas e referências). Todos os manuscritos deverão ser submetidos exclusivamente através do sistema eletrônico de submissão disponível no site [www.cm.utfpr.edu.br/rebrapa](http://www.cm.utfpr.edu.br/rebrapa).

Os autores devem eleger um autor responsável pela submissão, que conduzirá todo o processo de submissão. O autor responsável deve ter obtido permissão por escrito de todos os autores do artigo, devendo manter tal autorização sob sua custódia. Durante o processo de submissão online o autor responsável deverá aceitar as condições de submissão e a declaração de direitos autorais.

A REBRAPA aceita submissão de artigos em duas categorias:

**Artigos Originais:** Trabalhos que descrevam descobertas originais e de maior importância e devem ser escritos de maneira clara e sucinta.

**Artigos de Revisão:** Destinados à apresentação do progresso em uma área específica com o objetivo de dar uma visão crítica do ponto de vista do especialista altamente qualificado e experiente. É imprescindível que, na referida área, o autor tenha publicações que comprovem a sua experiência e qualificação. O Corpo Editorial da REBRAPA poderá, eventualmente, convidar pesquisadores qualificados para submeter artigo de revisão.

#### **Preparação dos manuscritos:**

Todas as páginas devem ser numeradas consecutivamente (canto inferior direito de cada página). A submissão deverá ser feita em arquivos do tipo DOC ou DOCX em formato A4. Para artigos submetidos em inglês ou espanhol, autores que não sejam fluentes na língua são encorajados a procurar ajuda na escrita do documento. Artigos submetidos em português devem ser redigidos em linguagem culta. Incorreções gramaticais levam inevitavelmente ao atraso no processo de avaliação e aceite do artigo.

Não incluir no manuscrito informações sobre os autores e suas respectivas filiações bem como e-mail de contato ou outros dados que possam identificar a autoria do trabalho. Tais informações serão incluídas no formulário de submissão e não serão enviadas para os avaliadores a fim de manter a revisão cega dos manuscritos.

**Texto:** deve ser utilizada a fonte Times New Roman tamanho 12 para o texto, parágrafos justificados com espaçamento duplo entre linhas. Todas as linhas do manuscrito devem ser numeradas consecutivamente utilizando o respectivo comando do editor de textos (*Layout de Página > Números de Linha > Contínuo*).

Para o processo de submissão, o manuscrito deve ser preparado na seguinte ordem:

- 1) Títulos do trabalho em português e inglês ou espanhol e inglês. O título (fonte tamanho 14) deve ser escrito de forma breve, concisa e clara e deve refletir de forma objetiva o tema do artigo;
- 2) Resumo na língua do manuscrito (máximo de 250 palavras). Este deve ser conciso, fornecendo o escopo do trabalho, objetivos, resultados significantes e conclusões.
- 3) Resumo em inglês, caso o manuscrito não seja escrito em inglês;
- 4) Palavras-chave (3 a 5) em português e inglês ou espanhol e inglês.
- 5) Texto principal. Será permitida alguma flexibilidade na apresentação do conteúdo, contudo deve ser respeitada uma sequência lógica (Introdução, Materiais e Métodos, Resultados e Discussão, Conclusão, Agradecimentos, Referências).

Na elaboração do texto principal, os seguintes pontos devem ser respeitados:

- Deixar a margem esquerda, direita, superior e inferior de 2,5 cm.
- Incluir figuras e tabelas nos locais onde estas devem aparecer no artigo após a publicação. As figuras e tabelas devem ser numeradas consecutivamente em algarismos arábicos (Exemplo: Figura 1: ...; Tabela 1: ...). Evite duplicar informações apresentando-as simultaneamente em gráficos e tabelas. Os textos das legendas de tabelas e figuras devem refletir seu conteúdo e conter toda a informação necessária para o seu entendimento.
- Imagens não podem ser melhoradas durante o processo de editoração, por isso a qualidade final da imagem depende da qualidade das imagens fornecidas pelos autores. Utilize apenas gráficos e imagens sem cor (preto e branco ou escalas de cinza).
- É preferível que as figuras e tabelas tenham o tamanho de uma coluna de texto (largura de 7cm).

- Abreviações, siglas e símbolos devem ser claramente definidos na primeira vez em que aparecem no texto.
- Notas de rodapé não são permitidas.
- Equações devem ser geradas por programas apropriados e identificadas no texto com algarismos arábicos entre parêntesis na ordem em que aparecem.
- As citações bibliográficas inseridas no texto devem ser indicadas dependendo do número de autores. Artigos com um, dois ou três autores, citam-se os sobrenomes separados por ponto e vírgula seguidos do ano de publicação; artigos com quatro ou mais autores, cita-se o sobrenome do primeiro autor, seguido da expressão “*et al.*” em itálico seguido do ano de publicação; se o nome do autor não é conhecido, cita-se a fonte de origem.

Exemplos:

“Como demonstrado por Silva, Souza e Costa (2008), as temperaturas...”;

“... relacionadas ao tipo de embalagem mais adequada ao seu acondicionamento (SANTOS; FIGUEIRÊDO; QUEIROZ, 2004).”

“De acordo com Silva *et al.* (2010), os fatores ...”

“... em uma determinada pressão e temperatura (LUZ *et al.*, 2006).

“... até atingir massa constante (AOAC, 1994).”

“... foram realizadas segundo metodologia descrita pela AOAC (1995).”

- Toda a literatura citada ou indicada no texto deverá ser listada em ordem alfabética nas Referências. Artigos em preparação ou submetidos à avaliação não devem ser incluídos nas referências. A formatação das referências deve seguir o padrão exemplificado a seguir.

*Livros:*

SILVA, D. B.; SILVA, J. A.; JUNQUEIRA, N. P. V.; ANDRADE, L. R. M. **Frutas do cerrado**. Brasília: EMBRAPA, 2001.

BORZANI, W.; SCHMIDELL, W.; LIMA, U. de A.; AQUARONE, E. **Biotecnologia industrial: Fundamentos**. São Paulo (SP): Edgard Blucher, 2001. V1.

*Artigos:*

LIMA, A.; SILVA, A. M. O.; TRINDADE, R. A.; TORRES, R. P.; MANCINI-FILHO, J. Composição química e compostos bioativos presentes na polpa e na amêndoa de pequi (*Caryocar brasiliense* Camb.). **Revista Brasileira de Fruticultura**, v. 29, n. 3, p. 695-967, 2007.

*Teses, Dissertações, Monografias e Trabalhos de Conclusão de Curso:*

LEIMANN, F. V. **Nanopartículas Híbridas de Polímero Natural (PHBV)/Polímero Sintético**. 133 f. Tese (Doutorado) - Universidade Federal de Santa Catarina, Florianópolis, 2011.

*Normas Técnicas:*

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 6023**: Informação e Documentação. Referências: Elaboração. Rio de Janeiro, 2002.

*Trabalhos Apresentados em Congressos:*

CLAROS, R. A. R.; PENZ JÚNIOR, A. M. Control de Calidad de los Diferentes Sistemas de Processado de la Soya. In: **III Seminário Internacional em Ciência Avícolas**. Santa Cruz, Bolívia: Anais, p. 25-32, 1997.

*Patentes e Marcas:*

EMBRAPA. Unidade de Apoio, Pesquisa e Desenvolvimento de Instrumentação Agropecuária (São Carlos). Paulo Estevão Cruvinel. **Medidor digital multisensor de temperatura para solos**. BR n. PI 8903105-9, 1995.

*Home Pages e Documentos Disponíveis Somente em Meio Eletrônico:*

SERVIÇO BRASILEIRO DE APOIO ÀS MICRO E PEQUENAS EMPRESAS. **Manual para implantação de incubadores de empresas: por que implantar**. Disponível em:

<[http://www.sebrae.com.br/br/parasuaempresa/incubadorasdeempresas\\_953.asp](http://www.sebrae.com.br/br/parasuaempresa/incubadorasdeempresas_953.asp)>. Acesso em: 12 mai. 2004.

- Segundo o conselho editorial da REBRAPA, artigos submetidos cujas referências bibliográficas estejam fora do padrão determinado ou com informações incompletas não serão publicados até que os autores tenham as referências totalmente adequadas às normas.
- Caso necessário a equipe editorial da REBRAPA pode requisitar o envio de arquivos separados contendo as tabelas e figuras com resolução adequada para publicação impressa.

### Condições para submissão

Como parte do processo de submissão, os autores são obrigados a verificar a conformidade da submissão em relação a todos os itens listados a seguir. As submissões que não estiverem de acordo com as normas serão devolvidas aos autores.

1. A contribuição é original e inédita, e não está sendo avaliada para publicação por outra revista; caso contrário, deve-se justificar em "Comentários ao editor".
2. O arquivo da submissão está em formato Microsoft Word (DOC, DOCX e RTF).
3. URLs para as referências foram informadas quando possível.
4. O texto está em espaço duplo; usa uma fonte de 12-pontos; emprega itálico em vez de sublinhado (exceto em endereços URL); as figuras e tabelas estão inseridas no texto, não no final do documento na forma de anexos.
5. O texto segue os padrões de estilo e requisitos bibliográficos descritos em [Diretrizes para Autores](#), na página Sobre a Revista.
6. Em caso de submissão a uma seção com avaliação pelos pares (ex.: artigos), as instruções disponíveis em [Assegurando a avaliação pelos pares cega](#) foram seguidas.

### Declaração de Direito Autoral

Autores que publicam nesta revista concordam com os seguintes termos:

- a. Autores mantêm os direitos autorais e concedem à revista o direito de primeira publicação, com o trabalho simultaneamente licenciado sob a [Licença Creative Commons Attribution](#) que permite o compartilhamento do trabalho com reconhecimento da autoria e publicação inicial nesta revista.
- b. Autores têm autorização para assumir contratos adicionais separadamente, para distribuição não-exclusiva da versão do trabalho publicada nesta revista (ex.: publicar em repositório institucional ou como capítulo de livro), com reconhecimento de autoria e publicação inicial nesta revista.
- c. Autores têm permissão e são estimulados a publicar e distribuir seu trabalho online (ex.: em repositórios institucionais ou na sua página pessoal) a qualquer ponto antes ou durante o processo editorial, já que isso pode gerar alterações produtivas, bem como aumentar o impacto e a citação do trabalho publicado (Veja [O Efeito do Acesso Livre](#)).

### Política de Privacidade

Os nomes e endereços informados nesta revista serão usados exclusivamente para os serviços prestados por esta publicação, não sendo disponibilizados para outras finalidades ou a terceiros.

ISSN (eletrônico): 2236-6563

ISSN (impresso): 2179-3174

## 5.3 ANEXO C

### 5.3.1 NORMAS PARA PUBLICAÇÃO NO PERIÓDICO FOOD CHEMISTRY

#### GUIDE FOR AUTHORS

##### Introduction

##### Types of paper

Original research papers; review articles; rapid communications; short communications; viewpoints; letters to the Editor; book reviews.

1. Research papers - original full-length research papers which have not been published previously, except in a preliminary form, and should not exceed 7,500 words (including allowance for no more than 6 tables and illustrations).

2. Review articles - will be accepted in areas of topical interest, will normally focus on literature published over the previous five years, and should not exceed 10,000 words (including allowance for no more than 6 tables and illustrations).

3. Rapid communications - an original research paper reporting a major scientific result or finding with significant implications for the research community, designated by the Editor.

4. Short communications - Short communications of up to 3000 words, describing work that may be of a preliminary nature but which merits immediate publication.

5. Viewpoints - Authors may submit viewpoints of about 1200 words on any subject covered by the Aims and Scope.

6. Letters to the Editor - Letters are published from time to time on matters of topical interest.

7. Book reviews

##### Page charges

This journal has no page charges.

##### Before You Begin

##### Ethics in publishing

For information on Ethics in publishing and Ethical guidelines for journal publication see <http://www.elsevier.com/publishingethics> and <http://www.elsevier.com/ethicalguidelines>.

The work described in your article must have been carried out in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans <http://www.wma.net/en/30publications/10policies/b3/index.html>; EU Directive 2010/63/EU for animal experiments [http://ec.europa.eu/environment/chemicals/lab\\_animals/legislation\\_en.htm](http://ec.europa.eu/environment/chemicals/lab_animals/legislation_en.htm).

##### Conflict of interest

All authors are requested to disclose any actual or potential conflict of interest including any financial, personal or other relationships with other people or organizations within three years of beginning the submitted work that could inappropriately influence, or be perceived to influence, their work. See also <http://www.elsevier.com/conflictsofinterest>.



## **Submission declaration and verification**

Submission of an article implies that the work described has not been published previously (except in the form of an abstract or as part of a published lecture or academic thesis or as an electronic preprint, see <http://www.elsevier.com/postingpolicy>), that it is not under consideration for publication elsewhere, that its publication is approved by all authors and tacitly or explicitly by the responsible authorities where the work was carried out, and that, if accepted, it will not be published elsewhere in the same form, in English or in any other language, including electronically without the written consent of the copyright-holder. To verify originality, your article may be checked by the originality detection service CrossCheck <http://www.elsevier.com/editors/plagdetect>.

## **Changes to authorship**

This policy concerns the addition, deletion, or rearrangement of author names in the authorship of accepted manuscripts:

Before the accepted manuscript is published in an online issue: Requests to add or remove an author, or to rearrange the author names, must be sent to the Journal Manager from the corresponding author of the accepted manuscript and must include: (a) the reason the name should be added or removed, or the author names rearranged and (b) written confirmation (e-mail, fax, letter) from all authors that they agree with the addition, removal or rearrangement. In the case of addition or removal of authors, this includes confirmation from the author being added or removed. Requests that are not sent by the corresponding author will be forwarded by the Journal Manager to the corresponding author, who must follow the procedure as described above. Note that: (1) Journal Managers will inform the Journal Editors of any such requests and (2) publication of the accepted manuscript in an online issue is suspended until authorship has been agreed.

After the accepted manuscript is published in an online issue: Any requests to add, delete, or rearrange author names in an article published in an online issue will follow the same policies as noted above and result in a corrigendum.

## **Copyright**

Upon acceptance of an article, authors will be asked to complete a 'Journal Publishing Agreement' (for more information on this and copyright see <http://www.elsevier.com/copyright>). Acceptance of the agreement will ensure the widest possible dissemination of information. An e-mail will be sent to the corresponding author confirming receipt of the manuscript together with a 'Journal Publishing Agreement' form or a link to the online version of this agreement.

Subscribers may reproduce tables of contents or prepare lists of articles including abstracts for internal circulation within their institutions. Permission of the Publisher is required for resale or distribution outside the institution and for all other derivative works, including compilations and translations (please consult <http://www.elsevier.com/permissions>). If excerpts from other copyrighted works are included, the author(s) must obtain written permission from the copyright owners and credit the source(s) in the article. Elsevier has preprinted forms for use by authors in these cases: please consult <http://www.elsevier.com/permissions>.

## **Retained author rights**

As an author you (or your employer or institution) retain certain rights; for details you are referred to: <http://www.elsevier.com/authorsrights>.

## **Role of the funding source**

You are requested to identify who provided financial support for the conduct of the research and/or

preparation of the article and to briefly describe the role of the sponsor(s), if any, in study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the article for publication. If the funding source(s) had no such involvement then this should be stated. Please see <http://www.elsevier.com/funding>.

### **Funding body agreements and policies**

Elsevier has established agreements and developed policies to allow authors whose articles appear in journals published by Elsevier, to comply with potential manuscript archiving requirements as specified as conditions of their grant awards. To learn more about existing agreements and policies please visit <http://www.elsevier.com/fundingbodies>.

### **Open access**

This journal offers you the option of making your article freely available to all via the ScienceDirect platform. To prevent any conflict of interest, you can only make this choice after receiving notification that your article has been accepted for publication. The fee of \$3,000 excludes taxes and other potential author fees such as color charges. In some cases, institutions and funding bodies have entered into agreement with Elsevier to meet these fees on behalf of their authors. Details of these agreements are available at <http://www.elsevier.com/fundingbodies>. Authors of accepted articles, who wish to take advantage of this option, should complete and submit the order form (available at <http://www.elsevier.com/locate/openaccessform.pdf>). Whatever access option you choose, you retain many rights as an author, including the right to post a revised personal version of your article on your own website. More information can be found here: <http://www.elsevier.com/authorsrights>.

### **Language (usage and editing services)**

Please write your text in good English (American or British usage is accepted, but not a mixture of these). Authors who feel their English language manuscript may require editing to eliminate possible grammatical or spelling errors and to conform to correct scientific English may wish to use the English Language Editing service available from Elsevier's WebShop <http://webshop.elsevier.com/languageediting/> or visit our customer support site <http://support.elsevier.com> for more information.

### **Submission**

Submission to this journal proceeds totally online and you will be guided stepwise through the creation and uploading of your files. The system automatically converts source files to a single PDF file of the article, which is used in the peer-review process. Please note that even though manuscript source files are converted to PDF files at submission for the review process, these source files are needed for further processing after acceptance. All correspondence, including notification of the Editor's decision and requests for revision, takes place by e-mail removing the need for a paper trail.

Authors must provide and use an email address unique to themselves and not shared with another author registered in EES, or a department.

### **Referees**

Authors are required to submit, with the manuscript, the names, addresses and e-mail addresses of 3 potential referees. Note that the editor retains the sole right to decide whether or not the suggested reviewers are used.

## Review Policy

A peer review system involving two or three reviewers is used to ensure high quality of manuscripts accepted for publication. The Managing Editor and Editors have the right to decline formal review of a manuscript when it is deemed that the manuscript is

- 1) on a topic outside the scope of the Journal;
- 2) lacking technical merit;
- 3) focused on foods or processes that are of narrow regional scope and significance;
- 4) fragmentary and providing marginally incremental results; or
- 5) is poorly written.

## Preparation

### Use of wordprocessing software

**General:** Manuscripts must be typewritten, double-spaced with wide margins on one side of white paper. Each page must be numbered, and lines must be consecutively numbered from the start to the end of the manuscript. Good quality printouts with a font size of 12 or 10 pt are required. The corresponding author should be identified (include a Fax number and E-mail address). Full postal addresses must be given for all co-authors. Authors should consult a recent issue of the journal for style if possible. An electronic copy of the paper should accompany the final version. The Editors reserve the right to adjust style to certain standards of uniformity. Authors should retain a copy of their manuscript since we cannot accept responsibility for damage or loss of papers. Original manuscripts are discarded one month after publication unless the Publisher is asked to return original material after use.

### Article structure

Follow this order when typing manuscripts: Title, Authors, Affiliations, Abstract, Keywords, Main text, Acknowledgements, Appendix, References, Vitae, Figure Captions and then Tables. Do not import the Figures or Tables into your text. The corresponding author should be identified with an asterisk and footnote. All other footnotes (except for table footnotes) should be identified with superscript Arabic numbers. The title of the paper should unambiguously reflect its contents. Where the title exceeds 70 characters a suggestion for an abbreviated running title should be given.

### Subdivision - numbered sections

Divide your article into clearly defined and numbered sections. Subsections should be numbered 1.1 (then 1.1.1, 1.1.2, ...), 1.2, etc. (the abstract is not included in section numbering). Use this numbering also for internal cross-referencing: do not just refer to 'the text'. Any subsection may be given a brief heading. Each heading should appear on its own separate line.

### Essential title page information

- **Title.** Concise and informative. Titles are often used in information-retrieval systems. Avoid abbreviations and formulae where possible.
- **Author names and affiliations.** Where the family name may be ambiguous (e.g., a double name), please indicate this clearly. Present the authors' affiliation addresses (where the actual work was done) below the names. Indicate all affiliations with a lower-case superscript letter immediately after the author's name and in front of the appropriate address. Provide the full postal address of each affiliation, including the country name and, if available, the e-mail address of each author.
- **Corresponding author.** Clearly indicate who will handle correspondence at all stages of refereeing and publication, also post-publication. **Ensure that telephone and fax numbers (with country and area code) are provided in addition to the e-mail address and the complete postal address. Contact details must be kept up to date by the corresponding author.**

• **Present/permanent address.** If an author has moved since the work described in the article was done, or was visiting at the time, a 'Present address' (or 'Permanent address') may be indicated as a footnote to that author's name. The address at which the author actually did the work must be retained as the main, affiliation address. Superscript Arabic numerals are used for such footnotes.

## **Abstract**

A concise and factual abstract is required. The abstract should state briefly the purpose of the research, the principal results and major conclusions. An abstract is often presented separately from the article, so it must be able to stand alone. For this reason, References should be avoided, but if essential, then cite the author(s) and year(s). Also, non-standard or uncommon abbreviations should be avoided, but if essential they must be defined at their first mention in the abstract itself.

The abstract should not exceed 150 words.

## **Highlights**

Highlights are mandatory for this journal. They consist of a short collection of bullet points that convey the core findings of the article and should be submitted in a separate file in the online submission system. Please use 'Highlights' in the file name and include 3 to 5 bullet points (maximum 85 characters, including spaces, per bullet point). See <http://www.elsevier.com/highlights> for examples.

## **Units**

Follow internationally accepted rules and conventions: use the international system of units (SI). If other units are mentioned, please give their equivalent in SI. Temperatures should be given in degrees Celsius. The unit 'billion' is ambiguous and should not be used.

## **Database linking**

Elsevier encourages authors to connect articles with external databases, giving their readers one-click access to relevant databases that help to build a better understanding of the described research. Please refer to relevant database identifiers using the following format in your article: Database: xxxx (e.g., TAIR: AT1G01020; CCDC: 734053; PDB: 1XFN). See <http://www.elsevier.com/databaselinking> for more information and a full list of supported databases.

## **Artwork**

### **Electronic artwork**

#### General points

- Make sure you use uniform lettering and sizing of your original artwork.
- Save text in illustrations as 'graphics' or enclose the font.
- Only use the following fonts in your illustrations: Arial, Courier, Times, Symbol.
- Number the illustrations according to their sequence in the text.
- Use a logical naming convention for your artwork files.
- Provide captions to illustrations separately.
- Produce images near to the desired size of the printed version.
- Submit each figure as a separate file.

A detailed guide on electronic artwork is available on our website:

<http://www.elsevier.com/artworkinstructions>

**You are urged to visit this site; some excerpts from the detailed information are given here.**

Formats

Regardless of the application used, when your electronic artwork is finalised, please 'save as' or convert the images to one of the following formats (note the resolution requirements for line drawings, halftones, and line/halftone combinations given below):

EPS: Vector drawings. Embed the font or save the text as 'graphics'.

TIFF: Color or grayscale photographs (halftones): always use a minimum of 300 dpi.

TIFF: Bitmapped line drawings: use a minimum of 1000 dpi.

TIFF: Combinations bitmapped line/half-tone (color or grayscale): a minimum of 500 dpi is required.

If your electronic artwork is created in a Microsoft Office application (Word, PowerPoint, Excel) then please supply 'as is'.

**Please do not:**

- Supply files that are optimised for screen use (e.g., GIF, BMP, PICT, WPG); the resolution is too low;
- Supply files that are too low in resolution;
- Submit graphics that are disproportionately large for the content.

Please insert the following text before the standard text - Photographs, charts and diagrams are all to be referred to as "Figure(s)" and should be numbered consecutively in the order to which they are referred. They should accompany the manuscript, but should not be included within the text. All illustrations should be clearly marked with the figure number and the author's name. All figures are to have a caption. Captions should be supplied on a separate sheet.

### **Color artwork**

Please make sure that artwork files are in an acceptable format (TIFF, EPS or MS Office files) and with the correct resolution. If, together with your accepted article, you submit usable color figures then Elsevier will ensure, at no additional charge, that these figures will appear in color on the Web (e.g., ScienceDirect and other sites) regardless of whether or not these illustrations are reproduced in color in the printed version. **For color reproduction in print, you will receive information regarding the costs from Elsevier after receipt of your accepted article.** Please indicate your preference for color: in print or on the Web only. For further information on the preparation of electronic artwork, please see <http://www.elsevier.com/artworkinstructions>.

Please note: Because of technical complications which can arise by converting color figures to 'gray scale' (for the printed version should you not opt for color in print) please submit in addition usable black and white versions of all the color illustrations.

### **Figure captions**

Ensure that each illustration has a caption. Supply captions separately, not attached to the figure. A caption should comprise a brief title (**not** on the figure itself) and a description of the illustration. Keep text in the illustrations themselves to a minimum but explain all symbols and abbreviations used.

### **Tables**

Number tables consecutively in accordance with their appearance in the text. Place footnotes to tables below the table body and indicate them with superscript lowercase letters. Avoid vertical rules. Be sparing in the use of tables and ensure that the data presented in tables do not duplicate results described elsewhere in the article.

### **References**

### **Citation in text**

Please ensure that every reference cited in the text is also present in the reference list (and vice versa). Any references cited in the abstract must be given in full. Unpublished results and personal communications are not recommended in the reference list, but may be mentioned in the text. If these references are included in the reference list they should follow the standard reference style of the journal and should include a substitution of the publication date with either 'Unpublished results' or 'Personal communication'. Citation of a reference as 'in press' implies that the item has been accepted for publication.

### **Web references**

As a minimum, the full URL should be given and the date when the reference was last accessed. Any further information, if known (DOI, author names, dates, reference to a source publication, etc.), should also be given. Web references can be listed separately (e.g., after the reference list) under a different heading if desired, or can be included in the reference list.

Example: CTAHR (College of Tropical Agriculture and Human Resources, University of Hawaii). Tea (*Camellia sinensis*) a New Crop for Hawaii, 2007. URL [http://www.ctahr.hawaii.edu/oc/freepubs/pdf/tea\\_04\\_07.pdf](http://www.ctahr.hawaii.edu/oc/freepubs/pdf/tea_04_07.pdf). Accessed 14.02.11.

All publications cited in the text should be presented in a list of references following the text of the manuscript. No more than 30 references should be cited in your manuscript. In the text refer to the author's name (without initials) and year of publication (e.g. "Steventon, Donald and Gladden (1994) studied the effects..." or "...similar to values reported by others (Anderson, Douglas, Morrison & Weiping, 1990)..."). For 2-6 authors all authors are to be listed at first citation. At subsequent citations use first author et al.. When there are more than 6 authors, first author et al. should be used throughout the text. The list of references should be arranged alphabetically by authors' names and should be as full as possible, listing all authors, the full title of articles and journals, publisher and year. The manuscript should be carefully checked to ensure that the spelling of authors' names and dates are exactly the same in the text as in the reference list.

### **Reference style**

Text: Citations in the text should follow the referencing style used by the American Psychological Association. You are referred to the Publication Manual of the American Psychological Association, Sixth Edition, ISBN 978-1-4338-0561-5, copies of which may be ordered from <http://books.apa.org/books.cfm?id=4200067> or APA Order Dept., P.O.B. 2710, Hyattsville, MD 20784, USA or APA, 3 Henrietta Street, London, WC3E 8LU, UK.

List: references should be arranged first alphabetically and then further sorted chronologically if necessary. More than one reference from the same author(s) in the same year must be identified by the letters 'a', 'b', 'c', etc., placed after the year of publication.

Examples:

Reference to a journal publication:

Van der Geer, J., Hanraads, J. A. J., & Lupton, R. A. (2010). The art of writing a scientific article. *Journal of Scientific Communications*, 163, 51–59.

Reference to a book:

Strunk, W., Jr., & White, E. B. (2000). *The elements of style*. (4th ed.). New York: Longman, (Chapter 4).

Reference to a chapter in an edited book:

Mettam, G. R., & Adams, L. B. (2009). How to prepare an electronic version of your article. In B. S. Jones, & R. Z. Smith (Eds.), *Introduction to the electronic age* (pp. 281–304). New York: E-Publishing Inc.

## Supplementary data

Elsevier accepts electronic supplementary material to support and enhance your scientific research. Supplementary files offer the author additional possibilities to publish supporting applications, high-resolution images, background datasets, sound clips and more. Supplementary files supplied will be published online alongside the electronic version of your article in Elsevier Web products, including ScienceDirect: <http://www.sciencedirect.com>. In order to ensure that your submitted material is directly usable, please provide the data in one of our recommended file formats. Authors should submit the material in electronic format together with the article and supply a concise and descriptive caption for each file. For more detailed instructions please visit our artwork instruction pages at <http://www.elsevier.com/artworkinstructions>.

## Submission checklist

The following list will be useful during the final checking of an article prior to sending it to the journal for review. Please consult this Guide for Authors for further details of any item.

### Ensure that the following items are present:

One author has been designated as the corresponding author with contact details:

- E-mail address
- Full postal address
- Telephone and fax numbers

All necessary files have been uploaded, and contain:

- Keywords
- All figure captions
- All tables (including title, description, footnotes)

Further considerations

- Manuscript has been 'spell-checked' and 'grammar-checked'
- References are in the correct format for this journal
- All references mentioned in the Reference list are cited in the text, and vice versa
- Permission has been obtained for use of copyrighted material from other sources (including the Web)
- Color figures are clearly marked as being intended for color reproduction on the Web (free of charge) and in print, or to be reproduced in color on the Web (free of charge) and in black-and-white in print
- If only color on the Web is required, black-and-white versions of the figures are also supplied for printing purposes

For any further information please visit our customer support site at <http://support.elsevier.com>.

## Additional information

Abbreviations for units should follow the suggestions of the British Standards publication BS 1991.

The full stop should not be included in abbreviations, e.g. m (not m.), ppm (not p.p.m.), % and '/' should be used in preference to 'per cent' and 'per'. Where abbreviations are likely to cause ambiguity or may not be readily understood by an international readership, units should be put in full.

Current recognised (IUPAC) chemical nomenclature should be used, although commonly accepted trivial names may be used where there is no risk of ambiguity.

The use of proprietary names should be avoided. Papers essentially of an advertising nature will not be accepted.

## After Acceptance

### Use of the Digital Object Identifier

The Digital Object Identifier (DOI) may be used to cite and link to electronic documents. The DOI

consists of a unique alpha-numeric character string which is assigned to a document by the publisher upon the initial electronic publication. The assigned DOI never changes. Therefore, it is an ideal medium for citing a document, particularly 'Articles in press' because they have not yet received their full bibliographic information. Example of a correctly given DOI (in URL format; here an article in the journal Physics Letters B):

<http://dx.doi.org/10.1016/j.physletb.2010.09.059>

When you use a DOI to create links to documents on the web, the DOIs are guaranteed never to change.

## Proofs

One set of page proofs (as PDF files) will be sent by e-mail to the corresponding author (if we do not have an e-mail address then paper proofs will be sent by post) or, a link will be provided in the e-mail so that authors can download the files themselves. Elsevier now provides authors with PDF proofs which can be annotated; for this you will need to download Adobe Reader version 7 (or higher) available free from <http://get.adobe.com/reader>. Instructions on how to annotate PDF files will accompany the proofs (also given online). The exact system requirements are given at the Adobe site: <http://www.adobe.com/products/reader/tech-specs.html>.

If you do not wish to use the PDF annotations function, you may list the corrections (including replies to the Query Form) and return them to Elsevier in an e-mail. Please list your corrections quoting line number. If, for any reason, this is not possible, then mark the corrections and any other comments (including replies to the Query Form) on a printout of your proof and return by fax, or scan the pages and e-mail, or by post. Please use this proof only for checking the typesetting, editing, completeness and correctness of the text, tables and figures. Significant changes to the article as accepted for publication will only be considered at this stage with permission from the Editor. We will do everything possible to get your article published quickly and accurately – please let us have all your corrections within 48 hours. It is important to ensure that all corrections are sent back to us in one communication: please check carefully before replying, as inclusion of any subsequent corrections cannot be guaranteed. Proofreading is solely your responsibility. Note that Elsevier may proceed with the publication of your article if no response is received.

## Offprints

The corresponding author, at no cost, will be provided with a PDF file of the article via e-mail (the PDF file is a watermarked version of the published article and includes a cover sheet with the journal cover image and a disclaimer outlining the terms and conditions of use). For an extra charge, paper offprints can be ordered via the offprint order form which is sent once the article is accepted for publication. Both corresponding and co-authors may order offprints at any time via Elsevier's WebShop (<http://webshop.elsevier.com/myarticleservices/offprints>). Authors requiring printed copies of multiple articles may use Elsevier WebShop's 'Create Your Own Book' service to collate multiple articles within a single cover (<http://webshop.elsevier.com/myarticleservices/offprints/myarticlesservices/booklets>).

## Author Inquiries

For inquiries relating to the submission of articles (including electronic submission) please visit this journal's homepage. For detailed instructions on the preparation of electronic artwork, please visit <http://www.elsevier.com/artworkinstructions>. Contact details for questions arising after acceptance of an article, especially those relating to proofs, will be provided by the publisher. You can track accepted articles at <http://www.elsevier.com/trackarticle>. You can also check our Author FAQs at <http://www.elsevier.com/authorFAQ> and/or contact Customer Support via <http://support.elsevier.com>.



