

**PROGRAMA DE PÓS-GRADUAÇÃO EM SISTEMAS E
PROCESSOS INDUSTRIAIS – MESTRADO ÁREA DE CONTROLE E
OTIMIZAÇÃO DE PROCESSOS INDUSTRIAIS**

Roberta Cristina Kaufmann

**AVALIAÇÃO DE EQUIPAMENTO ESPECTRÔMETRO NO INFRAVERMELHO
PRÓXIMO NO PROCESSO DE CLASSIFICAÇÃO DE TABACO**

Santa Cruz do Sul

2014

Roberta Cristina Kaufmann

**AVALIAÇÃO DE EQUIPAMENTO ESPECTRÔMETRO NO INFRAVERMELHO
PRÓXIMO NO PROCESSO DE CLASSIFICAÇÃO DE TABACO**

Dissertação apresentada ao Programa de Pós-Graduação em Sistemas e Processos Industriais – Mestrado, Área de Controle e Otimização de Sistemas Industriais, Linha de Pesquisa em Otimização de Processo, Universidade de Santa Cruz do Sul – UNISC, como requisito parcial para obtenção do título de Mestre em Sistemas e Processos Industriais.

Orientador: Prof. Dr. João Carlos Furtado

Co-orientador: Prof. Dr. Luciano Marder

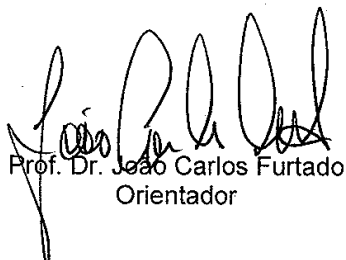
Santa Cruz do Sul

2014

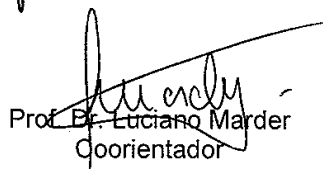
Roberta Cristina Kaufmann

AVALIAÇÃO DE EQUIPAMENTO ESPECTRÔMETRO NO INFRAVERMELHO
PRÓXIMO NO PROCESSO DE CLASSIFICAÇÃO DE TABACO

Esta Dissertação foi submetida ao Programa de Pós-Graduação em Sistemas e Processos Industriais – Mestrado – Universidade de Santa Cruz do Sul – UNISC, como requisito parcial para obtenção do título de Mestre em Sistemas e Processos Industriais.



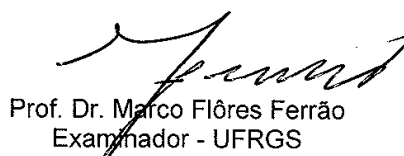
Prof. Dr. João Carlos Furtado
Orientador



Prof. Dr. Luciano Marder
Coorientador



Prof. Dr. Adilson Ben da Costa
Examinador - UNISC



Prof. Dr. Marco Flôres Ferrão
Examinador - UFRGS

AGRADECIMENTOS

Agradeço aos meus pais Jussara Kaufmann e Carmo José Kaufmann pelo incentivo e apoio; ao meu irmão Rodrigo Juliano Kaufmann pela assistência prestada a mim e a nossa família nos momentos em que não pude estar presente; aos professores e colegas do Programa de Pós-Graduação em Sistemas e Processos Industriais - Mestrado, pelos ensinamentos e pela amizade.

Também agradeço a Souza Cruz S.A pela concessão da bolsa de estudos.

RESUMO

Visando o aperfeiçoamento das medições de umidade e nicotina nos processos que envolvem o setor de recebimento de tabaco, a empresa Souza Cruz S.A. adquiriu o analisador de processo TM710e, espectrofotômetro no infravermelho próximo (NIR) com sistema de medição não invasivo visto que estas variáveis estão estreitamente relacionadas com a classificação do tabaco e influenciam diretamente o valor a ser pago ao fornecedor. A metodologia utilizada para planejar e definir os parâmetros de validação foi a *Orientação sobre Validação de Métodos Analíticos*, definida pelo INMETRO. Através desta orientação, os resultados do TM710e podem ser comparados com os resultados das medições clássicas utilizadas pelo controle de qualidade da organização, bem como verificar seus respectivos desvios. Através dos testes para identificar a seletividade e robustez do TM710e, encontraram-se variáveis como do inadequado tempo de resposta e local de instalação do equipamento, a absorção do espectro da corda e as diferentes cores do tabaco que contribuíam para o aparecimento de incertezas e de desvios nas medições. O local de instalação do equipamento foi modificado durante este trabalho de pesquisa, bem como reduzido o tempo de resposta, fatos que, em consoante com as periódicas calibrações, contribuíram para as significativas reduções do coeficiente de variação e aumento na precisão do equipamento, e, portanto, passíveis de otimização. Verificou-se através dos experimentos e do tratamento estatístico para determinar a precisão dos métodos analíticos envolvidos e compará-las, que as características de desempenho do equipamento estão compatíveis com o exigido pelo método em estudo. De forma a reduzir os desvios e melhorar a robustez do equipamento, a variação em virtude das diferentes cores de tabaco pode ser otimizada através da calibração para cada tipo de tabaco através da automatização por um sofisticado sensor de cor ou mesmo pelo tratamento de imagens, assim como para o auto-ajuste do Trim. Poder-se-ia evitar a absorção do espectro da corda; entretanto, como a leitura deste material é eventual e o desvio gerado por tal é baixo, entende-se que com a otimização das demais variáveis seria o suficiente para melhorar o desempenho do equipamento. Dado o exposto, o objetivo de avaliar o equipamento espectrofotômetro no infravermelho próximo TM710e através do seu monitoramento, análise dos processos de compra e classificação do tabaco, quantificação dos níveis de umidade e nicotina e análise exploratória dos dados visando investigar as variáveis que interferem no desempenho do equipamento, foi concluído. Sendo que, principalmente, os resultados dos testes de precisões servem de base para a empresa definir os critérios de aceitação e implementar este

controle de processo através da espectroscopia no infravermelho próximo em todas suas unidades de classificação de tabaco.

Palavras-chave: Espectroscopia no infravermelho próximo (NIR), controle de processo, Classificação do tabaco, umidade, nicotina.

ASBTRACT

The use of tools that allow strict control over product quality and have low operating costs can be the difference for an industry to remain the leader in its sector. Aiming at improving measurements of moisture and nicotine in processes involving the sector of receiving tobacco, the company Souza Cruz S.A. acquired the process analyzer TM710e, a near infrared spectrophotometer (NIR) with noninvasive measurement, as these variables are closely related to the classification of tobacco and directly influence the amount to be paid to the supplier. The methodology used to plan and define the validation parameters was the Guideline on Validation of Analytical Methods defined by INMETRO. Through this guidance, the results of TM710e can be compared with the results of classical measurements used in quality control in the organization, as well as verify their deviations. Through the tests to identify the selectivity and robustness of TM710e, variables were identified such as the inappropriate response time and place of installation of the equipment, the absorption spectrum of the rope and the different colors that tobacco that contributed to the appearance of uncertainty and deviations in the measurements. The place of installation of the equipment was modified during this research as well as a reduction in the response time, facts that, in compliance with the periodic calibrations, contributed to significant reductions in the coefficient of variation and increased the accuracy of the equipment, therefore enabling optimization. From the experiments and statistical analysis to determine the accuracy of the analytical methods involved and compare them to the performance, it was found that the characteristics of the equipment are compatible with the requirements of the method under study. In order to reduce the deviations and improve the robustness of the equipment, the variation due to different colors of tobacco could be optimized by calibration for each type of tobacco and by automating the exchange of parameterization by a sophisticated color sensor, or even by the treatment of images. Thus, the self-adjustment of the Trim could also be performed. It would be possible to prevent the absorption of the spectrum of the rope; however, as the reading of this material is possible and the diversion generated by such is low, it is understood that the optimization of other variables would be enough to improve the performance of equipment. Given the above, the purpose of evaluating the near infrared spectrophotometer equipment TM710e through its monitoring, analysis of purchasing processes and classification of tobacco, quantifying the levels of moisture and nicotine and exploratory data analysis in order to investigate the variables that affect the performance equipment was completed. The results obtained with precise testes could serve as the basis for

the company to define the acceptance criteria and to implement this process control through near infrared spectroscopy in all tobacco classification units.

Keywords: near infrared spectroscopy (NIR), process control, tobacco grading, moisture, nicotine.

LISTA DE ILUSTRAÇÕES

Figura 1 - Método de classificação do tabaco.	22
Figura 2 – Disposição de componentes e o trajeto ótico do TM 710e	30
Figura 3 - Arranjo da disposição óptica dos detectores.....	31
Figura 4. Decomposição dos dados originais em scores (t) e loadings (p) e a matriz de erros E.....	43
Figura 5. Tipos de pré-processamento.	44
Figura 6. Exemplo de dendograma HCA.	45

LISTA DE ABREVIATURAS E SIGLAS

AFUBRA – Associação dos Fumicultores do Brasil

CMD – Concentração média determinada

CQ – Controle de qualidade

CV – Coeficiente de variação

DP – Desvio padrão

DPP – Desvio padrão populacional

HCA – *Hierarchical clustering analysis*

L – Lemon

O – Orange

PCA – *Principal Componente Analysis*

R – Red

SINDITABACO – Sindicato das Industrias de Tabaco

X – Média

SUMÁRIO

1	INTRODUÇÃO	12
1.3	Hipótese	14
1.4	Justificativa	14
2	OBJETIVO GERAL	16
2.2	Objetivos específicos.....	16
3	REFERENCIAL TEÓRICO	17
3.1	Origem do tabaco.....	17
3.1.1	Nicotina 18	
3.1.2	Umidade 19	
3.2	O Tabaco e a economia brasileira	20
3.3	Classificação do tabaco.....	21
3.3.1	Grupo.....	22
3.3.2	Subgrupo	23
3.3.3	Classificação interna	23
3.4	Analisadores de processo por absorção de radiação	23
3.4.1	Espectroscopia no infravermelho	25
3.4.2	Analisadores de umidade e nicotina: Medição por reflexão de radiação infravermelha 26	
3.5	Espectroscopia no Infravermelho Aplicada ao Tabaco.....	28
3.6	Espectrofotômetro TM710e	29
3.6.1	Descrição geral.....	29
3.6.2	Disposição ótica do TM710e.....	30
3.7	Controle de qualidade	32
3.8	Validação	33
3.9	Métodos estatísticos	35
3.9.1	Dados estatísticos	36
3.9.2	Amostragem	37
3.9.3	Medidas estatísticas.....	38
3.9.4	Testes de hipóteses	39
3.10	Análise de componentes principais PCA (<i>Principal Component Analysis</i>)	42
3.11	Análise de agrupamentos hierárquicos (HCA)	45
3.12	Erros aleatórios e sistemáticos	46
	REFERÊNCIAS	49

1 INTRODUÇÃO

A disputa pela conquista de maiores fatias do mercado consumidor, o enfrentamento da concorrência, o volume de investimentos e sua adequada remuneração, a qualidade do serviço ou do produto, a melhoria do meio ambiente e social tornaram-se desafios constantes em todos os ramos da atividade econômica. Maximizar resultados e minimizar dispêndios, ou seja, otimizar a produção, torna-se cada vez mais importante (SILVA et al, 1998).

Neste cenário, também é de fundamental importância o controle no processo que garanta a qualidade do produto desenvolvido. A utilização de ferramentas que possibilitem um controle rigoroso sobre a qualidade do produto e tenham um baixo custo operacional, pode ser a diferença para uma indústria se manter líder no seu setor.

Na indústria de tabaco, é muito importante o monitoramento da qualidade e do estilo da produção, pois são estes fatores que irão determinar o teor de constituintes químicos incluindo açúcar, umidade, alcaloides totais entre outros, na folha ou no tabaco processado. Vários componentes podem ser identificados no tabaco, dos quais muitos contribuem para o sabor, aroma e efeitos fisiológicos que faz dele único entre outras plantas. Substâncias químicas determinam a qualidade e a fragrâncias dos estilos de tabacos. Muitas análises de componentes na folha de tabaco e no cigarro são significantes no controle de qualidade do processo. Estes componentes químicos são usados como índices para avaliar a qualidade do tabaco, mas suas detecções são demoradas na maioria dos casos e envolve um número de etapas que requer pessoas especializadas. E, como um poderoso método analítico, a espectroscopia no infravermelho próximo tem sido largamente aplicada no campo da indústria do tabaco devido às características de rapidez, simplicidade e de medidas não destrutivas (DUAN et al., 2012).

A empresa Souza Cruz S.A. adquiriu um analisador de processo denominado TM710e, espectrofotômetro com sistema de medição não invasivo, com o objetivo inicial de otimizar o setor de secagem e debulhamento, aperfeiçoando o controle destes processos e consequentemente a produção de *blends* através da rápida verificação da nicotina no tabaco.

Entretanto, as constantes dificuldades relacionadas ao atual processo de compra e classificação do tabaco, utilizado por mais de cinquenta anos nas indústrias que trabalham com esta atividade, idealizaram que o equipamento adquirido, baseado em um sistema para controle de processos e monitoramento de medições de produtos na linha de produção pudesse ser também utilizado como sistema analisador para o controle do processo de recebimento da matéria-prima. Instigou-se desta forma a investigação da propriedade do uso da espectroscopia no infravermelho próximo para a determinação da umidade e nicotina no setor de recebimento

do tabaco a fim de aperfeiçoar este processo na indústria, visto que estas variáveis estão estreitamente relacionadas com a classificação do tabaco e influenciam diretamente o valor a ser pago ao produtor rural.

Com a utilização de um equipamento que monitore de forma precisa tais variáveis no setor de recebimento, o Classificador poderá determinar as concentrações de umidade e nicotina instantaneamente e assim classificar e negociar a compra do tabaco com maior segurança, garantindo uma compra mais justa para o produtor e para a indústria.

Experimentos iniciais realizados pela empresa com o equipamento instalado no setor de recebimento do tabaco, apontaram a ocorrência de desvios significativos quando comparados aos valores de referência que são os resultados dos métodos de análises utilizados pelo controle de qualidade da organização para as determinações de umidade e nicotina. Assim, se fez necessário investigar o porquê da irregularidade dos resultados e adaptar o equipamento a estas linhas de produção industrial e/ou adaptar o próprio processo produtivo a este tipo de controle, bem como verificar se o equipamento de fato é capaz de realizar a tarefa de forma apropriada.

Firmou-se então a parceria entre esta indústria e o programa de Pós-Graduação em Sistemas e Processos Industriais da Universidade de Santa Cruz do Sul a fim de que fosse realizada uma avaliação do desempenho do equipamento no processo de classificação do tabaco, com base em métodos de pesquisa científica e por uma instituição imparcial tanto aos interesses da empresa quanto de seus fornecedores.

O trabalho apresenta relevância econômica e científica. Sob a ótica econômica, a comercialização do tabaco desempenha um importante papel na economia brasileira. Segundo a Fundação de Economia e Estatística (2013) as exportações de tabaco no Rio Grande do Sul aumentaram US\$ 108,0 milhões nos primeiros 6 meses de 2013, representando 12,91% dos embarques gaúchos.

Sob a ótica científica, investigar a funcionalidade do equipamento espectrofotômetro nesta aplicação, que utiliza o método não destrutivo, de fácil medição por reflexão no infravermelho, é relevante principalmente porque muitas destas metodologias podem ser adaptadas às linhas de produção industrial de outros produtos de origem vegetal (PASQUINI, 2003).

1.3 Hipótese

Diante deste cenário, espera-se com esse trabalho responder os seguintes questionamentos:

Quais variáveis contribuem para o aparecimento dos desvios nas medições dos parâmetros umidade e nicotina, apresentados pelo equipamento neste processo? As variáveis encontradas serão passíveis de otimização junto ao sistema de medição do equipamento? Será possível reduzir estes desvios? Como analisar os desvios e definir critérios de especificação para a avaliação do equipamento espectrofotômetro TM710e?

1.4 Justificativa

Devido às dificuldades apresentadas pelos processos que envolvem atualmente o setor de recebimento do tabaco, é oportuno estudar formas alternativas, automatizadas e independentes da avaliação humana de classificação. A proposta apresentada neste trabalho é fazer uso da espectroscopia no infravermelho próximo através do equipamento TM710e.

Espectrofotômetros vêm sendo largamente utilizados pelas indústrias com atividades voltadas ao processamento do tabaco no controle dos processos de beneficiamento, bem como nos laboratórios de controle de qualidade, e no monitoramento de processos de alimentos, farmacêuticos, polímeros, ambiental, combustíveis e mais recentemente, em bioprocessos (TREVISAN e POPPI, 2006).

Segundo Trevisan e Poppi (2006), é possível encontrar uma grande diversidade de analisadores de processos para as mais variadas aplicações. No entanto, a incompatibilidade das características destes dispositivos dificulta a escolha do sensor apropriado, como por exemplo, uma análise lenta e bastante precisa por outra menos precisa, mas com maior velocidade.

Há poucas décadas era considerada praticamente impossível a medição de umidade em sólidos, mais ou menos correta, em linha de processo. Entretanto, este desafio vem sendo vencido tendo atualmente muitas aplicações operando com êxito comprovado. Ainda assim, recomenda-se que o problema seja encarado com cautela e que se trabalhe com fornecedores especializados de comprovada experiência em aplicações similares (COHN, 2006).

Desta forma, idealizou-se a possibilidade de avaliar a adaptabilidade e o desempenho do equipamento espectrofotômetro de medição *non-invasive* TM710e, exemplar de última

geração da fabricante NDC *Infrared Engineering* (2012) para otimização dos processos que envolvem o recebimento do tabaco, como a classificação e compra, almejando garantir aos seus parceiros, fornecedores e clientes maior confiabilidade nos resultados destes processos.

A melhoria no procedimento de classificação se dará principalmente em função da medida instantânea do parâmetro nicotina, em função de cada classe possuir intervalo para essa variável. Tal fato resultará em uma classificação com maior segurança, diminuindo conflitos de interesses tanto por parte do fornecedor quanto do funcionário, denominado de classificador. E a medição imediata da umidade subsidiará o classificador a determinar o percentual de umidade do produto e subtraí-la do peso total do fardo, quando a umidade ultrapassar o limite de concentração estabelecido para cada tipo de tabaco.

Este avanço deverá resultar em um melhor controle das condições em que a matéria prima é comprada, reduzindo custos com a compra de produtos de menor qualidade e/ou em condições que adicionem peso ao produto garantindo uma classificação mais precisa. Podendo ainda fornecer informações de forma provisionada ao setor que planeja o processo de misturas (*blends*), possibilitando melhorar o beneficiamento e conseqüentemente, a qualidade do produto final. Bem como, aumentar a segurança na venda e negociação do produto por parte do fornecedor e auxiliar à a organização na tomada de decisões.

2 OBJETIVO GERAL

Este trabalho de pesquisa tem como objetivo geral avaliar o equipamento espectrofotômetro no infravermelho próximo (NIR) TM710e através da investigação do seu desempenho, na medição de umidade e nicotina nos processos que envolvem o setor de recebimento de tabaco.

2.2 Objetivos específicos

Os objetivos específicos deste trabalho são:

- a) Monitorar a atuação do equipamento TM710e no setor de recebimento do tabaco;
- b) Analisar os processos que envolvem o setor de recebimento do tabaco, especialmente a compra e a classificação do tabaco;
- c) Quantificar os níveis de nicotina e umidade identificando os desvios do TM710e, dos métodos aplicados pelo controle de qualidade e dos classificadores comparando-os entre si;
- d) Realizar análise exploratória dos dados visando identificar variáveis que interferem no desempenho do equipamento.

A opção de disponibilização da dissertação foi parte parcial do texto, onde o capítulo escolhido para apresentação foi o referencial teórico.

3 REFERENCIAL TEÓRICO

Nesse tópico será abordada de forma sucinta a origem do tabaco, bem como a história da indústria de fumo. Incluindo nesse contexto, uma base sobre os tipos de tabaco e do processo de classificação, dividindo-o este processo em duas etapas: externo e interno. A classificação externa é aqui entendida pela etapa de compra do tabaco por parte da empresa, enquanto que a classificação interna é um processo dependente de cada da empresa do setor e, portanto, muito variável e de caráter sigiloso para aqui ser discutido.

Também serão descritos os problemas que surgem neste processo de classificação, bem como uma descrição geral sobre sistemas analisadores em processos industriais, sobre o equipamento TM710e, uma revisão sobre espectroscopia no infravermelho, análise multivariada e tratamento estatístico.

3.1 Origem do tabaco

Segundo o Sindicato das Indústrias de Tabaco – SINDITABACO (2012) a maioria dos historiadores considera o tabaco como sendo de origem americana, onde teria sido cultivado pelos indígenas, tanto da América do Sul como do Norte. A hipótese mais provável é a de que a planta teria surgido nos vales orientais dos Andes Bolivianos, difundindo-se pelo território brasileiro através das migrações indígenas. Em 1530, plantas de tabaco teriam sido levadas pela tripulação de Cristóvão Colombo para a Europa e cultivadas pela família real portuguesa por seu aspecto ornamental e por sua função medicinal. Em 1560, o então Embaixador da França em Portugal, Jean Nicot, ao saber que a planta curava enxaquecas, a enviou para sua rainha, em Paris, que teria iniciado o hábito de pitar, sendo então imitada pelos nobres da sua corte, difundindo-se pelos demais países da Europa, o que teria originado o mercado de tabaco em pó, chamado rapé. Assim, em apenas um século o tabaco passou a ser conhecido e usado no mundo inteiro.

No início do século XVI, os primeiros portugueses a desembarcarem no Brasil já encontraram o cultivo de tabaco em quase todas as tribos indígenas. Para estes índios a planta era sagrada e de origem mística. Seu uso geralmente era limitado a ritos mágico-religiosos, como no evocar dos deuses, bem como para fins medicinais na cura de ferimentos, enxaquecas e dores de estômago, usado exclusivamente pelos pajés (feiticeiros). De planta mágico-religiosa, o tabaco passou a ser um produto comercial das colônias europeias, e mais

particularmente, das Antilhas, da Virgínia (a partir de 1612) e do Brasil (SINDITABACO, 2012).

Assim, rapidamente o cultivo e o comércio de tabaco no Brasil colonial passaram a ter importância destacada com o conhecimento de várias legislações e taxações já no século XVII, estando entre os principais produtos exportados durante o período do império. Esta importância encontra-se marcada até os dias atuais no brasão das Armas da República, onde o tabaco e o ramo de café constituem o coroamento deste símbolo da nacionalidade brasileira (SINDITABACO, 2012).

3.1.1 Nicotina

Dentre as mais de 300 substâncias químicas presentes no tabaco in natura (GARATTONI, 2008), segundo Ligocky (1987), a nicotina é o alcaloide vegetal mais importante, pois é encontrado principalmente nesta planta e em todas suas partes. Com concentrações que variam de 1 a mais de 10%, a nicotina foi descoberta e estudada por *Vaupelin* em 1822 e por *Posselt e Reimann* em 1828.

Embora a concentração varie em diferentes espécies de tabaco, as maiores concentrações estão em regiões altas e próximas ao talo da planta. Trata-se de um composto orgânico, uma amina terciária composta de anéis de piridina e pirrolidina, de forma molecular $C_{10}H_{14}N_2$. O composto é líquido à temperatura ambiente, incolor, inodoro e oleoso, e, quando exposto à luz ou calor, adquire uma coloração marrom e odor característico do tabaco. Em pequenas doses, a nicotina estimula especialmente o sistema nervoso vegetativo, favorecendo a liberação de adrenalina e, em grandes doses, paralisa o sistema nervoso autônomo provocando convulsões que podem levar à morte (AZNAR e MADURO, 2008; MADURO, 2005).

Segundo Ligocky (1987), há fatores que influenciam no teor de nicotina no tabaco, a saber:

- Fertilizantes: A grande maioria dos adubos fertilizantes para solo possui em sua composição NPK (nitrogênio, fósforo e cálcio). Sendo a nicotina um composto orgânico nitrogenado, quanto maior a assimilação de nitrogênio pela planta, maior a produção de nicotina.
- Água: Em função da produção de nicotina se dar nas raízes da planta, quando ocorre um período de seca a planta desenvolve novas raízes a fim de obter água em demais áreas do solo e com isso há uma sintetização maior de nicotina.

- Espaçamento: Para todas as culturas há um espaçamento mínimo entre uma planta e outra que deve ser observado. Para que haja maior possibilidade de desenvolvimento das raízes para que assim absorvam uma maior quantidade de nutrientes, aumentando assim o teor de nicotina.
- Maturação: O teor de alcaloides aumenta com a maturação da planta.
- Variedades: Existem variedades da planta do tabaco, como Virgínia, Burley, Amarelinho e Comum, onde algumas produzem naturalmente mais nicotina do que outras.
- Posição da Folha no pé: O teor de nicotina aumenta nas folhas baixas (mais próximas ao solo) para as pontas.

3.1.2 Umidade

Antes de passar pelo processo de cura o tabaco possui aproximadamente 80 a 85 por cento de água. No final do ciclo de cura, o tabaco é essencialmente zero por cento água. Nesta fase, o tabaco é demasiado frágil para manusear sem ruptura. Desta forma, a boa prática é de que alguma umidade possa ser colocada de volta ao tabaco, no final da cura, a fim de permitir uma boa manipulação. Muita umidade, no entanto, pode fazer com que o tabaco ao calor, escureça, e apodreça, e acabará por arruinar suas qualidades desejáveis. Como forma eficaz e eficiente o controle de umidade da folha curada sempre foi uma grande preocupação para os produtores e compradores (SMITH, 2005).

Segundo Smith (2005) o tabaco curado, como muitos materiais orgânicos, é higroscópico. Materiais higroscópicos possuem uma afinidade física pela umidade. No caso do tabaco, esta umidade é usualmente absorvida pelo vapor d'água do ar próximo a folha. A absorção da água pelo tabaco curado parte de um complexo processo que depende de muitos fatores biológicos e físicos. Entre os fatores biológicos estão inclusos as propriedades da folha das demais variedades, práticas agrícolas, posição na planta, e o clima. Os importantes fatores físicos incluem a ordenação de temperatura e umidade, velocidade do ar, a quantidade e disposição das folhas.

É bem conhecido que a taxa de absorção de umidade (geralmente expressa como uma porcentagem de aumento de umidade por hora), aumenta com o aumento da umidade relativa. Em umidades relativas mais altas, mais água está no ar e disponível para a absorção pelo tabaco. Talvez menos conhecido, o aumento da temperatura também aumenta drasticamente as taxas

de absorção de umidade. Por exemplo, com 80 % de umidade relativa, a taxa de absorção a 86 °F é mais do que o dobro da taxa a 68 °F. Com 140 ° F e 80 % de umidade relativa, a taxa pode se elevar a vários pontos percentuais por minuto. Além disso, a posição da folha na planta e a qualidade da folha afetam a taxa de absorção de água. Folhas baixas (próximas à raiz) ou finas e tabaco de má qualidade tem uma taxa de absorção mais rápida do que folhas mais espessas, ponteadas (na ponta superior do caule), ou de tabaco de melhor qualidade. Além disso, a taxa com que o tabaco absorve umidade também está relacionada com o teor de umidade atual da folha. Quanto maior a umidade na folha, mais lenta é a taxa de aumento. O tabaco com demasiada umidade está sujeito à fermentação e a decomposição por vários microorganismos. Os quatro fatores necessários para a decomposição são alimentos, água suficiente, ambiente adequado e inoculação. Se qualquer um destes está ausente, não pode ocorrer decomposição (SMITH, 2005).

3.2 O Tabaco e a economia brasileira

Segundo Associação dos Fumicultores do Brasil (AFUBRA, 2012), presente em 704 municípios do Rio Grande do Sul, Santa Catarina e Paraná, o tabaco é cultivado em 373 mil hectares, por 187 mil produtores integrados sendo assim indiscutível sua importância socioeconômica para a região Sul do Brasil.

Segundo Schunke (2014),

A qualidade de vida do produtor de tabaco expressa a importância econômica e social gerada pelo setor no Brasil. Com um desempenho histórico em 2012, nossa indústria foi responsável por 1,34% do total das exportações brasileiras no período e obteve representação significativa nos embarques gaúchos e catarinenses. Mas estes resultados não são novidade: somos há 20 anos líderes mundiais na exportação da folha, procurada por mais de cem países.

Além disso, o complexo agroindustrial de tabaco do Sul do Brasil é responsável por uma movimentação financeira que supera R\$ 10 bilhões ao ano, considerando-se as diversas etapas do processo produtivo e comercial. A elevada taxação imposta ao cigarro brasileiro, fez com que a indústria recolhesse em 2010 mais de 9,3 bilhões de reais ao Tesouro. Sendo que tal cifra não engloba outras obrigações fiscais como Previdência Social, Imposto de Renda, CPMF, e nem as taxas estaduais e municipais (AFUBRA, 2012).

3.3 Classificação do tabaco

A classificação do Tabaco segue a Instrução Normativa MAPA nº 10 DOU de 16 de abril de 2007, republicado em 17 de abril de 2007 (SINDITABACO, 2012). Portaria que delimita aspectos importantes e configuram a classificação de compra, para definir o valor a ser pago ao produtor bem como a forma de utilização na produção.

O processo de classificação, tanto externo quanto interno, é realizado por funcionários experientes treinados para a atividade com base na Instrução Normativa. A tarefa consiste em colher uma ou mais manocas de um fardo - conjunto de manocas agrupadas através de prensagem, com aproximadamente 50 kg – e através de visualização da cor, textura, comprimento e até mesmo odor é dada a classe a que pertence o produto. Toda esta tarefa depende da habilidade pessoal do funcionário e conseqüentemente tem gerado muitos problemas, a saber:

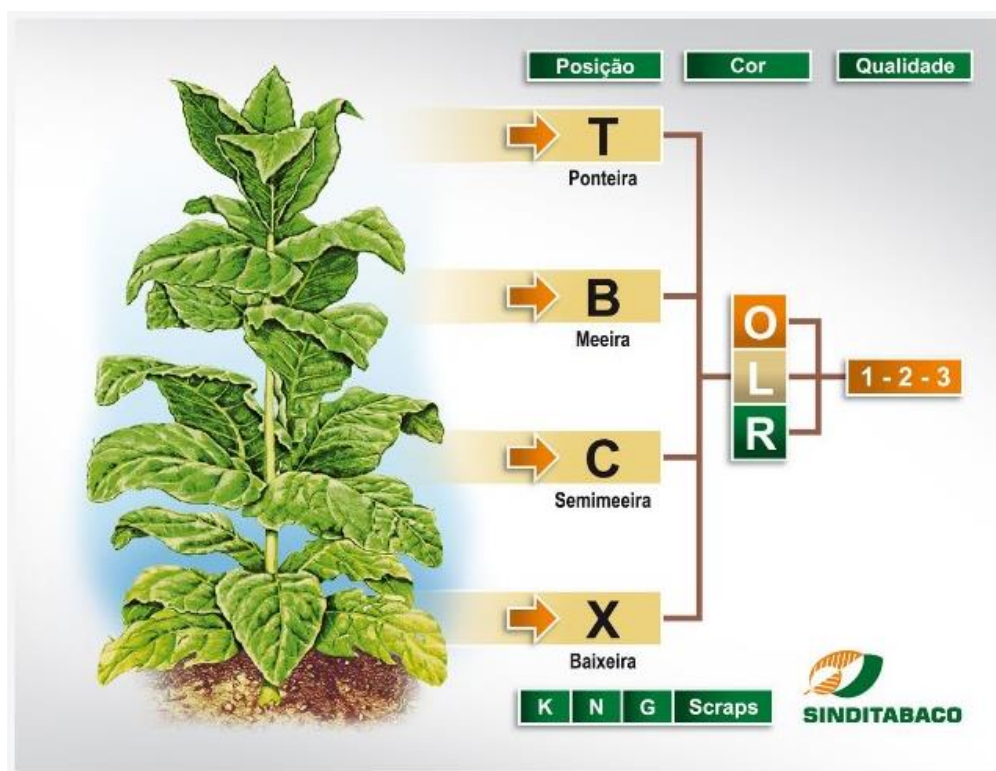
- Na classificação para a compra da matéria-prima, existe grande descontentamento por parte do produtor, que em alguns casos não acredita na isenção do funcionário da empresa e não concorda com a classificação atribuída ao tabaco e acaba por desistir da venda do produto. Isto acarreta interrupção na linha de produção, uma vez que os fardos que deixam de ser comercializados precisam ser retirados do processo antes mesmo de passar pela classificação interna e reembarcados para retorno a propriedade do produtor;

- Na classificação interna, o problema surge devido ao fato que os diferentes tipos de tabaco são insumos para a produção de cigarros de qualidade diferenciada. Desta forma, a má classificação provoca problemas nas etapas posteriores, no beneficiamento e fabricação, uma vez que não gera o cigarro com a qualidade desejada, ocasionando em muitas situações de retrabalho.

Outro aspecto importante a ser destacado e que provoca a classificação errônea pelo funcionário, é a fadiga no trabalho. Como a atividade é realizada ao longo de uma jornada de trabalho, é comum o trabalhador “perder” a sensibilidade na verificação das características do produto e desta forma cometer equívocos.

O tabaco em folha curado é classificado em grupos, subgrupos, classes, subclasses, tipos e subtipos, conforme o preparo, sua apresentação e arrumação, sua posição nas plantas, cor das folhas e sua qualidade, a Figura 1 ilustra claramente o que é observado no tabaco para a classificação de compra.

Figura 1 - Método de classificação do tabaco.



Fonte: SINDITABACO (2012, < <http://sinditabaco.com.br/>>).

3.3.1 Grupo

As empresas associadas ao SINDITABACO produzem na região Sul do Brasil tabaco em folhas proveniente da planta de espécie *Nicotiana tabacum L.*, submetidos à cura natural ou artificial, destinados à fabricação de cigarros e outras finalidades. O tabaco, produzido no Paraná, Santa Catarina e Rio Grande do Sul, é dividido em dois grupos (2012, < <http://sinditabaco.com.br/>>):

1) Tabaco de Estufa (TE) - Tabaco em folha submetido à cura (secagem) em estufas, com temperatura e umidade controladas (*Flue Cured*), incluindo-se neste grupo todas as cultivares das variedades de Virgínia e Amarelinho.

2) Tabaco de Galpão (TG) - Tabaco em folha submetido à cura (secagem) natural, à sombra ou em galpões (*Air Cured*), incluindo-se neste grupo todas as variedades de *Burley*, *Comum* e *Maryland*.

3.3.2 Subgrupo

O Tabaco em Folha Curado, independente do grupo a que pertencem, será classificado, segundo a sua apresentação e arrumação, em 2 (dois) subgrupos (2012, <<http://sinditabaco.com.br/>>):

1) Folhas Manocadas (FM) - Conjunto de folhas com no máximo 4,0 cm de diâmetro, cujas folhas se encontram juntadas e amarradas pela extremidade dos talos por uma folha da mesma classificação, devendo ser uniformes quanto ao tamanho, posição na planta, cor e qualidade.

2) Folhas Soltas (FS) - Conjunto de folhas a granel e com talo inteiro, devendo ser uniformes quanto ao tamanho, posição na planta, cor e qualidade.

3.3.3 Classificação interna

Uma vez comprado, o tabaco será novamente reclassificado de acordo com um sistema de classificação interna, onde outros aspectos de qualidade serão considerados para um melhor aproveitamento do mesmo. Este sistema de classificação interna visa a melhoria dos *blends* (misturas de tabaco) e varia de empresa para empresa. Sendo o acesso restrito a pessoas que trabalham na respectiva área em função do segredo industrial que envolvem essas misturas.

3.4 Analisadores de processo por absorção de radiação

Segundo Cohn (2006, pg. 37), “os analisadores por absorção de radiação encontram largo emprego tanto no laboratório como nos processos”, onde os primeiros grandes impulsos na técnica NIR foram dados por fatores econômicos na descoberta de que o conteúdo proteico de *commodities*, como milho e soja podia ser determinado rapidamente por espectroscopia NIR bastando que se tenham valores de referências.

Enquanto que em análises laboratoriais, amostras são manipuladas sob condições rígidas de controle podendo ainda ser pré-tratadas para promover um aumento na seletividade/ou sensibilidade. Instrumentos não ficam expostos a ambientes ou amostras corrosivas, possibilitando o emprego de técnicas analíticas clássicas. Medidas de controle de qualidade determinam somente a aceitabilidade ou não de um produto, mas não controlam o processo, nem reduzem custos na sua elaboração, já que a informação é geralmente obtida após o termino

do processo. Sensores de processos, aqui definidos como sistemas analisadores implementados na linha do processo, devem ser resistentes ao ambiente industrial submetido (TREVISAN e POPPI, 2006).

Segundo Honigs (1987, p. 19-48), citado por Trevisan e Poppi (2006, p. 1065),

Operações de amostragem, tratamento de amostras, medida, coleta de dados e processamento devem ser automatizados. O objetivo das análises em processos é eliminar ou reduzir as causas de variabilidade na linha de produção, aumentando a qualidade, produtividade e competitividade do produto.

Essas tecnologias objetivam possibilitar a obtenção de informações quantitativas e qualitativas sobre processos de forma que as mesmas possam ser usadas para otimizar sua eficiência no uso de energia, tempo e matéria-prima, contribuindo para a sustentabilidade e menor impacto ambiental além de apenas monitorar e controlar o processo. Onde “a rentabilidade na implementação de sensores em linha ocorre devido ao incremento na otimização e no controle do processo” (TREVISAN e POPPI, 2006).

Segundo Trevisan e Poppi (2006, p. 1068),

As principais técnicas espectroscópicas aplicadas em processos incluem a espectroscopia de absorção, espalhamento, reflexão e emissão nas regiões do ultravioleta, visível e infravermelho, além de espectrometria de massas e técnicas luminescentes.

As técnicas espectroscópicas apresentam a vantagem de serem mais rápidas, com a obtenção de espectros em poucos segundos e uma larga aplicação em processos industriais, tanto em sistemas *on-line*, *in-line*, como em *non-invasive*. No entanto, sistemas com mais de uma espécie apresentam uma alta sobreposição espectral. Para a resolução destes sistemas, é necessária a utilização de métodos de separação matemática, como as técnicas quimiométricas. Técnicas ópticas são capazes de realizar medidas quase instantâneas, empregando detectores com arranjo de diodos ou métodos multiplexados, como Fourier e Hadamard, devido à ausência de monocromadores móveis (TREVISAN e POPPI, 2006).

“Hoje, a espectroscopia NIR se aplica principalmente à medição da umidade e a medição de líquidos orgânicos” (COHN, 2006, p. 51).

O emprego da espectroscopia NIR para controle de matérias-primas em processos de manufaturamento é tão apropriado que pode ser considerado como ‘a sequência lógica’ para o monitoramento destes processos, chegando a ser uma técnica recomendado pela própria agência americana FDA (“U.S. Food and Drugs Administration”).

Em decorrência desta expansão da espectroscopia NIR como uma técnica alternativa de identificação e quantificação, Figuras Analíticas de Mérito têm sido propostas para estabelecer critérios de avaliação em relação aos métodos padrão de análise (BRAGA, 2004 apud TREVISAN & POPPI, 2006, p. 1069).

3.4.1 Espectroscopia no infravermelho

A espectroscopia no infravermelho é uma das técnicas analíticas mais importantes, pois sua área de aplicação envolve medicamentos, identificação de compostos orgânicos e inorgânicos, análise de misturas complexas, entre outras. Segundo Silverstein et al. (2007) “a radiação infravermelha (IR) corresponde a aproximadamente à parte do espectro situada entre as regiões do visível e das micro-ondas”. Onde, de acordo com Barbosa (2007) “a energia denominada de infravermelho corresponde à região do espectro eletromagnético situada na faixa de número de ondas entre 14290 e 200 cm^{-1} ”. Estes autores explicam que esta faixa de número de onda é comumente trabalhada e subdividida em outras três regiões, denominadas região do infravermelho próximo que está situada entre 14.290 cm^{-1} e 4000 cm^{-1} , região do infravermelho médio entre 4000 cm^{-1} e 400 cm^{-1} e do infravermelho distante na faixa de onda entre 700 cm^{-1} e 200 cm^{-1} .

A absorção da radiação no infravermelho está restrita a espécie molecular que apresenta pequena diferença de energia entre os diversos estados vibracionais ou rotacionais, já que não possui energia suficiente para promover transições eletrônicas como radiação ultravioleta ou visível. Portanto para que uma molécula absorva radiação no infravermelho é necessário que exista alteração no dipolo elétrico como consequência de seu movimento vibracional ou rotacional, resultando em absorção de energia radiante (SILVERSTEIN et al., 2007).

Silverstein et al. (2007, p. 70) comenta que “embora o espectro de infravermelho seja característico da molécula como um todo, certos grupos de átomos dão origem a bandas que ocorrem mais ou menos na mesma frequência, independentemente da estrutura da molécula”. Para o autor, a presença dessas bandas características de grupos que permite, através do exame do espectro e consulta a tabelas, a comparação das intensidades dos comprimentos de onda de absorção com as intensidades de comprimentos de onda de referência, para fazer a identificação de estruturas. Silverstein enfatiza que não dependemos somente do espectro para identificação dos compostos, mas sim a utilização deste em conjunto com outros dados espectrais para determinar a estrutura molecular.

Este trabalho de pesquisa se concentrará na região do espectro correspondente ao infravermelho próximo, pois é nesta região que o sensor do espectrofotômetro TM710e, equipamento que será utilizado neste trabalho, detecta os comprimentos de onda infravermelhos absorvidos pelo o tabaco.

3.4.1.1 Métodos de preparo das amostras

Os espectros no infravermelho podem ser obtidos em amostras líquidas, sólidas e gasosas. Entretanto, se discutirá sobre as sólidas em função de ser este o estado físico das amostras manipuladas neste trabalho. “A qualidade de um espectro no infravermelho depende muito do tipo de método de preparo da amostra e do tipo de acessório utilizado para sua obtenção” (BARBOSA, 2007, p. 30). O autor enfatiza a maneira de como as amostras são tratadas “o preparo correto da amostra é de fundamental importância para obtenção de dados de boa qualidade”, pois, tais manipulações podem afetar a aparência do espectro, assim como seu estado físico.

Entretanto, equipamentos analisadores de processos projetados para atuar em sistemas de medição *non-invasive*, como é o caso do espectrofotômetro TM710e da NDC *Infrared Engineering*, dispensam o tratamento das amostras para análises (TM710eTM, 2009). Segundo Trevisan e Poppi (2006), esta é a classe mais recente de analisadores, que apresentam as vantagens dos sensores *in-line*, de não necessitar de etapas de amostragem, além do fato que o sistema analisador não entra em contato com o processo. Nestes sistemas, o analisador não destrói e não entra em contato direto com a amostra, não provocando mudanças na sua composição ou gerando contaminação. Estas características fazem com que estes sensores sejam os mais apropriados para análise de processos.

3.4.2 Analisadores de umidade e nicotina: Medição por reflexão de radiação infravermelha

Quando se incide um feixe de radiação infravermelha na superfície de um sólido, a parte que é refletida depende das características da superfície do sólido e da sua umidade. Onde, “quanto maior a umidade maior a parte absorvida e menor a parte refletida” (COHN, 2006, p. 275).

Assim como os analisadores de umidade por reflexão de radiação, analisadores de nicotina também não requerem nenhuma habilidade especial de operação, em alguns

equipamentos *at-line* as amostras são coletadas e simplesmente inseridas no prato de amostras, o equipamento então reconhece automaticamente a presença do prato na plataforma giratória e começa sua sequência de medições em menos de 10 segundos os parâmetros são indicados claramente na tela (NDC, 2012).

Entretanto, Cohn (2006, p. 275) salienta que,

“como a reflexão depende de outros fatores, como a cor e a granulometria, a calibração deve ser feita caso a caso e será válida enquanto estas características do sólido não sofrerem alteração”.

Cohn (2006) comenta ainda que instrumentos modernos operam adicionalmente com um ou mais comprimentos de onda que não sofrem absorção pela água, servindo de referência, reduzindo o efeito dos outros fatores nas medições de umidade. Todavia, outro problema destes equipamentos se dá em função da pequena penetração pela radiação no material, camada onde ocorre a absorção, da ordem de milímetros e assim mede-se a umidade e a nicotina somente em uma camada superficial do material.

3.4.2.1 Calibração de analisadores

Segundo Cohn (2006), a umidade é a variável mais problemática e polêmica no que se refere à calibração. Paradoxalmente, por mais difícil que seja a medição de umidade nos sólidos em uma linha de processo, a calibração precisa é a mais simples, embora possa ser extremamente trabalhosa. “A vantagem é a simplicidade e a exatidão da medição no laboratório. As amostras são pesadas, secas e pesadas novamente, obtendo-se diretamente a umidade em peso”.

Cita-se como exemplo a curva de calibração do medidor de umidade no coque, onde são necessárias análises laboratoriais de dezenas de amostras, escolhidas sob um cuidadoso critério estatístico para que a quantidade seja representativa. Quando não for possível realizar calibrações, emitindo certificados e eventualmente também efetuando o ajuste, por não haver laboratório credenciado e fiscalizado pelo INMETRO, recomenda-se que um consultor independente, com os necessários conhecimentos realize uma auditoria para verificar se as calibrações estão sendo efetuadas da melhor maneira possível (COHN, 2006).

Entretanto, o sensor objeto deste estudo fabricado pela NDC é fornecido pré-calibrado. Na instalação é simplesmente ajustado para concordar com o método de referência local, no caso da empresa Souza Cruz S.A., o Controle de Qualidade - CQ. O *software* que acompanha

o equipamento *GaugeToolsXL* (2009) simplifica este processo, permitindo uma comparação dos valores apresentados pelo TM10e (recolhidos na linha de recebimento do tabaco), com os resultados de laboratório fornecidos pelo CQ.

O *software GaugeToolsXL* (2009) destina-se a receber esses dados de medição, analisando e aprovando esses dados. A análise irá gerar calibração dos parâmetros com base nos conjuntos de dados recebidos, apresenta recomendações quanto à sua adequação para uso. O *software* precisa ser registrado com a NDC para fazer todos os seus recursos disponíveis. Antes de os dados serem registrados, uma conexão de hardware entre o medidor e os comunicadores de *software* pode ser estabelecida.

Este *software* permite ainda que sejam ajustados os parâmetros de decisão da calibração, como os limites de outliers, os pontos necessários para um ajuste compensar, os pontos necessários para um ajuste ortogonal e o limite de confiança.

3.5 Espectroscopia no Infravermelho Aplicada ao Tabaco

A espectroscopia no infravermelho próximo vem sendo cada vez mais aplicada em pesquisas que envolvem o tabaco e produtos derivados como o cigarro.

Trabalhos como o de Garcia-Allende et al. (2008) já vem empregando a reflexão no NIR para detectar impurezas presentes no tabaco durante o controle de qualidade do processo industrial. Já as pesquisas apresentadas por Zhang et al. (2008), Jing et al. (2010) e Duan et al. (2012), apresentam modelos quantitativos empregando NIR para a determinação de constituintes no tabaco como por exemplo a nicotina, os açúcares totais, os açúcares redutores, metais pesados, entre outros.

Outros estudos importantes que empregam a espectroscopia NIR são na transferência de modelos de calibração multivariada entre distintos equipamentos (FAN et al., 2008).

No processo de classificação das folhas de tabaco o NIR tem sido recentemente empregado na China obtendo êxito e com isso direcionando esta técnica para classificação durante a seleção do tabaco para a produção dos *blends* utilizados na fabricação dos cigarros (NI et al., 2009).

Muitos estudos também vêm sendo desenvolvidos para a seleção da região espectral mais adequada para o desenvolvimento dos modelos quantitativos multivariados destacando-se a eliminação das variáveis não informativas (CAI et al., 2008), utilização do algoritmo das

projeções sucessivas (YE et al., 2008) e utilização do *random subspace* para rotinas de classificação e regressão multivariada (TAN et al., 2008).

Ainda entre os mais recentes trabalhos encontram-se aqueles que empregam a reflexão no infravermelho próximo para o desenvolvimento de métodos para a classificação de cigarros (MOREIRA et al., 2009; TAN et al., 2009).

3.6 Espectrofotômetro TM710e

O TM710 é um medidor que calcula todas as medições internamente. É uma rede baseada em um sistema para controle de processos. Como também pode ser configurado em diferentes formas com uma gama de dispositivos auxiliares para encontrar os requerimentos exatos (TM710ETM TECHNICAL REFERENCE MANUAL, 2009).

A fabricante do equipamento NDC *Infrared Engineering* possui quarenta anos de experiência de *design* e manufatura de instrumentos *on-line*, especialmente para indústria do tabaco. O TM710e é a oitava geração de instrumentos dedicados a medição exclusiva do tabaco (NDC TOBACCO, 2009).

Este analisador pode medir simultaneamente a umidade, a nicotina, a temperatura do produto e o açúcar. Podendo medir até baixas umidades de fumo expandido, umidade intermediária encontradas no processo primário e as umidades muito elevadas encontradas nos processos de preparação do caule. Pode ser instalado em quase todo o processo do tabaco (NDC, 2012):

- GLT ou Talo
- Processo Primário
- Planta de expansão
- Manufatura de lâmina

3.6.1 Descrição geral

O indicador TM710 *Gauge* é um dispositivo de alta precisão, baseado na tecnologia de infravermelhos. Ele é capaz de medir simultaneamente até quatro componentes no produto, das quais não são afetadas por variações de grau, tipo e mistura (PROFINETE IO, 2009).

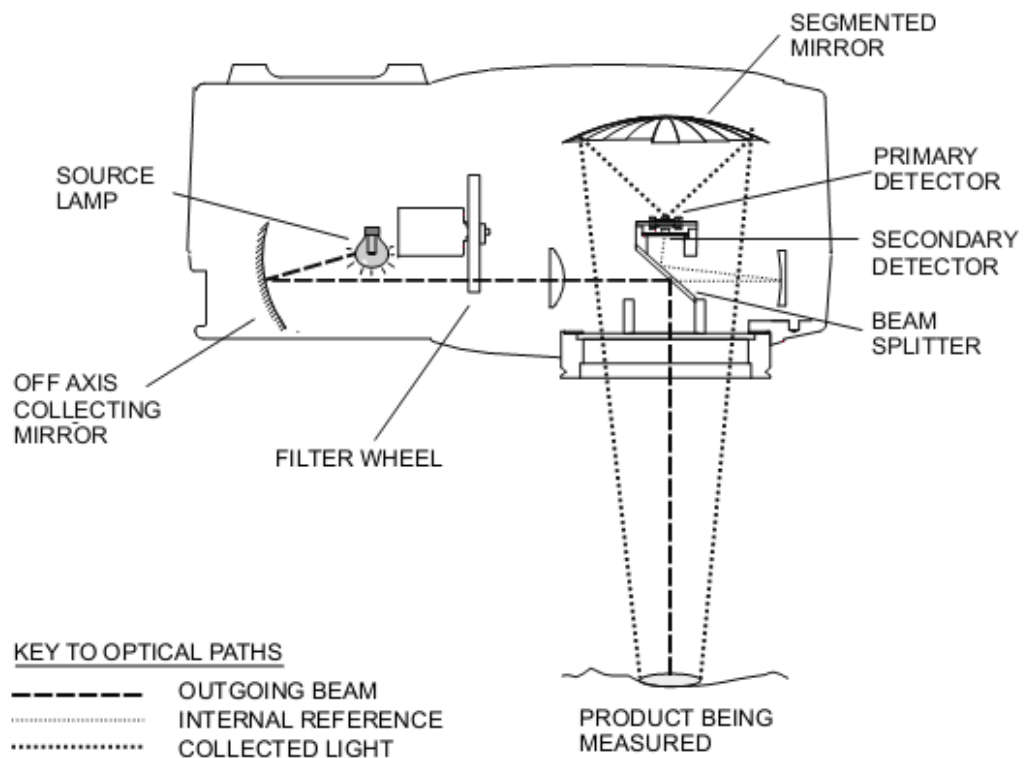
Projetado para operar em ambientes de fábricas típicos e não é afetado por variações na luz ambiente, umidade relativa, temperatura ambiente ou produto. O sensor é conhecido como

um tipo de retroespalhamento devido ao fato de que recolhe uma porção da luz infravermelha que é dispersa para trás do sensor depois reemerge a partir do produto. Toda a manipulação do equipamento (calibração e configuração) é através do serial ou do 710 *databus* a partir de uma interface de usuário. Este pode ser um micro computador executando um pacote de *software* adequado (tal como *software* de *Gauge Tools* NDC/IE que acompanha o equipamento), ou uma unidade de *hardware* independente (TM710E™ TECHNICAL REFERENCE MANUAL, 2009).

3.6.2 Disposição ótica do TM710e

Segundo TM710E™ *Technical Reference Manual* (2009), todos os elementos da ótica foram concebidos para fornecer módulos, os quais podem ser montadas em conjunto durante a fabricação, sem a necessidade de qualquer alinhamento óptico. O diagrama esquemático do equipamento é mostrado na Figura 2.

Figura 2 – Disposição de componentes e o trajeto ótico do TM 710e



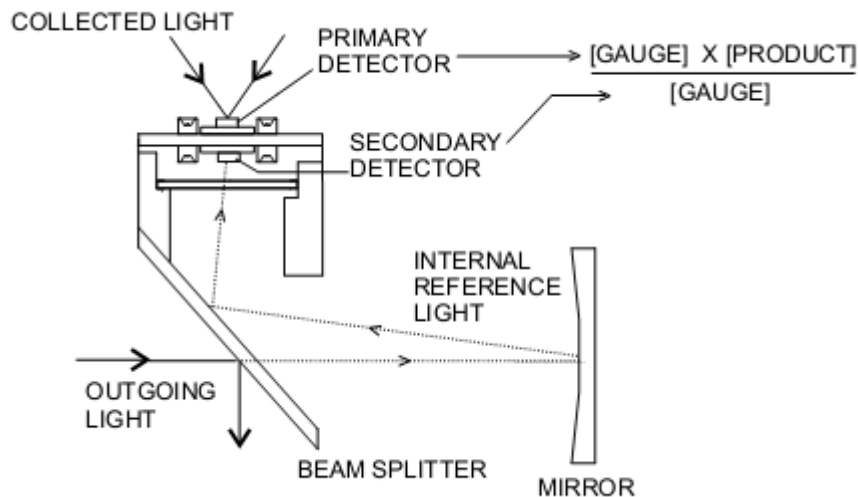
Fonte: TM710E™ Technical Reference Manual (2009, p. 24).

A luz do pré-alinhado, sob a execução da lâmpada de halogêneo de quartzo é refletida por um espelho fora do eixo de coleta que maximiza o sinal. A luz passa através do filtro de roda giratória, que gira em torno de 8000 rpm, proporcionando uma série de pulsos em cada um dos comprimentos de onda de medição. A luz passa então para o divisor de feixe. Este dispositivo projeta uma pequena parte do feixe de saída, de modo uniforme para o sistema óptico por trás dele. A vantagem deste dispositivo é que ele garante muito mais fidelidade no acompanhamento dos dois detectores o que é fundamental para o conceito do *Detector Referencing* para funcionar com precisão e conferir alta estabilidade no indicador. A luz que passa através do divisor de feixe é recolhida por outro espelho, retornando e centrando-se no detector (TM710E™ TECHNICAL REFERENCE MANUAL, 2009).

A maioria da luz passa do produto para fora do equipamento. O produto absorve parte dele. Uma pequena quantidade da luz (difusamente refletido) retrodifundida é recolhida pelo espelho. Este espelho tem 24 segmentos e é concebido para codificar os sinais infravermelhos para o detector e assim garantir que uma excelente altura de tolerância passe. Esta luz é focada para o detector primário.

Segundo TM710E™ Technical Reference Manual (2009), o algoritmo detector compara os dois sinais, comprimento de onda por comprimento de onda e fornece uma bancada estável. O princípio do detector duplo assegura elevada estabilidade a longo prazo do equipamento. Uma vista detalhada do arranjo da disposição óptica dos detectores é mostrada na Figura 3. A característica de calibre é continuamente avaliada pelo detector secundário e por razão este contra o detector primário, de comprimento de onda por comprimento de onda, a influência do sensor é eliminada.

Figura 3 - Arranjo da disposição óptica dos detectores



Fonte: TM710E™ Technical Reference Manual (2009, p. 24)

O detector secundário só identifica sinais de que pode ter sido influenciado pelo equipamento, por exemplo, os efeitos do envelhecimento da lâmpada, enquanto o detector primário detecta um sinal composto constituído pelas características do equipamento e do produto.

Para obter inequívoca informação, em longo prazo, estável sobre o produto, os dois sinais do detector são referenciados. Ao fazer referência a saída primária para a saída secundária produz uma saída independente da característica do equipamento, isto é, uma saída proporcional à medição única do produto. Os dois detectores são sempre combinados para ter as mesmas propriedades de infravermelho e são ambos montados num dissipador de calor para assegurar que eles operem à mesma temperatura.

3.7 Controle de qualidade

O Controle de Qualidade (CQ) é o setor que garante a qualidade dos produtos desde a compra da matéria prima até o produto acabado.

Neste trabalho, vamos nos deter a comentar sobre os métodos aplicados pelo CQ da empresa onde se realizou este trabalho para as medições de umidade e nicotina presentes no tabaco, utilizadas para monitorar estas variáveis já no recebimento da matéria prima coletando as amostras na célula de qualidade instalada dentro do setor de recebimento. Após, as amostras são enviadas ao Controle de Qualidade para que as mesmas sejam analisadas pelos métodos de Gravimetria (ou Forno Gravimétrico), que analisa através de Fornos BAT o percentual de

umidade presente na amostra, e o percentual de nicotina pelo Método de Determinação de Alcaloides Totais por Análise de Fluxo Contínuo no equipamento *Auto Analyzer*.

Com estas informações, detalhadas na Metodologia deste trabalho, o Controle de Qualidade pode treinar a sensibilidade dos Classificadores, de forma a aperfeiçoar o tato com relação a percepção dos níveis percentuais de umidade e nicotina presentes na folha de tabaco. Através deste treinamento e dos conhecimentos técnicos, os funcionários estão aptos a classificar o tabaco e a descontar do peso total do fardo o percentual de umidade que ultrapassar o objetivo traçado pela empresa, máximo de 17,5% de umidade para o tabaco tipo Virgínia e 18,5% para o *Burley*. Sendo que a Instrução Normativa MAPA nº 10 do Ministério da Agricultura determina que o desconto possa ser efetuado quando o tabaco estiver com umidade acima de 17%.

Os métodos e resultados das análises realizadas pelo Controle de Qualidade e pelos Classificadores atualmente serve de referência na comparação dos resultados apresentados pela leitura do sensor do espectrofotômetro TM710e, bem como no processo de validação do TM710e como controlador de processo no setor de recebimento da matéria-prima.

3.8 Validação

Segundo Cohn (2006), a validação significa simplesmente constatar se os resultados fornecidos pelo instrumento ou sistema são válidos dentro de um determinado critério de aceitação, que pode ser:

- Acordo entre as partes;
- Determinado por normas;
- Por comparação com análises efetuadas no laboratório da planta.

O autor cita como exemplo a Norma ASTM D3764 – *Standart Practice for Validation of Process Stream Analyzer Systems*. Da qual é voltada para área do petróleo, mas com conceitos de validade geral.

Cohn (2006) comenta que desta forma, o princípio tradicional de que o laboratório está sempre certo, tem-se mostrado muitas vezes incorreto. A própria norma ASTM D3764 recomenda que em casos de divergências, ambas análises (campo e laboratório) sejam reavaliadas.

A Norma ASTM D3764 (2009) descreve os procedimentos e metodologias baseadas nos princípios estatísticos da Norma D6708 - *Practice for Statistical Assessment and*

Improvement of Expected Agreement Between Two Test Methods that Purport to Measure the Same Property of a Material para validar se o grau de concordância entre os resultados produzidos por um sistema analisador total versus os resultados produzidos por um método de teste independente que se propõe a medir a mesma propriedade, atende aos requisitos especificados pelo usuário. Esta prática padrão para validação de sistema analisador de fluxo de processo pode ser utilizada para quantificar o desempenho de um sistema analisador de processo de fluxo ou o seu subsistema em termos de precisão e de viés em relação aos de um método de teste primário para a propriedade de interesse. Proporcionando aos desenvolvedores ou fabricantes de sistemas de processo analisador de fluxo procedimentos úteis para avaliar a capacidade dos sistemas recém-projetados para aplicações industriais que requerem previsão confiável de medições de uma propriedade específica de um método de ensaio primário de um componente ou produto que flui.

Entretanto, a Norma não recomenda a extrapolação dos resultados de validação além do tipo de material e faixa de imóvel utilizado para a obtenção desses resultados. Além disso, adverte os usuários de que para sistemas de medição que mostram matriz de dependências, informações através de determinado compostos puros ou misturas simples de compostos puros, podem não ser representativos do que o atingido em processo real ou amostras de produtos (ASTM D3764, 2009).

Segundo a Coordenação Geral de Acreditação do INMETRO, para se confirmar que os métodos são apropriados para o uso pretendido, deve validar (DOQ-CGCRE-008, 2011):

- Métodos não normalizados;
- Métodos criados/desenvolvidos pelo próprio laboratório;
- Métodos normalizados usados fora dos escopos para os quais foram concebidos;
- Ampliações e modificações de métodos normalizados.

O processo de validação de um método deve estar descrito em um procedimento, e os estudos para determinar os parâmetros de validação devem ser realizados com equipamentos e instrumentos dentro das especificações, funcionando corretamente e adequadamente calibrados (DOQ-CGCRE-008, 2011).

A Coordenação Geral de Acreditação do INMETRO sugere através do documento *Orientação sobre Validação de Métodos Analíticos* (DOQ-CGCRE-008, 2011) que o planejamento e execução da validação sigam a seguinte sequência de trabalho, a saber:

1. Definir a aplicação, objetivo e escopo do método;

2. Definir os parâmetros de validação e critérios de aceitação;
3. Verificar se as características de desempenho do equipamento estão compatíveis com o exigido pelo método em estudo;
4. Qualificar os materiais, por exemplo, padrões e reagentes;
5. Planejar os experimentos de validação, incluindo o tratamento estatístico, e
6. Fazer os experimentos de validação.

3.9 Métodos estatísticos

Os avanços na ciência e engenharia ocorrem em grande parte por meio de coleta e análise de dados (NAVIDI, 2012). Em geral, a estatística consiste na manipulação de um conjunto de dados com o objetivo de extrair informação sobre o comportamento de um processo ou produto. Utilizando a variabilidade presente nos dados para obter tal informação (2012, <<http://www.portaction.com.br/>>).

Segundo Navidi (2012), os métodos que tiram conclusões a partir de dados constituem o campo da estatística inferencial. O autor recomenda o estudo sobre os métodos de coleta de dados como amostragem e estatística descritiva para um melhor entendimento sobre a estatística inferencial.

Entretanto, Paradine (1974, p. 3) descreve que,

“o primeiro e adequado passo na análise de um grande conjunto de valores observados de uma variável consiste na sua classificação em termos de frequência com a qual ocorrem diferentes valores numéricos.”

Na análise estatística, uma amostra de valores observados é usada para estimar os parâmetros como média e desvio padrão, que caracterizam o conjunto de todos os valores possíveis da variável em questão. “É importante investigar como a precisão com que um parâmetro pode ser estimado depende do tamanho da amostra”. Pois, se os dados observados são apenas uma amostra de um conjunto bem maior, há uma possibilidade de que diferentes amostras produzam diferentes resultados. Esses problemas são tratados na Teoria das Probabilidades, que entra na teoria Estatística e em todo o campo da inferência científica. Consequentemente, antes de se trabalhar com amostras é necessário nos determos aos estudos da Teoria das Probabilidades (PARADINE, 1974).

“Além de sua aplicação em problemas de amostragem, a Teoria das Probabilidades conduz a certas “distribuições” de probabilidade que servem como modelos nos quais podem se encaixar as frequências de valores de uma variável” (PARADINE, 1974, p. 5).

Segundo Paradine (1974), esta matéria fornece um método de experimentação, particularmente adequado, quando há muitas causas de variação. Consiste na aplicação de um teste de significância para decidir se um conjunto de valores observados pode ser explicado com base na variabilidade inerente da quantidade média e no azar de amostragem, ou, sendo um efeito casual muito improvável, se as variações podem ser razoavelmente explicadas devido a alguma causa detectável.

3.9.1 Dados estatísticos

A estatística não centra seu interesse em objetos e sim na medida de propriedades que o identificam, caracterizam, descrevem, qualificam ou organizam o que se quer analisar (MILONE, 2004).

Os dados estatísticos consistem em um conjunto de valores de uma variável obtido através de uma mensuração (PARADINE, 1974).

Algumas variáveis apresentam-se como uma possível qualidade (atributo) do indivíduo ou objeto pesquisado e são denominadas de qualitativas. Outras variáveis assumem a forma de contagem ou mensuração e são chamadas de quantitativas (MORETTIN, 2011).

Segundo Morettin (2011), pode-se fazer distinção entre dois tipos de variáveis qualitativas. A nominal para a qual não existe nenhuma ordenação nas possíveis realizações e a variável qualitativa ordinal da qual existe uma ordem nos seus resultados.

As variáveis quantitativas são classificadas em quantitativas discretas, quando a mensuração assume a forma de contagem em números naturais, coisas ou unidades mínimas divisíveis. E denominadas de quantitativas contínuas quando podem ser indefinidamente subdivididas como por exemplo, o tempo, o comprimento e o peso (PARADINE, 1974; MORETTIN, 2011).

Segundo Paradine (1974), a variabilidade pode surgir tanto da diferença entre diferentes espécimes bem como da dificuldade de se efetuar medidas precisas, pois sabe-se que, devido a causas fora do nosso conhecimento e controle, observações repetidas da mesma quantidade produzem resultados discordantes (espera-se que essa discordância seja ligeira). O observador, em geral, calcula a média aritmética das leituras como a estimativa mais confiável do valor real.

Desta forma o autor sugere que as pequenas variações sejam atribuídas a causas desconhecidas (a menos que se opere alguma mudança definida no produto).

3.9.2 Amostragem

Segundo Navidi (2012), os métodos estatísticos consistem em analisar uma amostra obtida de uma população. Onde uma população é uma coleção inteira de peças ou resultados sobre os quais a informação é obtida. Já a amostra é um subconjunto de uma população que contém objetos ou resultados que são realmente observados. As populações que consistem em objetos físicos, como, clientes ou um lote de produtos, são denominadas de populações tangíveis. Estas são sempre finitas e após a amostragem o tamanho da população diminui.

Segundo Pocinho e Figueiredo (2008) a amostragem é o procedimento pelo qual um grupo de pessoas ou um subconjunto de uma população é escolhido com vista a obter informações relacionadas com um fenómeno, e de tal forma que a população inteira nos interessa esteja representada.

O Plano de Amostragem serve para descrever a estratégia a utilizar para selecionar a amostra. Este plano fornece os detalhes sobre a forma de proceder relativamente à utilização de um método de amostragem para determinado estudo (POCINHO & FIGUEIREDO, 2008).

Os melhores métodos de amostragem envolvem amostragem aleatória. Existem muitos métodos de amostragem aleatória, dos quais o mais básico é a amostragem aleatória simples, isto é, quando for igualmente provável a determinação do subconjunto, método frequentemente usado em muitas loterias (NAVIDI, 2012).

Entretanto, não for possível obter uma amostra aleatória, usa-se o método de amostra de conveniente. O problema é que este método pode diferir sistematicamente da população de alguma forma (NAVIDI, 2012).

Os itens de uma amostra são chamados de independentes se o conhecimento dos valores de alguns deles não ajudam a prever os valores de outros. Os itens de uma mostra aleatória simples podem ser tratados como independentes em muitas situações encontradas na prática. A exceção ocorre quando a população é finita e a amostra constitui uma fração substancial (mais do que 5%) da população (NAVIDI, 2012).

Na amostragem ponderada, para alguns itens são dadas chances maiores de serem selecionados do que outros, como uma loteria onde algumas pessoas tem mais apostas que outras (NAVIDI, 2012).

Na amostragem aleatória estratificada a população é dividida em subpopulações denominadas de estratos e uma amostra aleatória simples é retirada de cada estrato. A amostragem por grupos é utilizada quando a população é muito grande para que uma amostragem aleatória simples seja viável (NAVIDI, 2012).

Segundo BioEstat 5.3 (2007), amostras independentes: dois ou mais subconjuntos randômicos da população retirados de modo independente e de tamanhos iguais ou desiguais.

3.9.3 Medidas estatísticas

Para auxiliar a destacar as características importantes de uma amostra, são comumente usadas as medidas estatísticas de média amostral e o desvio padrão amostral. Onde a média fornece uma indicação do centro dos dados e o desvio padrão fornece uma indicação do quão dispersos são os dados (NAVIDI, 2012).

Também conhecida por média aritmética, ou simplesmente média, se trata da soma das observações divididas pelo número delas. (NAVIDI, 2012; MORETTIN, 2011; PARADINI, 1974).

Já a média geométrica é o centro de massa de um conjunto de dados sintetizável em uma progressão geométrica. Como a média aritmética, depende de todos os elementos e é sensível a quaisquer mudanças neles. Ao contrário dela, tende aos valores mais baixos e aos pontos de menor concentração. Por este motivo sua utilização é apropriada quando as mudanças relativas são mais importantes que as absolutas (NAVIDI, 2012).

Entretanto, segundo Morettin (2012), o resumo de um conjunto de dados por uma única medida representativa de posição central esconde toda a informação sobre a variabilidade do conjunto de observações.

Dentre outras medidas a raiz dos desvios quadrados médios da média, conhecida como desvio padrão e denotado por σ para população total e por s para uma amostra, é a medida mais adequada para se calcular a dispersão. Dentre as medidas de dispersão, o desvio padrão se destaca como a mais importante (PARADINI, 1974).

A variância e desvio padrão médio também são medidas de dispersão calculadas em relação à média das observações (MORETTIN, 2011).

Segundo Navidi (2012) a mediana, assim como a média é uma medida de centro. Entretanto é definida como o valor que divide o conjunto em dois grupos numericamente iguais (PARADINE, 1974).

Frequentemente a mediana é usada como uma medida de centro para amostras que contém valores atípicos (*outliers*). Pois, considere que uma amostra consiste dos valores 1, 2, 3, 4 e 20. A média é 6 e a mediana 3. Neste caso, pode se pensar que a mediana representa melhor o centro da amostra do que a média (NAVIDI, 2012).

Tanto a média como o desvio padrão podem não ser medidas adequadas para representar um conjunto de dados, pois: (a) são afetados, de forma exagerada, por valores extremos; (b) apenas com estes dois valores não temos ideia da simetria ou assimetria da distribuição dos dados (MORETTIN, 2011).

Uma medida de dispersão alternativa ao desvio padrão é a distância interquartil, definida como a diferença entre o terceiro e o primeiro quartis (MORETTIN, 2011).

Na prática, as medidas estatísticas descritas aqui são frequentemente calculadas usando um pacote de *software* de estatística (NAVIDI, 2012).

3.9.4 Testes de hipóteses

Segundo Barbetta (2010), normalmente os pesquisadores possuem ideias sobre o comportamento de uma variável, ou de uma possível associação entre variáveis. Nesses casos, deve-se planejar a pesquisa de forma que permita testar a veracidade de suas ideias sobre as populações estudadas. As ideias, ou hipóteses de pesquisa podem ser testadas por técnicas estatísticas denominadas de *testes de hipóteses* ou *testes de significância*.

Segundo Navidi (2012), pode-se realizar um teste de hipótese quando se tem amostras grandes e a intenção é de determinar se a diferença entre duas medias podem ser iguais a um valor especificado.

Nos casos em que se deseja estudar duas populações, precisa-se inicialmente saber como construir uma estatística de teste para então realizar testes de hipóteses e poder identificar se existe diferença entre duas médias, proporções ou variâncias (NAVIDI, 2012).

Barbetta (2010) comenta que para verificar estatisticamente a veracidade de uma hipótese precisa-se de um conjunto de dados e então formular as chamadas hipótese nula (H_0) e hipótese alternativa (H_1).

Onde a hipótese nula é a hipótese aceita como verdadeira, até prova estatística em contrário. Em geral, formulada em termos de igualdade entre parâmetros. Entretanto, quando os dados evidenciarem que H_0 é falsa, rejeita-se o teste, aceitando em seu lugar a hipótese alternativa, formulada em termos de desigualdades (BARBETTA, 2010).

Ou seja:

H_0 : população 1 = população 2

H_1 : população 1 \neq população 2

3.9.4.1 Teste *t de student*: dados amostrais

Segundo Navidi (2012), a distribuição *t* de Student foi descoberta em 1908 pelo estatístico William Sealy Gosset. Os gerentes da empresa em que trabalhava, o proibiram de publicar a descoberta em seu nome alegando que a informação era de propriedade da organização, para que assim seus concorrentes não pudessem perceber como os resultados poderiam ser úteis. Assim Gosset publicou seu trabalho usando o pseudônimo “Student”.

Segundo BioEstat 5.3 (2007), atualmente é um teste paramétrico de largo uso, sobretudo quando o tamanho das amostras – n_1 e n_2 – é igual ou inferior a 30 unidades e as variâncias paramétricas são desconhecidas. Não há impedimentos para valores amostrais superiores a 30.

Entretanto, segundo Navidi (2012), “a distribuição *t* de Student é apropriada sempre que a amostra vier de uma população que for aproximadamente normal”.

Os dados devem ser mensurados a nível intervalar ou de razões e as amostras podem ser de igual tamanho ou desiguais. As seguintes pressuposições devem ser observadas neste teste (BIOESTAT 5.3, 2007):

- a) amostras aleatórias de cada população investigada;
- b) as variâncias devem ser homogêneas;
- c) as variáveis das populações de onde as amostras foram selecionadas devem apresentar distribuição aproximadamente normal.

A questão das variâncias amostrais é importante no teste *t de Student*. Como as amostras podem ser de tamanho desiguais, é conveniente ponderar as duas variâncias mesmo que a pressuposição tenha sido feita de que elas não são heterogêneas. Outras vezes uma das variâncias é mais de quatro (4) vezes superior à outra ou, então, n_1 e n_2 são muito diferentes. O valor de *t* calculado para as duas médias amostrais – $1 X$ e $2 X$ – levará em consideração esses aspectos (BIOESTAT 5.3, 2007).

A estatística do teste toma como base a distribuição das diferenças entre as médias ($\mu_1 - \mu_2$), mas leva também em consideração o número de elementos em cada amostra e a variabilidade interna das amostras. Quanto maior o número de amostras, maior a evidencia de

uma possível diferença real. Todavia, “se há muita variabilidade entre os elementos de cada amostra, diferenças reais podem ficar nebulosas” (BARBETTA, 2010, p. 239).

3.9.4.2 ANOVA: um critério

A designação ANOVA deriva da expressão inglesa *ANalysis Of VAriance*. Teste estatístico cujo modelo de distribuição de probabilidades é o da variância para k amostras ou tratamentos independentes. As amostras podem ser do mesmo tamanho ou desiguais, no caso de k tratamentos (um critério), e devem ser iguais quando for o caso de k tratamentos e r blocos (dois critérios). Este teste é também conhecido como teste F , em homenagem Ronald Aylmer Fisher, e destina-se a comparar diferenças entre médias através das variâncias, cujos escores amostrais devem ser mensurados a nível intervalar ou de razões (BIOESTAT 5.3, 2007).

O teste F , assim como o teste t , necessita que as amostras sejam provenientes de populações normais. Ao contrário do teste t , o teste F para a comparação de variâncias é bastante sensível a essa suposição e por isso deve ser usado com cautela. Se o formato das populações forem muito diferente da curva normal, o teste F pode levar a resultados enganosos (NAVIDI, 2012). Segundo Barbetta (2010, p. 253) “essas suposições são mais fáceis de serem verificadas a partir dos resíduos (diferença entre valores observados e média dos grupos)”. Os valores dos resíduos podem ser apresentados em um gráfico de probabilidade normal dos resíduos, do qual valida a suposição de normalidade dos dados quando os pontos ficam distribuídos no entorno de uma linha.

A Análise de Variância compara a magnitude das variações de mais de duas amostras, decompondo a variância total em duas partes (BIOESTAT 5.3, 2007):

- 1^a) Entre as amostras, constituindo o chamado quadrado médio dos tratamentos;
- 2^a) Dentre cada tratamento, compondo o denominado quadrado médio do erro experimental.

O teste da análise da variância é a razão entre esses quadrados médios.

Segundo Navidi (2012), o ANOVA de um critério fornece uma forma de medir o espalhamento das várias amostras analisadas. Sendo as médias amostrais muito espalhadas, então é provável que as médias do tratamento sejam diferentes.

A designação “um critério” é pelo fato de se comparar somente as variações entre os tratamentos, cujo resultado é traduzido no valor do F . No caso de existir o interesse de verificar a magnitude da diferença, pode-se avaliar a diferença mínima significativa através do teste *Tukey*, comparando as diferenças entre as médias amostrais. Assim, as amostras podem ser do

mesmo tamanho ou desiguais. A probabilidade obtida na comparação entre as médias pelo teste de *Tukey* deve ser igual ou inferior a 0.01, para que a diferença seja estatisticamente significativa (BIOESTAT 5.3, 2007).

Em uma ANOVA de um critério, um teste *F* é usado para testar a hipótese nula de que todas as médias do tratamento são iguais. Se essa hipótese for rejeitada, podemos concluir que as médias do tratamento não são todas iguais. Mas o teste não nos diz quais são diferentes das restantes. Algumas vezes um pesquisador pode querer determinar todos os pares de médias que podem ser diferentes entre si. Neste caso, um tipo de procedimento, que é denominado método de comparação múltipla, tem que ser usado (NAVIDI, 2012, p. 433).

O método de *Tukey-Kramer* é o mais utilizado para as comparações múltiplas, envolve o cálculo de intervalos de confiança, denominados de intervalo de confiança simultâneos, para as diferenças entre cada par de médias do tratamento. Em função dos intervalos deste teste serem calculados ao nível $1 - \alpha$ e assim a probabilidade de todos os intervalos de confiança conterem simultaneamente seus valores verdadeiros é igual a $1 - \alpha$ (NAVIDI, 2012).

Estas técnicas estatísticas são utilizadas para identificar e quantificar efeitos aleatórios individuais em uma medição, de modo que possam ser apropriadamente levados em conta quando se avalia a incerteza do resultado da medição. Embora os métodos ANOVA sejam aplicáveis a uma ampla faixa de medições, por exemplo, a calibração de padrões de referência, tais como os padrões de tensão Zener e padrões de massa, e a certificação de materiais de referência, os métodos ANOVA por si só não podem identificar efeitos sistemáticos que possam estar presentes (Guia Para a Expressão da Incerteza de Medição – GUM, 2003, p.95).

Desta forma, calculando ANOVA para um critério, pode-se determinar através de um teste de hipóteses se há diferença significativa entre as médias dos tratamentos aplicados às amostras. Sendo que p-valor representa a probabilidade de ser obtida uma observação da distribuição F com $k-1$ e $k(n-1)$ graus de liberdade maior ou igual ao valor observado para F. Se p-valor for menor que α (alfa), rejeita-se H_0 . Ou seja, um p-valor pequeno indica uma forte evidência de que existem diferenças estatisticamente significativas entre as médias (DRUMOND et al., 1996).

3.10 Análise de componentes principais PCA (*Principal Component Analysis*)

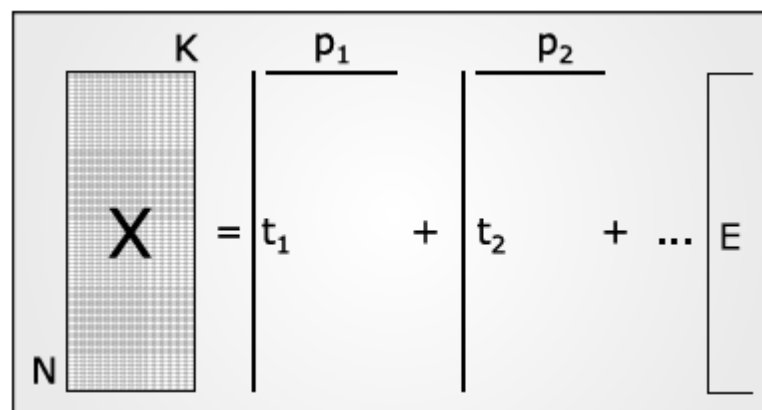
A Análise de Componentes Principais (PCA, do inglês, *Principal Component Analysis*) é uma ferramenta que permite extrair, de um determinado conjunto de dados, informações relevantes. Este conjunto de dados é organizado na forma de uma matriz (dados bidimensionais), onde as linhas podem ser amostras e as colunas variáveis. Neste sentido, com a PCA é possível efetuar uma simplificação, redução da

dimensão original dos dados, modelando, detecção de amostras anômalas (*outliers*), seleção de variáveis importantes em determinado sistema, classificação e previsão (WOLD et al., 1987 apud MATOS et al., 2003, p. 40).

Conforme Pasquini (2003), a PCA para dados bidimensionais é uma ferramenta quimiométrica bastante difundida e utilizada para diminuir a dimensão dos dados originais. Desta forma, uma gama infinita de informações pode ser organizada em uma matriz de dados X formada por N linhas e M colunas. As linhas podem ser, por exemplo, amostras de água de diferentes locais ou diferentes tratamentos, e as colunas (variáveis) podem ser características físico-químicas ou até mesmo concentrações de espécies químicas.

Na PCA, a dimensão dos dados originais é diminuída para um menor conjunto de dimensões chamadas de Componentes Principais – PC's. A partir dos PC's são gerados dois novos conjuntos de dados chamados de *scores* e *loadings*. Estes dois conjuntos trazem, respectivamente, informações sobre as amostras e as variáveis. A Figura 4 mostra a decomposição dos dados originais em *scores* (t) e *loadings* (p) e uma matriz de erros E. Ao se combinar os dados dos *scores* é possível efetuar um estudo mais criterioso dos dados originais sem perda de informações relevantes (PASQUINI, 2003).

Figura 4. Decomposição dos dados originais em scores (t) e loadings (p) e a matriz de erros E.



Fonte: Pasquini (2003).

Antes de aplicar a PCA a dados numéricos, é necessário efetuar algum tipo de pré-processamento nos dados originais. Os principais tipos de pré-processamento são o Centrado na Média – CM e o Autoescalamento – AS (2).

No CM calcula-se a média de cada variável, segundo a Equação 1:

$$\bar{x}_j = \frac{1}{n} \sum_{i=j}^n x_{ij} \quad (1)$$

Onde $\bar{\mathbf{X}}_j$ é a média dos dados \mathbf{X}_{ij} contidos em uma coluna (variáveis). Em seguida, subtrai-se os dados originais pela média calculada, Equação 2:

$$x_{ij(cm)} = x_{ij} - \bar{x}_j \quad (2)$$

No AS, calcula-se a variância dos dados S^2 conforme Equação 3:

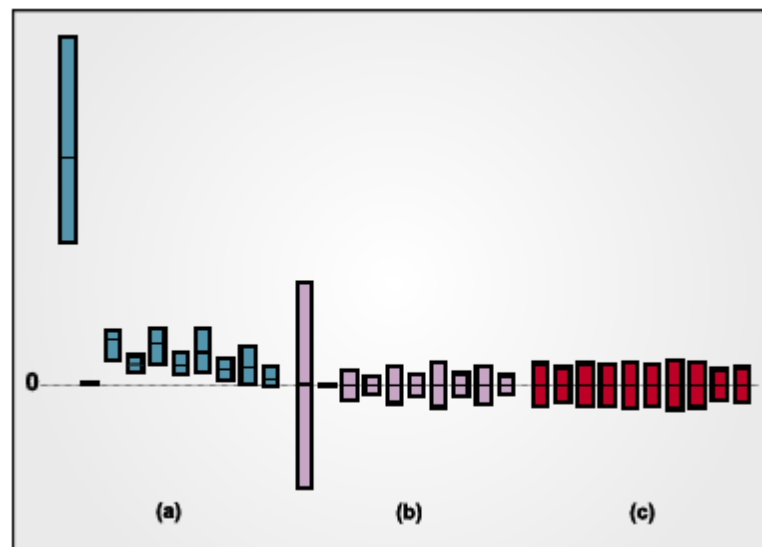
$$S_j^2 = \frac{1}{n-1} \sum_{i=j}^n (x_{ij} - \bar{x}_j)^2 \quad (3)$$

Em seguida, subtrai-se os dados originais pela média (ver equação 1), e divide-se pelo desvio padrão, segundo a Equação 4:

$$x_{ij(as)} = \frac{x_{ij} - \bar{x}_j}{S_j} \quad (4)$$

O pré-processamento CM é mais empregado para dados espectrais e o AS é empregado quando se quer dar a mesma importância para todas as variáveis. A Figura 5 mostra uma ilustração dos tipos de pré-processamento discutidos anteriormente. Na Figura 5a temos 10 variáveis (representadas por retângulos) e, como pode ser facilmente observado, elas possuem tamanhos muito diferentes. Ao se aplicar o pré-processamento centrado na média colocamos todas as médias em zero (ver Figura 5b). Já na Figura 5c temos os dados autoescalados e o tamanho das variáveis é basicamente o mesmo.

Figura 5. Tipos de pré-processamento.

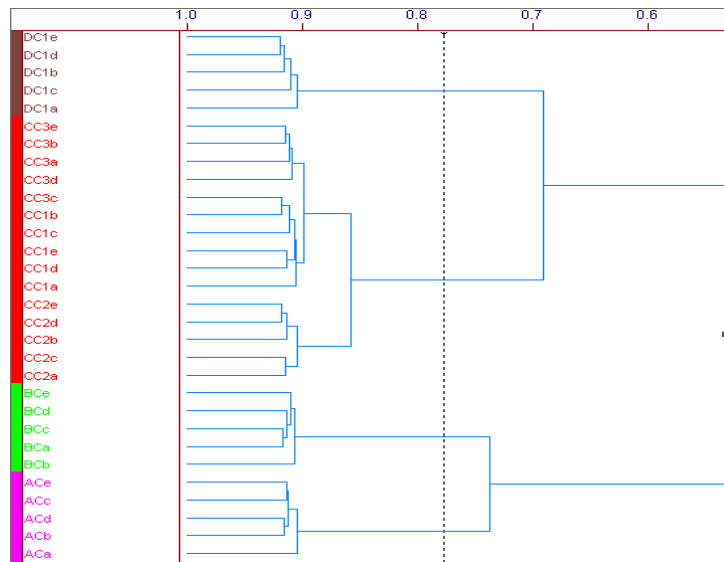


Fonte: Pasquini (2003).

3.11 Análise de agrupamentos hierárquicos (HCA)

Do inglês *Hierarchical clustering analysis* (HCA), é uma técnica de análise exploratória de dados multivariados. Tem por objetivo exibir os dados de forma que dê ênfase aos agrupamentos naturais presentes. Ao invés de tentar encaixar a informação em grupos pré-definidos, HCA procura descobri-los e evidenciá-los. O resultado da análise é exibido através de um dendrograma, onde a distância e parentesco entre as amostras representa o grau de similaridade entre elas. Quanto menor a distância e maior o parentesco, maior a semelhança. Um exemplo de dendrograma gerado pela análise HCA se encontra na Figura 6.

Figura 6. Exemplo de dendrograma HCA.



Fonte: Pasquini (2003).

Para calcular a similaridade entre as amostras, em geral, emprega-se o cálculo apresentado na Equação 5 da distância euclidiana entre elas.

$$d_{ab} = \left[\sum_j^m (x_{aj} - x_{bj})^2 \right]^{1/2} \quad (5)$$

Onde m é o número de variáveis que compõe uma amostra, a e b são amostras distintas e d é a distância. O próximo passo é transformar as distâncias em escalas de similaridade. Isto é realizado através da Equação 6:

$$similaridade_{ab} = 1 - \frac{d_{ab}}{d_{\max}} \quad (6)$$

Onde d_{ab} é a distância calculada entre as amostras a e b , e d_{max} é a maior distância calculada do conjunto de dados utilizado. Percebe-se então, que quanto maior a distância entre duas amostras, menor será a similaridade.

Após calculadas as similaridades, as amostras são então comparadas e agrupadas. O processo consiste em primeiro em criar um grupo com as duas amostras mais similares, ou seja, o par de amostras que possui a menor distância entre si. Após, a distância entre este novo grupo e as amostras é calculada, e amostras/grupos de maior semelhança são então agrupados. Isto é realizado até que todas as amostras tenham sido agrupadas. Esse processo é chamado de "linkagem".

3.12 Erros aleatórios e sistemáticos

Quando se relata o resultado de medição de uma grandeza física, é obrigatório que seja dada alguma indicação quantitativa da qualidade do resultado, de forma tal que aqueles que o utilizam possam avaliar sua confiabilidade. Sem essa indicação, resultados de medição não podem ser comparados, seja entre eles mesmos ou com valores de referência fornecidos numa especificação ou numa norma. É, portanto, necessário que haja um procedimento prontamente implementado, facilmente compreendido e de aceitação geral para caracterizar a qualidade de um resultado de uma medição, isto é, para avaliar e expressar sua incerteza (GALLAS, 2001, p. 1).

Quando todos os componentes de erro conhecidos ou suspeitos forem avaliados e as correções adequadas aplicadas, ainda permanece uma incerteza sobre quão correto é o resultado, isto é, uma dúvida acerca de quão corretamente o resultado da medição representa o valor da grandeza que está sendo medida. Gallas (2001) enfatiza as definições:

- **Incerteza de medição:** é o parâmetro associado ao resultado de uma medição, que caracteriza a dispersão dos valores que podem ser atribuídos ao mensurando (quantidade particular submetida a medição);
- **Incerteza padrão:** caracteriza-se pela incerteza do resultado de uma medição expressa como um desvio padrão;
- **Incerteza padrão combinada:** é quando o resultado é obtido por meio dos valores de várias outras grandezas, sendo igual a raiz quadrada positiva de uma soma de termos, que constituem as variâncias e covariâncias destas outras grandezas, ponderadas de acordo com quanto o resultado da medição varia com mudanças nestas grandezas.

Segundo o Guia para a Expressão da Incerteza de Medição (GUM, 2003), existem muitas fontes possíveis, incluindo:

- a) Definição incompleta do mensurando;
- b) Realização imperfeita da definição do mensurando;

- c) Amostragem não representativa – a amostra medida pode não representar o mensurando definido;
- d) Conhecimento inadequado dos efeitos das condições ambientais sobre a medição ou medição imperfeita das condições ambientais;
- e) Erro de tendência pessoal na leitura de instrumentos analógicos;
- f) Resolução finita do instrumento ou limiar de mobilidade;
- g) Valores inexatos dos padrões de medição e materiais de referência;
- h) Valores inexatos de constantes e de outros parâmetros obtidos de fontes externas e usados no algoritmo de redução de dados;
- i) Aproximações e suposições incorporadas ao método e procedimento de medição;
- j) Variações nas observações repetidas do mensurando sob condições aparentemente idênticas.

Essas fontes não são necessariamente independentes e algumas das fontes como as descritas de a) a i) podem contribuir para a fonte de incerteza descrita em j). Naturalmente, um efeito sistemático não reconhecido não pode ser levado em consideração na avaliação da incerteza do resultado de uma medição, porém contribui para seu erro (GUM, 2003).

Uma medição tem imperfeições que dão origem a um erro no resultado. Tradicionalmente, um erro é visto como tendo dois componentes: um componente aleatório e um componente sistemático. Destaca-se que o erro é um conceito idealizado, dos quais não podem ser conhecidos exatamente (GUM, 2001).

Conforme o Guia para a Expressão da Incerteza de Medição (GUM, 2003, p.15), “o erro aleatório presumivelmente se origina de variações temporais ou espaciais, estocásticas ou imprevisíveis, de grandezas de influência”. Os efeitos de tais variações, denominados efeitos aleatórios, são a causa de variações em observações repetidas do mensurando. Embora não seja possível compensar o erro aleatório de um resultado de medição, ele pode geralmente ser reduzido aumentando-se o número de observações; sua esperança ou valor esperado zero.

O desvio padrão experimental da média aritmética ou média de uma série de observações não é o erro aleatório da média, mas sim uma medida da incerteza da média devida a efeitos aleatórios. O valor exato do erro na média, que se originam destes efeitos, não pode ser conhecido (GUM, 2003).

O erro sistemático ou efeito sistemático, como o erro aleatório, não pode ser eliminado, porém ele também, frequentemente, pode ser reduzido. Se um erro sistemático se origina de um efeito reconhecido de uma grandeza de influência em um resultado de medição, o efeito pode

ser quantificado e, se for significativo com relação à exatidão requerida da medição, uma correção ou fator de correção pode ser aplicado para compensar o efeito. Supõe-se que, após esta correção, a esperança ou valor esperado do erro provocado por um efeito sistemático seja zero.

A incerteza de uma correção aplicada a um resultado de medição, para compensar um efeito sistemático, não é o erro sistemático no resultado de medição. Este efeito sistemático é frequentemente denominado tendência. “É uma medida da incerteza do resultado devido ao conhecimento incompleto do valor requerido da correção. O erro originado da compensação imperfeita de um efeito sistemático não pode ser exatamente conhecido” (GUM, 2003, p. 15).

Segundo Barbetta (2010), os erros associados a um teste de hipótese se dão em função de se tomar uma decisão considerando as evidências nas amostras analisadas, mas nunca se saberá com certeza sobre a realidade da população. Entretanto, estabelecidas as hipóteses e o nível de significância α , tem-se as possibilidades apresentadas na Tabela 1.

Tabela 1 - Erros associados a decisão do teste de hipóteses.

Realidade (desconhecida)	Decisão do teste	
	Aceita H_0	Rejeita H_0
H_0 verdadeira	Decisão correta (probabilidade = $1 - \alpha$)	Erro tipo I (probabilidade = α)
H_0 falsa	Erro tipo II (probabilidade = β)	Decisão correta (probabilidade = $1 - \beta$)

Fonte: Barbetta (2010).

Como α é fixado a priori, se o teste rejeita H_0 em favor de H_1 ($p \leq \alpha$), o risco de estarmos tomando a decisão errada (erro tipo I) fica limitado pelo nível de significância α adotado. Desta forma, temos certa garantia da veracidade de H_1 . Em outras palavras quando o teste rejeita H_0 , podemos afirmar (considerando o nível de significância do teste) que H_1 é verdadeira (Barbetta, 2010, p. 206).

Todavia, quando o teste aceita H_0 ($p > \alpha$) não se tem muito controle do risco de estar tomando a decisão errada (erro tipo II) em função da probabilidade β , em geral, não ser conhecida. Ou seja, H_0 pode não ser realmente a hipótese verdadeira, mas os dados não estão mostrando evidências suficientes para rejeitá-la (BARBETTA, 2010).

REFERÊNCIAS

- AFUBRA. *Associação dos Fumilcutores do Brasil*. 2012. Apresenta informações gerais sobre a instituição. Disponível em: <2012, <http://www.afubra.com.br/>>. Acesso em: 10 de jul. de 2012.
- ASTM D3764 – *Standart Practice for Validation of Process Stream Analyzer Systems*. 2009. Disponível em: <<http://www.astm.org/Standards/D3764.htm>>. Acesso em: 31 de out. de 2012.
- BARBOSA, Luiz Cláudio de Almeida. *Espectroscopia no infravermelho na caracterização de compostos orgânicos*. Viçosa: Universidade Federal de Viçosa, p. 189, 2007.
- BIOESTAT 5.3. Aplicações Estatísticas nas Áreas das Ciências Biomédicas. Pará, Brasil, p. 1-364, 2007. Disponível em: <http://www.mamiraua.org.br/pt-br/downloads/programas/bioestat-versao-53/>. Acesso em: 20 de set. de 2013.
- CAI, W., LI, Y., SHAO, X. *A variable selection method based on uninformative variable elimination for multivariate calibration of near-infrared spectra*. *Chemometrics and Intelligent Laboratory Systems* 90 (2), p. 188-194, 2008.
- COHN, ESTÉFANO PEDRO. *Analisadores Industriais no processo, na area de utilidades, na supervisão da emissão de poluentes e na segurança*. Rio de Janeiro: Interciência: IBP, p. 788, 2006.
- DRUMOND, F. B., WERKEMA, M. C. C., AGUIAR, S. *Análise de variância: comparação de várias situações*. Belo horizonte, MG: Fundação Christiano Ottoni, Escola de Engenharia da UFMG, p, 276, 1996.
- DUAN, J., HUANG, Y., LI, Z., ZHENG, B., LI, Q., XIONG, Y., WU, L., MIN, S.; *Determination of 27 chemical constituents in Chinese southwest tobacco by FT-NIR spectroscopy*. *Industrial Crops and Products*, v. 40, p. 21-26, 2012.
- FAN, W., LIANG, Y., YUAN, D., WANG, J. *Calibration model transfer for near-infrared spectra based on canonical correlation analysis*. *Analytica Chimica Acta* 623 (1), p. 22-29, 2008.
- GARATTONI, Bruno. *O novo cigarro*. Revista Super Interessante, Editora Abril S.A., v. 253, 2008. Disponível em: <<http://super.abril.com.br/tecnologia/novo-cigarro-447541.shtml>>. Acesso em: 10 mai. 2013.
- GARCIA-ALLENDE, P.B., CONDE, O.M., MIRAPEIX, J., COBO, A., LOPEZ-HIGUERA, J.M. *Quality control of industrial processes by combining a hyperspectral sensor and Fisher's linear discriminant analysis*. *Sensors and Actuators, B: Chemical* 129 (2), p. 977-984, 2008.
- GAUGETOOLSXL. *User Guide. NDC Infrared Engineering LTD*, Publication Reference: 105/14033-01SA Issue 5. United Kingdom, p. 1-73, 2009.

GUM - *Guia para a Expressão da Incerteza De Medição*. 3ª edição brasileira em língua portuguesa – Rio de Janeiro: ABNT, INMETRO, p. 131, 2003.

JING, M., CAI, W., SHAO, X. *Multiblock partial least squares regression based on wavelet transform for quantitative analysis of near infrared spectra*. *Chemometrics and Intelligent Laboratory Systems* 100 (1), p. 22-27, 2010.

LIGOCKY, Ingrid. *Controle das características do fumo através de métodos analíticos*. 1987. 53 f. Dissertação (Pós-Graduação em Química) - Faculdade Integradas de Santa Cruz do Sul, Santa Cruz do Sul, 1987.

MATOS, G.D., PEREIRA FILHO, E.R., POPPI, R.J., et al. *Análise exploratória em Química Analítica com emprego de quimiometria: PCA e PCA de imagens*. *Revista Analytica*, nº6, p. 38-50, Universidade estadual de Campinas – UNICAMP, 2003.

MEDCALC: *easy-to-use statistical software*. *Ostend: MedCalc Software bvba*. v. 12.7.0, 1993-2013. Disponível em: <<http://www.medcalc.org/manual/youdenplot.php>>. Acesso em: 19 jun. 2013.

MOREIRA, E.D.T., PONTES, M.J.C., GALVÃO, R.K.H., ARAÚJO, M.C.U. *Near infrared reflectance spectrometry classification of cigarettes using the successive projections algorithm for variable selection*, *Talanta* 79 (5), pp. 1260-1264, 2009.

NDC TOBACCO. *710e The Measure of Quality™*. NDC Infrared Engineering LTD 00/14221-01 Iss 3. United Kingdom, p. 1-6, jan. 2009.

NI, L.-J., ZHANG, L.-G., XIE, J., LUO, J.-Q. *Pattern recognition of Chinese flue-cured tobaccos by an improved and simplified K-nearest neighbors classification algorithm on near infrared spectra*. *Analytica Chimica Acta* 633 (1), p. 43-50, 2009.

PASQUINI, C. *Near infrared spectroscopy: fundamentals practical aspects and analytical applications*. *J. Braz. Chem. Soc.* v.14, Brasil, p. 138-219, 2003.

PROFINETE IO. *Industrial Communication Protocol In 710e Gauges*. NDC Infrared Engineering LTD, Publication Reference: 105/14728-01 Issue 2, United Kingdom, p. 1-25, 2009.

POCINHO, M.; Figueiredo, J. P. *Estatística e Bioestatística*. 2008. Disponível em: <<http://biocistron.blogspot.com/search/label/Bioestat%C3%ADstica>>. Acesso em 23 fev. 2008.

PORTAL ACTION. Copyright: 1997 - 2011, Estatcamp. Disponível em: <<http://www.portalaction.com.br/>>. Acesso em: 06 de out. 2012.

SCHUNKE, Iro. *No passo da Tecnologia, a mudança de perfil*. SindiTabacoNews. Julho, agosto e setembro de 2013. Disponível em: <<http://sinditabaco.com.br/sinditabaco-news-julhoagostoseptembro-2013/>>. Acesso em: 20 de jan. 2014.

SINDITABACO. *Sindicato dos fumicultores do Brasil*. 2012. Apresenta informações gerais sobre a instituição. Disponível em: <<http://sinditabaco.com.br/>>. Acesso em: 10 jul. 2012.

SILVA, E. M. et al. *Pesquisa operacional: programação linear*. 3 ed. São Paulo: Atlas, 1998.

SILVERSTEIN, R. M.; WEBSTER, F. X.; KIEMLE, D. J. *Identificação Espectrométrica de Compostos Orgânicos*. 7 ed. Rio de Janeiro: LTC, 2007.

SMITH, David W. (Org.). *Flue-Cured Tobacco Information*. North Carolina Cooperative Extension Service: Phillips Brothers Printing, Springfield, Illinois, 2005.

TAN, C., LI, M., QIN, X. *Random subspace regression ensemble for near-infrared spectroscopic calibration of tobacco samples*. Analytical Sciences 24 (5), p. 647-653, 2008.

TAN, C., QIN, X., LI, M. *Comparison of chemometric methods for brand classification of cigarettes by near-infrared spectroscopy*. Vibrational Spectroscopy 51 (2), p. 276-282, 2009.

TM710e™. *Technical Reference Manual*. NDC. Publication Reference: 105/13050-01SA, Infrared Engineering LTD, United Kingdom, p. 1-200, 2009.

TREVISAN, M. G.; POPPI, R. J. *Química Analítica de Processos*. Quim. Nova, Vol. 29, Nº 5, pp. 1065-1071, 2006.

YE, S., WANG, D., MIN, S. *Successive projections algorithm combined with uninformative variable elimination for spectral variable selection*. Chemometrics and Intelligent Laboratory Systems 91 (2), pp. 194-199, 2008.

ZHANG, Y., CONG, Q., XIE, Y., JINGXIUYANG, ZHAO, B. *Quantitative analysis of routine chemical constituents in tobacco by near-infrared spectroscopy and support vector machine*. Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy 71 (4), pp. 1408-1413, 2008.

