

UNIVERSIDADE FEDERAL DO MARANHÃO
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA

MARCELO JOSÉ CASTRO DOS SANTOS

ESTIMATIVA DA MASSA ESPECÍFICA EM ETANOL COMBUSTÍVEL POR
MODELOS DE REDES NEURAIS ARTIFICIAIS E DE REGRESSÃO POR
MÍNIMOS QUADRADOS PARCIAIS

SÃO LUÍS - MA

2013

UNIVERSIDADE FEDERAL DO MARANHÃO
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
DEPARTAMENTO DE TECNOLOGIA QUÍMICA
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA

ESTIMATIVA DA MASSA ESPECÍFICA EM ETANOL COMBUSTÍVEL POR
MODELOS DE REDES NEURAIS ARTIFICIAIS E DE REGRESSÃO POR
MÍNIMOS QUADRADOS PARCIAIS

MARCELO JOSÉ CASTRO DOS SANTOS

Dissertação apresentada ao Programa de Pós-Graduação em Química da Universidade Federal do Maranhão, como um dos requisitos para obtenção do título de Mestre em Química, na área de concentração: Química Analítica.

Aldaléa Lopes Brandes Marques

Orientadora

Edmar Pereira Marques

Co-Orientador

SÃO LUÍS - MA

2013

Santos, Marcelo José Castro dos.

Estimativa da massa específica em etanol combustível por modelos de redes neurais artificiais e de regressão por mínimos quadrados parciais/ Marcelo José Castro dos Santos – São Luís, 2013.

52 f.

Orientadora: Aldaléa Lopes Brandes Marques.

Co-orientador: Edmar Pereira Marques.

Dissertação (Mestrado) – Universidade Federal do Maranhão, Curso de Pós-graduação em Química, 2013.

1. Etanol. 2. Redes Neurais. 3. Biocombustível. I. Título.

CDU 547.262

ESTIMATIVA DA MASSA ESPECÍFICA EM ETANOL COMBUSTÍVEL POR
MODELOS DE REDES NEURAIS ARTIFICIAIS E DE REGRESSÃO POR
MÍNIMOS QUADRADOS PARCIAIS

MARCELO JOSÉ CASTRO DOS SANTOS

Dissertação apresentada ao Programa de Pós-Graduação em Química da Universidade Federal do Maranhão, como um dos requisitos para obtenção do título de Mestre em Química, na área de concentração: Química Analítica.

BANCA EXAMINADORA

Prof^a Dr^a. Aldaléa Lopes Brandes Marques – UFMA/DETQI (Orientadora)

Prof. Dr. Edmar Pereira Marques – UFMA/DEQUI (Co-Orientador)

Prof. Dr. Heronides Adonias Dantas Filho – UFPA/ICEN/FQ

AGRADECIMENTOS

Agradeço a Deus pela sua infinita misericórdia e graça na minha vida.

À minha amada namorada Lene pelo amor, incentivo, paciência e por acreditar em mim sempre.

Agradeço a minha família por sempre me apoiar e o incentivo na busca dos meus sonhos.

Agradeço àqueles que contribuíram de forma direta nesse trabalho, ao Alex e a Isabelle que me ajudaram, ensinaram e têm um papel fundamental neste trabalho.

Agradeço a todos os amigos que contribuíram indiretamente neste trabalho, especialmente dos laboratórios LPQA E LAPQAP pelo incentivo e em busca do sucesso acadêmico.

Ao exemplo de profissional, minha orientadora Prof^a Dr^a Aldaléa, pelas valiosas discussões, confiança, orientação, incentivo e amizade. Agradeço ainda por me proporcionar realizar outras atividades que contribuíram para o meu desenvolvimento profissional e pessoal.

Aos laboratórios LPQA E LAPQAP pela infraestrutura oferecida.

SUMÁRIO

1. INTRODUÇÃO.....	1
1.1. Proálcool.....	2
1.2. Etanol Combustível.....	4
1.3. Controle de Qualidade e Especificações do Etanol.....	5
1.4. A Matriz Energética e as Questões Ambientais.....	7
1.5. Monitoramento da Qualidade de Etanol Combustível no Estado do Maranhão ...	8
1.6. Métodos Alternativos voltados para a Qualidade do Etanol	10
1.7. Redes Neurais Artificiais	11
1.7.1. <i>Redes Neurais Artificiais de Múltiplas Camadas</i>	12
1.7.2. Retropropagação de Erros ou “Backpropagation”	13
1.8. Regressão por Mínimos Quadrados Parciais (PLS).....	14
1.9. RNA e PLS em Combustíveis - Revisão Bibliográfica.....	16
2. OBJETIVOS	19
2.1. Objetivo Geral	19
2.2. Objetivos Específicos	19
3. METODOLOGIA	20
3.1. Massa Específica.....	20
3.2. Teor Alcoólico.....	21
3.3. Condutividade Elétrica.....	21
3.4. Potencial Hidrogeniônico (pH)	22
3.5. Programação.....	23
4. RESULTADOS E DISCUSSÃO.....	24
4.1. Estudo Estatístico Preliminar	24
4.2. Estimativa da Massa Específica por RNAs de Múltiplas Camadas	27
4.3. Validação – RNAs de Múltiplas Camadas.....	31
4.4. Regressão por Mínimos Quadrados Parciais (PLS).....	32
4.5. Validação - PLS.....	34
5. CONCLUSÃO.....	36
6. REFERÊNCIAS BIBLIOGRÁFICAS.....	37

LISTA DE FIGURAS

Figura 1: Diminuição de não conformidades de etanol hidratado combustível nos anos de 2002 a 2012.....	10
Figura 2: Rede neural de múltiplas camadas com quatro variáveis na camada de entrada, duas camadas intermediárias contendo três neurônios em cada uma e uma camada de saída.....	12
Figura 3: Densímetro digital ANON PAAR DMA 4500.....	21
Figura 4: Condutivímetro DIGIMED DM 32.....	22
Figura 5: pHmetro METROHM 827 lab.....	23
Figura 6: Comportamento da massa específica de 2002 ao 1º trimestre de 2013.....	24
Figura 7: Comportamento de pH (6,0 a 8,0) de 2002 ao 1º trimestre de 2013.....	25
Figura 8: Comportamento de condutividade elétrica de 2002 ao 1º trimestre de 2013.....	26
Figura 9: Topologia da RNA obtida para estimar a massa específica do etanol hidratado combustível.....	27
Figura 10: Valores estimados pela rede neural artificial de múltiplas camadas e os valores medidos da massa específica (g/cm^3) na fase de treinamento.....	28
Figura 11: Valores reais de massa específica versus valores estimados pelo modelo neural de múltiplas camadas na fase de treinamento.....	28
Figura 12: Valores dos resíduos do modelo neural na fase de treinamento.....	29
Figura 13: Valores estimados pela rede neural artificial de múltiplas camadas e os valores medidos da massa específica (g/cm^3) na fase de teste.....	29
Figura 14: Valores dos resíduos obtidos na fase de teste do modelo com as estimativas da massa específica (g/cm^3) do etanol hidratado combustível.....	30
Figura 15: Valores estimados pela rede neural artificial de múltiplas camadas na fase de validação e os valores medidos da massa específica (g/cm^3).....	31
Figura 16: Valores dos resíduos das estimativas do modelo de rede neural de múltiplas camadas na fase de validação.....	32
Figura 17: Valores estimados da massa específica (g/cm^3) por PLS versus valores medidos.....	33
Figura 18: Valores dos resíduos das estimativas do modelo PLS no período de 2002 a 2011.....	33

Figura 19: Valores estimados pela análise de regressão PLS versus valores medidos da massa específica (g/cm^3) medida.....	35
Figura 20: Valores dos resíduos estimados por PLS da massa específica (g/cm^3) para o ano de 2013 (validação).....	35

LISTA DE TABELA

Tabela 1: Variâncias explicadas e acumuladas nas VLS.....	33
---	----

LISTA DE SIGLAS E SÍMBOLOS

ANN – ARTIFICIAL NEURAL NETWORK

ABNT – ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS

ANP – AGÊNCIA NACIONAL DE PETRÓLEO, GÁS NATURAL E BIOCOMBUSTÍVEIS

ASTM – AMERICAN SOCIETY OF TESTING AND MATERIALS

ECS – EUROPEAN COMMITTEE STANDARDIZATION

EHC – ETANOL HIDRATADO COMBUSTÍVEL

EN – NORMA EUROPÉIA

FSADU – FUNDAÇÃO SOUSÂNDRADE EM APOIO AO DESENVOLVIMENTO DA UFMA

GEE – GASES DO EFEITO ESTUFA

LAQPAP – LABORATÓRIO DE ANÁLISE E PESQUISA EM QUÍMICA ANALÍTICA DE PETRÓLEO E BIOCOMBUSTÍVEIS

NBR – NORMAS BRASILEIRAS

NC – NÃO CONFORMIDADE

pH – POTENCIAL HIDROGENIÔNICO

PLS – PARTIAL LEAST SQUARE

PMQC – PROGRAMA NACIONAL DE MONITORAMENTO DE QUALIDADE DE COMBUSTÍVEIS

PROÁLCOOL – PROGRAMA BRASILEIRO DO ÁLCOOL

SIMCA – SOFT INDEPENDENT MODELLING OF CLASS ANALOGY

UNICA – UNIÃO DA INDÚSTRIA DE CANA-DE-AÇÚCAR

RESUMO

O etanol tem alcançado crescente interesse em muitos países, principalmente, no Brasil devido ao programa PROÁLCOOL. A determinação experimental das propriedades deste biocombustível e de outros combustíveis por meio de métodos oficiais é muito demorada, bem como é considerado um tedioso processo. A estimativa dessas propriedades com a ajuda de ferramentas computacionais pode ser de grande utilidade. No presente trabalho, os métodos de regressão por mínimos quadrados parciais (PLS) e redes neurais artificiais de múltiplas camadas (RNA) foram usados para estimar uma das mais importantes propriedades do etanol combustível, massa específica, utilizando parâmetros de qualidade oficiais de etanol, oriundos de análises realizadas no laboratório LAPQAP/UFMA, durante 12 anos (período: 2002-2013). Inicialmente, uma análise cuidadosa dos dados foi realizada a fim de selecionar um conjunto de variáveis e dados que melhor representasse um desempenho satisfatório dos dois modelos estudados. As estimativas de ambas as abordagens foram comparadas e validadas. A capacidade preditiva da rede neural obtida foi considerada muito boa para os parâmetros estudados, e compatível com a precisão das medidas experimentais. O baixo erro quadrático médio, a aleatoriedade, a média nula e a variância constante, obtida para os resíduos, evidenciaram a adequabilidade dos modelos usados, sugerindo a utilização destes modelos para estimar (predizer) a massa específica do etanol. Resultados indicaram que o modelo de RNA foi adequado, sendo o valor de NMSE (erro quadrático médio normalizado) de 0,0012, valor este, muito inferior ao modelo de PLS de 0,2221. Este resultado alcançado é inferior aos valores da faixa de incerteza de medição do equipamento responsável pelo ensaio experimental da massa específica, comprovando que o modelo utilizado possui desempenho considerado muito bom.

Palavras Chaves: etanol, biocombustível, predição, massa específica, redes neurais, método de mínimos quadrados parciais.

ABSTRACT

The ethanol has continuously gained interests in many countries including Brazil due to the PROÁLCOOL program. The experimental determination of properties of ethanol and other fuels through official methods is very time consuming as well as tedious process. The estimation of these properties with the help of computational tools can be very useful. In the present work, the methods of partial least squares regression (PLS) and artificial neural network multilayer (ANN) were used to estimate one of the most important properties of fuel ethanol, density, using official quality parameters for ethanol, collected from LAPQAP/UFMA laboratory corresponding to 12 years (period: 2002-2013) of analyzes. A careful analysis of the data was performed to obtain a set of variables and data that best represents satisfactory performance of the two models. The estimates of both approaches were compared and validated. The predictive ability of the network obtained was very good for the parameters studied, consistent with the accuracy of the experimental measurements. The low mean square error, the randomness, the zero mean and the constant variance, obtained for the residues, indicated the suitability of the models, suggesting their use to estimate (predict) the density of ethanol. Results indicated that the model ANN was adequate, and the value of NMSE (normalized mean square error) of 0.0012, less than the PLS model of 0.2221. The result achieved is less than the range of measurement uncertainty of the equipment responsible for testing the density proving that the model used has satisfactory performance.

Keywords: ethanol, biofuel, predict, density, artificial neural networks, partial least squares regression.

1. INTRODUÇÃO

O etanol é frequentemente usado como matéria-prima para as bebidas, cosméticos, produtos farmacêuticos e na indústria química. Devido à variedade de recursos renováveis disponíveis no Brasil, o etanol é utilizado como um combustível de veículo automotivo e pode ser encontrado na forma anidra (misturado com a gasolina) e como etanol hidratado combustível.

O Brasil ocupa a posição de maior produtor mundial de etanol de cana-de-açúcar, além da posição de liderança na tecnologia de sua produção. A destacada produtividade com custos de produção bem inferiores aos dos concorrentes internacionais é devido aos avanços tecnológicos alcançados nacionalmente. Este sucesso é, em grande parte, devido a liderança e competitividade alcançados e ao trabalho de muitos anos desenvolvidos pelos grupos de pesquisa das instituições públicas e privadas, que tem resultado em alto e valioso grau de conhecimento e tecnologia sobre a cana, seus derivados, qualidade dos produtos e sobre o processo de fabricação do etanol de cana.

A produção do etanol combustível é feita sob as diretrizes da Agenda 21 que busca constante sinergia baseada, tanto no setor sucroenergético como nos propósitos de sustentabilidade ambiental estabelecida pela Agenda 21. As práticas agrícolas de superação da queima, pré-colheita, manejo de solos, recuperação de áreas de proteção permanente, reutilização de resíduos industriais, entre outras, são uma realidade constante de parte expressiva das unidades de produção do setor, formalizada em protocolos ambientais e mecanismos de certificação como as ISOs ¹.

Para garantir a comercialização do etanol no País é importante determinar limites de contaminantes presentes para uma melhor qualidade do produto final. Para avaliar a qualidade desse combustível a agência reguladora ANP tem estabelecido resoluções e especificações para alguns parâmetros físico-químicos do etanol. A massa específica é um dos parâmetros mais utilizados para o controle de qualidade de etanol hidratado, uma vez que está diretamente relacionada ao teor de água contido no produto. O teor de etanol também é um parâmetro importante para a avaliação da qualidade, uma vez

que está relacionado com a pureza do etanol combustível e a possíveis contaminações com água, hidrocarbonetos, metanol e alcoóis superiores ².

Outros parâmetros como condutividade elétrica e pH têm sua importância e todos estes ensaios experimentais avaliam a conformidade do etanol.

Através do Programa Brasileiro de Álcool (PROÁLCOOL), criado em 1975, que o etanol começou a substituir de forma significativa o uso da gasolina, cujos preços se elevaram subitamente durante a crise mundial de petróleo. Atualmente, cerca de 90% dos veículos comercializados no País são do tipo “*flex fuel*”, que permitem o uso de gasolina e etanol em qualquer proporção ².

O objetivo central desta dissertação envolve o estudo da aplicação de ferramentas estatísticas e computacionais na predição de um importante parâmetro de qualidade do etanol combustível, que é a massa específica.

Antes dos resultados obtidos pelo estudo desenvolvido, algumas considerações são apresentadas acerca do padrão e especificação deste biocombustível, bem como o estado da arte dos aspectos teóricos e metodológicos sobre a qualidade do etanol combustível.

1.1. Proálcool

O Brasil é o maior produtor mundial de etanol de cana-de-açúcar segundo a União da Indústria de Cana-de-Açúcar (ÚNICA), sendo responsável por 20% da produção mundial e 20% das exportações mundiais ³. Mas para chegar a esse patamar foi desenvolvido um programa para atender a produção de etanol anidro combustível proveniente de fonte renovável, automóveis movidos a etanol e posteriormente a tecnologia “*flex fuel*”.

O PROÁLCOOL foi um programa criado pelo governo brasileiro em 1975 como resposta a 1ª crise do petróleo. O álcool que era utilizado apenas

como regulador do mercado de açúcar passava a ser visto como uma alternativa energética ao petróleo.

O programa teve um forte caráter intervencionista na economia e editou uma série de medidas de incentivo a uso e produção do álcool combustível. Dentre essas medidas, pode-se citar: subsídios para instalação de destilarias anexas às usinas de açúcar, obrigatoriedade da adição de álcool à gasolina, incentivos à produção de carros a álcool, medidas protecionistas em relação ao álcool importado, controle dos preços e incentivos fiscais, e a pesquisa e desenvolvimentos nos campos de produção e uso de álcool combustível ⁴.

Em 1979, o Governo Federal decidiu reformular aspectos do PROÁLCOOL para estimular a produção do etanol hidratado, que viria a ser usado diretamente em motores de veículos de passageiros especialmente desenvolvidos para este fim, mas só em meados daquele ano o governo federal através do Decreto 83.700 viabilizou a segunda etapa do Proálcool, em 5 de julho de 1979, com a produção de etanol hidratado combustível (EHC) para ser utilizado no recém lançado “carro a álcool”; que tratava-se de automóveis com motores ciclo Otto que foram modificados para receber 100% de álcool hidratado ⁵.

Com o aumento da produção de etanol hidratado, a diminuição de importação de petróleo do Brasil as vendas de carros a álcool seguiu uma trajetória ascendente até 1986. Coube ao governo estimular a demanda de carros à álcool, mas surgiram alguns problemas na primeira geração de carros a álcool ⁶. O preço da queda do petróleo, aumento da procura da gasolina, a redução do teor de etanol anidro na gasolina foram alguns fatores que desencadearam uma crise no setor sucroalcooleiro.

De 1986 a 2002 praticamente passou-se a produzir o etanol anidro adicionado a gasolina A que ao longo desses anos o programa estava em certa dormência no setor industrial devido a importação de metanol, a diminuição das vendas de veículos a álcool e a alta dos preços do açúcar no mercado internacional.

No ano de 2004, inicia-se com força o terceiro e atual ciclo do etanol combustível; consolidando uma nova revolução tecnológica no mercado, os veículos com motores “*flex fuel*” que utilizam gasolina e/ou etanol puro ou em qualquer teor de mistura. Essa inovação tecnológica genuinamente nacional, desenvolvida ao longo de anos de estudos, com a contribuição do período de amadurecimento e incentivos do PROÁLCOOL ¹.

Em razão do sucesso dessa nova tecnologia a ascensão do mercado de veículos “*flex-fuel*”, preços competitivos com a gasolina e uma acentuada produção de etanol combustível a partir de 2009 passa a ser consumido em volume próximo à gasolina, considerando o volume comercializado de etanol anidro e hidratado.

Com novas exigências de mercado em formação para uma comercialização eficiente e investimentos em pesquisa e desenvolvimento houve a necessidade da criação da Agência Nacional de Petróleo (ANP) em agosto de 1997 com o objetivo de especificar a qualidade do etanol combustível, seja aquele adicionado a gasolina (anidro) ou aquele revendido diretamente nos postos de abastecimento (hidratado) e regular as atividades de distribuição e revenda de combustíveis ⁷.

Atualmente as normas da ANP que regulam o etanol são as Resoluções ANP nº 7/2011 que estabelece a especificação do etanol anidro combustível e hidratado e a ANP nº 5/2006 que regula a distribuição para comercialização do etanol ^{8,9}.

1.2. Etanol Combustível

O uso de bicomcombustíveis em veículos motorizados tem sido reconhecido como uma alternativa de grande viabilidade em diversas matrizes energéticas de vários países do mundo ¹⁰. O etanol se destaca pelas duas formas de uso, tanto como uso direto na forma hidratada como também na forma de aditivo antioxidante (anidro) na gasolina. A avaliação da qualidade

deste biocombustível se reveste de enorme importância, tanto para os processos de produção, quanto para os mercados consumidores.

A qualidade do etanol é definida por meio de padrões e especificações técnicas elaboradas pelas agências reguladoras nacionais e internacionais, onde a Agência Nacional do Petróleo e Biocombustíveis (ANP) é a agência reguladora da qualidade deste produto no Brasil. As especificações nos diversos países envolvem a utilização de métodos de análises, que utilizam varias técnicas instrumentais. A evolução das especificações é decorrente de vários fatores, tais como, o desempenho dos motores e a preservação do meio ambiente, que exigem cada vez mais o desenvolvimento de novos métodos de análise.

1.3. Controle de Qualidade e Especificações do Etanol

O azeótropo formado no processo de produção do etanol brasileiro via fermentação de açúcares por ação de leveduras da espécie *Saccharomyces cerevisiae*, é responsável pela formação do Etanol Hidratado Combustível (EHC), cujo teor mínimo de etanol deve ser de 94,5% v/v, e do Etanol Anidro Combustível (EAC), cujo teor mínimo de etanol deve ser de 98% v/v. No Brasil, o EHC é utilizado diretamente como combustível, já o EAC é utilizado para adição na gasolina, em substituição a aditivos poluentes como chumbo tetraetila e MTBE.

A qualidade dos combustíveis é definida por um conjunto de características físicas e químicas previstas nas Normas Brasileiras (NBR) da Associação Brasileira de Normas Técnicas (ABNT) e de normas da American Society for Testing and Materials (ASTM). A especificação estabelecida pela ANP, conforme a Lei nº 9.478/1997 determina valores-limites para essas características, de modo a assegurar o desempenho adequado dos combustíveis.

A adulteração dos combustíveis se caracteriza pela adição irregular de qualquer substância, sem recolhimento de impostos, com vistas à obtenção

de lucro, sendo que a adulteração mais comum do álcool hidratado é a adição de mais água à fórmula, o que resulta num produto fora de especificação, portanto inadequado ao uso. Outra forma de adulteração é a adição de álcool anidro ao álcool hidratado. O anidro é o tipo de álcool destinado exclusivamente para ser adicionado à gasolina. Para coibir a adulteração do álcool hidratado com o anidro, a ANP aprovou, em 2005, a Resolução nº 36, que determina a adição de um corante de cor laranja ao álcool etílico anidro. (ANP, 2009). Testes específicos são realizados para revelar se os produtos estão de acordo com as especificações estabelecidas pela ANP e, também se houve adição ilegal de solventes.

O controle de qualidade dos biocombustíveis tão bem como os combustíveis fósseis é de capital importância para a comercialização e aceitação dos mesmos no mercado nacional e internacional ¹¹.

A nível internacional, no sentido de assegurar o controle da qualidade do etanol combustível, há importantes agências reguladoras responsáveis pela aplicabilidade das especificações técnicas e normas oficiais ¹².

Também, para estimular o comércio internacional de biocombustíveis, foi elaborado, pelos órgãos regulamentadores brasileiros, coordenados pelo Ministério de Relações Exteriores em parceria com equipes da União Européia e dos Estados Unidos, o documento "White Paper on Internationally Compatible Biofuel Standards" ^{12, 14} o qual classifica as especificações adotadas nestas regiões em três categorias (Categoria A, B e C) de acordo com a equivalência entre elas e estabelece metas para torná-las equivalentes.

Neste contexto, torna-se necessário um controle rígido dos parâmetros que afetam a qualidade desse produto, tais como, acidez, pH, sódio, sulfato dentre outros pois eles podem intensificar a capacidade de corrosão do etanol, causando a aceleração do processo de deterioração dos componentes metálicos presentes nos motores. A presença de traços de cobre, ferro e cloretos pode comprometer a mistura combustível gasolina-álcool, pela formação de gomas e produtos corrosivos, sendo, então, especificados limites

também para estes elementos. Já a presença de enxofre deve ser minuciosamente controlada, pois, após a combustão, esse contaminante é convertido em SO_x sendo liberado para a atmosfera e tornando-se um grande problema ambiental.

Tendo em vista a relevância de se garantir o cumprimento das especificações técnicas exigidas no controle de qualidade do etanol combustível, faz-se necessário o estabelecimento de métodos de análises oficiais que apresentem confiabilidade, norteando todas as condições analíticas, como por exemplo, tipo de pré-tratamento da amostra, procedimentos de calibração, técnica analítica, entre outros, a fim de atender a especificações nacionais e internacionais.

Por isso, o controle da qualidade do etanol combustível requer métodos analíticos cada vez mais sensíveis para análise de seus contaminantes inorgânicos e orgânicos e, também, para acompanhar a possível modificação química e/ou física desse combustível após o processo de produção. Embora estes contaminantes muitas vezes estejam presentes em quantidades muito pequenas no etanol produzido, o conhecimento e o controle dessas impurezas assumem papel fundamental nas diversas etapas de seu emprego com o intuito de: evitar “envenenamento” de catalisadores automotivos, cada vez mais sofisticados e caros, minimizar reações químicas secundárias em motores do tipo “*flex fuel*” desenvolvidos com alta tecnologia, cuja emissão de poluentes acarreta prejuízos ao meio ambiente e evitar outros danos ao motor e às peças adjacentes no que se refere aos processos de corrosão ou lixiviamento, acarretados por teores de contaminantes acima do valor permitido ou estabelecido.

1.4. A Matriz Energética e as Questões Ambientais

O Brasil possui uma matriz energética privilegiada comparada com o restante do mundo, diversificada e em grande parte, a energia produzida e consumida é renovável. A diversificação das fontes de energia primária tem extraordinária relevância na segurança da oferta do abastecimento,

complementaridade dessas fontes, e numa salutar competitividade rumo ao desenvolvimento tecnológico e eficiência econômica.

A matriz veicular no Brasil é também referência, especialmente pela participação dos biocombustíveis. Em 2008 o consumo de etanol no Brasil foi superior ao da gasolina. Foram comercializados nesse período 19.583.791.750 de litros de etanol para 18.881.087.250 de litros de gasolina. O etanol produzido da cana-de-açúcar é uma das melhores opções para mitigar as emissões de gases de efeito estufa pela queima de combustíveis fósseis. O balanço energético (unidade de energia renovável extraída por unidade de energia fóssil inserida) é extremamente favorável, atingindo 8,9:1, com a redução de emissões de 2,86 e 2,16 t CO₂ eq./m³, para o etanol anidro e hidratado respectivamente ¹³.

O uso do etanol combustível caracteriza um excepcional desempenho da matriz energética, evitando emissões de GEE (Gases do Efeito Estufa) equivalentes a 13% das emissões de todo o setor de energia no Brasil, tomando por base o ano de 1994, hoje o cenário é muito melhor. Para cada 100 milhões de toneladas de cana adicionais, nos próximos anos, poderiam ser evitadas emissões de 12,6 mt (milhões de toneladas) CO₂ eq., com etanol, bagaço e com energia elétrica excedente adicional ¹⁴.

1.5. Monitoramento da Qualidade de Etanol Combustível no Estado do Maranhão

O monitoramento da qualidade do etanol combustível no Estado do Maranhão é realizado acerca de 12 anos através do Programa Nacional de Monitoramento da Qualidade de Combustíveis (PMQC) com o intuito de fiscalizar e controlar o percentual de não conformidades.

O PMQC foi instituído pela Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP) e atualmente regulamentado pela Resolução ANP nº 8 de 10/02/2011, de modo a dar cumprimento às atribuições da ANP, especialmente ao disposto no artigo 8º da Lei 9.478/97, que trata da proteção

dos interesses dos consumidores quanto à qualidade dos derivados do petróleo e dos biocombustíveis.

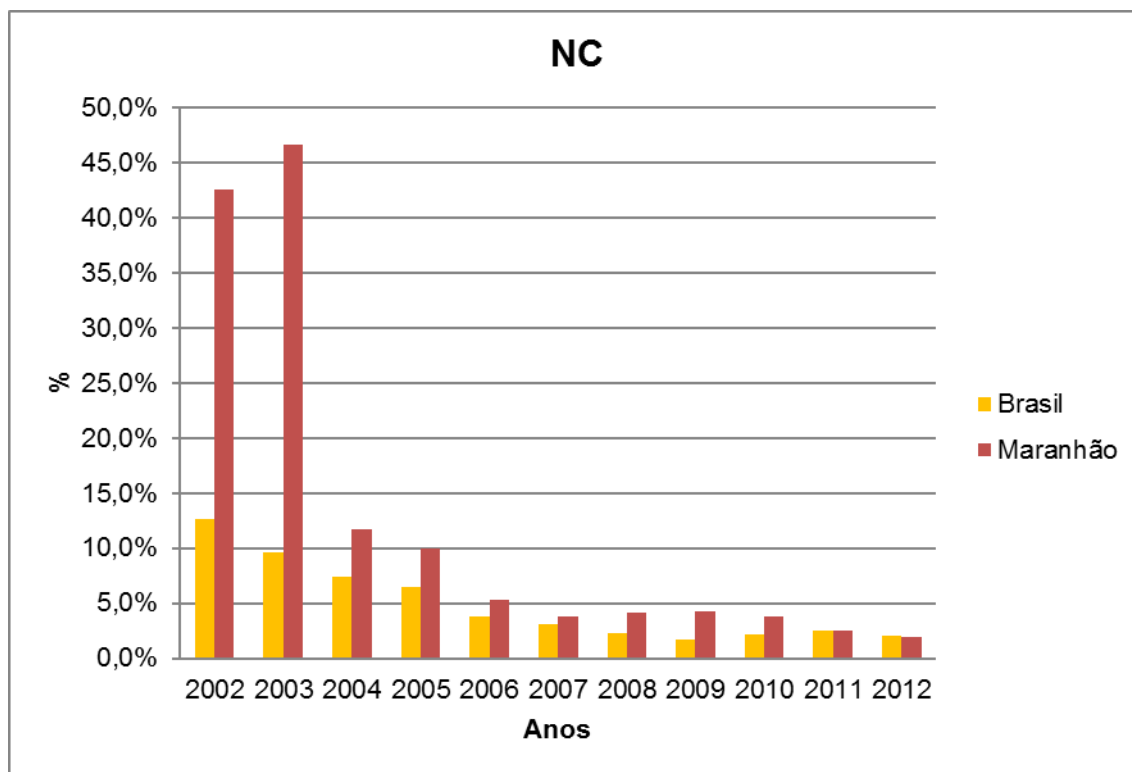
A metodologia de coleta de etanol combustível é realizada pelo Laboratório de Análises e Pesquisa em Química Analítica de Petróleo e Biocombustível (LAPQAP) através do convênio ANP/UFMA/FSADU para realização das análises, tratamento das amostras e envio de resultados.

Logo, no início do Programa no Estado do Maranhão houve índices de não conformidades no etanol combustível e uma das principais causas era a adição indevida de água no álcool hidratado (que contém normalmente entre 6,2 e 7,4 % de água). Essa não-conformidade pode ser constatada por ensaios de massa específica e teor alcoólico ¹⁵.

O etanol é coletado diretamente nos postos revendedores e os resultados são enviados diretamente para a ANP, que realiza um estudo estatístico e os mesmos são publicados mensalmente no Boletim da Qualidade dos Combustíveis Líquidos Automotivos Brasileiros. Esse boletim pode ser consultado pelos consumidores no site da ANP.

Ao longo dos anos houve uma diminuição significativa quanto ao índice de não conformidades relacionados no Estado do Maranhão e o Brasil conforme a Figura 1.

Figura 1: Diminuição de não conformidades de etanol hidratado combustível nos anos de 2002 a 2012.



Fonte: ANP.

1.6. Métodos Alternativos voltados para a Qualidade do Etanol

Os novos métodos analíticos descritos na literatura ¹⁶ para avaliação da qualidade do etanol combustível representam atualmente um número muito maior do que os métodos oficiais utilizados para aferir a qualidade do etanol combustível, mostrando a existência de uma grande preocupação com a qualidade deste biocombustível.

Entre os novos métodos analíticos desenvolvidos para aferir a qualidade do etanol combustível existe uma grande predominância dos métodos espectroscópicos, mas observa-se um real crescimento dos métodos eletroanalíticos ¹⁷ para a determinação de espécies inorgânicas, principalmente metais, ao passo que em relação às espécies orgânicas existe uma maior incidência de métodos cromatográficos.

As técnicas instrumentais utilizadas nestes novos métodos mostram a existência de um número maior de técnicas isoladas em relação às técnicas acopladas ¹⁸, tanto para espécies orgânicas quanto para espécies inorgânicas no desenvolvimento dos novos métodos para a avaliação da qualidade do etanol combustível. As perspectivas de desenvolvimento de novos métodos analíticos apontam para a determinação de outras espécies químicas além daquelas normalmente avaliadas pelas agências reguladoras e para a busca de limites menores do que os atualmente utilizados para a avaliação da qualidade do etanol combustível.

No desenvolvimento destes novos métodos analíticos prevalece uma maior ocorrência de determinações de espécies inorgânicas, principalmente de íons metálicos, em comparação com as determinações de espécies orgânicas, diminuição esta, possivelmente, decorrente da complexidade biológica da matriz estudada.

1.7. Redes Neurais Artificiais

As redes neurais artificiais são abordagens que tentam imitar o complexo processamento paralelo e não linear do cérebro contidos em conjunto de dados.

Possuem uma arquitetura formada por camadas contendo neurônios que permitem mapear entradas e saídas segundo algoritmos de aprendizagem.

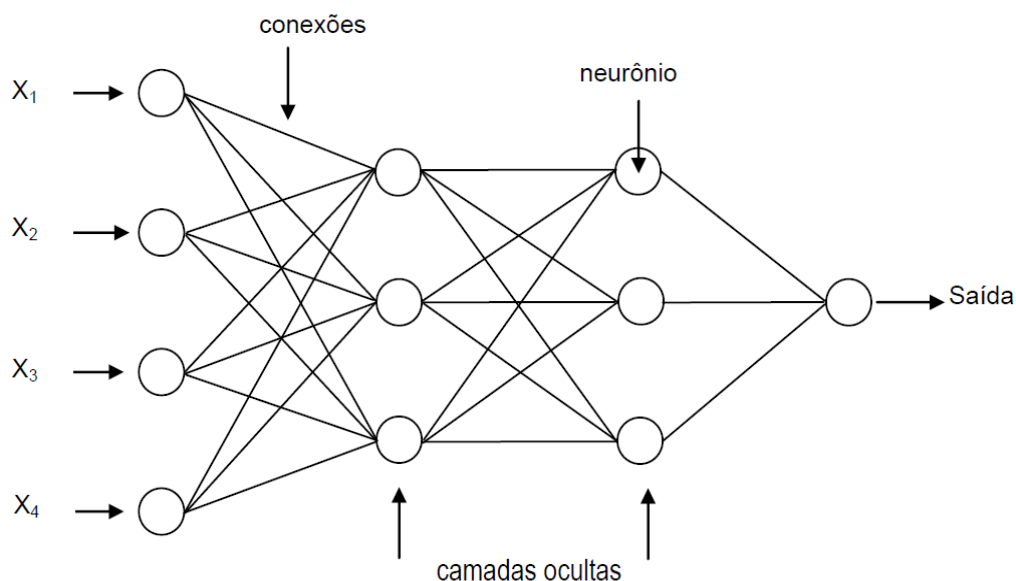
São modelos computacionais inspirados na estrutura de redes neurais biológicas consistindo de neurônios interligados entre si através de conexões ponderadas que associadas a pesos, denominados de pesos sinápticos armazenam o conhecimento da rede e ponderam as entradas recebidas pelos neurônios ¹⁹.

1.7.1. Redes Neurais Artificiais de Múltiplas Camadas

A arquitetura ou topologia de uma rede (Figura 2) é constituída de três partes:

- Camada de entrada: responsável pela entrada de informações. Aqui os padrões são inicialmente apresentados à rede.
- Camada intermediária ou oculta: responsável pela extração de características do processo ao qual está se inferindo.
- Camada de saída: contém as estimativas da variável resposta processadas segundo as estruturas definidas internamente.

Figura 2: Rede neural de múltiplas camadas com quatro variáveis na camada de entrada, duas camadas intermediárias contendo três neurônios em cada uma e uma camada de saída com uma única saída.



A arquitetura ou topologia de uma RNA compreende o número de neurônios em cada camada, tipo de conexão entre os neurônios e o número de camadas existentes ¹⁹.

Existem outros tipos de redes e com várias camadas ocultas para resolução de problemas mais complexos denominada Redes Neurais Multilayer Perceptron (MLP) que é a mais utilizada atualmente.

Quanto ao treinamento, existem diversos métodos desenvolvidos para treinamento de redes que podem ser separados em aprendizado supervisionado e aprendizado não supervisionado.

No treinamento supervisionado alimenta-se a rede com dados de entrada e de saída desejados. O objetivo é ajustar os parâmetros da rede de forma a encontrar uma correlação entre as entradas e saídas. O exemplo mais utilizado de rede com treino supervisionado é o “backpropagation” ou retropropagação.

No treinamento não supervisionado o resultado desejado não é fornecido, permitindo que a rede encontre grupos ou classes por associação estatística dos dados de entrada.

As redes neurais de múltiplas camadas são redes supervisionadas uma vez que a estimativa é sempre comparada com um valor medido. A rede neural adquire conhecimento por treinamento o qual é passado aos pesos sinápticos. O método de treinamento mais utilizado é o método conhecido por retropropagação do erro (backpropagation). Até que o resíduo (diferença entre o valor estimado e valor medido) não satisfaça uma condição ele retorna à rede modificando os pesos sinápticos até que a condição seja aceita e a estimativa da resposta calculada.

1.7.2. Retropropagação de Erros ou “Backpropagation”

A arquitetura de uma RNA por retropropagação é representada pelos números de neurônios em cada camada e conexões entre elas. A camada de entrada corresponde ao número de variáveis de entrada de um processo, enquanto que a de saída fornece as variáveis a serem previstas de um processo. A camada oculta possui um número variável de neurônios os quais fornecem a estrutura não linear que conecta os dados de entrada e saída

O conjunto de dados geralmente é dividido em duas partes. Uma parte com 75% dos dados é utilizada na fase de treinamento da rede que

através da passagem das entradas por todas as camadas e neurônios, determinará os valores dos pesos que irão compor o modelo neural. Os 25% restantes serão utilizados na fase de teste para avaliar o desempenho da rede na fase anterior. É importante validar o modelo neural com outros conjuntos de dados. Este procedimento garante a qualidade do modelo.

Normalmente, a complexidade do modelo de redes neurais é referida ao número de parâmetros, ou seja, o número de pesos e o número de bias a serem ajustados. Análises devem ser conduzidas para examinar a eficiência e adequação do modelo construído. O parâmetro de avaliação abordado neste trabalho foi o erro quadrático médio normalizado (NMSE).

O Erro Quadrático Médio Normalizado é definido por:

$$NMSE = \frac{\sum_t (observação_t - predição_t)^2}{\sum_t (observação_t - observação_{t-1})^2} \quad \text{Equação 1}$$

Onde é calculado o somatório do número de dados experimentais menos os valores estimados ou preditos.

1.8. Regressão por Mínimos Quadrados Parciais (PLS)

A regressão por PLS é baseada na decomposição simultânea dos dados da matriz (X) e da propriedade a ser estimada (Y). Cada componente principal do modelo sofre uma pequena modificação para buscar a máxima covariância entre X e Y e passa a receber a terminologia de variável latente (VL). Não requer um conhecimento exato de todos os componentes presentes nas amostras podendo realizar a previsão de amostras mesmo na presença de interferentes, desde que estes também estejam presentes por ocasião da construção do modelo ²⁰.

Neste método, as matrizes X e Y são decompostas em matrizes menores de acordo com as equações 1 e 2 ²¹:

$$X = TP^t + E \quad \text{Equação 1}$$

$$Y = UQ^t + F$$

Equação 2

onde:

X e Y são as matrizes que serão decompostas;

T e U são matrizes de escores;

P é a matriz de pesos de X;

E é o erro (resíduo) de X;

Q é a matriz de pesos de Y;

F é o erro (resíduo) de Y.

Após a decomposição de X e Y, uma relação linear é estabelecida entre os escores t e u, conforme a Equação 3:

$$u_h = b_h + t_h$$

Equação 3

onde:

b é o vetor dos coeficientes de regressão,

h é o número de variáveis latentes usadas no modelo PLS.

O termo escores está relacionado com a classificação das amostras, e os loadings relacionam-se com as variáveis.

Uma maneira mais simples de interpretar o PLS é que ele forma novas variáveis X (estimativa de variáveis latentes), através da combinação linear das antigas variáveis e usa esta matriz para o cálculo das estimativas de Y.

Para otimização de modelos PLS, pode combinar-se validação cruzada com testes de validação externa. A validação cruzada é usada para determinação do modelo e do respectivo número de variáveis latentes, e o teste de validação externa para testar o desempenho final do modelo.

A validação cruzada é uma técnica de validação baseada nos dados de calibração sendo usada para avaliar a capacidade de previsão dos modelos

PLS para o conjunto de amostras. Esta validação cruzada é efetuada testando-se vários submodelos, ou seja, nesta técnica eliminam-se sucessivamente amostras do próprio conjunto de calibração.

Primeiramente, a primeira amostra dos dados de calibração é eliminada. Seguidamente, a calibração é efetuada usando as restantes amostras antes de ser testada na primeira amostra por comparação do y real com o previsto pelo modelo. A primeira amostra é então colocada no conjunto de calibração e a segunda amostra é eliminada, repetindo-se o procedimento. Este procedimento continua até cada uma das amostras terem sido objeto de teste, sendo retido o modelo que revelou melhor desempenho (erro mínimo) ²².

A eficiência do modelo por mínimos quadrados parciais pode ser avaliada pelo cálculo dos valores de Erro Quadrático Médio Normalizado (NMSE). Esses valores expressam a exatidão do modelo, ou seja, a proximidade entre o valor calculado pelo modelo e o valor verdadeiro ou obtido por um método de referência.

1.9. RNA e PLS em Combustíveis - Revisão Bibliográfica

Considerável atenção tem sido dada nos últimos anos para o desenvolvimento de técnicas sofisticadas para explorar os conjuntos de dados. Uma dessas técnicas compreendem as RNAs que tentam imitar a forma elementar do cérebro humano. Uma área que tem vindo a ganhar interesse na utilização dessas técnicas está focada na qualidade de combustíveis, como gasolina, biodiesel. etc. Pouco ou quase nada se encontrou na literatura sobre a aplicação dessas técnicas em etanol combustível, fato este que motivou o presente trabalho. Métodos RNAs também têm sido utilizados nas áreas de predição em séries temporais e classificação ²³. Aqui as RNAs serão utilizadas como análise de regressão, ou seja, estimar saídas.

Os procedimentos estatísticos e computacionais têm sido aplicados com sucesso, incluindo a análise de regressão linear tradicional, utilizando o princípio dos mínimos quadrados, e RNAs ²⁴. A análise de regressão pode ser

realizada para relacionar e prever dados que apresentam uma alta correlação entre valores experimentais e preditos ²⁵. Algumas das ferramentas estatísticas tradicionais aplicadas para a previsão e diagnóstico em muitas disciplinas são a análise discriminante ^{26, 27}, a regressão logística ²⁷⁻²⁹, abordagem " Bayesian" ^{30, 31} e Regressão Linear Múltipla ³²⁻³⁵. Estes modelos têm sido provados ser muito eficazes, no entanto, as variáveis que geram as respostas devem ser independentes. Os modelos PLS e RNAs admitem a estrutura de dependência contida no conjunto de variáveis.

Atualmente o campo de pesquisa em RNAs e PLS é muito grande por causa da correlação e predição de composições e dos parâmetros físico-químicos de combustíveis.

Podemos destacar a determinação do teor de álcool e glicerol no biodiesel, através da calibração multivariada, utilizando espectroscopia na região do NIR e do visível realizada por DORADO ³⁶. FELIZARDO e colaboradores usaram espectros de infravermelho na região do NIR para prever a quantidade de metanol e água em biodiesel ³⁷.

A adição de etanol à gasolina foi estudada por DELGADO; ARAUJO e FERNANDES JR. ³⁸ para avaliar propriedades físico-químicas (curva de destilação, número de octanas, massa específica, pH e condutividade) de misturas com 20, 40, 60 e 80% de etanol (hidratado) em gasolina comum. Quando o etanol foi adicionado, ocorreu o aumento da massa específica, não sendo tão desejável, pois uma grande massa específica injeta maior quantidade de combustível no motor. Ao avaliar todas as propriedades, os autores recomendam a utilização de etanol hidratado em uma faixa de 40 a 60% para motores que funcionam com tecnologia flex.

Diversas propriedades físico-químicas (densidade, MON, RON, IAD, curva de destilação e resíduo, teor de benzeno, olefínicos e aromáticos) foram estudadas por OLIVEIRA ³⁹, através de análise do programa SIMCA (Soft Independent Modelling of Class Analogy) para classificação dos dados e montou um modelo com 5 componentes principais e 34 amostras de gasolina no treinamento. Todas as 16 amostras de gasolina que fazem parte do conjunto de teste foram classificadas corretamente pelo modelo gerado, porém

4 amostras do conjunto de calibração que eram conformes, foram consideradas pelo modelo como não conformes. Os autores concluíram que os parâmetros físico-químicos usados para montagem do modelo não foram suficientes para classificar de maneira correta todas as amostras de treinamento.

2. OBJETIVOS

2.1. Objetivo Geral

A proposta deste trabalho é estimar a massa específica em etanol hidratado combustível.

2.2. Objetivos Específicos

Estudar a influência da massa específica e outros parâmetros do etanol hidratado combustível utilizando banco de dados por meio de histogramas.

Uso do banco de dados para processar os modelos de estimativas por RNAs de múltiplas camadas e PLS da massa específica.

Treino e validação de um modelo de rede neural artificial de múltiplas camadas para regressão com os parâmetros identificados.

Construção do modelo de estimativa da massa específica utilizando regressão por mínimos quadrados parciais (PLS).

Comparação dos modelos RNAs de múltiplas camadas e PLS quando aplicados para estimar a massa específica do etanol hidratado combustível.

3. METODOLOGIA

As análises físico-químicas (parâmetros físico-químicos) do etanol hidratado combustível foram realizadas pelo LAPQAP no período de 2002 ao 1º trimestre de 2013, em um total de 6566 amostras, submetidas aos seguintes ensaios: massa específica, teor alcoólico, condutividade elétrica e pH. Todas as análises foram realizadas seguindo métodos padronizados e normas estabelecidas pela Resolução nº 7 de 14 de abril de 2011 da ANP. No período anterior a 2011, outras normatizações vigentes no período, foram usadas.

3.1. Massa Específica

A massa específica a 20 °C do etanol hidratado combustível foi determinada pela norma ABNT NBR 14065 ⁴⁰ (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS). O equipamento utilizado para estas determinações foi o densímetro digital (Figura 3) modelo DMA 4500 da ANTON PAAR.

A amostra é introduzida, com auxílio de uma seringa, em um tubo capilar em forma de U, vibrando com frequência definida. A modificação da frequência, provocada pela introdução da amostra, em conjunto com os parâmetros de calibração determina a massa específica da amostra, na temperatura escolhida ⁴¹.

Figura 3: Densímetro digital ANON PAAR DMA 4500.



3.2. Teor Alcoólico

O teor alcoólico do etanol hidratado combustível foi determinado pela norma ABNT NBR 5992 ⁴² com o uso de uma tabela de conversão de valores de massa específica obtidas pelo densímetro digital.

3.3. Condutividade Elétrica

A condutividade elétrica do etanol hidratado combustível foi determinada pela norma ABNT NBR 10547 ⁴³. O equipamento utilizado foi um condutímetro DIGIMED (Figura 4) modelo DM-32.

Figura 4: Condutímetro DIGIMED DM 32.



O ensaio é realizado através de uma célula de condutividade eletrolítica platinizada para medição da quantidade de cargas transportadas pelos íons presentes em uma solução ⁴⁴.

3.4. Potencial Hidrogeniônico (pH)

O potencial hidrogeniônico do etanol hidratado combustível foi determinado pela norma ABNT NBR 10891 ⁴⁵. O equipamento utilizado foi um pHmetro METROHM (Figura 5) modelo 827 lab.

O ensaio é realizado através de um eletrodo de pH na medição de atividade de íons H^+ .

Figura 5: pHmetro METROHM 827 lab.



3.5. Programação

Os cálculos de estimativas realizados no presente trabalho foram feitos pelos módulos de RNA e de PLS, do software Statistica 10, através do qual foram obtidos os conjuntos de parâmetros para cada rede treinada. A ordem de entrada dos parâmetros na rede foram teor alcoólico, pH e condutividade elétrica.

O programa utiliza dois tipos de função de transferência, linear e hiperbólica combinando as camadas para otimização da arquitetura. A camada de saída continha um neurônio, a massa específica. O algoritmo de aprendizagem utilizado foi a retropropagação e a correção de pesos foi baseada na soma dos erros quadráticos da rede e foi realizada após a apresentação de cada exemplo de treinamento.

Para o modelo PLS, os dados foram centrados na média e o número de variáveis latentes foi determinado utilizando o gráfico de variância capturada (%) versus número de variáveis latentes.

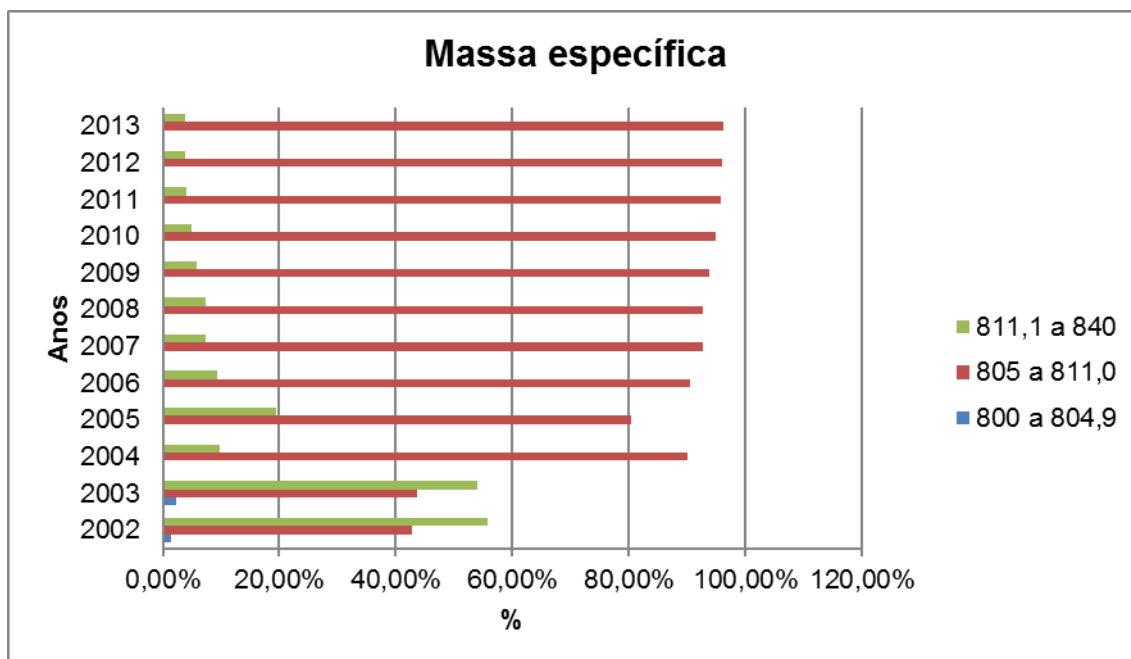
A eficiência dos modelos foi avaliada pelos valores de Erro Quadrático Médio Normalizado e pelo gráfico de valores observados *versus* valores estimados.

4. RESULTADOS E DISCUSSÃO

4.1. Estudo Estatístico Preliminar

Inicialmente, foi feita uma análise no banco de dados de 6566 amostras do etanol hidratado combustível, referente a todos os parâmetros estabelecidos (massa específica, condutividade, pH e teor alcoólico). O estudo teve como objetivo detectar qual o parâmetro que mais influenciaria nos demais e, com isso, determinar qual a meta do estudo em termos de predição ("target") e quais os parâmetros a serem usados como dados de entrada. Desta forma, o estudo preliminar indicou tanto a viabilidade de aplicação de redes neurais, como também de PLS. Para isso foi realizada a construção de histogramas destes dados, com a finalidade de se extrair informações sobre o comportamento de cada parâmetro.

Figura 6: Comportamento da massa específica de 2002 ao 1º trimestre de 2013.

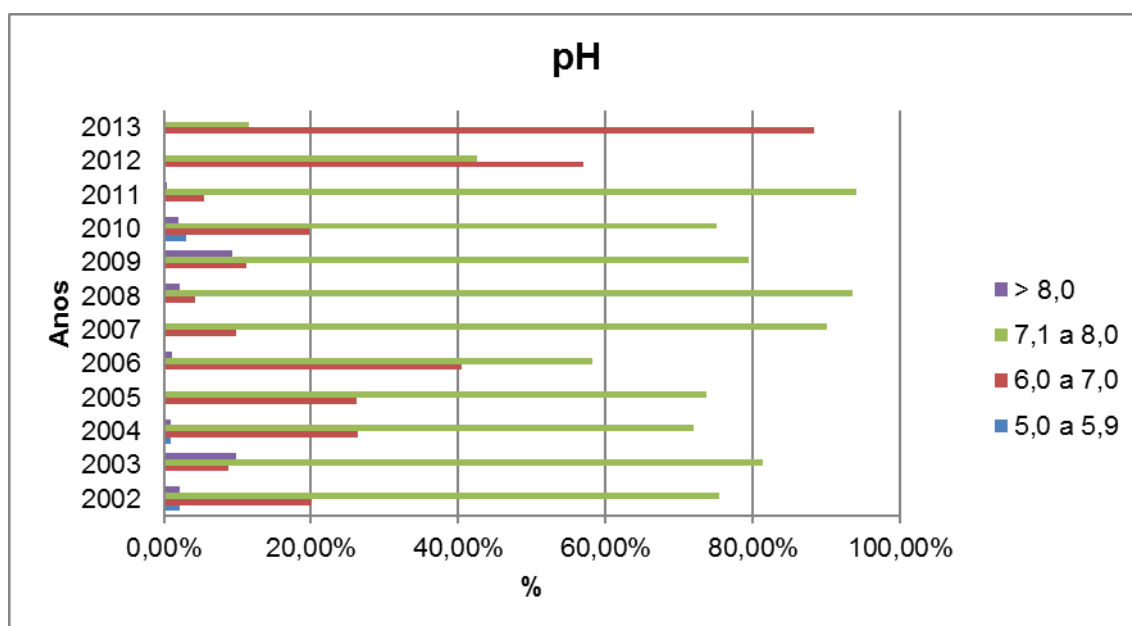


Entre os anos de 2002 e 2003, observa-se que algumas amostras de etanol hidratado combustível apresentam valores de massa específica abaixo de 805,0 kg/m³ (Figura 6), o que caracteriza um etanol mais leve, com alto teor de alcoólico e teor de água baixo ou contaminação por hidrocarbonetos de baixo peso molecular como gasolina entre outros, e desta forma menos estável às condições ambientais.

Observa-se grande percentual de etanol com massa específica acima de $811,0 \text{ kg/m}^3$, que já era naquele período considerado fora de especificação, este aumento na massa deve-se possivelmente associar à contaminação por água, pois a maioria das amostras nos anos de 2002 e 2003 apresentava resultados de massa altos, o que pode ser comprovado avaliando o gráfico de condutividade elétrica (Figura 8), onde exatamente nestes períodos observou-se a maior condutividade de todo o período avaliado.

A partir de 2004 observou-se um crescimento na quantidade de amostras conformes e a expressiva diminuição da não conformidade, pois não se observou amostras abaixo de $805,0 \text{ kg/m}^3$, ao mesmo tempo em que o percentual de amostras com massa acima de $811,0 \text{ kg/m}^3$ teve uma diminuição progressiva, indicando melhoria da qualidade do etanol combustível.

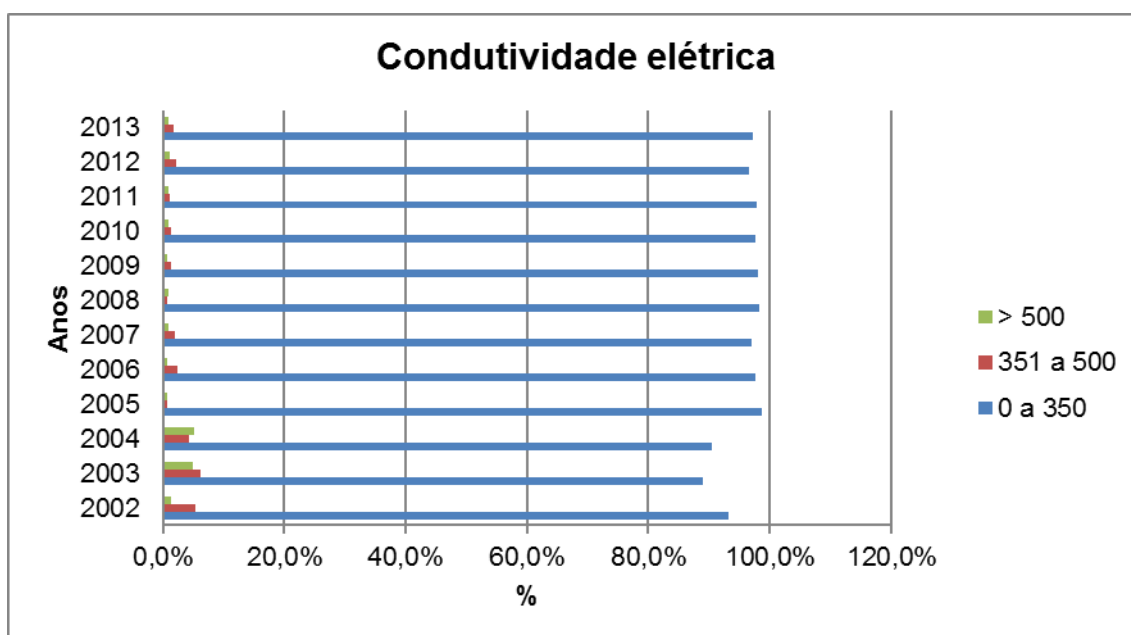
Figura 7: Comportamento de pH (6,0 a 8,0) de 2002 ao 1º trimestre de 2013.



O potencial hidrogeniônico indica a atividade dos íons H^+ em solução, e determina a acidez ou basicidade do etanol, valores abaixo de 6,0; indicam amostras não-conformes, e que o etanol pode ter sofrido reações biológicas com produção de aldeídos, ácido acético e cetonas, o que provavelmente acontece quando o mesmo é submetido a condições assépticas no transporte ou armazenamento e misturas mesmo em pequena proporção

com materiais ou substâncias contaminadas, este caso é muito pouco observado, representa casos isolados com frequência apenas nos anos de 2002, 2004 e 2010 em pequeno percentual; entre os anos de 2002 a 2011, observou-se que o etanol hidratado combustível na sua maioria apresentava pH na faixa de 7,1 à 8,0 o que é considerado conforme.

Figura 8: Comportamento de condutividade elétrica de 2002 ao 1º trimestre de 2013.



Verifica-se que nos anos de 2002 a 2004 quantidades ainda relevantes de amostras de etanol hidratado combustível com condutividade acima de 500 $\mu\text{S}/\text{cm}$ (Figura 8) sugestivo de adulteração com água sendo confirmado pelo aumento da massa específica apresentada na Figura 6, neste período o valor máximo aceito pela legislação era de 500 $\mu\text{S}/\text{cm}$, e já se observava o percentual de cerca de 90% apresentando condutividade entre 0 e 350 $\mu\text{S}/\text{cm}$, o que já era indicativo de que a legislação poderia ser modificada para ajustar o valor de condutividade máxima em 350 $\mu\text{S}/\text{cm}$.

Entre 2005 e 2013 quase 100% de todo o etanol comercializado no Maranhão já apresentava valores de condutividade abaixo de 350 $\mu\text{S}/\text{cm}$, e foi só a partir de janeiro de 2012 que houve mudança na legislação ANP fixando o máximo de 350 $\mu\text{S}/\text{cm}$, e dessa forma os demais resultados acima de 351 $\mu\text{S}/\text{cm}$ foram considerados fora de conformidade.

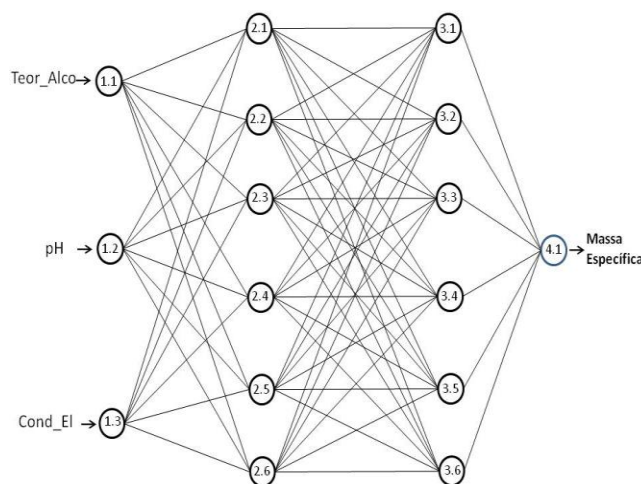
4.2. Estimativa da Massa Específica por RNAs de Múltiplas Camadas

Todas as amostras de etanol hidratado combustível foram submetidas à predição utilizando uma rede neural de múltiplas camadas.

A topologia de uma rede neural constou de duas camadas ocultas com seis neurônios cada uma.

Na fase de treinamento foram utilizados 5847 registros e na fase de teste 500 registros. A rede foi alimentada com as seguintes variáveis: teor alcoólico, pH e condutividade elétrica tendo como saída a estimativa da massa específica. A Figura 9 mostra um gráfico ilustrando a topologia utilizada.

Figura 9: Topologia da RNA obtida para estimar a massa específica do etanol hidratado combustível.



Inicialmente os valores das estimativas não foram convincentes e após eliminar o ano de 2012 o modelo neural apresentou um ótimo desempenho. (ver Figuras 10 e 11).

Figura 10: Valores estimados pela rede neural artificial de múltiplas camadas e os valores medidos da massa específica (g/cm^3) na fase de treinamento.

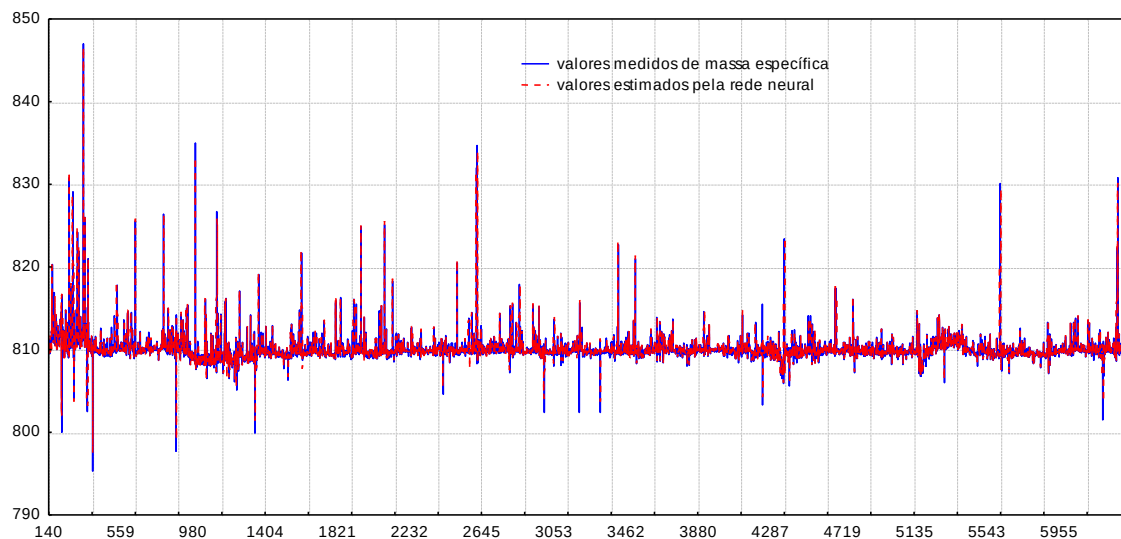
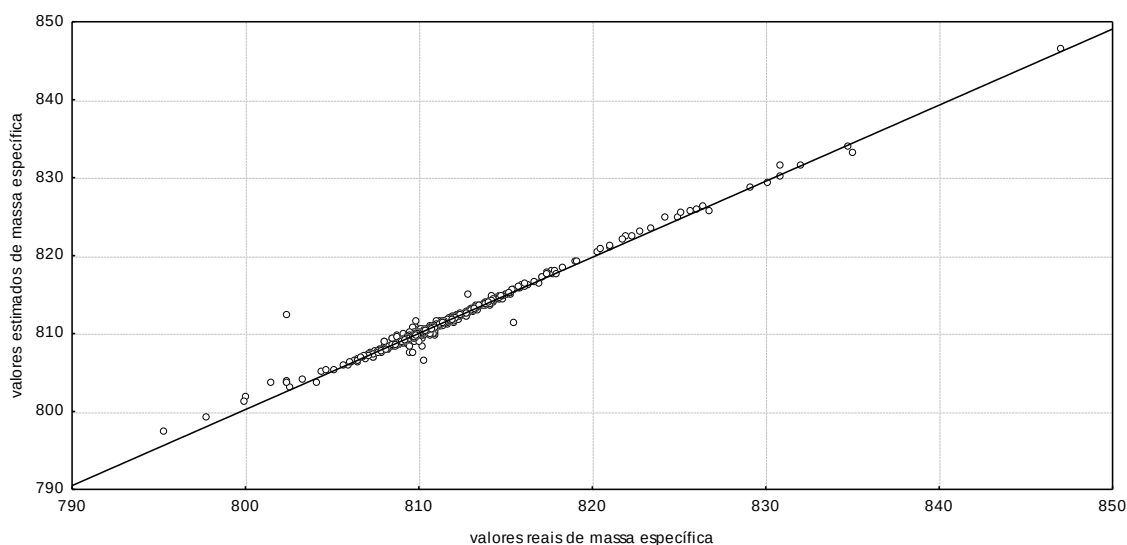
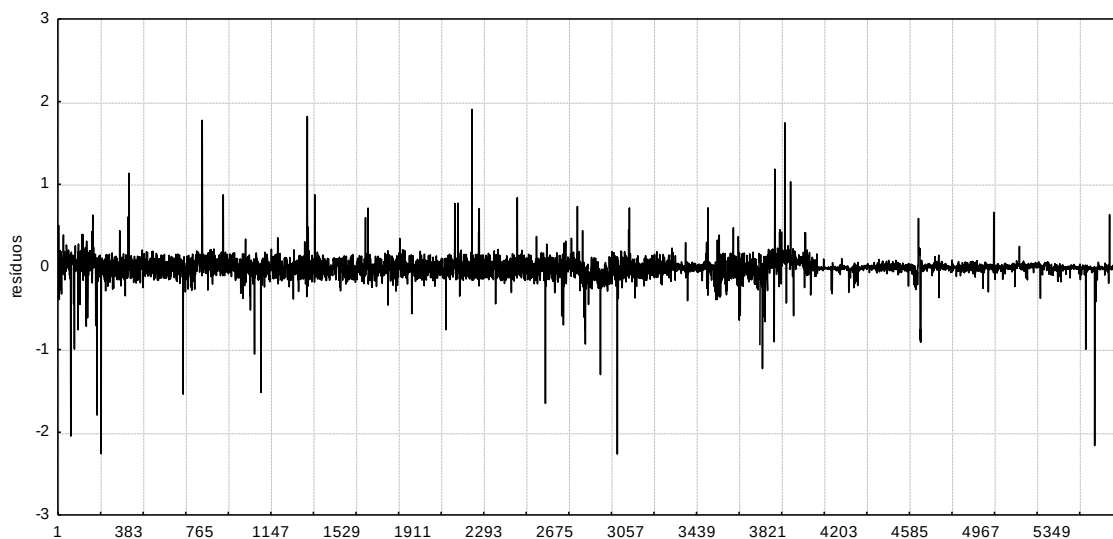


Figura 11: Valores reais de massa específica (g/cm^3) versus valores estimados pelo modelo neural de múltiplas camadas na fase de treinamento.



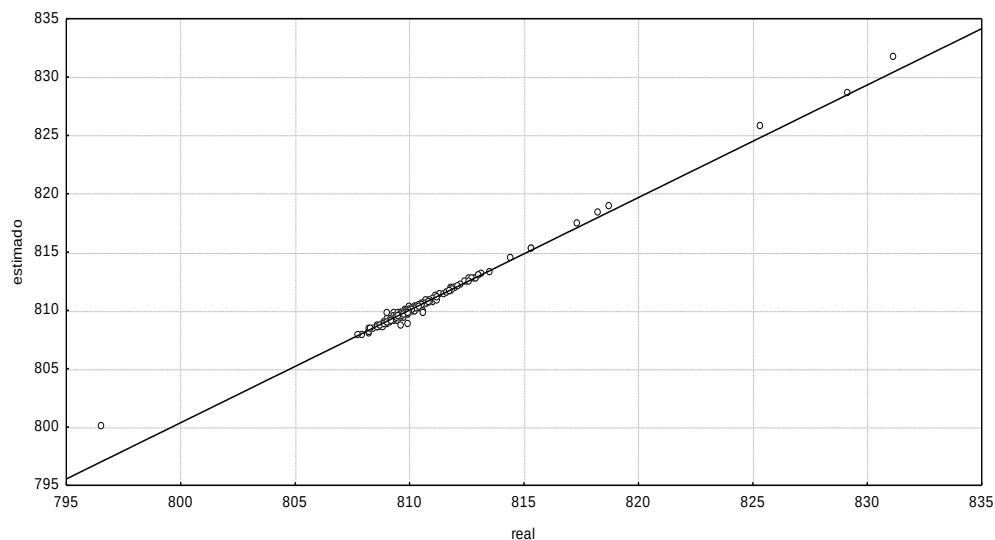
A análise dos resíduos mostrou a adequabilidade da rede neural na estimativa da massa específica uma vez que os erros apresentaram distribuição normal, com média nula e variância constante. Os valores dos resíduos estimados na fase de treinamento constam na Figura 12.

Figura 12: Valores dos resíduos do modelo neural de múltiplas camadas na fase de treinamento.



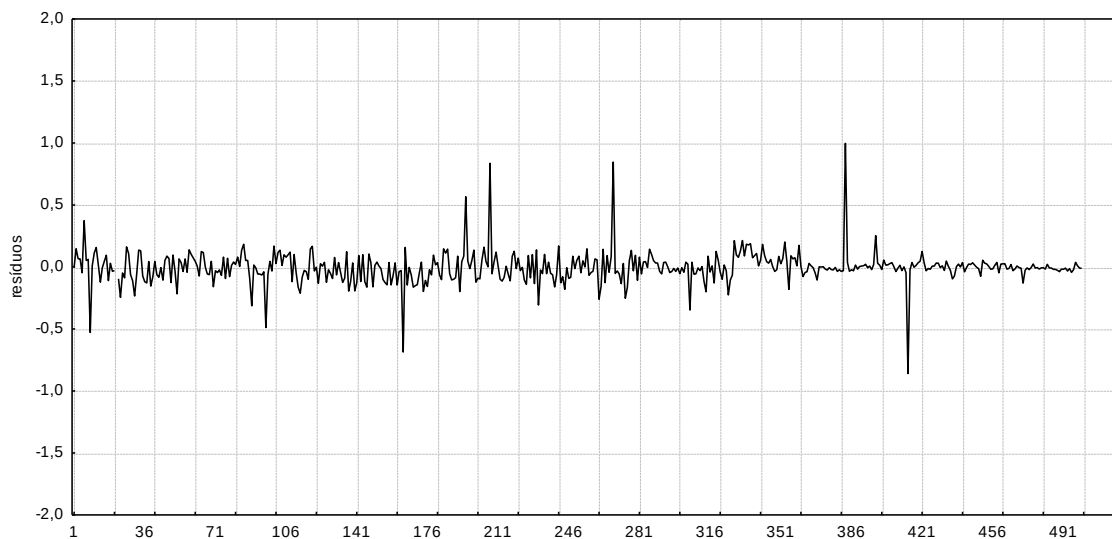
Na Figura 13 o parâmetro massa específica foi testada retratando a eficácia do modelo na predição da massa específica.

Figura 13: Valores estimados pela rede neural artificial de múltiplas camadas e os valores medidos da massa específica (g/cm^3) na fase de teste.



Os resíduos na Figura 14 mostram adequabilidade do ajuste do modelo de redes neurais de múltipla camada na fase de teste uma vez que mostram média nula, aleatoriedade e a variância constante.

Figura 14: Valores dos resíduos obtidos na fase de teste do modelo com as estimativas da massa específica do etanol hidratado combustível.



O Erro Médio Quadrático Normalizado (Normalized Mean Squared Error - NMSE) é utilizado para avaliar a complexidade do modelo, a qual será também avaliada a fim de obter um modelo menos complexo. Por se tratar de uma rede de aprendizagem supervisionada, cujo objetivo do processo de aprendizagem é ajustar os parâmetros livres (pesos sinápticos e bias) da rede para minimizar a energia do erro quadrado, o NMSE é um dos principais parâmetros de avaliação utilizado.

A rede neural de múltiplas camadas com a estrutura contendo duas camadas ocultas e seis neurônios em cada uma mostraram eficiência na estimativa da massa específica com erros médios nas estimativas de 1,0 e 0,8 % nas fases de treinamento e teste, respectivamente. As Equações 3 e 4 foram utilizadas para calcular os erros quadráticos médios nas fases de treinamento e teste do modelo RNA de múltiplas camadas.

$$NMSE = \frac{\sum_t (observação_t - predição_t)^2}{\sum_t (observação_t - observação_{t-1})^2} = 0,010371 \quad \text{Equação 3}$$

$$NMSE = \frac{\sum_t (observação_t - predição_t)^2}{\sum_t (observação_t - observação_{t-1})^2} = 0,008305 \quad \text{Equação 4}$$

4.3. Validação – RNA de Múltiplas Camadas

Foi realizada a validação cruzada com o objetivo de verificar a capacidade de generalização da rede, ou seja, a capacidade da rede neural de múltiplas camadas utilizada de estimar a massa específica de padrões que não foram apresentados durante a fase de treinamento e teste. Para a validação a rede foi processada utilizando os dados obtidos no ano de 2013.

A Figura 15 mostra os valores experimentais e os valores estimados pela rede neural de múltiplas camadas na fase de validação do modelo, ilustrando a sua adequabilidade, enquanto que a Figura 16 mostra os valores dos resíduos com aleatoriedade, média nula e variabilidade constante.

Figura 15: Valores estimados pela rede neural artificial de múltiplas camadas na fase de validação e os valores medidos da massa específica (g/cm^3).

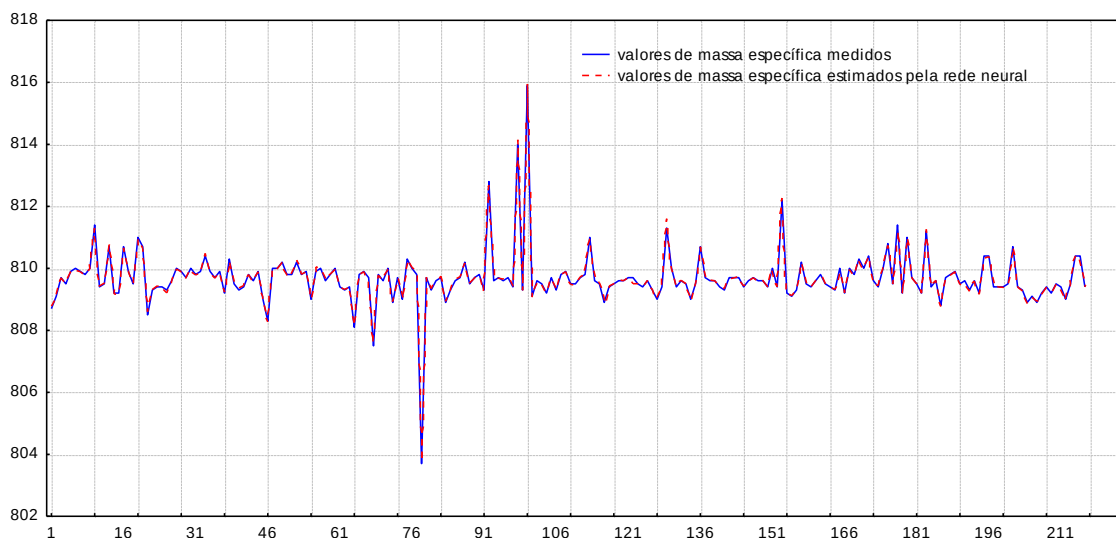
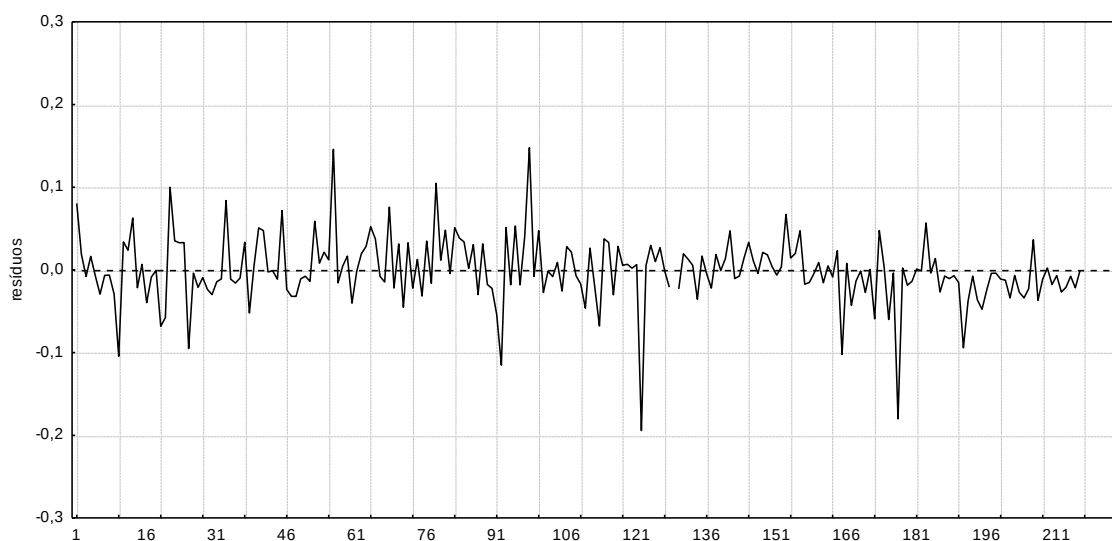


Figura 16: Valores dos resíduos do modelo neural na fase de validação.



O valor de NMSE para as estimativas da massa específica obtidas com a rede neural de múltiplas camadas na fase de validação foi de 0,001271 o que comprova a adequabilidade do modelo.

A Equação 5 foi utilizada para calcular o erro quadrático médio na fase de validação do modelo RNA de múltiplas camadas.

$$NMSE = \frac{\sum_t (observação_t - predição_t)^2}{\sum_t (observação_t - observação_{t-1})^2} = 0,001271 \quad \text{Equação 5}$$

4.4. Regressão por Mínimos Quadrados Parciais (PLS)

As variáveis latentes são as novas variáveis que serão utilizadas para representar os dados. Como as variáveis originais estão sendo trocadas por novas variáveis, deve-se avaliar até que ponto a troca é representativa do conjunto de dados, além de avaliar quantas variáveis latentes devem ser utilizadas para representar o conjunto de dados.

Na construção do modelo é apresentada a porcentagem de variância capturada, ou seja, a quantidade de dados originais que está sendo explicada em função do número de variáveis latentes. Na Tabela 1 são apresentados os resultados das variâncias explicadas e acumuladas das variáveis latentes obtidas.

Tabela 1: Variâncias explicadas e acumuladas das VLs.

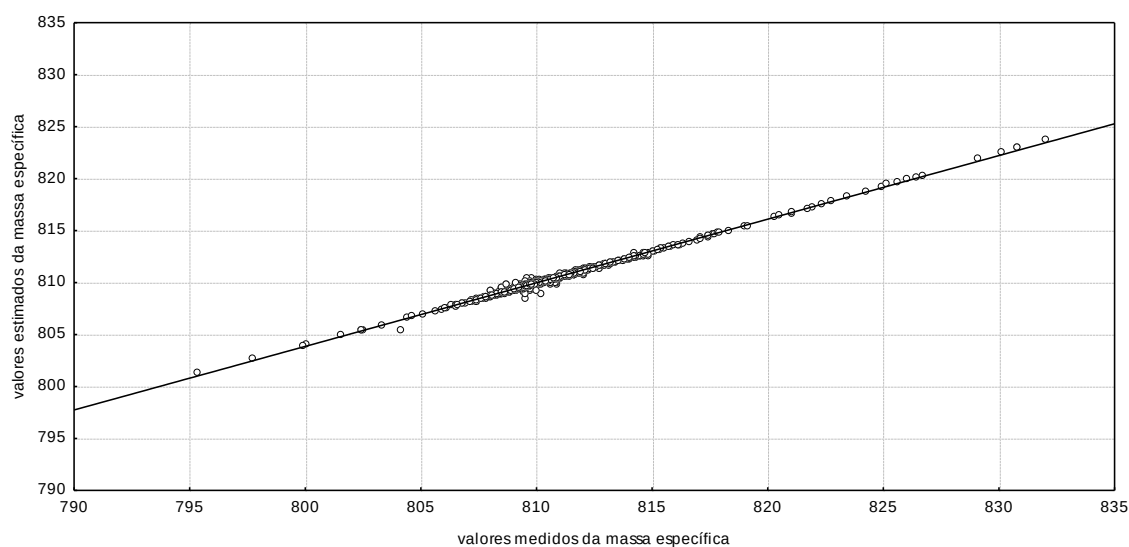
VLs	Variância explicada de Y	Variância acumulada de Y	Variância explicada de X	Variância acumulada de X
1	0,940316	0,940316	0,393094	0,393094
2	0,053289	0,993605	0,269956	0,663050
3	0,000113	0,993718	0,336950	1,000000

Foram utilizadas três variáveis latentes no modelo, mas apenas duas foram escolhidas, pois o ganho a partir da terceira variável latente é muito pequeno (0,0113 %), não justificando o uso de demais variáveis.

Por fim, com a seleção das variáveis, o conjunto de dados com três variáveis latentes (VLs) e 6566 amostras foi reduzido a duas variáveis latentes (VLs) e 6566 amostras, formando assim o conjunto de dados utilizados no presente trabalho.

Os conjuntos de dados referentes ao período 2002 a 2011 foram utilizados na construção do modelo enquanto que o ano de 2013 foi utilizado para validação do modelo PLS.

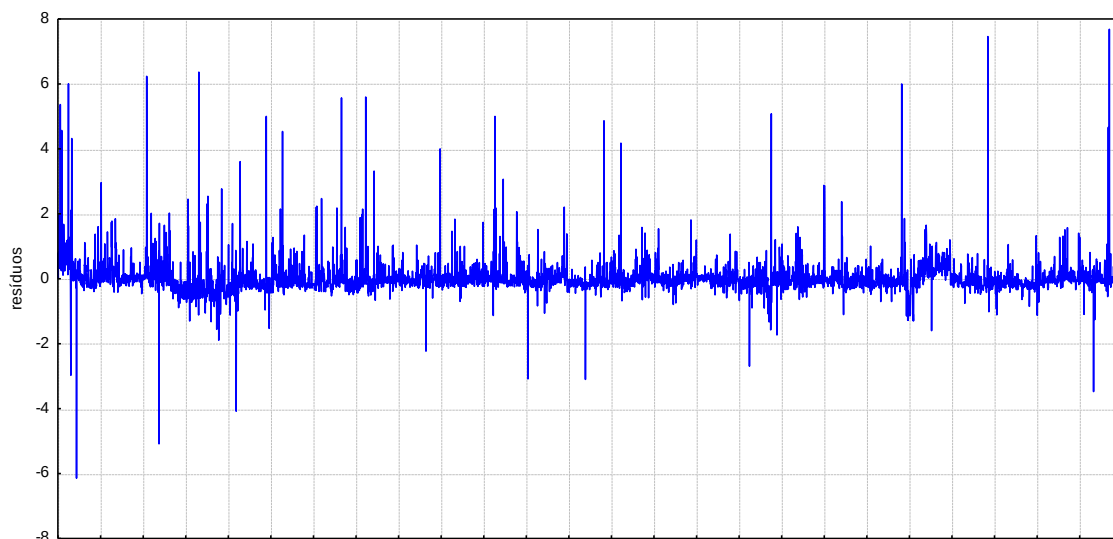
Na Figura 17 estão os valores estimados *versus* os valores previstos pelo modelo de calibração desenvolvido (PLS).

Figura 17: Valores estimados da massa específica (g/cm^3) por PLS versus valores medidos.

O modelo PLS desenvolvido apresentou erro superior aos erros obtidos com RNAs de múltiplas camadas. O valor do erro quadrático médio para o modelo PLS foi de 0,093881 conforme a Equação 6. Pelas Figuras 17 e 18 podemos admitir a adequabilidade do modelo PLS. O modelo resultante foi: $Ma_Esp = 0,008699 - 0,788400 (Teor_alco) + 0,006210 (pH) + 0,000336 (Cond_EI)$.

$$NMSE = \frac{\sum_t (observação_t - predição_t)^2}{\sum_t (observação_t - observação_{t-1})^2} = 0,093881 \quad \text{Equação 6}$$

Figura 18: Valores dos resíduos das estimativas do modelo PLS no período de 2002 a 2011.



A Figura 18 mostra os valores dos resíduos das estimativas obtidas para o período de 2002 a 2011 do treino. Qualitativamente, este gráfico pode indicar se os dados seguem ou não um comportamento linear. A distribuição aleatória desses resíduos é um indicativo de comportamento linear.

4.5. Validação – PLS

As Figuras 19 e 20 mostram os valores experimentais versus os estimados ilustrando a eficácia da validação e os resíduos apresentando média nula e variância constante.

Figura 19: Valores estimados pela análise de regressão PLS versus valores medidos da massa específica (g/cm³) medida.

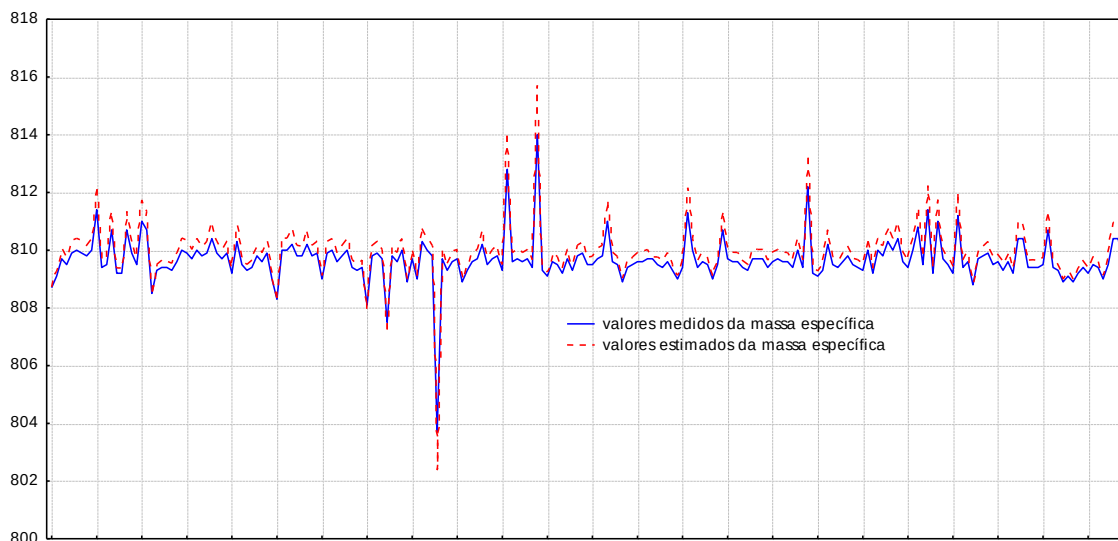
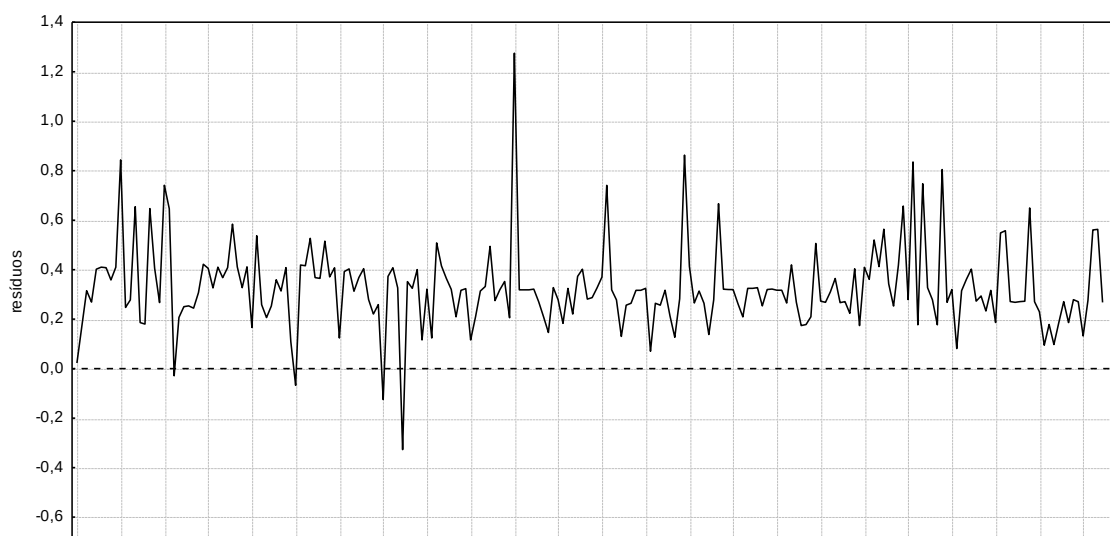


Figura 20: Valores dos resíduos estimados por PLS da massa específica para o ano de 2013 (validação).



Os erros com as estimativas obtidas com o modelo PLS na validação foi superior aos erros das estimativas por RNAs. A Figura 20 mostra que a média não foi nula, porém pode-se admitir aleatoriedade e variância constante. A Equação 7 mostra um erro superior apresentando pouca exatidão.

$$NMSE = \frac{\sum_t (observação_t - predição_t)^2}{\sum_t (observação_t - observação_{t-1})^2} = 0,222156$$

Equação 7

5. CONCLUSÃO

No levantamento estatístico preliminar, aplicado ao banco de dados do etanol hidratado combustível, observou-se que a água influencia na massa específica e com o decorrer do tempo houve uma melhora na qualidade do etanol hidratado combustível.

O modelo de previsão proposto, baseado nos mínimos quadrados parciais apresentou valores de erro quadrático médio elevado (22 %). E o modelo de rede neural proposto mostrou bons resultados (0,1 %).

A aplicação destes modelos aos dados de etanol hidratado combustível indicou que os modelos de RNAs de múltiplas camadas são mais eficientes do que os modelos de regressão por mínimos quadrados parciais no cálculo das estimativas da massa específica do etanol. Os modelos de RNAs de múltiplas camadas alcançaram os melhores ajustes para massa específica quando comparados aos modelos de regressão por mínimos quadrados parciais.

Estudos futuros devem ser feitos para aprimoramento dos ajustes dos modelos RNAs de múltiplas camadas e modelos de regressão por mínimos quadrados parciais.

6. REFERÊNCIAS BIBLIOGRÁFICAS

1. NEVES, F. N. C. Regulação da indústria do etanol combustível no Brasil. Dissertação de Mestrado, Faculdade de Tecnologia e Ciências, 2011.
2. GONÇALVES, M. A., *et al.* Avaliação de laboratórios brasileiros na determinação de alguns parâmetros de qualidade de combustíveis. *Química Nova*, v. 36, n 3, 393-399, 2013.
3. Maior produtor mundial de cana-de-açúcar. UNICA. Disponível em <http://www.unica.com.br/faq/>. Acesso em 17/07/2013.
4. Álcool combustível derivado da cana-de-açúcar e o desenvolvimento sustentável. Disponível em www.feagri.unicamp.br/energia/agre2004/Fscommand/PDF/Agrener/Trabalho%20110.pdf. Acesso em 17/07/2013.
5. Análise prospectiva do álcool combustível no Brasil – Cenários 2004 a 2024. Tese de Doutorado, Universidade Estadual de Campinas, 2005.
6. MORAES, M. A. F. D. de. A Desregulamentação do Setor Sucroalcooleiro do Brasil – Caminho Editorial, Americana, SP, 2000.
7. Etanol: Situação atual, desafios e perspectivas. Disponível em <http://www2.camara.leg.br/atividadelegislativa/comissoes/comissoespermanentes/cme/banners/etanol/Etanol-web.pdf>. Acesso em 10/07/2013.
8. ANP – Agência Nacional de Petróleo, Gás Natural e Biocombustíveis. Resolução nº 7/2011. Disponível em: www.anp.gov.br.
9. ANP – Agência Nacional de Petróleo, Gás Natural e Biocombustíveis. Resolução nº 5/2006. Disponível em: www.anp.gov.br.
10. PAULILLO, L. F. *et al.* Álcool Combustível e Biodiesel no Brasil: *quo vadis?*. *Rev. Econ. Sociol. Rural*, v. 45, nº 3, 531-565, 2007.
11. LUCON, O. Aspectos ambientais na cadeia de bicomcombustíveis. Apresentado no workshop “Aspectos ambientais da cadeia do etanol de cana-de-açúcar”, Projeto diretrizes de políticas públicas para a agroindústria canavieira do Estado de São Paulo, São Paulo, 2008.

12. MACEDO, I. C. *et al.* Balanço das emissões de gases do efeito estufa na produção e no uso do etanol no Brasil. São Paulo: Secretaria do Meio Ambiente, Governo de São Paulo, 2007.
13. SANTOS, A. L.; TAKEUCHI, R. M.; MUÑOZ, R. A. A.; ANGNES, L.; STRADIOTTO, N. R. Electrochemical Determination of Inorganic Contaminants in Automotive Fuels. *Electroanalysis*, 24, 1681-1691, 2012.
14. White paper on internationally compatible biofuel standards. Tripartite task force: Brazil, European Union & United States of America, 2007. Disponível em: www.inmetro.gov.br/painelsetorial/biocombustiveis/whitepaper.pdf
15. SANTOS, A. L. *et al.* Novos métodos analíticos para avaliação da qualidade do bioetanol combustível. In: LEMOS, E. G. M.; STRADIOTTO, N. R. (Org.). Bioenergia Desenvolvimento, Pesquisa E Inovação. 1 ed. São Paulo: Editora UNESP, 945-978, 2012.
16. Entendendo a adulteração de combustíveis. Ministério Público Federal, 3ª Edição Revista e Atualizada, 2007.
17. MARTINIANO, L. C. *et al.* Direct simultaneous determination of Pb(II) and Cu(II) in biodiesel by anodic stripping voltammetry at a mercury-film electrode using microemulsions. *Fuel*, 1164-1167, 2012.
18. GHISI, M. *et al.* Simple method for the determination of Cu and Fe by electrothermal atomic absorption spectrometry in biodiesel treated with tetramethylammonium hydroxide. *Microchemical Journal*, 98, 62-65, 2011.
19. CÔCCO, L. C. Aplicação de Redes Neurais Artificiais para Previsão de Propriedades da Gasolina a partir de sua Composição Química. Dissertação de Mestrado, Universidade Federal do Paraná, 2003.
20. ARYEE, A. N. A., VAN DE VOORT, F. R., SIMPSON, B. K. FTIR determination of free fatty acids in fish oils intended for biodiesel production. *Process Biochemistry*, 44 (4), p. 401-105, 2009.
21. ALEME, H. G. Determinação de Parâmetros Físico-Químicos do Óleo Diesel a partir de Curvas de Destilação utilizando técnicas quimiométricas. Tese de Doutorado, Universidade Federal de Minas Gerais, 2011.

22. NAES, *et al.* A User Friendly Guide to Multivariate Calibration and Classification, NIR Publications, 2002.
23. BUNTINE, W. L.; WEIGEND, A. S. Bayesian back-propagation. *Complex Systems*, 5, 1991. 603-643.
24. DUDA, R. O.; HART, P. E. Pattern classification and scene analysis. New York: Wiley, 1973.
25. MENARD, S. Applied logistic regression analysis, series: Quantitative applications in the social sciences. Thousand Oaks: CA: Sage, 1993.
26. SNEDECOR, G. W.; COCHRAN, W. G. Statistical methods. 7th ed., 1980.
27. YANG, H.; GRIFFITHS, P. R.; TATE, J. D. Comparison of partial least squares regression and multi-layer neural networks for quantification of nonlinear systems and application to gas phase Fourier transform infrared spectra. *Anal. Chim. Acta*, 489, 2003. 125–136.
28. BERTRAN, E. *et al.* Handling intrinsic non-linearity in near-infrared reflectance spectroscopy. *Chemometr. Intell. Lab.*, 49, 1999. 215-224.
29. SEKULIC, S. *et al.* Nonlinear multivariate calibration methods in analytical chemistry. *Anal. Chem.*, 65, n. 19, 1993. 835–845.
30. GELADI, P.; HADJIISKI, L. M.; HOPKE, P. Multiple regression for environmental data: nonlinearities and prediction bias. *Chemometr. Intell. Lab.*, 47, 1999. 165–173.
31. LI, Y.; BROWN, C. W.; LO, S. C. Near infrared spectroscopic determination of alcohols-solving non-linearity with linear and non-linear methods. *J. Near Infrared Spec.*, 7, 1999. 55–62.
32. BALABIN, R. M.; SAFIEVA, R. Z.; LOMAKINA, E. I. Comparison of linear and nonlinear calibration models based on near infrared (NIR) spectroscopy data for gasoline properties prediction. *Chemometr. Intell. Lab.*, 88, 2007. 183-188.
33. BAPTISTA, P. *et al.* Multivariate near infrared spectroscopy models for predicting the iodine value, CFPP, kinematic viscosity at 40°C and density at 15°C of biodiesel. *Talanta*, 77, 2008. 144-151.
34. SCHALKOFF, R. Artificial Neural Networks. McGraw-Hill: Clemson University, 1997.

35. ANDERSON, J. A. Cognitive and psychological computation with neural models. *IEEE Trans. Sys. Man Cybern.*, 13, n. 5, 1983. 799-814
36. DORADO, M. P.; BALLESTEROS, E.; LÓPEZ, F. J.; MITTELBACH, M. Optimization of Alkali-Catalyzed Transesterification of Brassica Carinate Oil for Biodiesel Production; *Energy & Fuels*, 18, 77-83, 2004.
37. FELIZARDO, P.; BAPTISTA, P.; MENEZES, J. C.; CORREIA, M. J. N. Multivariate Near Infrared Spectroscopy Models for Predicting Methanol and Water Content in Biodiesel; *Analytica Chimica Acta*, 595, 107-113, 2007.
38. DELGADO, R. C. O. B.; ARAUJO, A. S.; FERNANDES JR., V. J. Properties of Brazilian gasoline mixed with hydrated ethanol for flex-fuel technology. *Fuel Processing Technology*. v. 88, p. 365 – 368, 2007.
39. OLIVEIRA F. C. C. *et al.* A escolha da faixa espectral no uso combinado de métodos espectroscópicos e quimiométricos. *Química Nova*, v. 27, n.2, p. 218-225, 2004.
40. Associação Brasileira de Normas Técnicas. NBR 14065 – Destilados de petróleo e óleos viscosos – Determinação da massa específica e da densidade relativa pelo densímetro digital, 2013
41. DMA 4500 / 5000 – Instruction Handbook, Viena: ANTOON PAAR GmbH. Versão 4.510d, 2001.
42. Associação Brasileira de Normas Técnicas. NBR 5992 - Determinação da massa específica e do teor alcoólico do álcool etílico e suas misturas com água, 1980.
43. Associação Brasileira de Normas Técnicas. NBR 10547 – Etanol combustível – Determinação da condutividade eletrolítica, 2013
44. Comparação interlaboratorial de Condutividade Eletrolítica em Etanol Combustível. INMETRO, 2010.
45. Associação Brasileira de Normas Técnicas. NBR 10891 – Etanol hidratado combustível – Determinação de pH – Método Potenciométrico, 2013.