

AVALIAÇÃO METROLÓGICA DA CURVA ANALÍTICA PARA DETERMINAÇÃO DO TEOR DE ENXOFRE EM BAIXÍSSIMAS CONCENTRAÇÕES EM ÓLEO DIESEL S-10Shirley L. de Farias^{a,b}, Lívio G. da Costa^a, Rodrigo F. Calili^b, Ramon M. Rios^c e Elcio C. de Oliveira^{a,b,*}^aPetrobras Transporte S.A., 20091-060 Rio de Janeiro – RJ, Brasil^bDepartamento de Metrologia, Pontifícia Universidade Católica do Rio de Janeiro, 22453-900 Rio de Janeiro – RJ, Brasil^cPetróleo Brasileiro S.A., 20271-000 Rio de Janeiro – RJ, Brasil

Recebido em 07/01/2015; aceito em 06/03/2015; publicado na web em 28/04/2015

METROLOGICAL EVALUATION OF THE ANALYTICAL CURVE FOR DETERMINATION OF VERY LOW CONCENTRATIONS OF SULPHUR CONTENT IN DIESEL OIL S-10. The increasing search for a reduction in air emissions has provided a challenging analytical scenario. This paper discusses metrologically, the best practices relating to analytical curves for the determination of very low concentrations of sulfur content in diesel S-10 by monochromatic wavelength dispersive X-ray fluorescence spectrometry (MWDXRF). For weighted linear regression, a new strategy to correlate the weights with the output variable is proposed, which is lacking in the literature. The analytical curve 0-500 mg kg⁻¹, built in compliance with international standards, featured heteroscedastic behavior. The results showed that weighted linear regression is the most suitable choice for assessing concentrations at the beginning of an analytical curve, such as 10 mg kg⁻¹ (product specification limit) rather than simple linear regression which holds that the measured results must be closest to the centroid of the regression line. Finally, it was observed that, when a shorter analytical curve is built, it loses both in accuracy and in precision for very low concentrations, as compared to simple and weighted linear regression.

Keywords: metrological evaluation of the analytical curve; simple linear regression; weighted linear regression; very low concentrations of sulfur content in diesel S-10; MWDXRF.

INTRODUÇÃO

No Brasil, o principal modal para transporte de óleo diesel é por dutos. O custo anual desta atividade é de bilhões de dólares, uma vez que grandes quantidades devem ser transportadas por longas distâncias em torno de um país ou mesmo entre países da América do Sul. Embora o investimento necessário para preparar os sistemas de polidutos seja alto, os custos operacionais são muito baixos quando comparados com outros meios de transporte, como trens ou caminhões.¹

A arquitetura mais simples de um duto é composta por uma origem, um destino e um único tipo de produto a ser transportado. Em uma rede mais complexa e realista, dutos podem ter múltiplos destinos e podem transportar vários produtos, tais como nafta, querosene, gasolina e diesel.² Quando os produtos são transportados sequencialmente, os dutos são chamados de polidutos. Isto conduz, inevitavelmente, à formação de uma interface de mistura destes produtos que pode ser um volume muito grande, dependendo da vazão, do diâmetro do duto, da distância durante a qual os produtos vão estar em contato e das características dos mesmos. O produto gerado pela mistura dos derivados é chamado de interface de mistura ou transmix. Transmix significa uma interface que não atende as especificações para que um combustível possa ser usado ou comercializado.³ O envio do transmix junto ao derivado pode causar desacordo com o cliente.

O principal desafio em sistemas de polidutos é o planejamento da sequência em que os produtos são transportados da refinaria para o terminal, e, em seguida, para o cliente. Isso torna difícil detectar o ponto de corte exato em que os produtos entregues ainda estão em conformidade com os requisitos estabelecidos pela agência reguladora. Dependendo do grau de contaminação, os produtos misturados tendem a perder as suas características originais.

O aumento das redes de polidutos e o fato de que novos produtos, especificamente, diesel de baixíssimo teor de enxofre, estão sendo adicionados aos sistemas de dutos, certamente aumenta este desafio. Este cenário ainda fica potencializado pelo fato de cada propriedade físico-química do derivado estar otimizada; ou seja, estar próxima ao limite que o “tira de especificação”.

Assim sendo, o objetivo deste trabalho é avaliar metrologicamente a análise de enxofre total em óleo diesel rodoviário S10, cujo limite de especificação máximo é 10 mg kg⁻¹ e qualquer divergência analítica pode alterar a sua situação de conformidade à especificação.

MOTIVAÇÃO

O conhecimento do teor de enxofre em óleo diesel é importante para a previsão e controle de problemas operacionais de processo, tais como corrosão, envenenamento de catalisador, que não permite a redução das emissões de NOx e hidrocarbonetos, degradação de produtos provenientes de misturas e impacto ambiental.

O elevado teor de enxofre em óleo diesel comercializado na indústria automotiva no Brasil, até recentemente, ocasionava os principais problemas com relação à emissão de material particulado.⁴

Segundo dados do Ministério do Meio Ambiente (2014), houve uma redução drástica da quantidade de material particulado desde a fase inicial do Programa de Controle da Poluição por Veículos Automotores (PROCONVE) até a fase atual, de 0,660 g kWh⁻¹ a 0,0122 g kWh⁻¹ (caminhões pesados) e 0,0144 g kWh⁻¹ (ônibus rodoviários), respectivamente, atingindo índices praticados na Europa.⁵

A participação relativa da emissão de material particulado por caminhões pesados pretende passar de 45% em 2009 para 53% em 2020 e o percentual de emissão de material particulado do óleo diesel deve diminuir de 96% para 92%.⁶ Sendo assim, os limites de emissão definidos e restritivos no PROCONVE são de fácil observação na

*e-mail: elciooliveira@petrobras.com.br

emissão de material particulado, razão pela qual se justifica todo o investimento na tecnologia, logística e análise do óleo diesel S10.

METODOLOGIA

Um método analítico instrumental é geralmente calibrado pela observação das repostas (y) aos diferentes níveis do analito (x). Na prática, uma curva analítica é obtida a partir da quantidade medida y de amostras de concentração x conhecida, provenientes de um conjunto de padrões. Na calibração univariada clássica, n pontos de calibração y definem a curva analítica ($y = f(x)$), e a quantidade desconhecida (x_0) é determinada pela solução da equação ($y_0 = f(x_0)$), na qual y_0 é a resposta para uma concentração desconhecida. Na maioria dos casos, assume-se que esta relação é linear ($y = a + bx$), quando o metrologicamente correto é uma avaliação por ANOVA.⁷ Uma curva analítica, que nada mais é do que um modelo estatístico, é, então, utilizada para obter a concentração do analito de uma amostra (x_{previsto}) que produz uma resposta observada ($y_{\text{observada}}$). Os coeficientes de ajuste da curva são obtidos por meio de regressão linear simples, ou ponderada, de um conjunto de n pares de valores (x_i, y_i). Uma regressão linear convencional (não ponderada), ou simples, possui duas premissas principais:

- Os valores da variável independente x utilizados na construção da curva analítica são considerados com incerteza desprezível.
- A variável de resposta y possui erros (resíduos) aleatoriamente distribuídos com desvios padrão constantes – homocedasticidade.

A segunda proposição nem sempre é atendida, pois é muito comum que a precisão de y varie com a concentração e o mais indicado nesses casos seria empregar uma regressão ponderada pela variância, o que significa dar peso maior aos pontos de menor concentração e maior precisão.

Teste de homogeneidade dos resíduos

Muitas vezes, os erros da regressão linear não possuem variância constante, chamado comportamento heterocedástico. Desse modo, os parâmetros da curva obtidos por meio de uma regressão linear simples, na qual todos os pontos possuem a mesma influência, não são os mais adequados. A hipótese de homocedasticidade pode ser verificada tanto pela avaliação visual dos resíduos quanto por meio de testes estatísticos. O teste empregado nesse trabalho para diagnóstico da homocedasticidade dos erros, raramente discutido em Química Analítica, é o teste de Goldfeld-Quandt.⁸

No presente trabalho, a causa da heterocedasticidade está na dependência da precisão com a concentração, em outras palavras, a variância em y aumenta com a concentração. Os métodos mais utilizados para verificar a hipótese de igualdade das variâncias são os testes de Cochran, Bartlett e Levene. Dentre esses, o mais comum é o teste de Cochran,⁹ que é aplicado apenas quando todos os desvios padrão são obtidos a partir de um mesmo número de resultados em condições de repetibilidade. O teste de Cochran é um teste unicaudal, aceito depois de verificar a uniformidade da regressão e comparar a maior variância com as demais. O teste consiste em calcular $C_{\text{calculado}}$ e comparar com C_{tabelado} , Equação (1). Se a repetibilidade das medidas é independente do valor da concentração, esta condição de variância uniforme é chamada homocedástica, e quando não uniforme, heterocedástica.

$$C_{\text{calculado}} = \frac{s_{\text{máx}}^2}{\sum_{i=1}^p s_i^2} \quad (1)$$

na qual: $s_{\text{máx}}^2$ = maior variância; s_i^2 = variância amostral.

Os testes de Bartlett e Levene também são largamente utilizados. Esses testes são testes bicaudais e o método proposto por Bartlett tem um comportamento melhor quando os dados seguem uma distribuição normal.

Regressão linear simples

Nos casos onde os erros presentes no eixo x são desprezíveis em relação aos do eixo y , busca-se o melhor ajuste que minimiza os desvios no eixo y entre os pontos experimentais e os valores previstos pelo modelo. Uma vez que alguns desses desvios (tecnicamente conhecidos como os resíduos em y) são positivos e outros negativos, é sensato minimizar a soma dos quadrados dos resíduos, uma vez que estes quadrados passam a ser positivos. Isso explica o uso frequente do termo método dos mínimos quadrados para o procedimento. A reta necessária é calculada sobre esse princípio e deve passar pelo centroide dos pontos, (\bar{x}, \bar{y}) .

Os coeficientes angular e linear da regressão podem ser calculados pelas Equações (2) e (3), respectivamente:

$$b = \frac{\sum_i [(x_i - \bar{x})(y_i - \bar{y})]}{\sum_i (x_i - \bar{x})^2} \quad (2)$$

$$a = \bar{y} - b\bar{x} \quad (3)$$

É importante observar que os valores experimentais e os valores previstos pelo modelo de regressão não são exatamente iguais. O desvio padrão da concentração associado à regressão pode ser calculado pela Equação (4):⁹

$$s_{x_0} = \frac{s_{y/x}}{b} \sqrt{\frac{1}{m} + \frac{1}{n} + \frac{(y_0 - \bar{y})^2}{b^2 \sum_i (x_i - \bar{x})^2}} \quad (4)$$

$s_{y/x} = \sqrt{\frac{\sum_i (y_i - y_i)^2}{n-2}}$: desvio padrão da regressão; \hat{y}_i : valores previstos pelo modelo de regressão; x_i e y_i : valores experimentais; m : número de medições para determinar x_0 ; n : número de medições para a construção da curva analítica.

O intervalo de confiança, IC , para x_0 pode ser calculado pela Equação (5):

$$IC = x_0 \pm t_{(n-2)} \times s_{x_0} \quad (5)$$

Regressão linear ponderada

Entretanto, em alguns casos, o desvio padrão dos valores de y , sinal, aumenta com a concentração, Figura 1, comportamento heterocedástico. Estas curvas são melhores ajustadas por uma regressão linear ponderada.

Os coeficientes da curva analítica devem ser estimados para dar peso adicional aos pontos onde as barras de erro são menores, ou seja, é mais importante que a reta ajustada passe mais próxima aos pontos de menor erro. Com isso, os pontos iniciais da curva, menores desvios padrão, ficam mais bem ajustados do que os pontos finais da curva, maiores desvios padrão. Assim sendo, o peso atribuído a cada ponto diminui com a concentração e deve ser inversamente proporcional a sua variância, s_i^2 , ou seja, o peso será maior quando a variância for menor. Se os pontos individuais da curva analítica são considerados (x_1, y_1) , (x_2, y_2) , etc., e os desvios

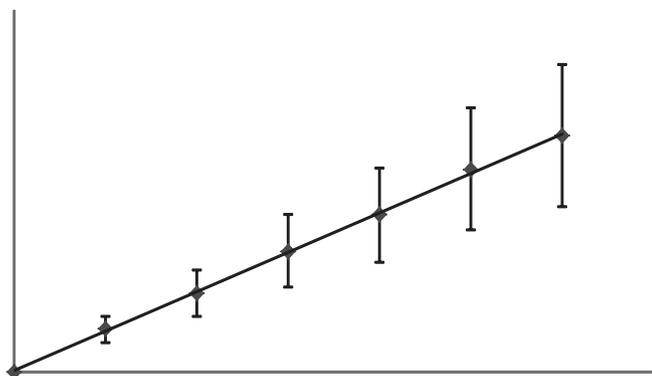


Figura 1. Ponderação de erros em um cálculo de regressão

padrão correspondentes são s_1, s_2, \dots , então os pesos individuais, w_1, w_2, \dots , são dados pela Equação (6):

$$w_i = \frac{s_i^{-2}}{\sum_i s_i^{-2}/n} \quad (6)$$

onde $s_i^{-2} = 1/s_i^2$.

Ao utilizar o divisor n no denominador da Equação (6), os pesos foram dimensionados de modo a que a sua soma seja igual ao número de pontos no gráfico.

A utilização da equação (6), baseada nas variâncias experimentais, seria excelente caso as variâncias fossem conhecidas ou obtidas de forma bastante confiável, porém, isso requer uma quantidade muito grande de repetições, para minimizar o efeito do caráter não linear. Uma estratégia para minimizar este efeito, não disponível na literatura, é encontrar uma função que relacione a variância com os valores de y . A função indicada para modelar variâncias é uma função do tipo potência, genericamente $y = ax_i^b$, que para este caso específico toma a forma da Equação (7), por deixar resíduos menores:

$$s_i^2 \approx \kappa_1 \bar{y}_i^{\kappa_2} \quad (7)$$

A aplicação do logaritmo lineariza a função potência, Equações (8) e (9):

$$\ln(s_i^2) = \ln(\kappa_1 \bar{y}_i^{\kappa_2}) \quad (8)$$

$$\ln(s_i^2) = \ln(\kappa_1) + \kappa_2 \ln(\bar{y}_i) \quad (9)$$

nas quais os coeficientes da regressão são: $a = \kappa_1$; $b = \kappa_2$.

Desse modo, os pesos w_i são estimados por meio do ajuste da equação (9) e posterior utilização da equação (6).

Os coeficientes angular e linear da curva de regressão ponderada são, então, dados pelas Equações (10) e (11), respectivamente:

$$b_w = \frac{\sum_i w_i x_i y_i - n \bar{x}_w \bar{y}_w}{\sum_i w_i x_i^2 - n \bar{x}_w^2} \quad (10)$$

$$a_w = \bar{y}_w - b_w \bar{x}_w \quad (11)$$

Nestas equações \bar{y}_w e \bar{x}_w representam coordenadas do centroide ponderado,¹⁰ através do qual a linha de regressão ponderada deve passar. Estas coordenadas podem ser calculadas por $\bar{x}_w = \sum_i w_i x_i / n$ e $\bar{y}_w = \sum_i w_i y_i / n$.

O desvio padrão de uma concentração prevista por esta regressão ponderada, $s_{x_{0w}}$, pode ser calculado pela Equação (12):¹¹

$$s_{x_{0w}} = \frac{s_{(y/x)_w}}{b} \left[\frac{1}{m \times w_0} + \frac{1}{n} + \frac{(y_0 - \bar{y}_w)^2}{b^2 \left(\sum_i w_i x_i^2 - n \bar{x}_w^2 \right)} \right]^{1/2} \quad (12)$$

$$s_{(y/x)_w} = \left(\frac{\sum_i w_i (y_i - y_i)^2}{n - 2} \right)^{1/2} \quad (\text{desvio padrão da regressão}); w_0 \text{ é o peso apropriado para o valor de } y_0.$$

O intervalo de confiança, IC , para x_{0w} pode ser calculado pela Equação (13):

$$IC = x_{0w} \pm t_{(n-2)} \times s_{x_{0w}} \quad (13)$$

A forma geral dos intervalos de confiança para uma concentração calculada é mostrada na Figura 2.

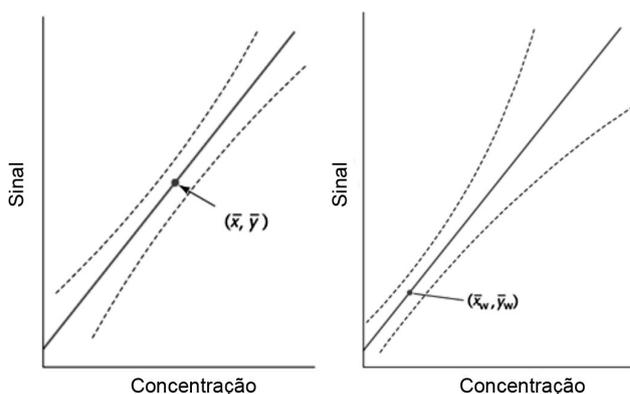


Figura 2. Intervalos de confiança a partir de uma regressão: (a) simples e (b) ponderada

Os desvios padrão (erro padrão) dos coeficientes (inclinação, s_{b_w} e intercepto, s_{a_w}) para a regressão ponderada podem ser calculados pelas Equações (14) e (15), respectivamente:¹¹

$$s_{b_w} = \frac{s_{(y/x)_w}}{\sqrt{\sum (x_i - \bar{x})^2}} \quad (14)$$

$$s_{a_w} = s_{(y/x)_w} \sqrt{\frac{\sum x_i^2}{n \sum (x_i - \bar{x})^2}} \quad (15)$$

Os intervalos de confiança para estes coeficientes, também, são calculados analogamente pela Equação (13).

PARTE EXPERIMENTAL

O teor de enxofre de 8 materiais de referência certificados, Tabela 1, na matriz óleo mineral, em 6 repetições, foi analisado no Terminal Terrestre de Guarulhos em junho de 2014. Os experimentos foram realizados em ordem crescente de concentração, a fim de evitar contaminação na célula e em repetições verdadeiras.

O experimento foi realizado no equipamento XOS SINDIE GEN 3, que se baseia no padrão ASTM D7039-13,¹² cuja técnica analítica é a fluorescência de Raios X com dispersão monocromática de comprimento de onda (MWDXRF).

Método de ensaio

Um feixe de raios-X monocromático com comprimento de onda

Tabela 1. Experimento para análise do teor de enxofre

Concentração, mg kg ⁻¹	Contagem					
0,00	177	182	172	182	164	199
5,04	922	1085	927	987	991	919
10,29	1842	1841	1791	1962	1843	1897
15,16	2562	2606	2566	2650	2720	2687
25,18	4023	4246	4387	4168	4016	4206
50,48	8100	8408	8404	8419	8400	8306
250,00	40056	39499	40523	41364	41326	40303
495,90	80046	82206	79524	82594	82243	81706

adequado para excitar os elétrons da camada *K* do enxofre é emitido em uma amostra contida em uma célula de amostra. A radiação fluorescente *Kα*, em 0,5373 nm (5,373 Å), emitida pelo enxofre é receitada por um monocromador fixo. A intensidade (contagens por segundo) dos raios X de enxofre é medida usando um detector apropriado e convertida para a concentração de enxofre (mg kg⁻¹) na amostra de teste a partir da curva analítica. A excitação por raios X monocromáticos reduz a radiação de fundo (background), oriunda de espalhamentos sem interesse, simplifica a correção do efeito de matriz e aumenta a relação sinal / radiação de fundo em comparação com excitação policromática utilizada em técnicas convencionais de fluorescência de raios X.

RESULTADOS E DISCUSSÃO

Os dados foram tratados de duas maneiras distintas, a fim de avaliar a abordagem mais adequada, voltada à concentração de interesse: 10 mg kg⁻¹.

A primeira abordagem segue a orientação da norma ASTM D7039-13, com a construção da curva analítica de 0 a 500 mg kg⁻¹.

A segunda abordagem propõe a construção de uma curva analítica de menor extensão, de 0 a 50 mg kg⁻¹.

Nas duas abordagens, este trabalho considera a regressão simples e ponderada.

Abordagem 1: Curva analítica com os pontos recomendados pela ASTM D7039-13

Os resultados da construção da curva analítica, de acordo com a norma ASTM D7039-13, estão apresentados na Tabela 2.

Tabela 2. Curva analítica de 0 a 500 mg kg⁻¹

Concentração, mg kg ⁻¹	Contagem					
0,00	177	182	172	182	164	199
5,04	922	1085	927	987	991	919
15,16	2562	2606	2566	2650	2720	2687
50,48	8100	8408	8404	8419	8400	8306
250,00	40056	39499	40523	41364	41326	40303
495,90	80046	82206	79524	82594	82243	81706

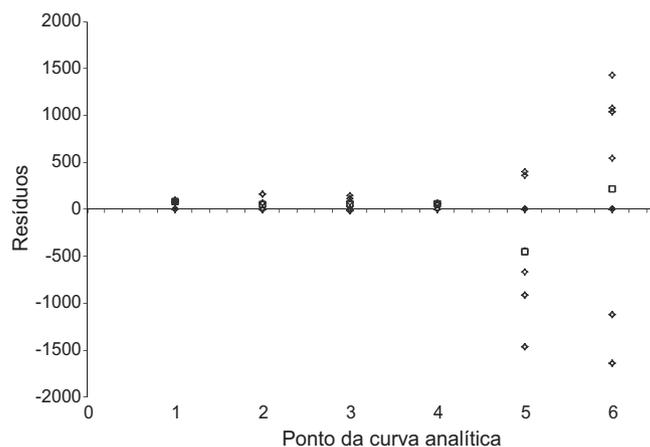
Os valores em negrito foram excluídos do conjunto de dados por serem considerados pontos anômalos segundo o teste de Grubbs para duplas e os resíduos apresentam comportamento independente, segundo o teste de Durbin-Watson.¹¹

A ANOVA da regressão indica que o ajuste linear é o mais adequado, Tabela 3.

Tabela 3. ANOVA da regressão, abordagem 1

Fonte de Variação	Soma dos Quadrados	Grau de Liberdade	Média Quadrática	
Regressão	31668348489	1	31668348489	
Resíduo	12561890,14	32	392559,067	
Falta de ajuste	1621718,0599	4	405429,515	
Erro experimental	10940172,0833	28	390720,4315	
Total	31680910380	33		
<i>O modelo é adequado</i>				
MQ_{Reg}/MQ_{Res}	80671,550	>	$F_{Crítico}$	4,149
<i>O ajuste do modelo é adequado</i>				
MQ_{FA}/MQ_{EE}	1,038	<	$F_{Crítico}$	2,714

Entretanto, quando os resíduos da regressão linear simples são avaliados visualmente, observa-se comportamento heterocedástico, pois a variância das contagens aumenta com a concentração, Figura 3. A mesma conclusão é obtida utilizando o teste de Goldfeld-Quandt, pois o P-valor encontrado é $2,72E-13 < 0,05$, rejeitando a hipótese de homocedasticidade.

**Figura 3.** Análise dos resíduos, abordagem 1

O teste de Goldfeld-Quandt⁸ consiste inicialmente em ordenar as observações de acordo com a variável explicativa. Em seguida, é escolhido um número central de observações, *k*, a ser omitida de tal forma que contenha aproximadamente 20% dos dados e que os outros dois grupos, parte 1 e parte 3, tenham quantidades semelhantes de observações. Quando o número de observações restantes é ímpar, a parte 3 fica com uma observação a mais. Esses dois grupos restantes são ajustados pelo método dos mínimos quadrados ordinários, separadamente. Dado que S_1 e S_3 sejam as somas dos quadrados dos resíduos obtidos nas regressões, parte 1 e parte 3, respectivamente, testa-se a hipótese de que a variância dos erros em ambas as regressões sejam iguais. Para isso, calcula-se a estatística *R*, Equação (14):

$$R = \frac{S_3}{S_1} \quad (14)$$

Com os seguintes graus de liberdade: ($n_3 - (m + 1)$; $n_1 - (m + 1)$), na qual: *n* = número de observações = 34; n_3 = número de observações na parte 3 = 14; n_1 = número de observações na parte 1 = 13; *m* = número de variáveis explicativas, somente a concentração = 1.

Com um nível de significância de 0,05 %, rejeita-se a hipótese nula, se S_3 for maior que S_1 .

Abordagem 1:

Com 34 observações, o valor de k é aproximado para 7, $34 \times 0,20 = 6,8$, deixando a parte 1 com 13 observações e a parte 3 com 14. A estatística R é: $R = 438/0,86 = 508$

Para 12 e 11 graus de liberdade, o valor p obtido é 2,75E-13. Portanto, rejeita-se a hipótese nula com 0,05% de confiança.

Abordagem 2:

Com 34 observações, o valor de k é aproximado para 7, deixando a parte 1 com 13 observações e a parte 3 com 14. A estatística R é: $R = 5,70/0,87 = 6,53$

Com 12 e 11 graus de liberdade, o valor p obtido é 0,0020. Portanto, rejeita-se a hipótese nula com 0,05% de confiança.

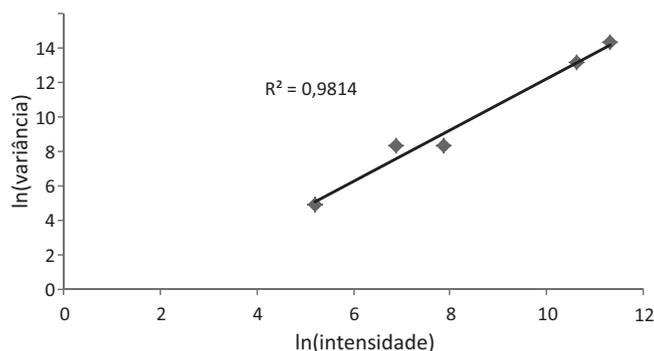
A mesma conclusão é obtida por meio dos testes estatísticos descritos na Tabela 4, na qual P-valores menores do que o nível de significância de 0,05 indicam rejeição da hipótese de homogeneidade das variâncias. Assim sendo, a regressão linear ponderada pela variância é o modelo mais adequado.

Tabela 4. Testes para homogeneidade da variância, abordagem 1

Teste estatístico*	P-valor
Cochran	<0,05
Bartlett	4,5E-15
Levene	0,005

*Os testes foram realizados sem o padrão de 50,48 mg kg⁻¹ para que as variâncias fossem obtidas a partir do mesmo número de resultados.

A Equação (9) foi utilizada para estimar a variância em função da intensidade da fluorescência e, posteriormente, gerar os pesos atribuídos a cada ponto, Figura 4. A variância do padrão de 50,48 mg kg⁻¹ não foi incluída nesse ajuste, pois, após a exclusão dos pontos anômalos, somente restaram quatro resultados para essa concentração e com um número tão pequeno o erro associado à medida da variância é muito alto.

**Figura 4.** Modelo utilizado para estimar a variância dos dados da Tabela 2

Os coeficientes para a regressão simples e ponderada pela variância e seus respectivos erros padrão (bem menores, mais confiáveis, para a regressão ponderada) estão descritos na Tabela 5.

Tabela 5. Coeficientes da regressão, abordagem 1

	Coeficiente angular		Coeficiente linear	
	Estimador	Erro padrão	Estimador	Erro padrão
Regressão simples	163,48	0,576	96,34	134,7
Regressão ponderada	162,53	0,026	176,23	6,178

A partir de então, é possível calcular os intervalos de confiança para as regressões simples, Equações (4) e (5), e ponderada, Equações (12) e (13), para concentrações em torno de 10 mg kg⁻¹, utilizando-se os valores de contagem (sinal) da Tabela 1. O material de referência certificado (MRC) utilizado tem concentração de 10,29 mg kg⁻¹.

A fim de se entender melhor e ratificar a pesquisa bibliográfica a respeito destas duas abordagens, decidiu-se realizar esta mesma análise para outra concentração, mais elevada, de 250 mg kg⁻¹. Foram realizadas seis repetições para cada MRC e não foram encontrados valores anômalos para esses conjuntos segundo o teste de Grubbs, então, os cálculos para o desvio padrão e o intervalo de confiança da previsão foram efetuados a partir de suas médias, com $m = 6$ e $n = 34$.

Estes resultados estão disponíveis na Tabela 6.

Tabela 6. Intervalos de confiança, abordagem 1

Valor do MRC	Regressão simples	Regressão ponderada
10,29 mg kg ⁻¹	(10,81 ± 3,58) mg kg ⁻¹	(10,37 ± 0,39) mg kg ⁻¹
250,00 mg kg ⁻¹	(247,2 ± 3,5) mg kg ⁻¹	(248,2 ± 4,3) mg kg ⁻¹

Conforme previsto pela teoria, o intervalo de confiança para a regressão simples é praticamente constante ao longo da faixa; entretanto, para a regressão ponderada, o intervalo de confiança aumenta consideravelmente com a concentração. O fato do intervalo de confiança aumentar com a concentração na regressão ponderada é, inclusive, muito mais coerente e realista.

Como o principal interesse deste estudo é voltado a concentrações próximas a 10 mg kg⁻¹, observa-se que ambos intervalos de confiança contêm o valor verdadeiro de 10,29 mg kg⁻¹, porém, a regressão ponderada apresenta um intervalo de confiança mais preciso para a previsão.

Abordagem 2: Curva analítica de menor extensão, de 0 a 50 mg kg⁻¹

Utilizar uma faixa de concentração mais estreita para a construção de uma curva analítica poderia trazer vantagens para previsões próximas a 10 mg kg⁻¹, em outras palavras, poderia produzir intervalos de confiança mais precisos para a previsão.

Os resultados da construção da curva analítica, com faixa reduzida, estão apresentados na Tabela 7.

Tabela 7. Curva analítica de 0 a 50 mg kg⁻¹

Concentração, mg kg ⁻¹	Contagem					
0,00	177	182	172	182	164	199
5,04	922	1085	927	987	991	919
10,29	1842	1841	1791	1962	1843	1897
15,16	2562	2606	2566	2650	2720	2687
25,18	4023	4246	4387	4168	4016	4206
50,48	8100	8408	8404	8419	8400	8306

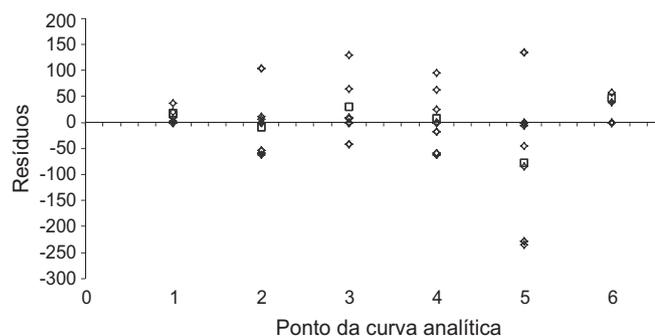
Os valores em negrito foram excluídos do conjunto de dados por serem considerados pontos anômalos segundo o teste de Grubbs para duplas e os resíduos apresentam comportamento independente, segundo o teste de Durbin-Watson.¹¹

A ANOVA da regressão indica que o ajuste linear é o mais adequado, Tabela 8.

Como já era esperado, quando os resíduos são avaliados visualmente, Figura 5, e ratificado pelo teste de Goldfeld-Quandt, observa-se comportamento heterocedástico, pois o P-valor encontrado é $0,002 < 0,05$.

Tabela 8. ANOVA da regressão, abordagem 2

Fonte de Variação	Soma dos Quadrados		Grau de Liberdade	Média Quadrática
Regressão	203563089,1		1	203563089,1
Resíduo	211933,6106		32	6622,925332
Falta de ajuste	52481,1940		4	13120,29849
Erro experimental	159452,4167		28	5694,729167
Total	203775022,7		33	
MQ_{Reg}/MQ_{Res}	30736,129	>	$F_{Crítico}$	4,149
<i>O modelo é adequado</i>				
MQ_{FA}/MQ_{EE}	2,304	<	$F_{Crítico}$	2,714
<i>O ajuste do modelo é adequado</i>				

**Figura 5.** Análise dos resíduos, abordagem 2

O diagnóstico de heterogeneidade também é obtido quando se utilizam os testes estatísticos para avaliar a variância dos dados, conforme Tabela 9, na qual P-valores menores do que o nível de significância de 0,05 indicam rejeição da hipótese de homogeneidade das variâncias. Assim sendo, a regressão linear por mínimos quadrados ponderados é mais adequada.

Tabela 9. Testes para homogeneidade da variância, abordagem 2

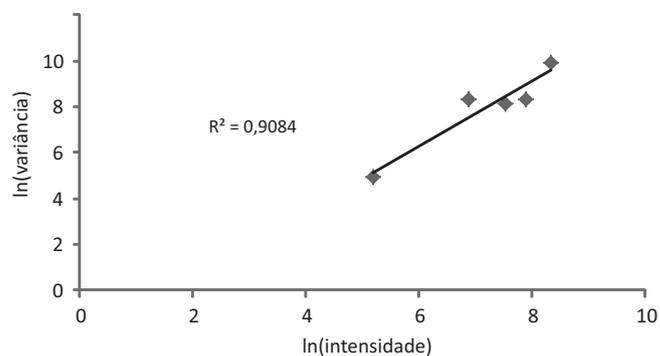
Teste*	P-valor
Cochran	<0,05
Bartlett	0,0007
Levene	0,03

*Os testes foram realizados sem o padrão de 50,48 mg kg⁻¹ para que as variâncias fossem obtidas a partir do mesmo número de resultados.

A equação (9) foi utilizada para estimar a variância em função da intensidade da fluorescência e posteriormente gerar os pesos atribuídos a cada ponto, Figura 6. A variância do padrão de 50,48 mg kg⁻¹ não foi incluída nesse ajuste pelo mesmo motivo descrito anteriormente.

Os coeficientes para a regressão simples e ponderada e seus respectivos erros padrão (bem menores, mais confiáveis, para a regressão ponderada) pela variância estão descritos na Tabela 10.

Novamente, é possível calcular os intervalos de confiança para as regressões simples, Equações (4) e (5), e ponderada, Equações (12) e (13), para concentrações em torno de 10 mg kg⁻¹, utilizando-se os valores de contagem (sinal) da Tabela 1. Além do MRC de 10,29 mg kg⁻¹, outro de 25 mg kg⁻¹ também foi avaliado pelos dois modelos. Foram realizadas seis repetições para cada MRC e não foram encontrados valores anômalos para esses conjuntos segundo o teste de Grubbs, portanto, os cálculos para o desvio padrão da previsão e o intervalo de confiança foram efetuados a partir de suas médias, com m = 6 e n = 32. Esses resultados estão disponíveis na Tabela 11.

**Figura 6.** Modelo utilizado para estimar a variância dos dados da Tabela 6**Tabela 10.** Coeficientes da regressão, abordagem 2

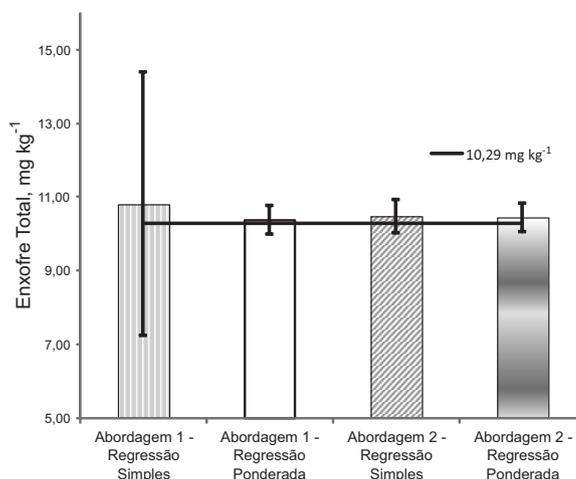
	Coeficiente angular		Coeficiente linear	
	Estimador	Erro padrão	Estimador	Erro padrão
Regressão simples	162,43	0,9265	161,76	20,20
Regressão ponderada	161,16	0,3331	178,34	7,26

Tabela 11. Intervalos de confiança

Valor do MRC	Regressão simples	Regressão ponderada
10,29 mg kg ⁻¹	(10,47 ± 0,46) mg kg ⁻¹	(10,45 ± 0,39) mg kg ⁻¹
25,18 mg kg ⁻¹	(24,70 ± 0,46) mg kg ⁻¹	(24,79 ± 0,72) mg kg ⁻¹

Novamente observamos que todos os intervalos de confiança contêm o valor verdadeiro e que o intervalo de confiança para a regressão simples é praticamente constante ao longo da faixa, enquanto que para a regressão ponderada o intervalo de confiança aumenta com a concentração. Dessa vez, porém, a regressão ponderada apresenta um intervalo de confiança para o MRC de 10,29 mg kg⁻¹, apenas ligeiramente mais preciso do que o da regressão simples. Isso ocorre porque na regressão linear simples de 0 a 50 mg kg⁻¹ não há padrões de alta concentração e grande variância que afetem negativamente a precisão do modelo como ocorreu na primeira abordagem.

Na Figura 7 é possível comparar os intervalos de confiança das duas abordagens e podemos concluir que não há vantagem em construir um modelo com uma faixa mais estreita de concentração com o intuito de melhorar as previsões para a faixa de 10 mg kg⁻¹. Pelo contrário, além de não melhorar o intervalo de confiança para as previsões, quando comparado com a regressão ponderada, perde-se a oportunidade de estimar concentrações entre 50 e 500 mg kg⁻¹, uma

**Figura 7.** Intervalos de confiança para as previsões na faixa de 10 mg kg⁻¹

vez que é fortemente recomendado que os resultados sejam obtidos apenas por interpolação.

De uma maneira geral, a comparação dos resultados dos cálculos de regressão simples e ponderada é muito instrutiva. Os efeitos do processo de ponderação são claros. O centroide ponderado é muito mais próximo da origem do gráfico que o centroide não ponderado devido ao peso maior atribuído aos pontos mais próximos à origem (pontos de menor variância). A inclinação e o intercepto da curva com ajuste ponderado são muito semelhantes aos da curva com ajuste não ponderado, por isso os resultados das concentrações são obviamente similares entre as duas abordagens. O que diferencia um modelo do outro é a precisão dos intervalos de confiança, principalmente para baixas concentrações e a principal vantagem do modelo linear ponderado é a obtenção de intervalos mais precisos para baixas concentrações. Apesar de envolver uma matemática um pouco mais complexa, a regressão linear ponderada é uma ferramenta muito útil e seu uso deve ser encorajado.

CONCLUSÕES

O presente trabalho avaliou metrologicamente a análise de enxofre total em óleo diesel rodoviário S10 por fluorescência de Raios X com dispersão monocromática de comprimento de onda.

Os resultados baseados na metodologia ASTM D7039-13 são perfeitamente satisfatórios, visto que os valores certificados estão dentro do intervalo de confiança proposto pela regressão linear ponderada.

Não se justifica a criação de outra curva analítica, mais estreita, diferente do padrão proposto.

O presente estudo mostra que dados de uma curva analítica, com comportamento heterocedástico, quando ponderados pela variância, apresentam resultados mais satisfatórios.

Como trabalho futuro, devido à carência, curta validade e alto custo dos materiais de referência certificados utilizados, pretende-se estudar a construção destas curvas analíticas com materiais de referência não certificados, i.e., com erros em ambas variáveis.¹³

AGRADECIMENTOS

Agradecemos aos técnicos do laboratório do Terminal de Guarulhos que realizaram as análises, especificamente C. C. P. Cunha.

REFERÊNCIAS

1. Herrán, A., De la Cruz, J. M., De Andrés, B.; *Comput. Chem. Eng.* **2010**, *34*, 401.
2. Sasikumar, M.; Prakash, P. R.; Patil, S. M.; Ramani, S.; *Knowledge-Based Systems* **1997**, *10*, 169.
3. Petroleum Argus In *Oil Trading Manual*, Product Markets – United States, Supplement 5, Woodhead Publishing Limited: Cambridge, 1998.
4. Braun, S., Appel, L.G., Schmal, M.; *Quim. Nova* **2004**, *27*, 472.
5. CONAMA; *Resolução CONAMA 018/1986*. Dispõe sobre a criação do Programa de Controle de Poluição do Ar por veículos Automotores – PROCONVE. Brasília – DF: CONAMA, 06 de maio de 1986.
6. Ministério do Meio Ambiente. *Primeiro Inventário Nacional de Emissões Atmosféricas por Veículos Automotores Rodoviários*. Relatório Final. Secretaria de Mudanças Climáticas - Gerência de Qualidade do Ar. Janeiro/2011.
7. Oliveira, E. C.; Monteiro, M. I. C.; Pontes, F. V. M.; Almeida, M. D.; Carneiro, M. C.; Silva, L. I. D.; Neto, A. A.; *J. AOAC Int.* **2012**, *95*, 560.
8. Goldfeld, S. M.; Quandt, R. E.; *J. Am. Stat. Assoc.* **1965**, *60*, 539.
9. Miller, J. N.; Miller J. C.; *Statistics and Chemometrics for Analytical Chemistry*, 6th ed., Pearson: Harlow, 2010.
10. Brasil, B.; da Silva, R. J. N. B.; Camões, M. F. G. F. C.; Salgueiro, P. A. S.; *Anal. Chim. Acta* **2013**, *804*, 287.
11. Vandeginste, B. G. M.; Massart, D. L.; Buydens, L. M. C.; De Jong, S.; Lewi, P. J.; Smeyers-Verbeke, J.; *Handbook of Chemometrics and Qualimetrics: Part a*, Elsevier: Amsterdam, **1998**.
12. American Society for Testing and Material – ASTM D7039: *Standard Test Method for Sulfur in Gasoline, Diesel Fuel, Jet Fuel, Kerosine, Biodiesel, Biodiesel Blends, and Gasoline-Ethanol Blends by Monochromatic Wavelength Dispersive X-ray Fluorescence Spectrometry*, **2013**.
13. Oliveira, E. C., Aguiar, P. F.; *Quim. Nova* **2013**, *36*, 885.