

7 Resultados e Discussão

7.1. Detecção de Amostras Aberrantes

Observando-se o conjunto de espectros obtidos das misturas de solventes (figura 13) e gasolinas (figura 14), verifica-se a presença de algumas amostras com comportamento diferenciado das demais, principalmente nas regiões do espectro relativas às bandas de água (1600cm^{-1} e $3650\text{-}3750\text{cm}^{-1}$). Foi constatado, posteriormente, que a anormalidade apresentada por esses espectros era devido à umidade presente na célula túnel, que havia sido lavada com água antes da realização dessas análises.

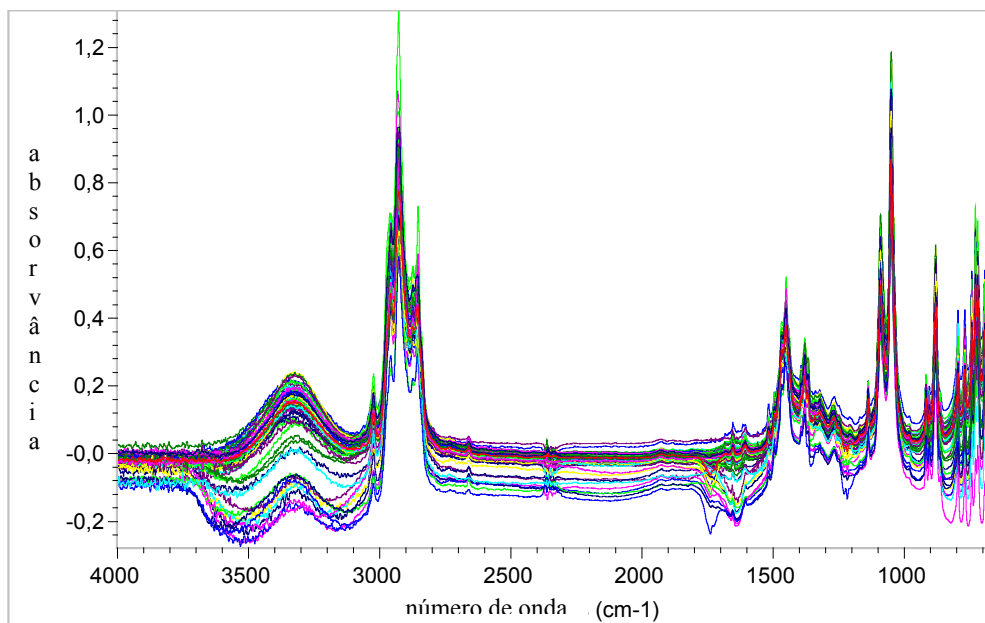


Figura 13- Espectros no infravermelho de 67 misturas de solventes, na região de 4000 a 650cm^{-1} , para verificação da presença de amostras anômalas.

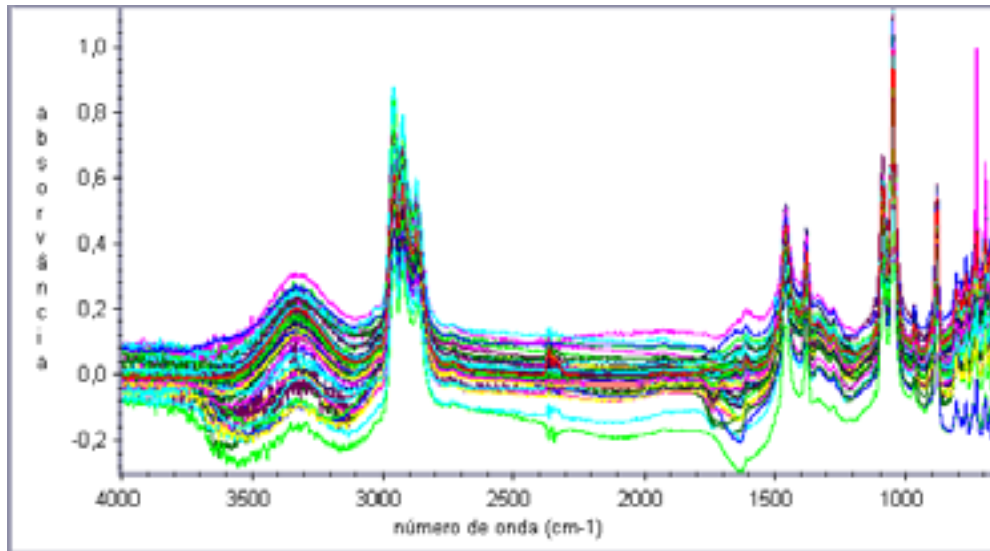


Figura 14- Espectros no infravermelho de 126 gasolinas, na região de 4000 a 650 cm^{-1} , para verificação da presença de amostras anômalas.

Uma amostra considerada aberrante, também chamada “*outlier*”, apresenta-se tão diferente das demais que pode não ser bem descrita pelo modelo ou influenciá-lo demais. Assim sendo, pelo menos um dos componentes principais do modelo pode concentrar-se apenas em descrever como essa amostra é diferente das outras, mesmo que isso seja irrelevante para modelar as características mais importantes presentes nas demais amostras. Usando a análise dos componentes principais, as amostras aberrantes detectadas visualmente podem ser confirmadas através do gráfico dos “*scores*” e dos valores dos resíduos e “*leverages*”, em geral combinados num gráfico de influência (46).

- O gráfico dos “*scores*” mostra o perfil das amostras em relação a um, dois ou três componentes principais. Com esse gráfico é fácil identificar uma amostra que se situe longe das demais, indicando a possibilidade de ser considerada aberrante.
- Os resíduos são uma medida de o quanto as amostras ajustam-se ao modelo. Uma amostra com resíduo alto é pobremente descrita pelo modelo, o qual descreve bem as demais amostras. Esse tipo de amostra é considerado como um elemento estranho em relação

ao conjunto de amostras bem descritas pelo modelo, podendo ser considerada uma amostra aberrante.

- “*Leverage*” é uma medida de quanto um ponto de um conjunto de dados é extremo, comparado com a maioria. Em PCA ou PLS, “*leverage*” pode ser interpretado como a distância entre o ponto projetado e o centro do modelo. Pontos com alto “*leverage*” têm uma alta influência sobre o modelo e podem ou não ser amostras aberrantes. Uma amostra com alto “*leverage*” e alto resíduo é considerada um caso mais crítico e pode ser detectada usando-se o gráfico de influência.

A análise dos componentes principais, com os dados centrados e padronizados, confirma que aquelas amostras podem ser classificadas como aberrantes. Os gráficos dos “*scores*” e de influência estão apresentados, respectivamente, nas figuras 15 e 16 para as misturas, e nas figuras 17 e 18 para as gasolinas. Nesses gráficos, os pontos envolvidos com um círculo referem-se às amostras que tiveram seus espectros coletados usando-se a célula túnel previamente lavada com água e que, por isso, foram descartados. Também se observa, através desses gráficos, que outras amostras poderiam ser consideradas aberrantes, pelos mesmos critérios. No entanto, como não foi verificada nenhuma anormalidade em seus espectros, elas foram mantidas.

As misturas de solventes cujos espectros foram descartados foram então preparadas e analisadas novamente e o conjunto de espectros válidos estão apresentados na figura 19. No caso das gasolinas, o conjunto de espectros válidos constitui-se das 110 amostras restantes após a retirada das aberrantes.

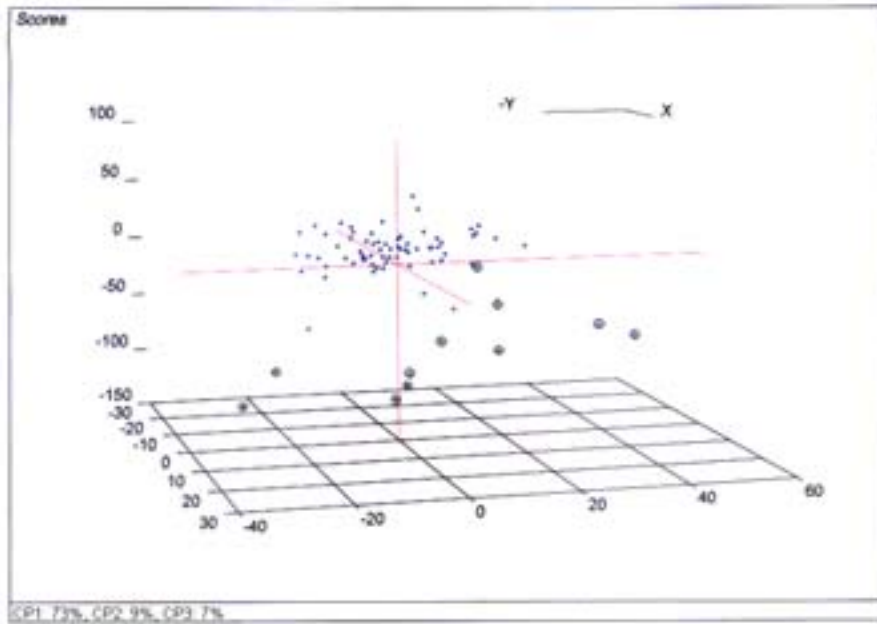


Figura 15- Gráfico dos “scores” nos três primeiros componentes principais dos espectros das misturas apresentados na figura 13.

Os pontos marcados com um círculo (⊙) representam os espectros que foram descartados.

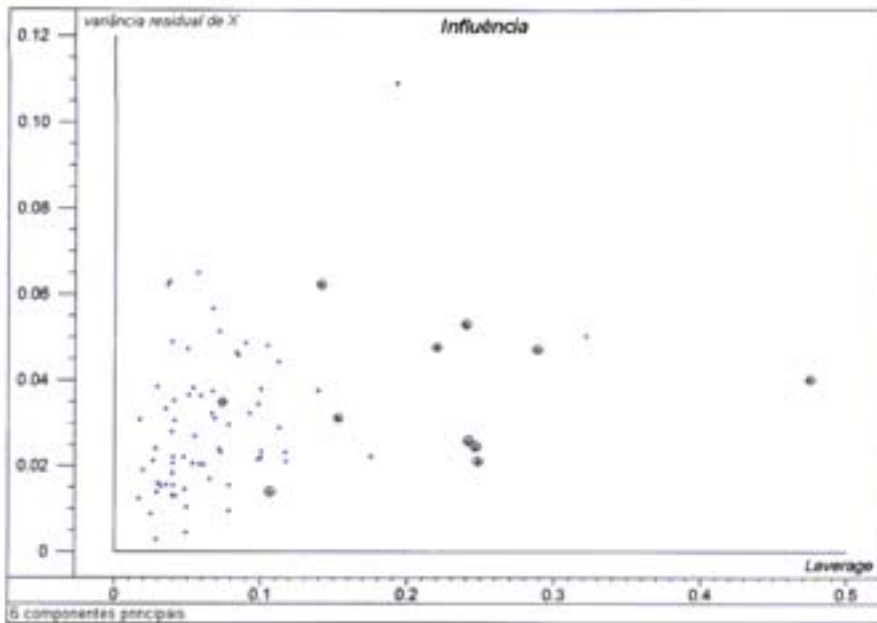


Figura 16- Gráfico de influência dos espectros das misturas apresentados na figura 13.

Os pontos marcados com um círculo (⊙) representam os espectros que foram descartados.

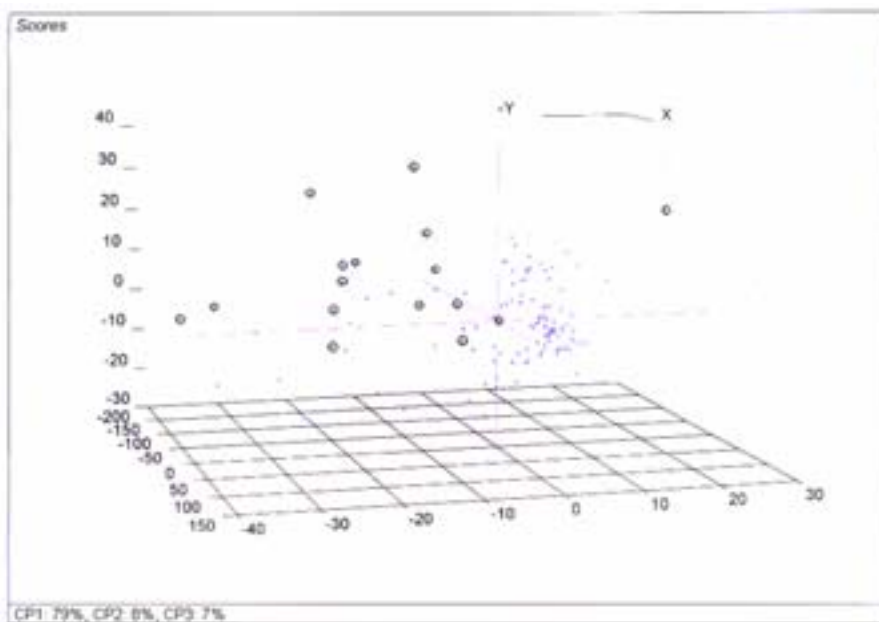


Figura 17- Gráfico do “scores” nos três primeiros componentes dos espectros das gasolinas apresentados na figura 14.

Os pontos marcados com um círculo (⊙) representam os espectros que foram descartados.

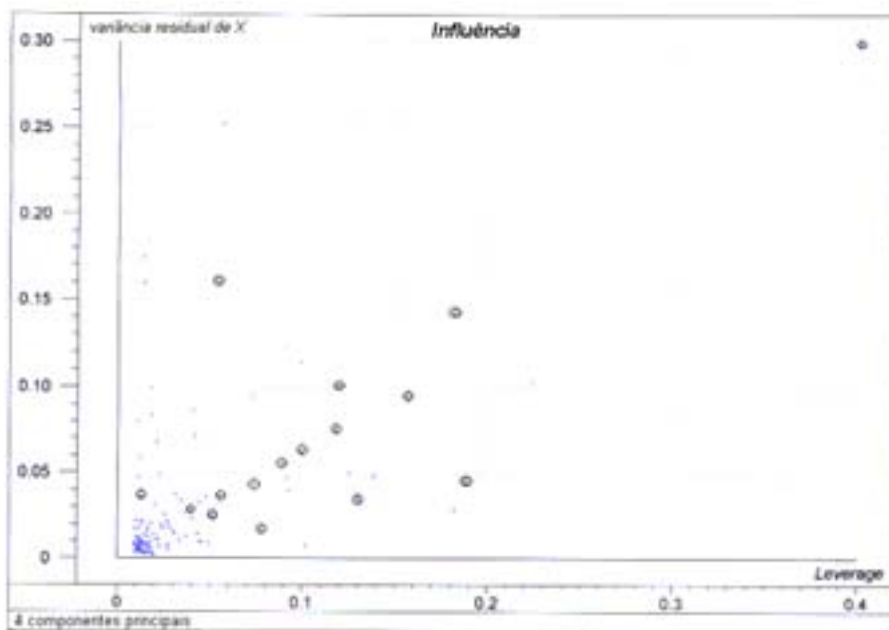


Figura 18- Gráfico de influência dos espectros das gasolinas apresentados na figura 14.

Os pontos marcados com um círculo (⊙) representam os espectros que foram descartados.

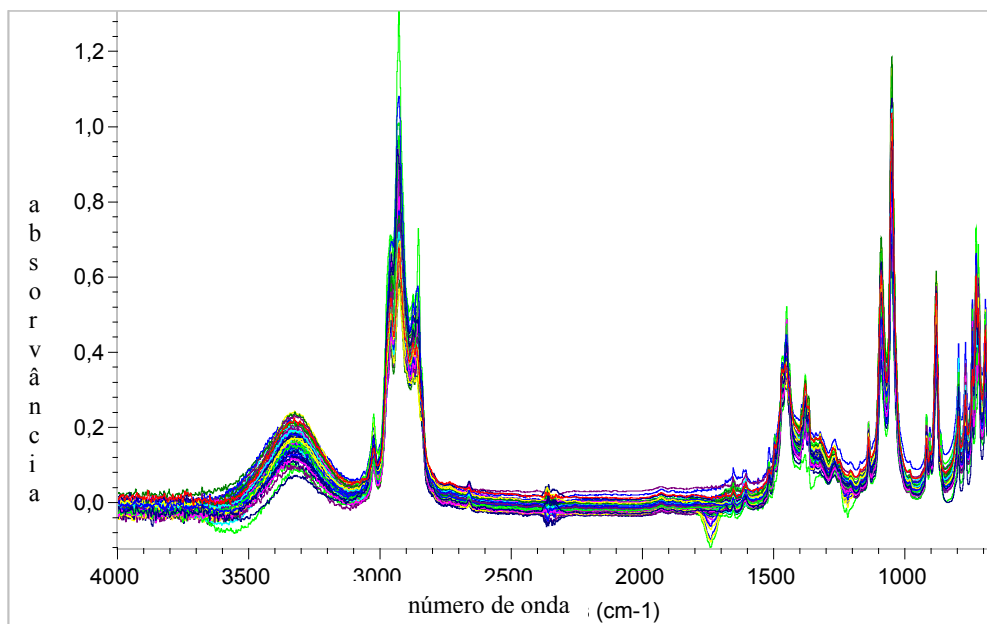


Figura 19- Espectros no infravermelho das 66 misturas de solventes, na região de 4000 a 650 cm^{-1} , utilizados na calibração, validação e teste do modelo PLS.

7.2.

Seleção de Variáveis

A utilização de todas as variáveis X (comprimentos de onda) disponíveis muitas vezes pode não ser o mais indicado, causando, por exemplo, tempo computacional longo, tendência a superajuste (*overfitting*), colinearidades, alta complexidade nos modelos, etc., Por causa disso, três critérios de seleção de variáveis foram utilizados:

1. utilização de todas as principais regiões do espectro, 650-1650 e 2600-3100 cm^{-1} , chamada aqui de “espectro inteiro”, descartando-se as regiões onde visualmente só há ruído e a banda referente ao estiramento O-H do etanol (3100-3600 cm^{-1}), por ser altamente afetada pela presença de água;
2. utilização de um subconjunto de variáveis diretamente relacionadas com a propriedade a ser predita, selecionado de acordo com conhecimentos químicos.

3. utilização de um subconjunto de variáveis selecionadas de acordo com dois gráficos fornecidos pelo programa estatístico utilizado: coeficientes de regressão e cargas das variáveis X (*X-loadings weights*).

Os coeficientes de regressão são usados principalmente para verificar a importância das diferentes variáveis X em prever Y. Valores absolutos altos de coeficientes de regressão indicam variáveis de grande importância (significância) e valores pequenos (perto de 0) indicam variáveis sem importância. O valor do coeficiente indica o aumento médio em Y quando a variável X correspondente é aumentada de uma unidade, mantidas todas as outras variáveis constantes (46).

O gráfico das cargas das variáveis X, para um determinado componente principal, contra o número da variável também é usado como um dos meios de descobrir quais variáveis X são importantes para prever Y. Esse gráfico mostra a relação entre o componente especificado e as diferentes variáveis X. Se uma variável tiver um grande valor de carga, positivo ou negativo, isto significa que a variável é importante para o componente. Variáveis com valores de carga altos nos primeiros componentes são aquelas que explicam a maior variação em Y (46).

Para ilustrar como são utilizados aqueles gráficos, a figura 20 apresenta os coeficientes de regressão e a figura 21 as cargas das variáveis X do modelo PLS construído a partir dos espectros inteiros das misturas de solventes, para predição do teor de benzeno, com os dados centrados em torno da média e padronizados. Observa-se que os valores mais altos de coeficientes de regressão e as maiores cargas nos componentes principais 1 e 2 pertencem às variáveis na região de 650 a 750 cm^{-1} .

Para o modelo construído com os espectros inteiros, obteve-se RMSEP igual a 0,107794 para as amostras de validação, utilizando-se 11 componentes principais. Selecionando a região do espectro de 650 a 750 cm^{-1} , obteve-se RMSEP igual a 0,054932 para as mesmas amostras, utilizando-se 4 componentes principais. Portanto, a seleção de variáveis proporcionou uma queda de 51% no erro médio da predição, além de diminuir o número de componentes principais utilizados.

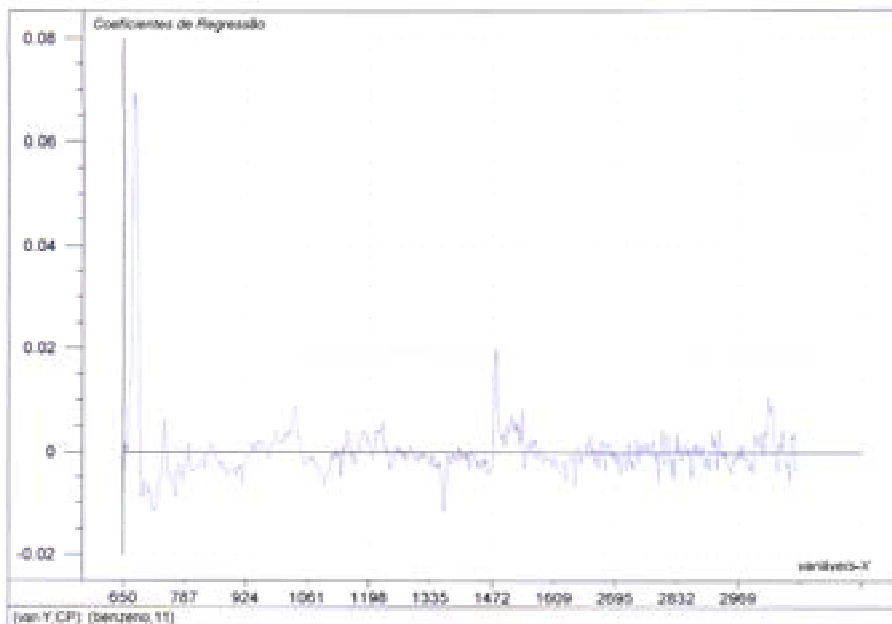


Figura 20- Coeficientes de regressão do modelo PLS construído com os espectros inteiros (650-1650 e 2600-3100 cm^{-1}) das misturas de solventes para a predição de benzeno.

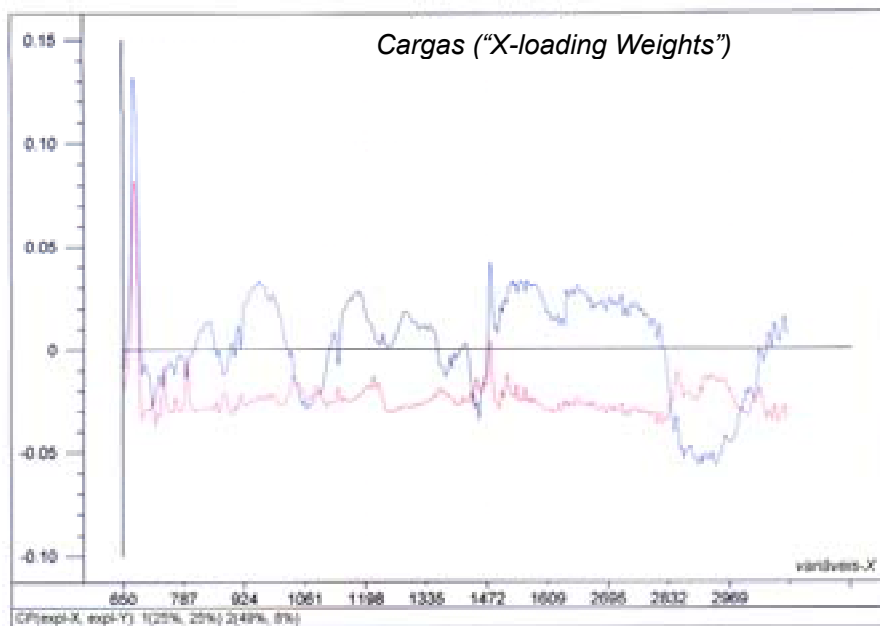


Figura 21- Cargas das variáveis X nos componentes principais 1 e 2 do modelo PLS construído com os espectros inteiros (650-1650 e 2600-3100 cm^{-1}) das misturas de solventes para a predição do teor de benzeno.

PC1: (—) e PC2: (—)

Para todos os modelos PLS construídos, tanto para misturas quanto para gasolinas, foi necessário algum tipo de seleção de variáveis. Essa seleção foi feita, na maioria dos casos, combinando-se subconjuntos de variáveis selecionados pelos critérios 2 e 3 mencionados acima. As variáveis utilizadas nos modelos finais para as propriedades das misturas e das gasolinas estão apresentadas nas tabelas 9 e 10, respectivamente.

7.3. Os Modelos de Regressão PLS

Para todas as propriedades, foram obtidos diferentes modelos PLS, utilizando-se:

1. espectro inteiro, sem pré-tratamentos;
2. espectro inteiro, com cada um dos pré-tratamentos já mencionados;
3. subconjuntos de variáveis, sem pré-tratamentos;
4. subconjuntos de variáveis, com cada um dos pré-tratamentos.

A habilidade preditiva de cada modelo foi avaliada utilizando-se o RMSEP, que representa uma medida da diferença média entre os valores fornecidos pelo modelo e os valores de referência, nos estágios de validação e predição. O RMSEP pode ser interpretado como o erro médio da predição, expresso na mesma unidade da propriedade que está sendo medida, e é bastante sensível à presença de amostras com erros elevados (17).

Como já foi dito, durante o processo de validação, seja com um conjunto de amostras independentes da calibração ou por validação cruzada, um passo crítico é estabelecer o número ótimo de componentes principais que serão utilizados no modelo. Para exemplificar, a figura 22 apresenta o gráfico do RMSEP *versus* o número de componentes principais utilizado, do modelo construído para a predição do teor de aromáticos nas misturas. Verifica-se que a utilização de 9 componentes principais fornece o menor valor de RMSEP. Os componentes principais utilizados nos modelos finais para as propriedades das misturas e das gasolinas estão citados nas tabelas 9 e 10, respectivamente.

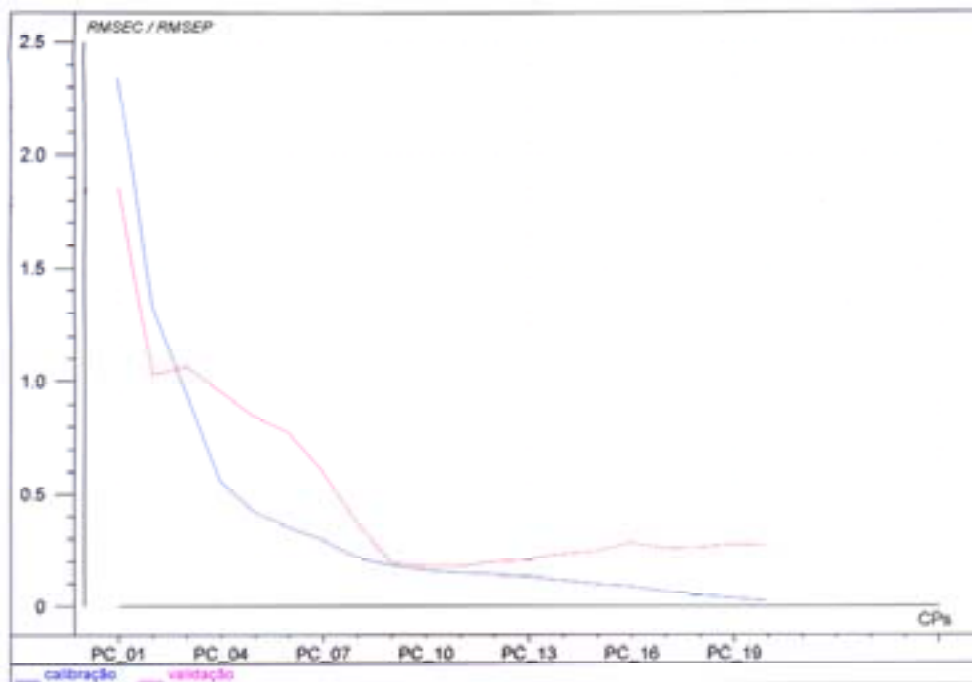


Figura 22- Gráfico RMSEP *versus* o número de componentes principais, do modelo PLS construído para a determinação do teor de aromáticos nas misturas.

Com relação aos pré-tratamentos testados, os modelos que forneceram resultados com menores valores de RMSEP, tanto para as misturas quanto para as gasolinas, foram aqueles construídos a partir dos espectros centrados em torno da média e padronizados. As tabelas 9 e 10 apresentam os modelos de regressão selecionados para a predição das propriedades das misturas e das gasolinas, respectivamente, mostrando as regiões do espectro (variáveis X) e o número de componentes principais utilizados, e os valores de RMSEP obtidos nas etapas de calibração, validação e teste. Uma discussão mais detalhada sobre os resultados das misturas e das gasolinas encontra-se nos itens subsequentes.

Por agora, fazendo uma comparação entre as regiões do espectro utilizadas para prever as mesmas propriedades nas misturas e nas gasolinas, observa-se que há bastante coincidência entre elas, confirmando o que já era esperado. No caso do etanol, por exemplo, foram utilizadas as regiões 800-1550 e 2850-3050 cm^{-1} para as misturas e as regiões 650-1500 e 2700-3050 cm^{-1} para as gasolinas, no caso do etanol determinado pelo método da proveta. A figura 23 apresenta o espectro do etanol puro, com aquelas regiões destacadas, as quais

correspondem a praticamente toda a região que foi chamada de espectro inteiro (650-1650 e 2600-3100 cm^{-1}).

Além disso, observa-se também que, para quase todas as propriedades em comum entre misturas e gasolinas, o número de componentes principais utilizados nos modelos foi menor para as misturas do que para as gasolinas, provavelmente devido ao fato de estas últimas serem amostras de maior complexidade, com um número elevado de componentes.

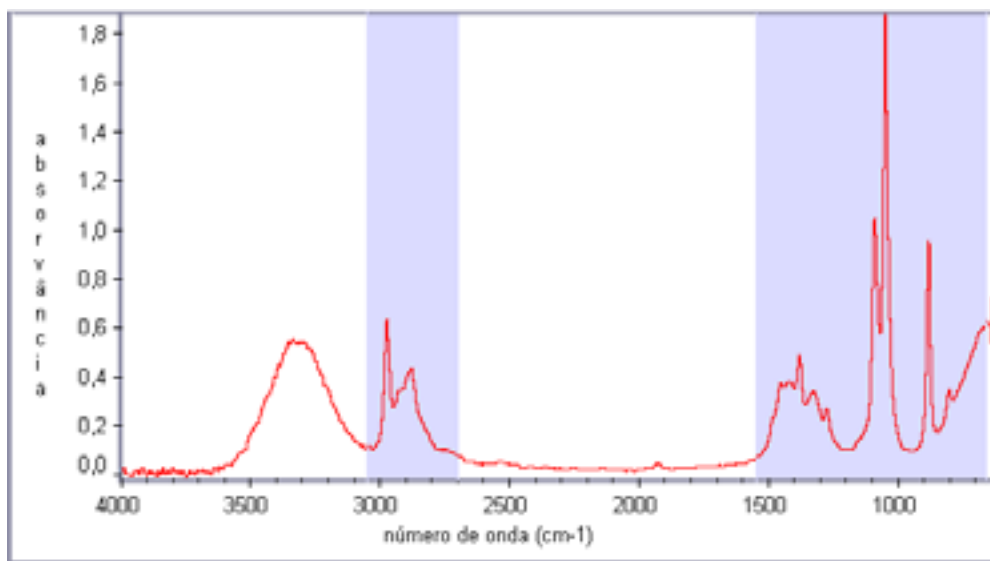


Figura 23- Espectro do etanol com as regiões 650-1500 e 2700-3050 cm^{-1} destacadas.

Tabela 9- Modelos PLS selecionados para predição das propriedades das misturas de solventes.

Propriedade	Região do Espectro (cm ⁻¹)	Nº de CP's	RMSEP			Média*
			cal	val	teste	teste
aromáticos	650-1500	10	0,207	0,384	0,139	15,9
benzeno	650-800	7	0,119	0,156	0,079	0,99
ciclo-hexano	1000-1100	10	0,299	0,656	0,341	7,29
	650-900					
	1020-1280					
ciclo-hexeno	1370-1590	10	0,232	0,519	0,745	22,8
	650-1550					
	2700-3100					
etanol	800-1550	8	0,273	0,483	0,182	23,9
isooctano	2850-3050	14	0,089	0,250	0,180	9,89
	650-1500					
n-heptano	2800-3100	10	0,372	0,632	0,256	20,2
	700-760					
saturados	800-1550	7	0,494	0,862	0,471	37,4
	700-1600					
tolueno	650-780	7	0,141	0,214	0,143	4,69
	1000-1100	7	0,152	0,992	0,095	3,49
	1350-1600					
650-930						
o-xileno	980-1140	7	0,149	0,206	0,132	3,50
	650-800					
m-xileno	2840-2950	10	0,078	0,142	0,054	3,21
	750-1240					
p-xileno	1350-1500	10	0,171	0,269	0,283	10,2
	680-800					
xilenos	1090-1650					

*Valor médio da propriedade no conjunto teste.

Tabela 10- Modelos PLS selecionados para predição das propriedades das gasolinas.

Propriedade (IROX)	Região do Espectro (cm ⁻¹)	N° de CP's	RMSEP			Média*
			cal	val	teste	teste
benzeno	650-800	8	0,020	0,027	0,030	0,30
	970-1045					
	1470-1510					
aromáticos	700-1650	8	0,523	0,629	0,916	14,2
saturados	800-1700	14	1,029	2,102	3,246	40,7
	2600-2850					
	2900-3000					
olefinas	700-1400	12	1,084	1,682	2,358	21,3
	1500-1700					
etanol	700-1500	16	0,207	0,713	1,345	23,4
	2730-3050					
MON	650-1240	11	0,299	0,486	0,553	82,5
	1390-1650					
	2800-3100					
RON	650-950	8	0,366	0,449	0,746	95,5
	1000-1500					
	300-3100					

Propriedade método-padrão	Região do Espectro (cm ⁻¹)	N° de CP's	RMSEP			Média*
			cal	val	teste	teste
PEI	700-1700	3	1,893	0,218		
PE 10%	700-1050	11	0,628	1,020	0,886	56,8
	1340-1510					
	1620-1670					
PE 50%	700-1200	18	0,092	0,377	0,391	73,1
	1340-1700					
PE 90%	740-890	16	0,860	2,208	2,499	173,4
	930-1280					
	1340-1460					
	1500-1650					
PEF	760-840	15	1,505	3,984	3,024	212,1
	930-1280					
	1340-1460					
	1500-1600					
densidade	650-1245	12	0,0009	0,0016	0,0023	0,7556
	1350-1700					
etanol	650-1500	12	0,305	0,500	0,727	23,9
	2700-3050					

*Valor médio da propriedade no conjunto teste.

7.3.1. Misturas de Solventes

O presente estudo realizado com as misturas de solventes, que também podem ser chamadas de “gasolinas sintéticas”, serviram de base para o trabalho com as gasolinas comerciais. Qualquer resultado negativo encontrado com as misturas, indicaria mau resultado também com as gasolinas.

Para avaliar a qualidade dos resultados, além do RMSEP, foram também calculados os erros relativos (equação 15) para o conjunto teste de cada modelo. Esses erros para as misturas estão apresentados na tabela 11, onde os valores em vermelho destacam as amostras que apresentaram erros relativos superiores a 10%.

$$\text{Erro\%} = [(y_{\text{predito}} - y_{\text{referência}}) / y_{\text{referência}}] \times 100 \quad \text{equação 15}$$

Como mencionado no item 5.1.1, o teor de ciclo-hexano foi previsto tanto através de uma calibração própria quanto por diferença (100% menos os teores dos outros componentes). Os melhores resultados foram obtidos a partir da calibração, apresentada na tabela 9, e o cálculo dos erros relativos para o conjunto teste está baseado nos resultados desta calibração. Vale mencionar, em termos de comparação, que o RMSEP encontrado usando o cálculo da diferença foi 0,841 contra 0,341 para os resultados da calibração.

Tabela 11- Erros relativos do conjunto teste das misturas de solventes.

	aromáticos	benzeno	c-hexano	c-hexeno	etanol	isooctano	n-heptano
T01	-0,63	3,0	3,8	-1,6	0,54	-1,3	2,3
T02	-0,63	4,9	3,6	-1,6	0,23	-0,57	-0,83
T03	-0,20	6,8	0,85	-0,31	-0,78	0	1,5
T04	-0,47	7,3	-0,69	-0,17	0,26	0,3	-0,38
T05	-1,3	2,1	2,4	2,0	-0,53	-2,6	1,7
T06	-0,19	2,1	1,3	-0,92	0,034	-0,53	1,4
T07	0,23	3,1	-0,20	1,0	0,24	-0,65	0,52
T08	0,66		-2,3	1,0	-1,1	-1,5	0,74
T09	0,71	5,7	-3,8	0,51	0,57	0,10	0,043
T10	0,34	5,0	-16,5	0,51	1,1	-1,6	0,17
T11	-0,02		5,0	-2,7	-1,2	-2,5	-3,0
T12	0,78		4,8	0,35	-0,66	-0,53	-1,0
T13	-2,2	8,8	-7,7	-1,0	0	2,9	0,58
T14	0,99	6,4	0,77	2,5	-1,6	-4,2	1,1
T15	-0,66	5,2	28,2	11,4	-0,022	-2,7	-0,078
T16	-1,2		48,4	-0,30	0,57	-1,9	1,8

	saturados	tolueno	o-xileno	m-xileno	p-xileno	xilenos
T01	0,32	3,1	-0,44	0,33	-2,4	-1,1
T02	0,75	-1,4	1,4	-0,57	-1,5	-0,82
T03	0,56	3,1	0,62	4,5	0,030	1,2
T04	0,55	-0,31	1,0	0	-0,66	1,0
T05	0,18	-6,9	-3,1	-2,6	0,12	3,9
T06	1,6	-0,16	1,5	1,4	-0,57	-0,64
T07	-0,55	-0,85	4,8	9,9	-2,9	0,64
T08	-1,3	1,9	-3,1	1,0	0	1,4
T09	-1,1	-0,18	1,7	-0,65	1,0	0,83
T10	-2,4	2,4	-2,8	-3,1	3,3	0,13
T11	0,08	2,1	0,088	-1,2	-0,54	1,0
T12	-2,7	4,7	-1,0	-6,3	-1,0	0,25
T13	1,6	-4,8	-3,3	1,9	-2,8	-6,0
T14	-0,38	-0,14	8,8	0,81	-3,3	-2,4
T15	-1,4	-9,6	0,71	7,0	1,5	5,6
T16	-1,5	1,7	2,3	0,59	-1,0	-0,31

Observa-se na tabela acima que os erros relativos encontrados para o conjunto teste foram quase todos menores que 10%, com exceção da amostra T15 para o ciclo-hexeno (11,36%) e das amostras T10, T15 e T16 para o ciclo-hexano (-16,5%, 28,2% e 48,4%, respectivamente). Os valores não mencionados referem-se a amostras que não contém benzeno, não sendo possível, portanto, o cálculo do erro relativo.

Os altos valores de erro para o ciclo-hexano mencionados acima podem ser explicados pelo fato de os teores de ciclo-hexano naquelas amostras serem muito baixos (T10: real = 1,93%, estimado = 1,62%; T15: real = 1,09%, estimado = 1,40%; T16: real = 0,86%, estimado = 1,28%).

As figuras 24 a 33 apresentam os gráficos de correlação entre os valores de referência e os valores preditos pelos modelos FTIR-PLS para os teores de benzeno, aromáticos, saturados, olefinas e etanol, as propriedades realmente importantes para as gasolinas.

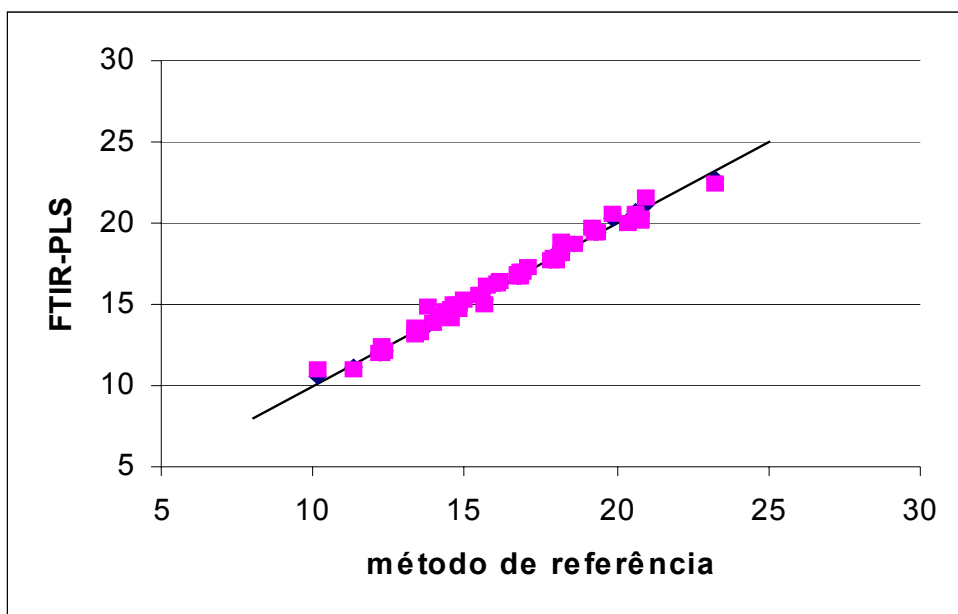


Figura 24- Teor de aromáticos (%m) nas misturas de solventes.

◆ calibração ($R^2 = 0,9949$); ■ validação ($R^2 = 0,9826$); — teórico.

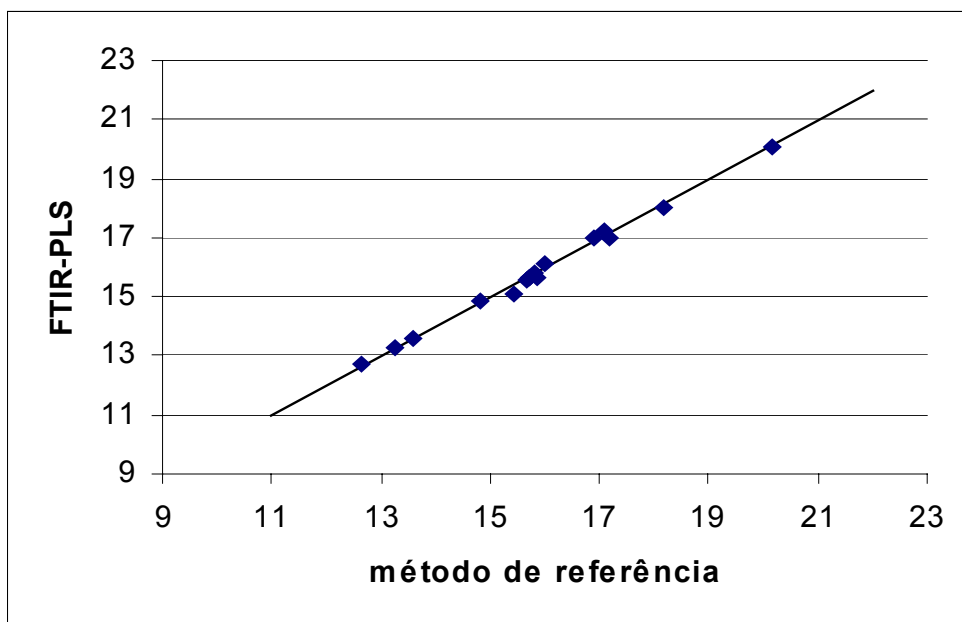


Figura 25- Teor de aromáticos (%m) nas misturas de solventes.

◆ conjunto teste ($R^2 = 0,9935$); — teórico.

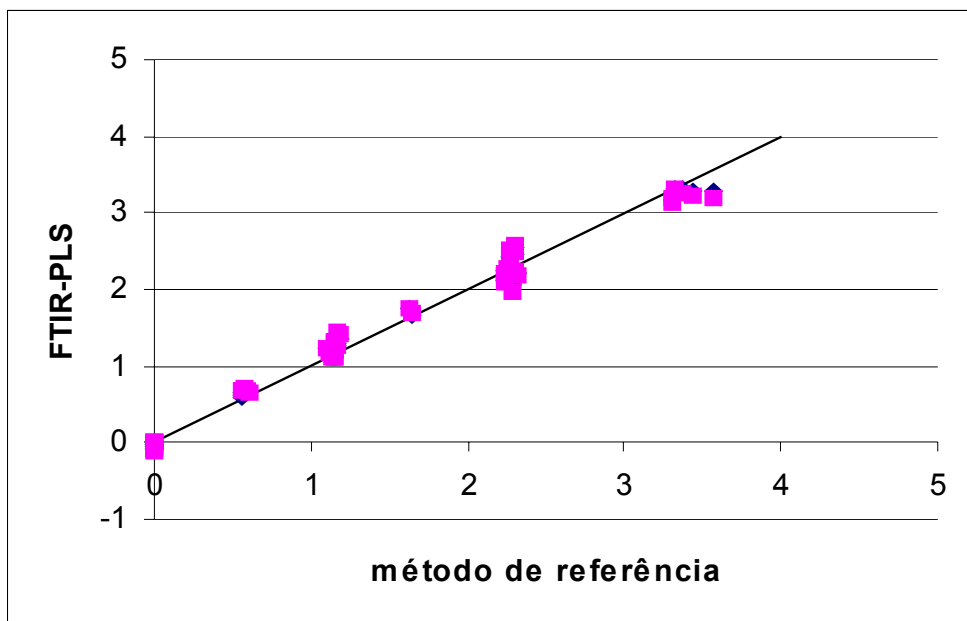


Figura 26- Teor de benzeno (%m) nas misturas de solventes.

◆ calibração ($R^2 = 0,9905$); ■ validação ($R^2 = 0,9854$); — teórico.

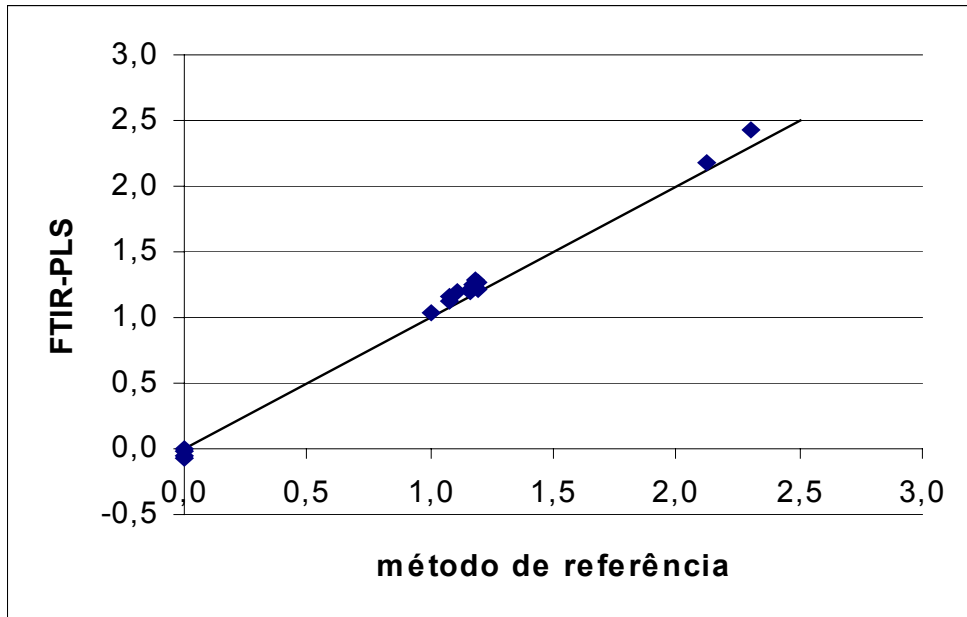


Figura 27- Teor de benzeno (%m) nas misturas de solventes.

◆ conjunto teste ($R^2 = 0,9982$); — teórico.

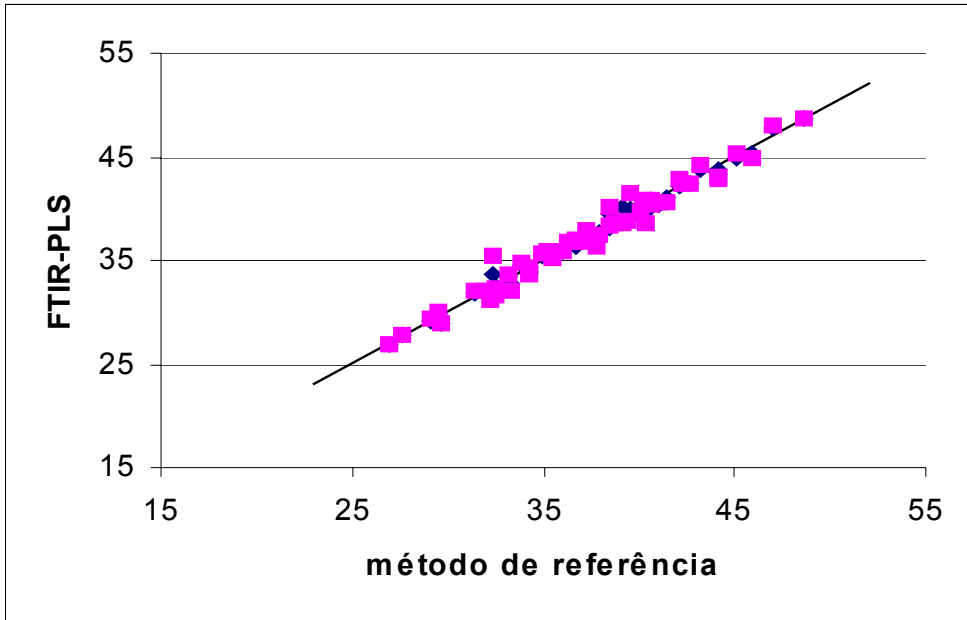


Figura 28- Teor de saturados (%m) nas misturas de solventes.

◆ calibração ($R^2 = 0,9906$); ■ validação ($R^2 = 0,9713$); — teórico.

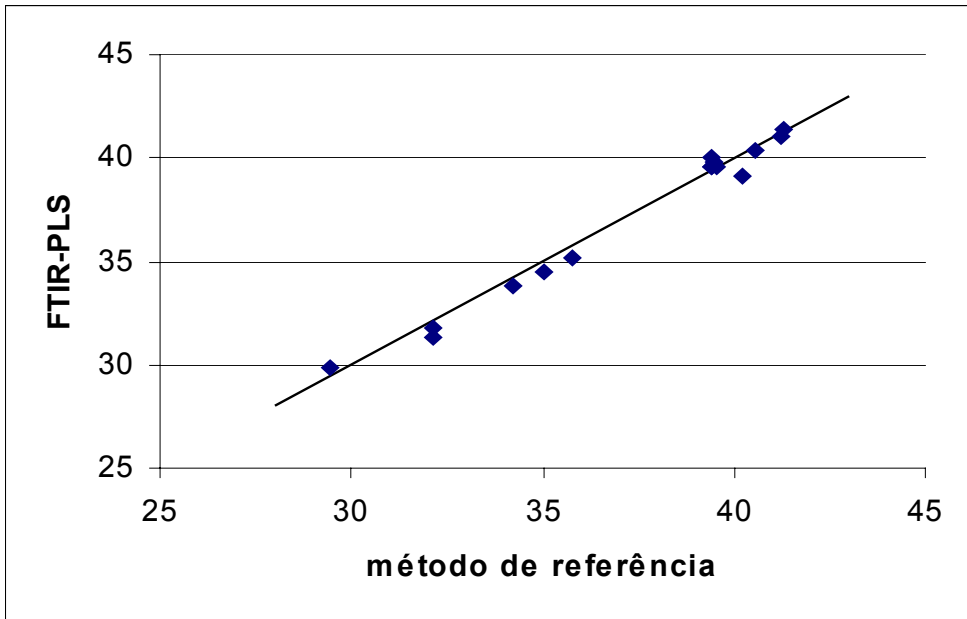


Figura 29- Teor de saturados (%m) nas misturas de solventes.

◆ conjunto teste ($R^2 = 0,9857$); — teórico.

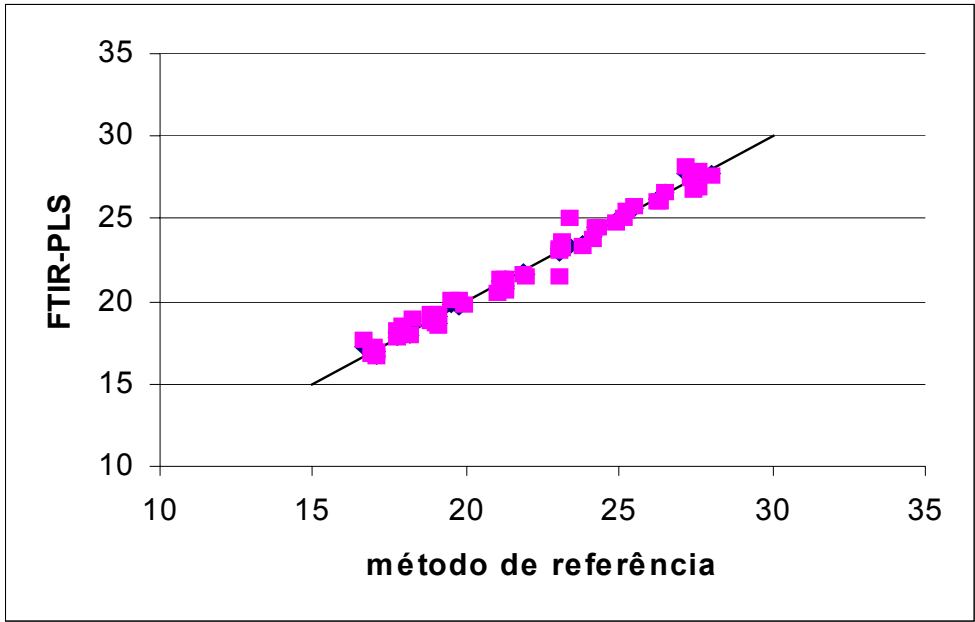


Figura 30- Teor de ciclo-hexeno (%m) nas misturas de solventes.

◆ calibração ($R^2 = 0,9955$); ■ validação ($R^2 = 0,9773$); — teórico.

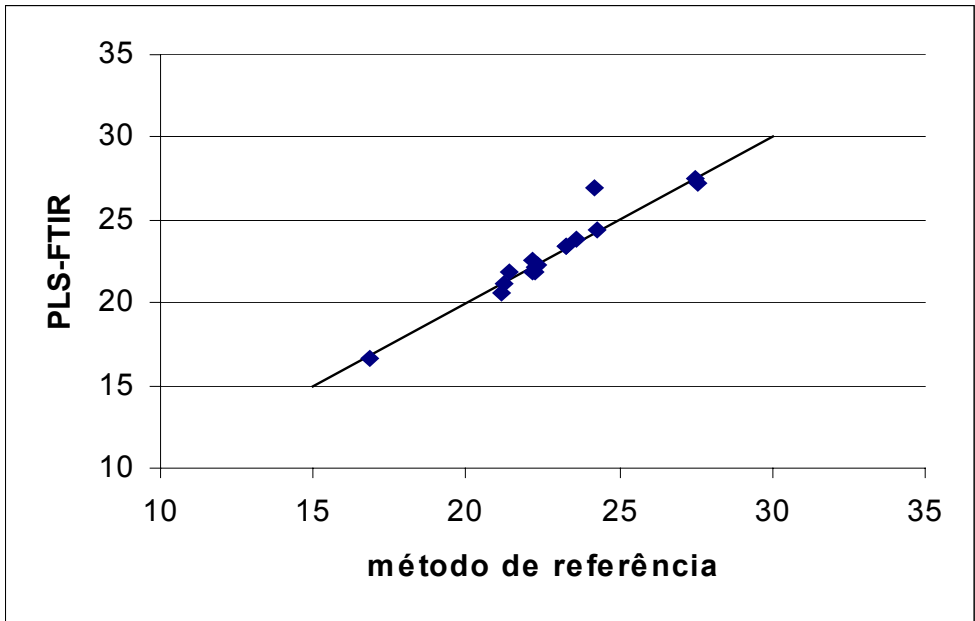


Figura 31- Teor de ciclo-hexeno (%m) nas misturas de solventes.

◆ conjunto teste ($R^2 = 0,9271$); — teórico.

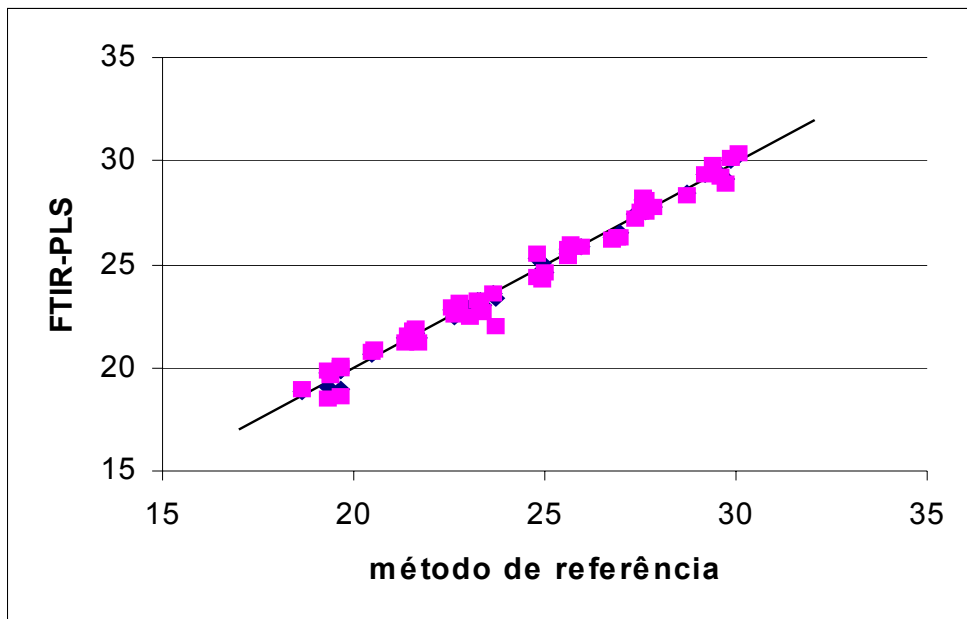


Figura 32- Teor de etanol (%m) nas misturas de solventes.

◆ calibração ($R^2 = 0,9933$); ■ validação ($R^2 = 0,9795$); — teórico.

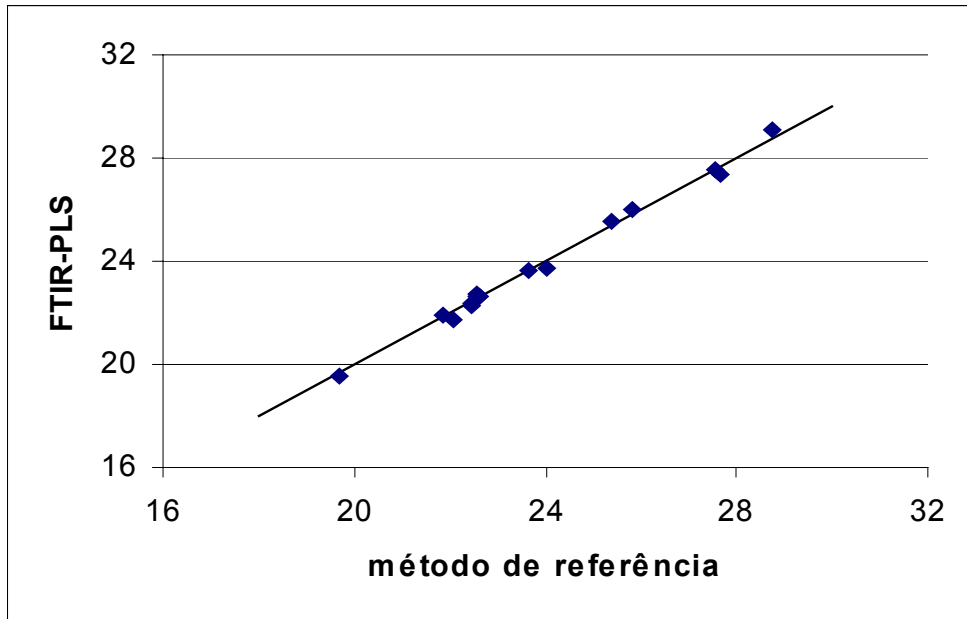


Figura 33- Teor de etanol (%m) nas misturas de solventes.

◆ conjunto teste ($R^2 = 0,9952$); — teórico.

7.3.2. Gasolinas

Assim como para as misturas, foram calculados os erros relativos de cada propriedade das gasolinas para o conjunto teste. Esses erros estão apresentados nas tabelas 14 e 15, onde os valores em vermelho destacam as amostras que apresentaram erro relativo superior a 10%. A tabela 14 apresenta os modelos que tiveram como valores de referência aqueles fornecidos pelo equipamento IROX e a tabela 15 os modelos que tiveram como valores de referência aqueles fornecidos por métodos ASTM ou ABNT.

Tabela 12- Erros relativos do conjunto teste das gasolinas para os modelos cujo método de referência foi o IROX.

Erro%	benzeno	etanol	aromáticos	olefinas	saturados	MON	RON
gT01	-22,4	8,55	-12,9	-24,0	-9,85	0,0802	1,58
gT02	0,571	-5,48	-2,23	-11,4	-3,12	-1,14	-1,24
gT03	14,7	-3,42	-3,02	-3,13	0,0506	0,395	0,397
gT04	9,52	-3,53	2,16	-27,1	-18,1	-0,378	-1,94
gT05	-6,21	-5,04	3,05	10,7	6,97	-1,46	-1,05
gT06	17,3	6,94	6,60	-2,05	-5,90	-0,099	0,781
gT07	7,04	-0,873	4,23	0,0251	0,264	-0,466	0,217
gT08	-15,0	-2,63	8,13	15,1	4,77	0,364	0,124
gT09	6,50	-7,69	-6,99	6,14	7,54	0,443	0,772
gT10	-1,79	-9,79	-8,79	9,29	12,8	-0,0856	-0,192
gT11	5,14	-2,49	-2,91	12,6	6,73	-1,89	-0,948
gT12	6,67	1,69	-5,02	4,18	2,01	-0,765	-0,204
gT13	21,0	4,26	9,99	-17,2	-12,3	-0,406	0,295
gT14	6,49	-1,42	-1,66	-3,22	-1,69	-0,0802	-0,316
gT15	-12,6	-5,73	0,429	25,2	21,5	0,904	-0,286
gT16	4,24	-3,22	10,0	-1,98	-1,28	0,372	-0,0617
gT17	9,03	5,00	11,0	6,83	-3,22	1,03	0,355
gT18	-6,88	-8,99	-6,94	8,63	15,7	0,750	0,0935
gT19	4,67	5,24	-6,21	6,60	4,15	0,540	0,941
gT20	3,00	6,95	-10,5	8,42	4,99	0,0803	1,10
gT21	4,14	2,67	-0,541	11,5	4,08	0,201	0,587
gT22	-11,1	2,97	8,03	8,73	4,33	0,680	0,348
gT23	2,42	3,28	8,63	2,97	-1,04	0,499	0,0157
gT24	3,75	5,94	-1,80	9,23	0,52	0,144	-0,326
gT25	-19,7	4,02	-4,30	-2,08	1,02	-0,270	-0,490
gT26	5,60	-0,688	-11,9	3,61	6,48	0,880	1,73
gT27	7,67	0,991	0,667	-0,94	0,37	0,414	0,119
gT28	-2,19	-6,83	4,35	0,81	1,38	0,425	0,0938
gT29	-8,18	-8,40	-3,77	0,47	6,88	-0,865	-0,873
gT30	-5,88	-8,18	2,33	-0,68	3,51	-0,0642	-0,220
gT31	-8,38	-10,5	-5,47	-6,00	5,64	-0,385	-1,22
gT32	14,1	-2,70	-2,18	-1,45	2,94	-0,485	-0,161
gT33	-0,714	-7,82	11,9	16,4	11,1	0,313	0,511

Tabela 13- Erros relativos do conjunto teste das gasolinas para os modelos cujos métodos de referência foram ASTM ou ABNT.

Erro%	etanol	densidade	PE10%	PE50%	PE90%	PEF
gT01	0,848	-0,265	-0,617	-0,0135	1,47	0,662
gT02	-0,200	0,263	1,71	-0,187	1,29	2,01
gT03	1,98	-0,132	-0,558	-0,470	-0,0892	1,40
gT04	-2,07	0,265	0,574	0,585	-1,37	-0,043
gT05	1,14	0,132	-0,513	-0,413	-1,69	-0,008
gT06	1,64	0,000	-1,39	-0,604	1,34	1,91
gT07	0,552	0,132	-2,70	-0,322	-1,11	-0,563
gT08	0,939	0,265	-0,0643	-0,00544	-0,422	-0,915
gT09	-1,26	-0,266	-1,16	0,676	-0,224	0,780
gT10	-1,80	-0,528	0,0194	-0,191	-1,06	-7,61
gT11	3,88	0,000	-3,11	-0,297	-1,11	-2,69
gT12	1,69	0,264	-1,65	0,130	0,222	0,805
gT13	-1,15	0,132	0,696	0,453	1,96	-0,113
gT14	2,66	0,264	2,00	0,741	1,16	1,97
gT15	-3,41	0,000	-2,35	-0,683	0,466	0,278
gT16	-6,73	-0,132	-2,37	0,296	-0,0155	0,249
gT17	2,09	0,132	-0,542	-0,593	-0,204	-0,849
gT18	-2,70	0,000	-1,70	0,143	-0,277	-3,17
gT19	-3,77	0,264	0,410	0,618	0,688	0,793
gT20	-2,42	0,000	-0,639	0,110	-0,271	-0,408
gT21	-4,30	-0,266	-1,20	-0,659	-0,576	-1,03
gT22	5,15	0,264	2,47	0,844	1,69	0,741
gT23	-4,29	0,132	-0,774	-0,252	0,448	3,33
gT24	0,0792	0,266	-1,08	-0,172	-0,247	0,0420
gT25	-1,76	0,000	1,46	0,195	-0,828	-1,81
gT26	-4,73	-1,32	2,57	0,191	2,89	1,36
gT27	-3,64	0,000	0,267	0,559	2,77	1,25
gT28	-2,73	0,000	-1,46	-0,0150	0,706	-0,926
gT29	-3,06	0,132	-1,88	-0,535	-0,761	-0,904
gT30	-4,60	0,000	-0,733	0,097	0,690	0,362
gT31	-1,48	0,133	-0,416	-1,82	-2,47	-0,424
gT32	1,08	-0,132	-1,19	-0,141	-0,220	-0,486
gT33	-3,75	0,813	2,91	2,99	5,12	3,03

Observa-se nas tabelas acima que os modelos construídos a partir de valores determinados pelo IROX forneceram diversos valores de erro relativo maiores que 10% para as amostras do conjunto teste, principalmente para os teores de benzeno, aromáticos, olefinas e saturados. Já para MON e RON, todas as amostras apresentaram erros relativos menores que 2%, e para o teor de álcool apenas uma amostra apresentou erro relativo maior que 10% (-10,5%).

Em termos do RMSEP, os melhores resultados obtidos foram para MON ($82,5 \pm 0,5$) e RON ($95,5 \pm 0,7$). No entanto, apesar de os erros relativos também terem sido considerados satisfatórios para essas propriedades, os coeficientes de correlação entre os valores estimados pelo método FTIR-PLS e os estimados pelo IROX (referência), para as amostras do conjunto teste, apresentaram valores muito baixos, 0,2343 para MON e 0,1229 para RON. Isso se deve, principalmente, a pouca variação nos valores dessas propriedades, mascarando as respostas, já que em todas as gasolinas analisadas, MON apresentou valores entre 81,4 e 83,8 e RON entre 93,0 e 97,0.

O mesmo aconteceu para o teor de álcool, cujos valores em todas as amostras variou entre 21,9 e 25,7, com apenas uma amostra com 30% de etanol que incluída por comportar-se como amostra aberrante. O coeficiente de correlação entre os teores de álcool estimados pelo método FTIR-PLS e os estimados pelo IROX, para as amostras do conjunto teste, foi 0,0402. As figuras 34 a 47 apresentam os gráficos de correlação entre os valores preditos pelos modelos FTIR-PLS e os valores fornecidos pelo IROX para as propriedades das gasolinas determinadas por esse equipamento.

Já para os modelos construídos a partir de valores determinados por métodos ASTM ou ABNT, valores de erro relativo menores que 10% foram obtidos para todas as propriedades. A única exceção foi o ponto de ebulição inicial (PEI), para o qual sequer foi possível a construção de um modelo de predição a partir das 77 amostras de gasolinas do conjunto calibração/validação.

Para testar se um conjunto maior de amostras melhoraria os resultados de PEI, todas as 110 amostras de gasolinas disponíveis foram utilizadas na etapa de calibração. No entanto, ainda assim o modelo obtido foi considerado sem sensibilidade. Esse fato pode ser explicado pela perda dos componentes mais voláteis presentes na gasolina no intervalo de tempo entre a análise pelo método de referência e a coleta dos espectros no equipamento FTIR.

Apesar de os erros relativos para aquelas propriedades não serem considerados altos, calculando-se os coeficientes de correlação entre os valores estimados pelo método FTIR-PLS e os de referência, para as amostras do

conjunto teste, valores menores que 0,70 foram obtidos. As figuras 48 a 60 apresentam os gráficos de correlação entre os valores de referência e os valores preditos pelos modelos FTIR-PLS para aquelas propriedades das gasolinas.

No caso do teor de álcool determinado pelo método da proveta, por exemplo, os valores de referência eram 23%, 24% ou 25%, com apenas uma amostra com 26% (g047) e outra com 31% (g094) de álcool, as quais não foram incluídas na construção do modelo por comportarem-se como amostras anômalas. Portanto, com apenas três níveis de variação em uma faixa tão estreita, o coeficiente de correlação encontrado (0,4473) também não foi considerado satisfatório.

O mesmo aconteceu para o ponto de ebulição a 50% da destilação, cujos valores de referência variaram entre 71,7 e 74,3 para todas as amostras e o coeficiente de correlação encontrado para o conjunto de amostras teste foi 0,3861.

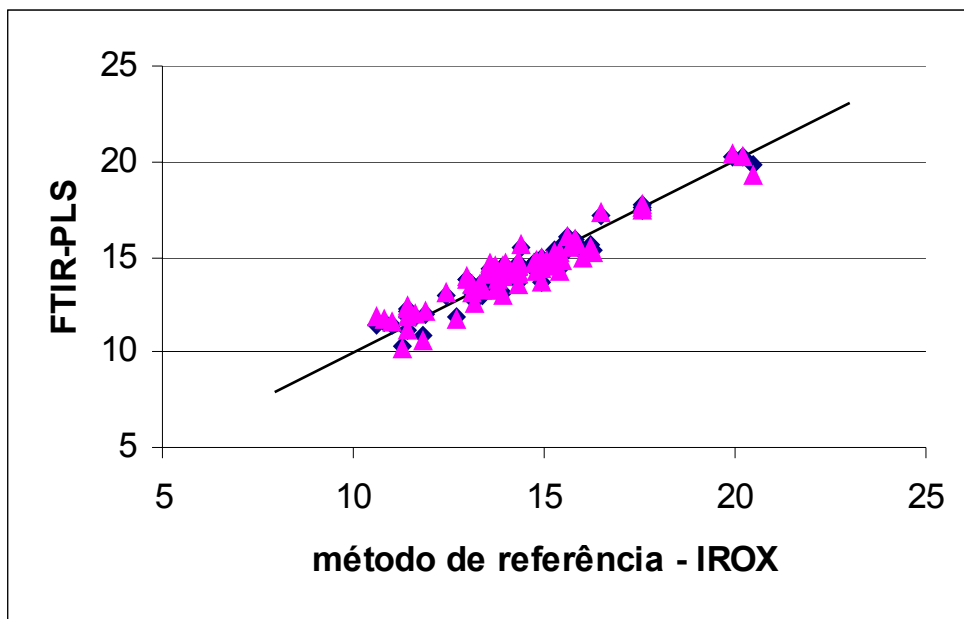


Figura 34- Teor de aromáticos (%v) nas gasolinas.

◊ calibração ($R^2 = 0,9283$); ▲ validação ($R^2 = 0,8961$); — teórico.

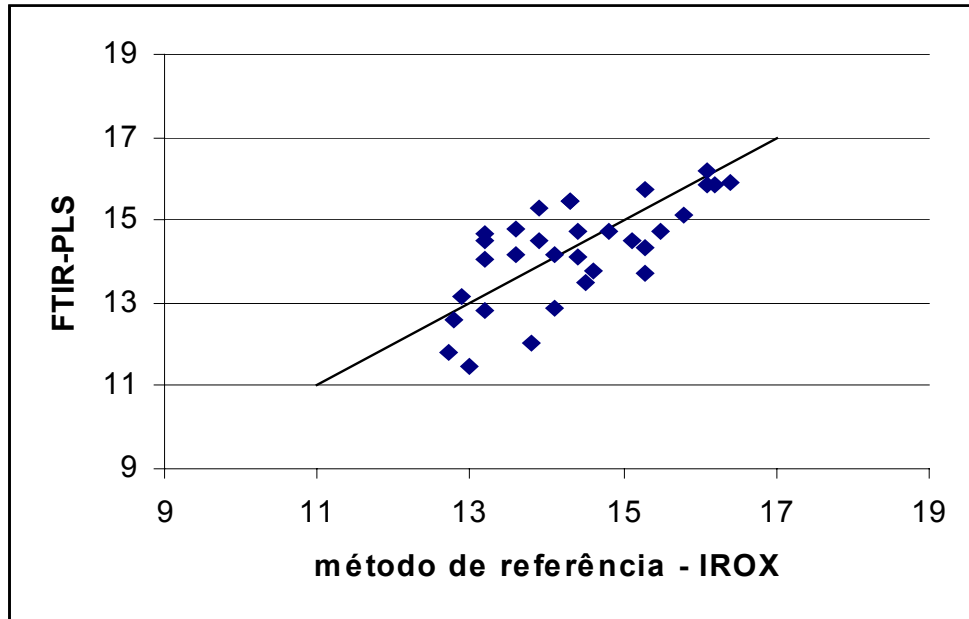


Figura 35- Teor de aromáticos (%v) nas gasolinas.

◊ conjunto teste ($R^2 = 0,4793$); — teórico.

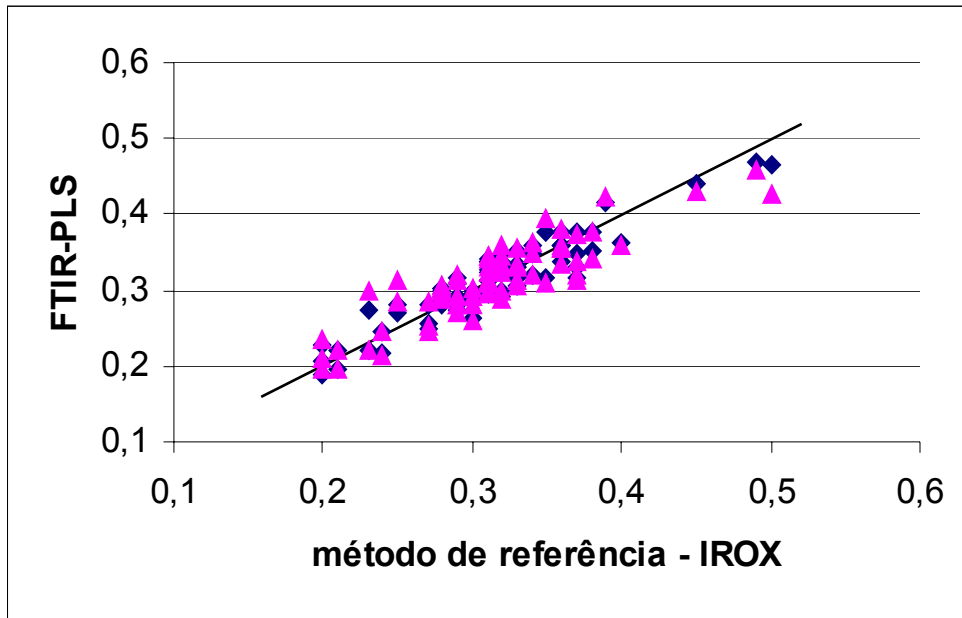


Figura 36- Teor de benzeno (%v) nas gasolinas.

◊ calibração ($R^2 = 0,8745$); ◀ validação ($R^2 = 0,7817$); — teórico.

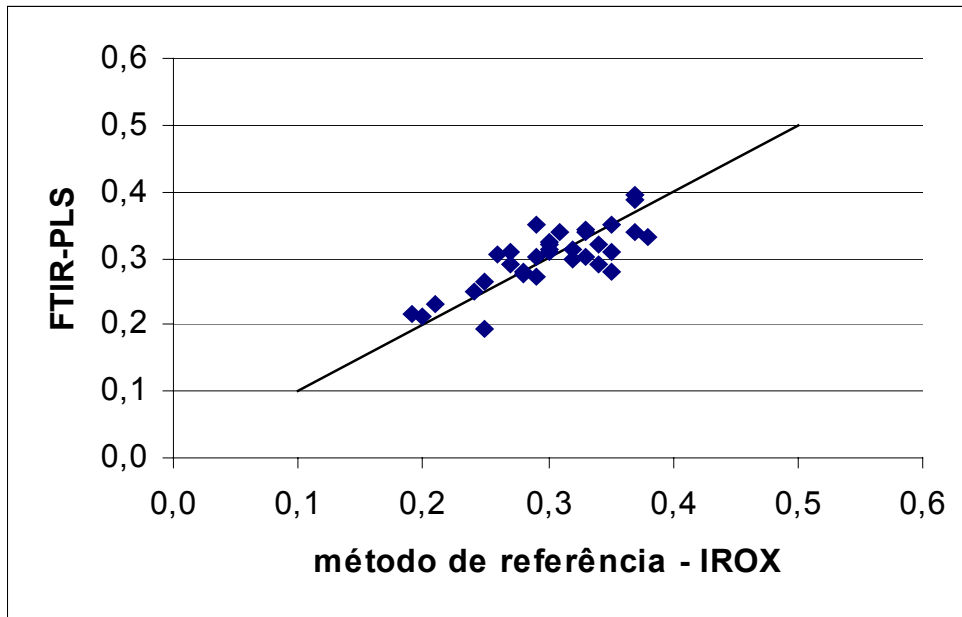


Figura 37- Teor de benzeno (%v) nas gasolinas.

◊ conjunto teste ($R^2 = 0,6386$); — teórico.

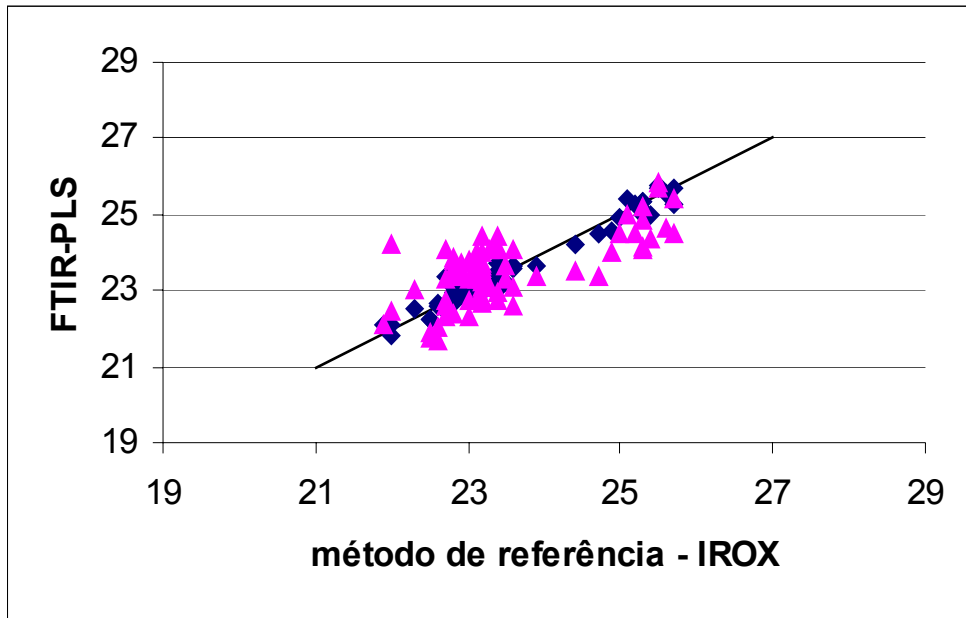


Figura 38- Teor de álcool (%v) nas gasolinas.

◊ calibração ($R^2 = 0,9562$); ◀ validação ($R^2 = 0,5116$); — teórico.

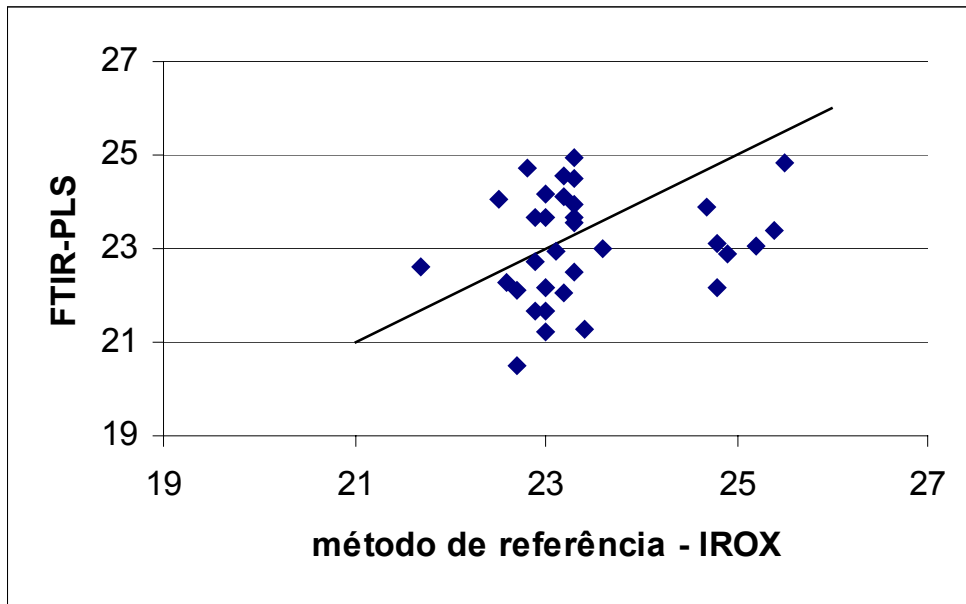


Figura 39- Teor de álcool (%v) nas gasolinas.

◊ conjunto teste ($R^2 = 0,0402$); — teórico.

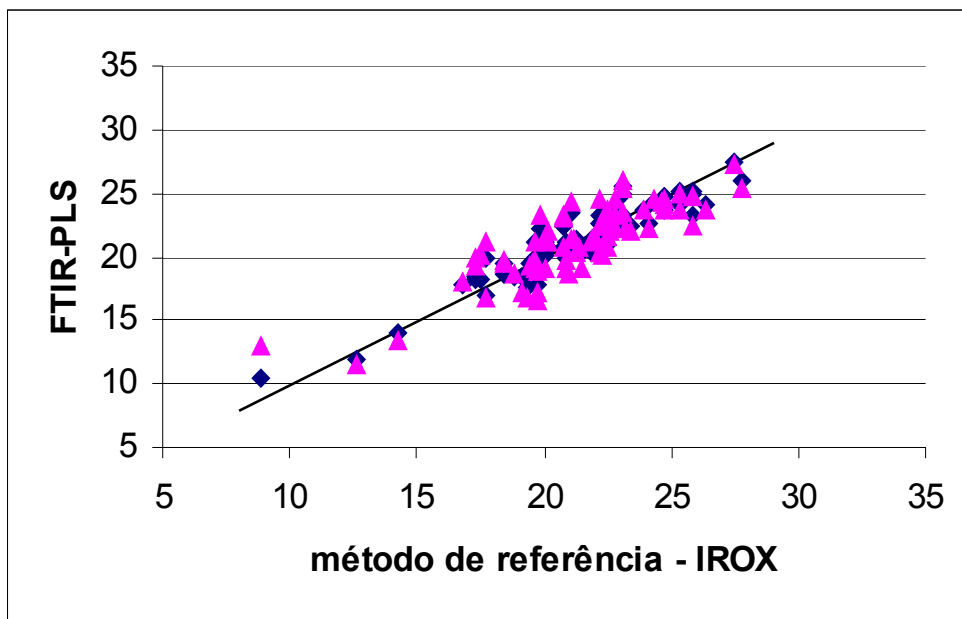


Figura 40- Teor de olefinas (%v) nas gasolinas.

◊ calibração ($R^2 = 0,8783$); ▲ validação ($R^2 = 0,7173$); — teórico.

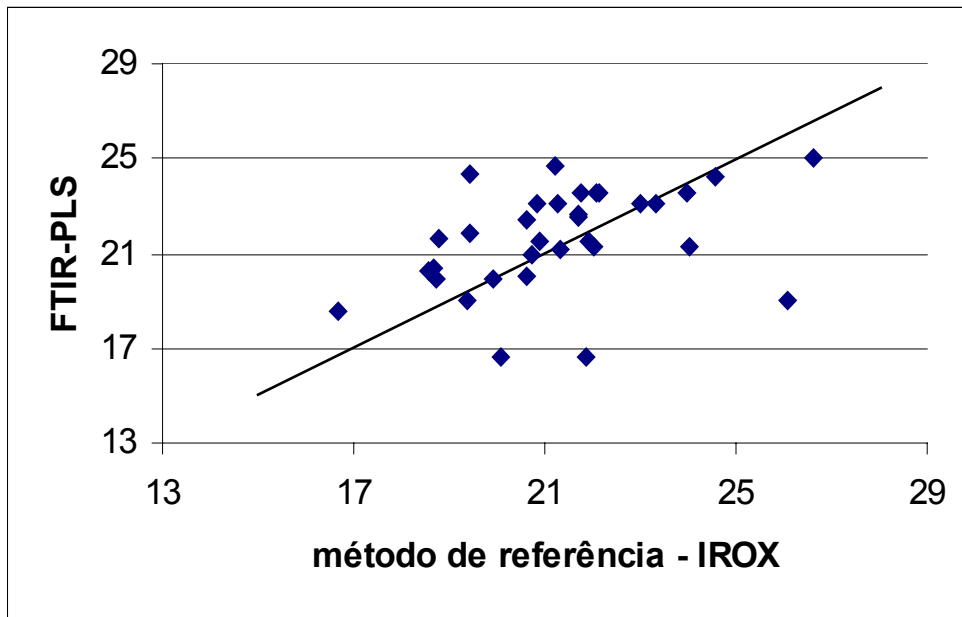


Figura 41- Teor de olefinas (%v) nas gasolinas.

◊ conjunto teste ($R^2 = 0,1550$); — teórico.

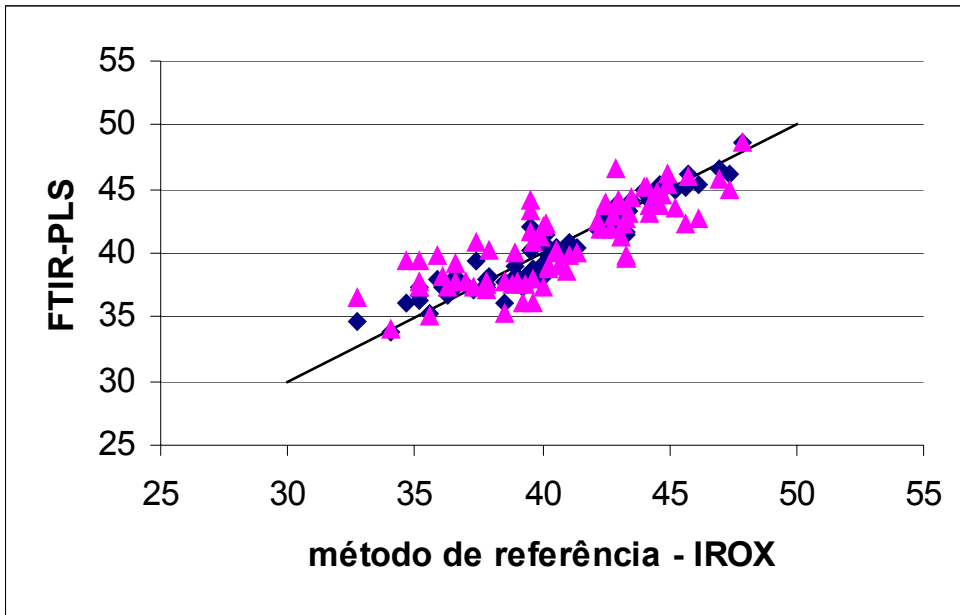


Figura 42- Teor de saturados (%v) nas gasolinas.

◊ calibração ($R^2 = 0,9109$); ◀ validação ($R^2 = 0,6422$); — teórico.

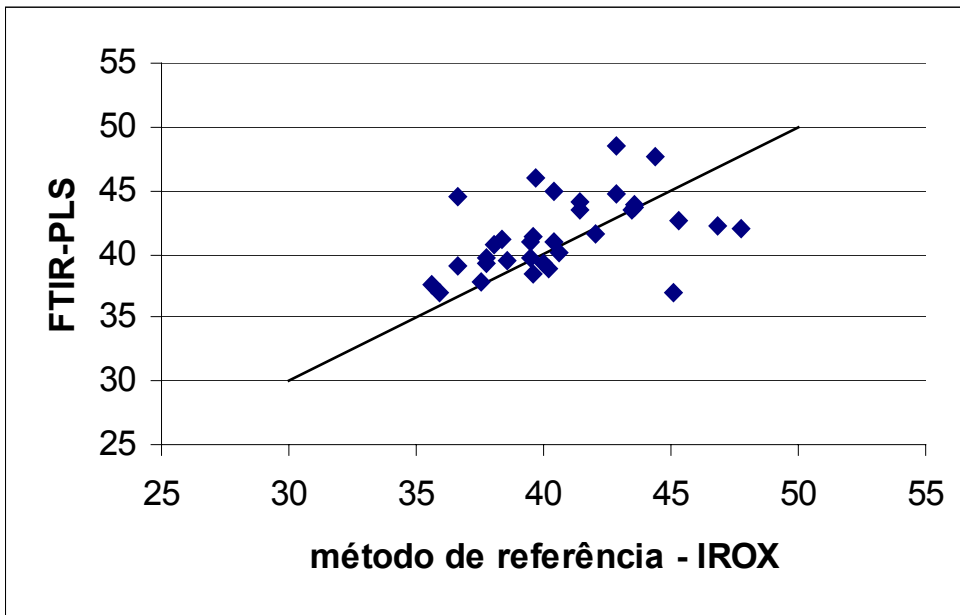


Figura 43- Teor de saturados (%v) nas gasolinas.

◊ conjunto teste $R^2 = 0,2052$; — teórico.

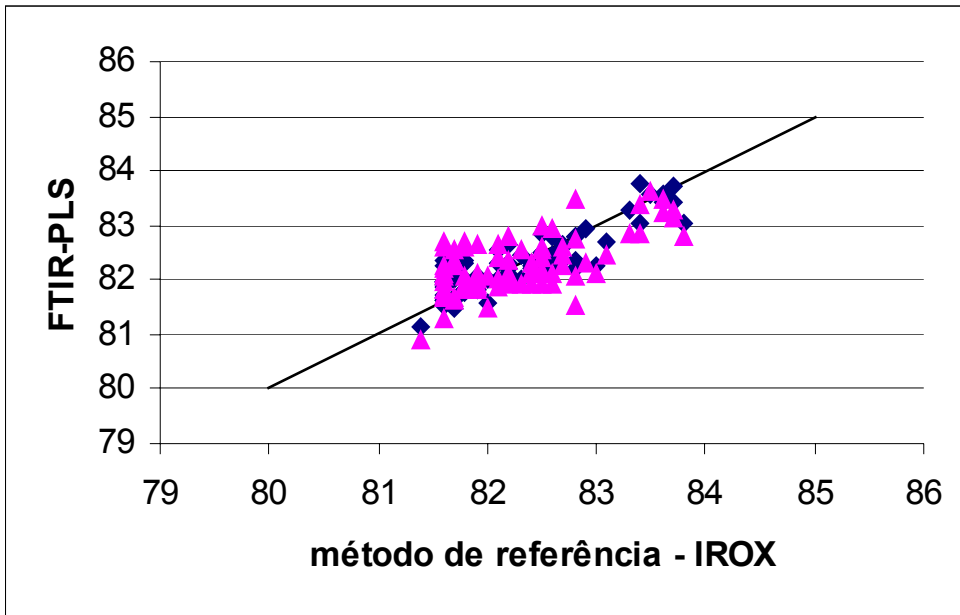


Figura 44- Valores de MON nas gasolinas.

◊ calibração ($R^2 = 0,7682$); ◻ validação ($R^2 = 0,4217$); — teórico.

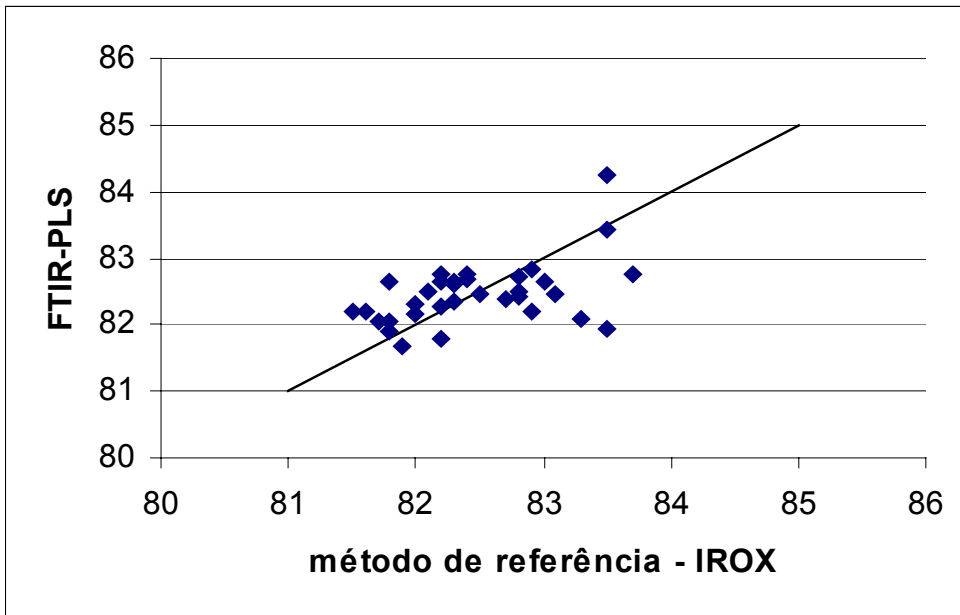


Figura 45- Valores de MON nas gasolinas.

◊ conjunto teste ($R^2 = 0,2343$); — teórico.

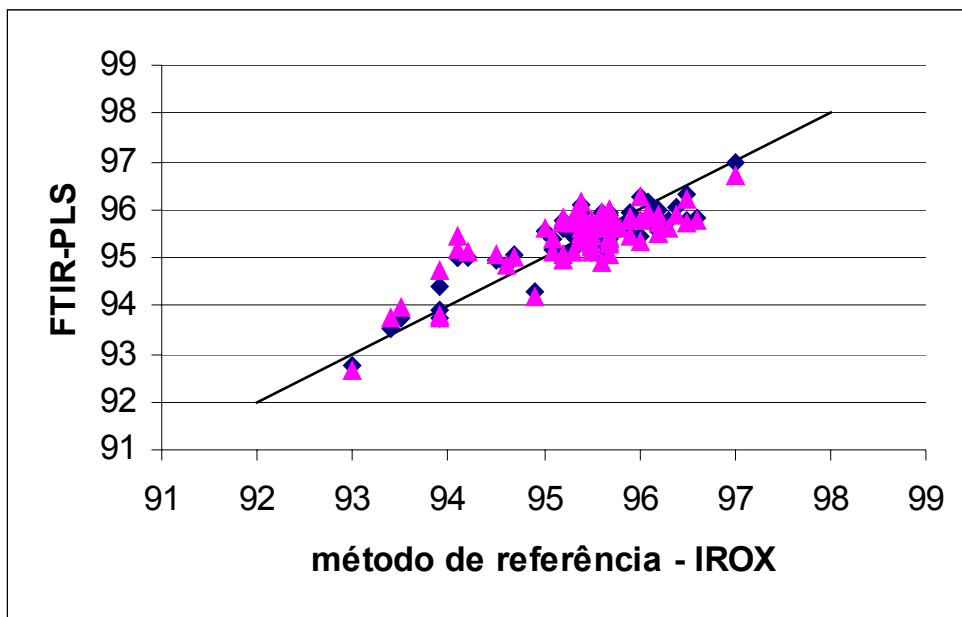


Figura 46- Valores de RON nas gasolinas.

◊ calibração ($R^2 = 0,7601$); ◀ validação ($R^2 = 0,6407$); — teórico.

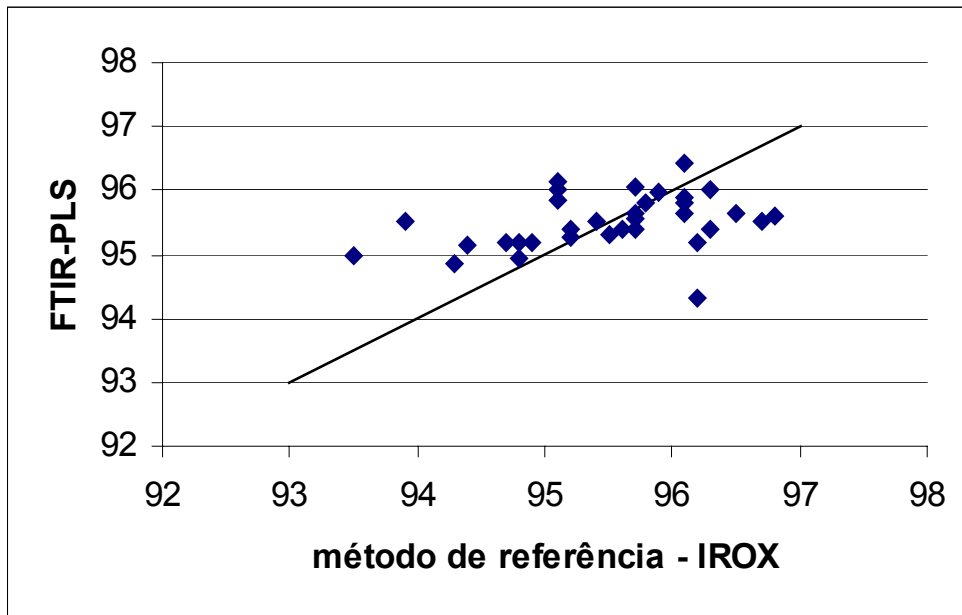


Figura 47- Valores de RON nas gasolinas.

◊ conjunto teste ($R^2 = 0,1229$); — teórico.

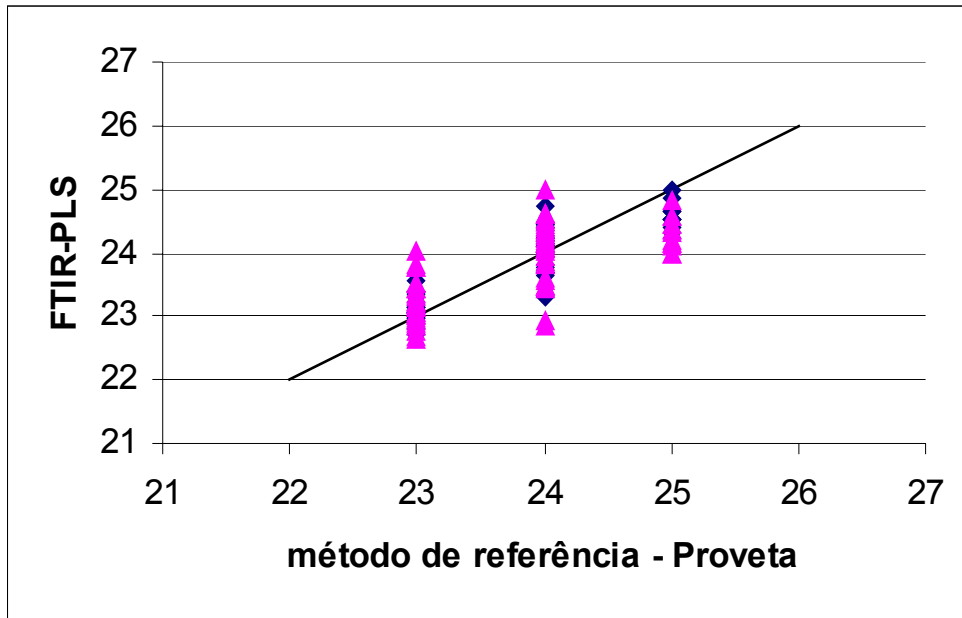


Figura 48- Teor de álcool (%v) nas gasolinas.

◊ calibração ($R^2 = 0,8009$); ▾ validação ($R^2 = 0,4901$); — teórico.

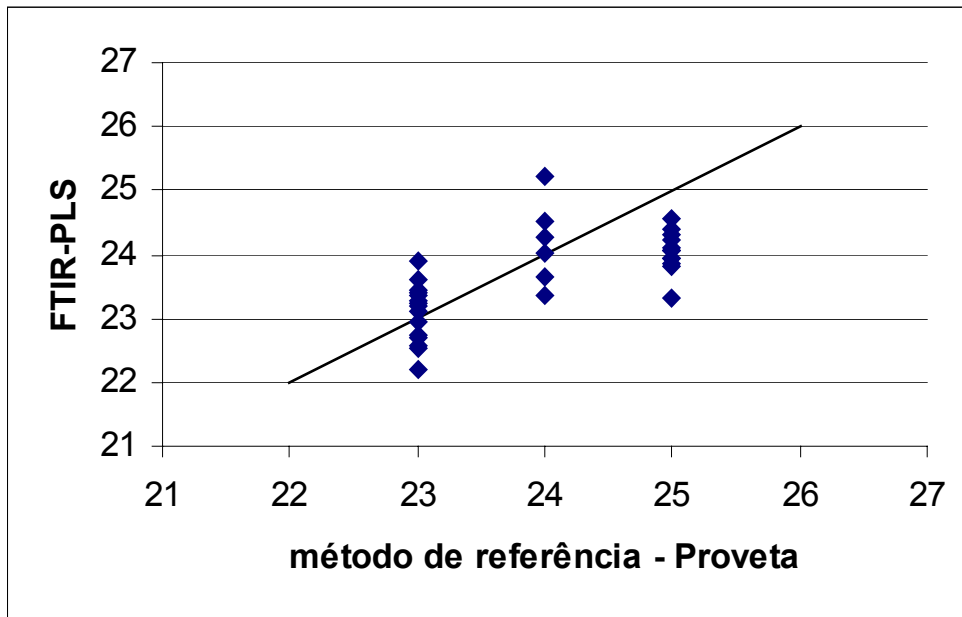


Figura 49- Teor de álcool (%v) nas gasolinas.

◊ conjunto teste ($R^2 = 0,4473$); — teórico.

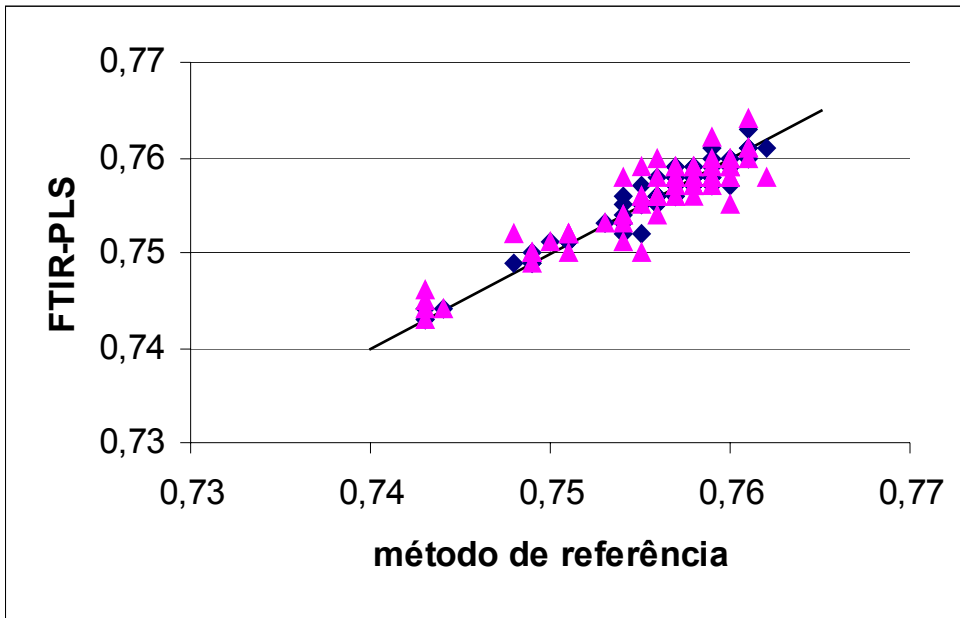


Figura 50- Valores de densidade das gasolinas.
 ◊ calibração ($R^2 = 0,9466$); ◡ validação ($R^2 = 0,8531$); — teórico.

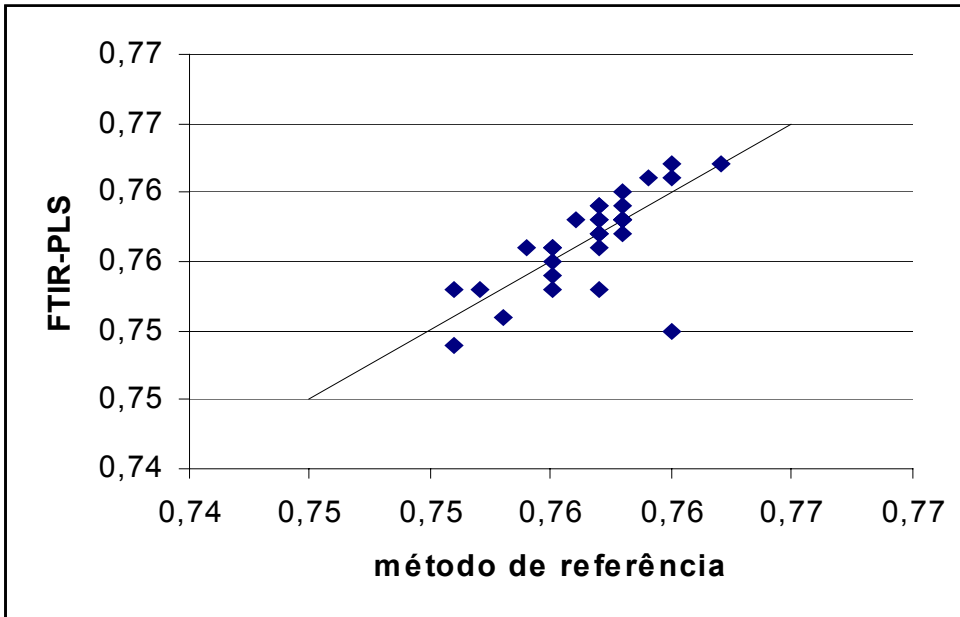


Figura 51- Valores de densidade nas gasolinas.
 ◊ teste ($R^2 = 0,4981$); — teórico.

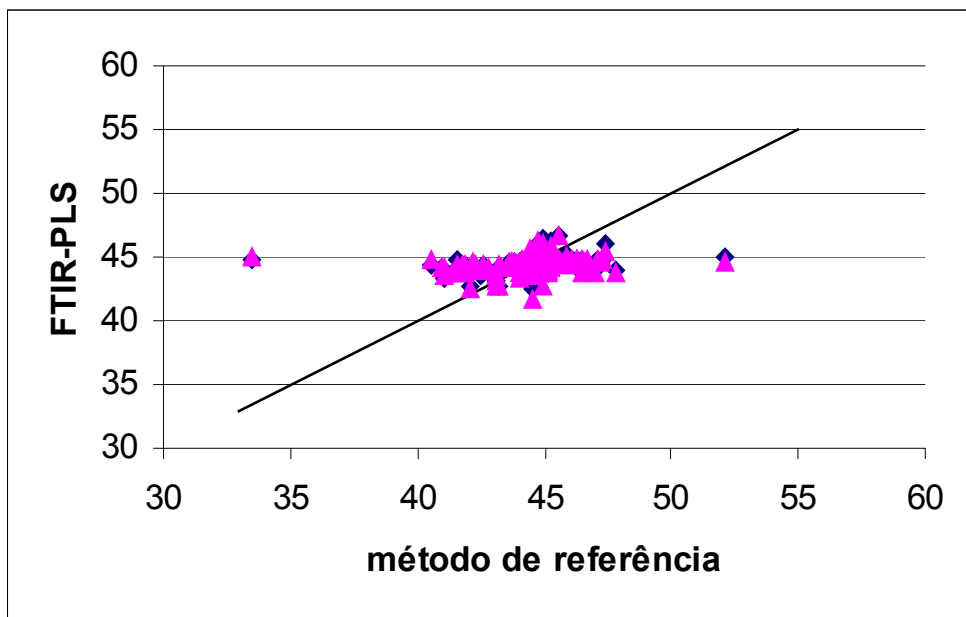


Figura 52- Valores de ponto de ebulição inicial nas gasolinas.

○ calibração ($R^2 = 0,1126$); ▲ validação ($R^2 = 0,0474$); — teórico.

Foram utilizadas todas as 110 amostras para validação cruzada.

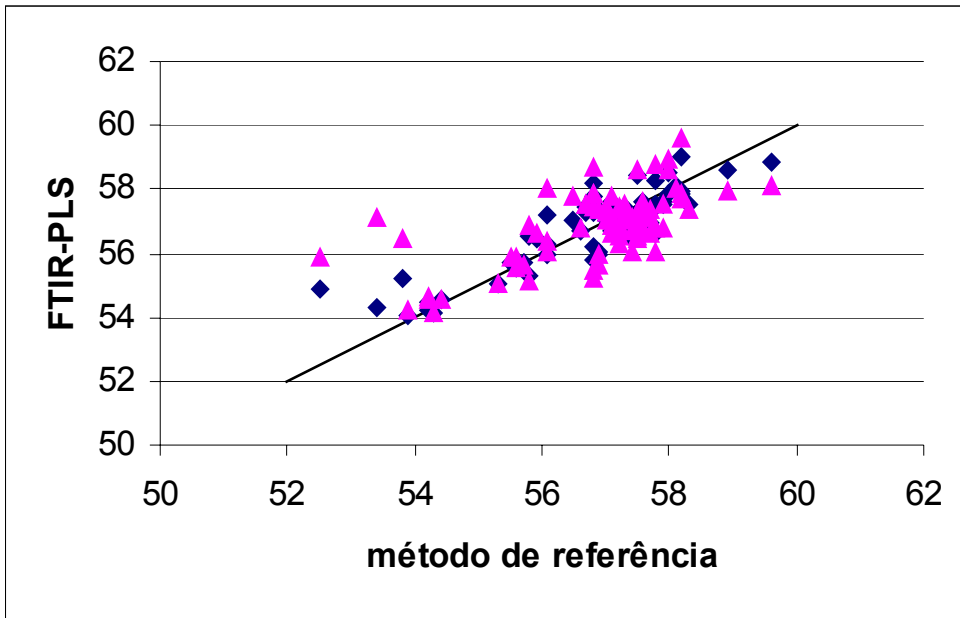


Figura 53- Valores de PE 10% nas gasolinas.

◊ calibração ($R^2 = 0,7590$); ◈ validação ($R^2 = 0,4143$); — teórico.

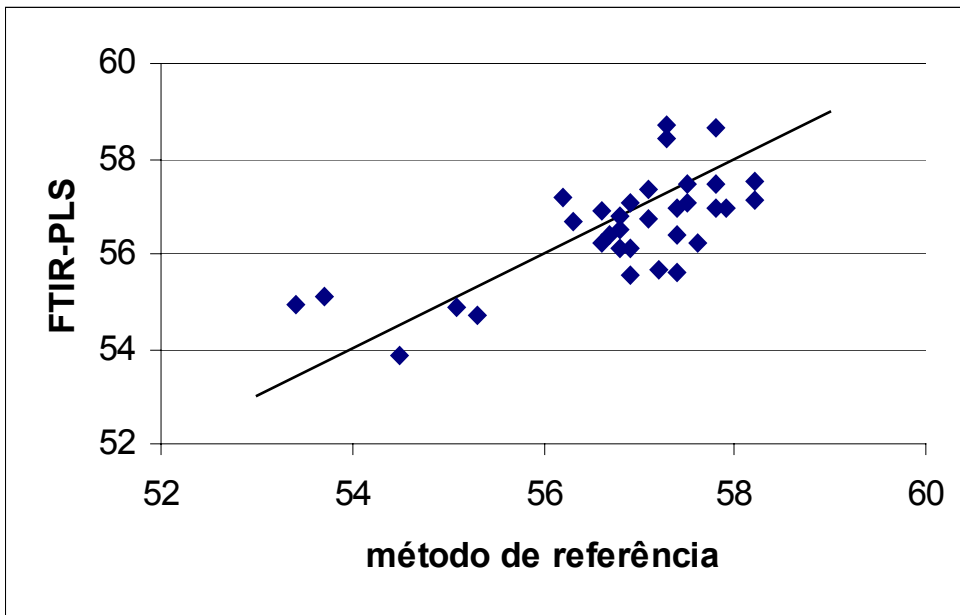


Figura 54- Valores de PE 10% nas gasolinas.

◊ conjunto teste ($R^2 = 0,5057$); — teórico.

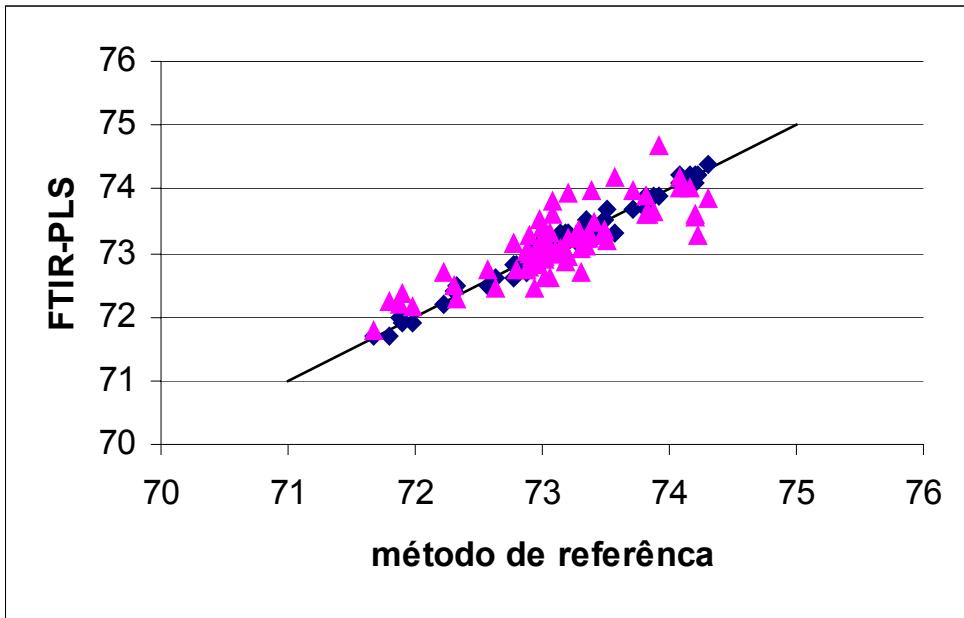


Figura 55- Valores de PE 50% nas gasolinas.

◊ calibração ($R^2 = 0,9733$); ▲ validação ($R^2 = 0,5802$); — teórico.

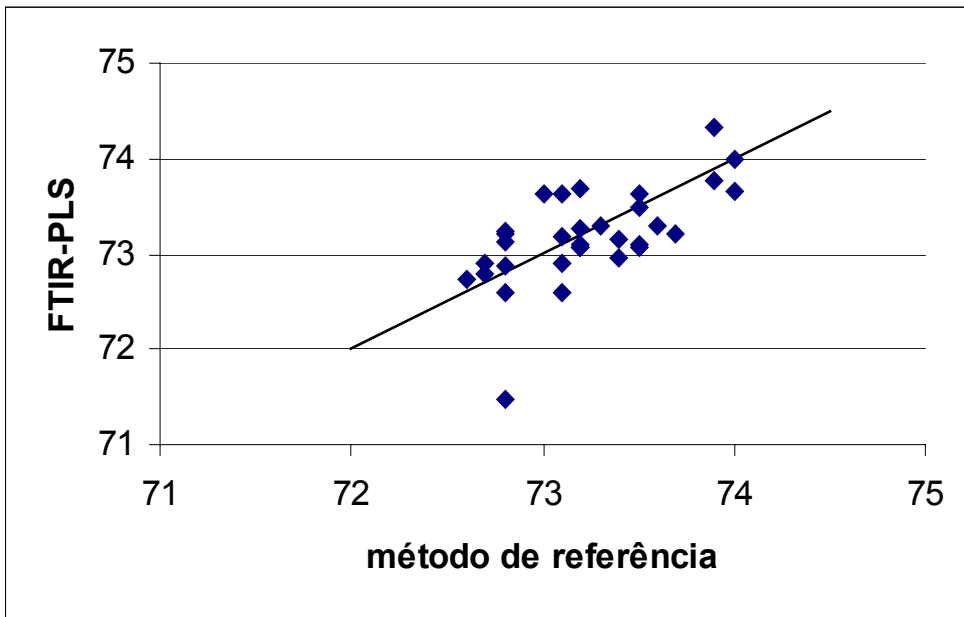


Figura 56- Valores de PE 50% nas gasolinas.

◊ conjunto teste ($R^2 = 0,4067$); — teórico.

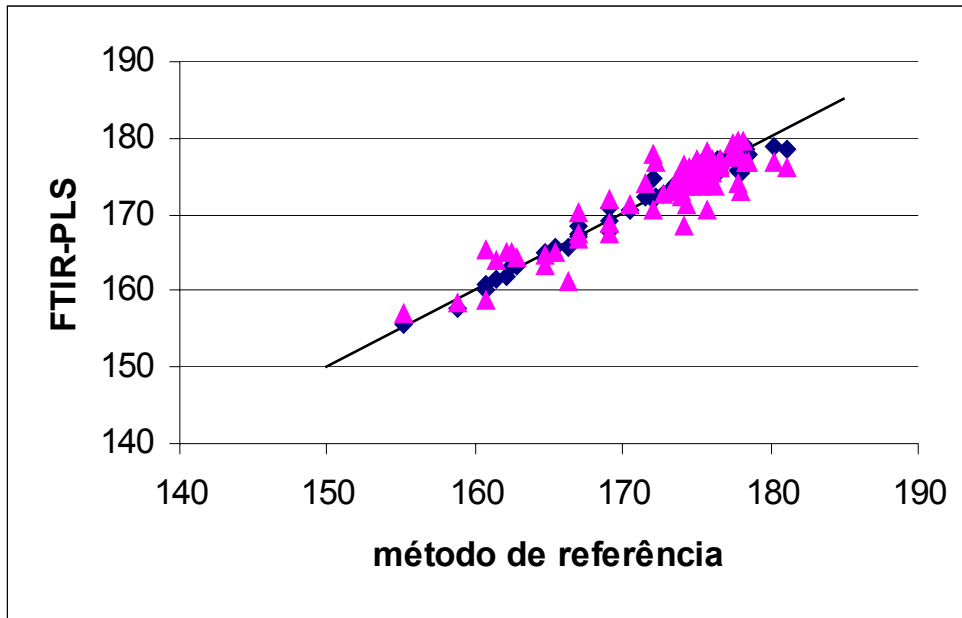


Figura 57- Valores de PE 90% nas gasolinas.

◊ calibração ($R^2 = 0,9746$); ▲ validação ($R^2 = 0,8348$); — teórico.

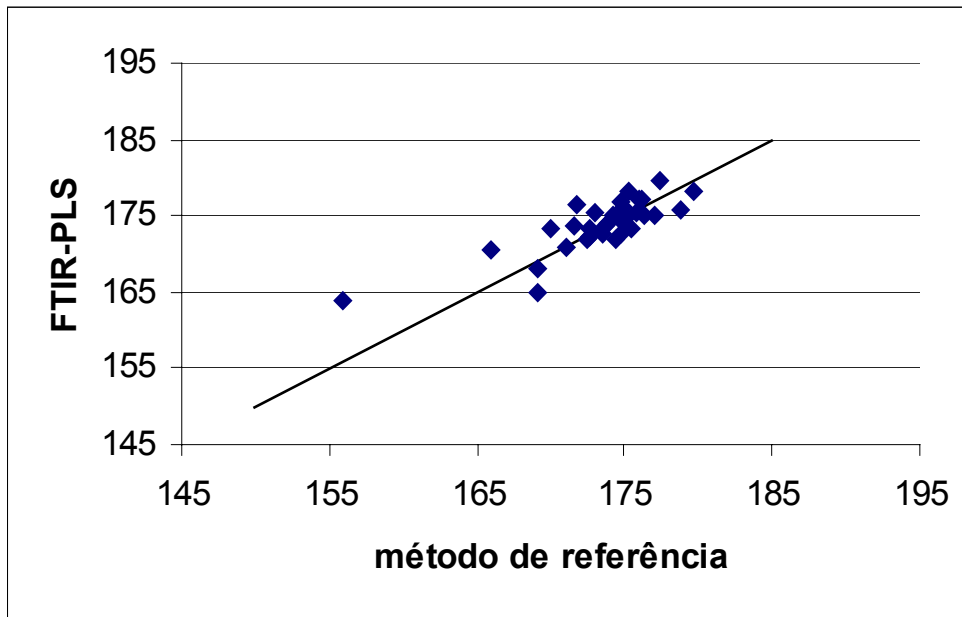


Figura 58- Valores de PE 90% nas gasolinas.

◊ teste ($R^2 = 0,6516$); — teórico.

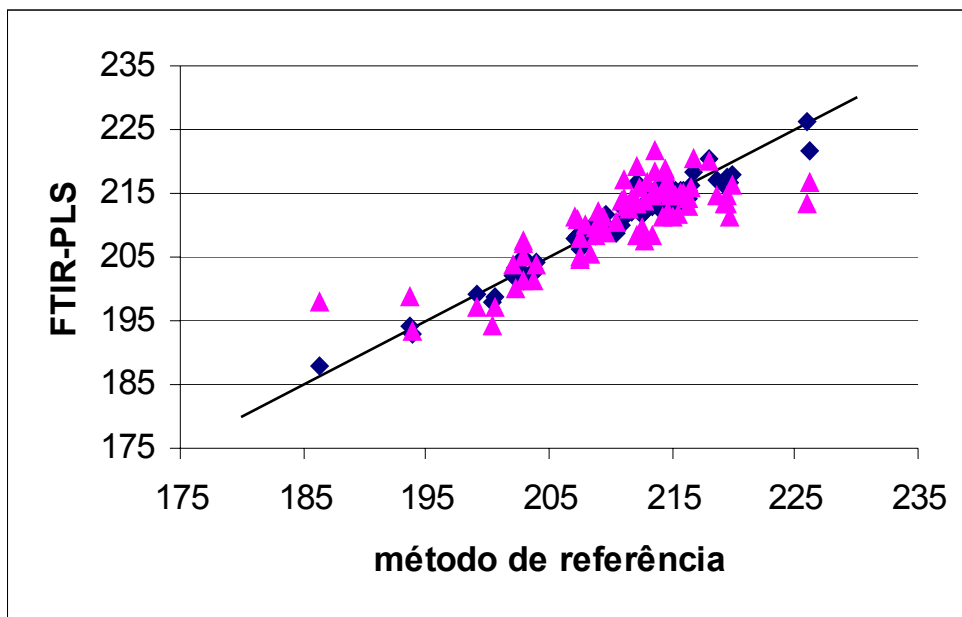


Figura 59- Valores de PEF nas gasolinas.

◊ calibração ($R^2 = 0,9504$); ▲ validação ($R^2 = 0,6630$); — teórico.

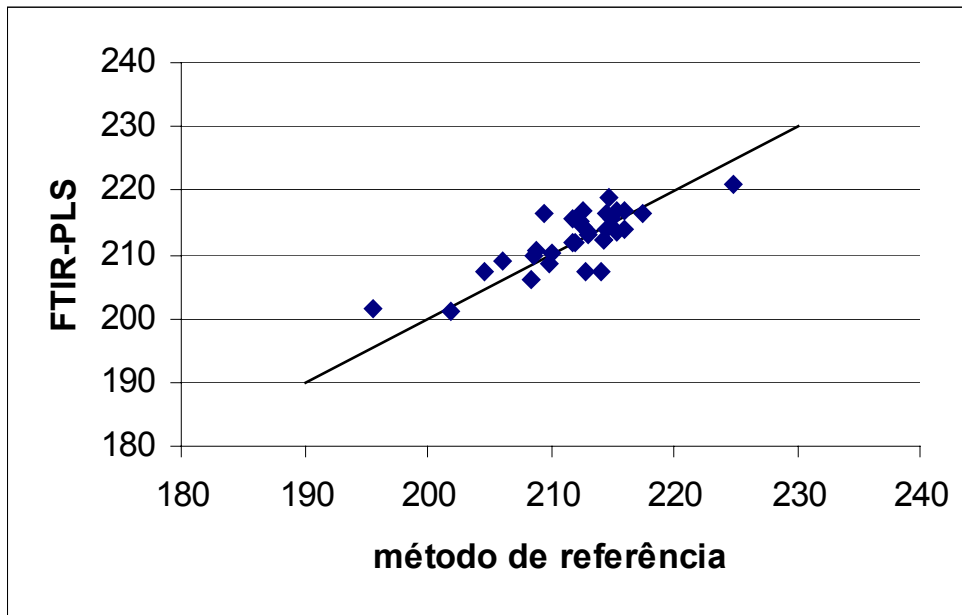


Figura 60- Valores de PEF nas gasolinas.

◊ conjunto teste ($R^2 = 0,6570$); — teórico.

7.3.3. Precisão do Método FTIR-PLS

Para avaliar a predição do método FTIR-PLS foram calculados os valores de repetibilidade (r) e reprodutibilidade (R). Seguindo o critério IUPAC, r foi calculado como o desvio padrão (DP) de 20 resultados independentes de uma mesma amostra obtidos num mesmo dia, e R foi calculado como o desvio padrão de 10 resultados independentes de uma mesma amostra obtidos em dias diferentes. A tabela 16 apresenta o valor médio de cada propriedade para os dois conjuntos de dados utilizados, junto com o desvio padrão e o coeficiente de variação ($CV\%$), calculado como mostra a equação 16. O coeficiente de variação foi incluído para possibilitar uma melhor comparação entre as diferentes propriedades através de valores percentuais.

$$CV\% = (DP * 100) \div \bar{y} \quad \text{equação 16}$$

Tabela 14- Repetibilidade e reprodutibilidade do método FTIR-PLS para predição de propriedades de gasolinas comerciais.

PROPRIEDADE	REPETIBILIDADE (r)			REPRODUTIBILIDADE (R)		
	média	DP	$CV\%$	média	DP	$CV\%$
IROX*						
% aromáticos	17,0	0,2	1,1	17,1	0,4	2,1
% benzeno	0,33	0,02	6,8	0,34	0,0	5,2
% etanol	23,3	0,5	2,0	22,4	0,8	3,3
% olefinas	18,1	0,7	3,9	17,6	1,0	5,7
% saturados	42,3	0,7	1,6	43,1	0,9	2,1
MON	81,7	0,2	0,2	81,9	0,3	0,3
RON	93,9	0,1	0,1	94,0	0,1	0,1
ASTM/ABNT*						
% etanol	23,1	0,3	1,1	23,1	0,3	1,2
PEI	43,4	0,2	0,4	43,5	0,1	0,2
PE 10%	57,2	0,3	0,5	56,9	0,3	0,6
PE 50%	73,2	0,1	0,2	73,2	0,2	0,2
PE 90%	168,1	0,8	0,5	167,6	0,9	0,5
PEF	205,8	1,5	0,7	205,0	0,9	0,4
dens	0,7554	0,0010	0,1	0,7550	0,0012	0,2

*Método de referência utilizado para determinar as propriedades listadas.

Os valores de r e R obtidos para todas as propriedades, principalmente aquelas cujos valores de referência foram obtidos por métodos-padrão, foram considerados satisfatórios, indicando boa precisão do método FTIR-PLS. As propriedades medidas pelo IROX apresentaram maiores desvios, como era esperado, chegando a um valor superior a 6% para o teor de benzeno. Já as propriedades obtidas pelos métodos-padrão ASTM ou ABNT, apresentaram desvios inferiores a 1% para todas as propriedades, com exceção do teor de álcool, cujos valores foram 1,1% para repetibilidade e 1,2% para reprodutibilidade.

7.3.4. Gasolinas versus Misturas

Uma outra abordagem para a aplicação do método FTIR-PLS seria a utilização das calibrações construídas com as misturas de solventes para a predição da composição das gasolinas em termos de benzeno, aromáticos, saturados, olefinas e etanol. No entanto, os resultados assim obtidos apresentaram erros muito altos. Duas razões podem ser mencionadas para justificar esse fato:

- os valores de referência para a composição das gasolinas não foram obtidos por um método-padrão de referência, mas sim através do equipamento IROX, que não foi calibrado com gasolinas brasileiras para essas propriedades;
- as gasolinas são misturas muito mais complexas que as misturas de solventes utilizadas, havendo, portanto, diferenças consideráveis entre os espectros infravermelho desses dois grupos de amostras.

A análise dos componentes principais (PCA), feita com os espectros de todas as amostras de gasolinas e misturas, deixa bem evidente como estes espectros dividem-se em dois grupos distintos, indicando a impossibilidade de usar a calibração obtida com um conjunto para predizer as propriedades do outro. A figura 61 apresenta os “scores” obtidos nos três primeiros componentes principais e o conjunto em destaque refere-se às misturas de solventes.

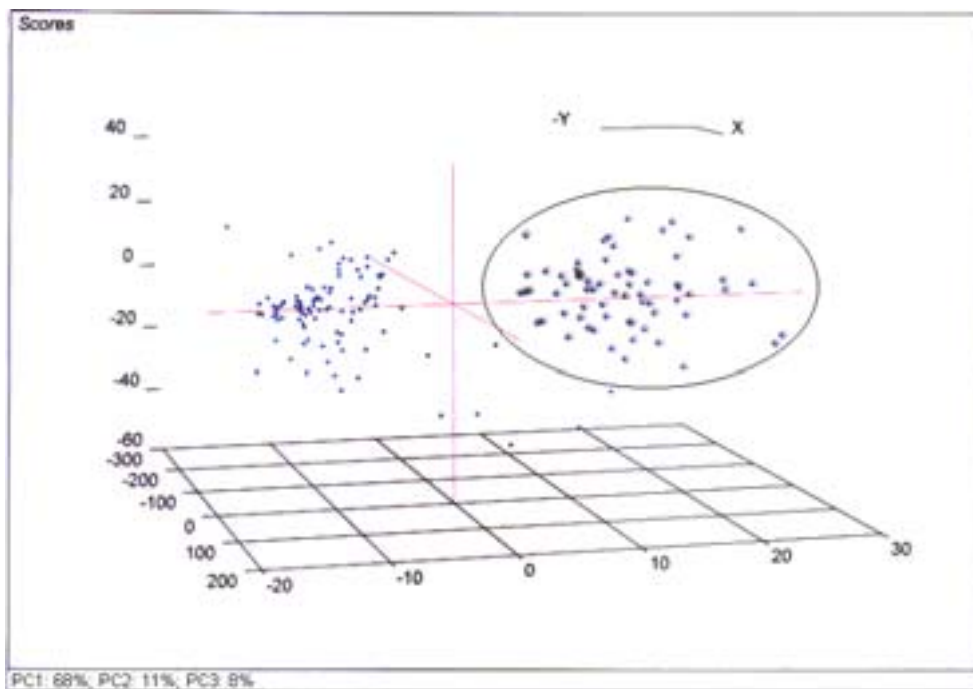


Figura 61- Análise dos componentes principais (PCA) de todas as amostras de gasolinas e misturas de solventes.