

**UNIVERSIDADE FEDERAL DE MINAS GERAIS**  
**Faculdade de Farmácia**  
**Programa de Pós-Graduação em Ciência de Alimentos**

Matheus Mendes Malaquias Friaça Glória

**ESTUDO DO PERFIL DE CONTAMINAÇÃO DE ALIMENTOS POR DIOXINAS  
QUANTO À POSSÍVEL ORIGEM UTILIZANDO MODELOS QUIMIOMÉTRICOS  
DISCRIMINANTES**

Belo Horizonte

2020

Matheus Mendes Malaquias Friaça Glória

**ESTUDO DO PERFIL DE CONTAMINAÇÃO DE ALIMENTOS POR DIOXINAS  
QUANTO À POSSÍVEL ORIGEM UTILIZANDO MODELOS QUIMIOMÉTRICOS  
DISCRIMINANTES**

Dissertação apresentada ao Programa de Pós-graduação em Ciência de Alimentos da Faculdade de Farmácia da Universidade Federal de Minas Gerais como requisito parcial à obtenção de grau de Mestre em Ciência de Alimentos.

Orientador: Prof. Dr. Leandro Soares Oliveira

Colaborador: M.e.Rafael Pissinatti

Área de concentração: Ciência de Alimentos

Belo Horizonte

2020

G562e

Glória, Matheus Mendes Malaquias Friaça.

Estudo do perfil de contaminação de alimentos por dioxinas quanto à possível origem utilizando modelos quimiométricos discriminantes / Matheus Mendes Malaquias Friaça Glória. – 2020.

122 f. : il.

Orientador: Leandro Soares Oliveira.

Colaborador: Rafael Pissinatti.

Dissertação (mestrado) - Universidade Federal de Minas Gerais, Faculdade de Farmácia, Programa de Pós-Graduação em Ciência de Alimentos.

1. Dioxinas – Teses. 2. Furanos – Teses. 3. Alimentos – Contaminação – Teses. 4. Gordura – Teses. 5. I. Oliveira, Leandro Soares. II. Pissinatti, Rafael. III. Universidade Federal de Minas Gerais. Faculdade de Farmácia. IV. Título.

CDD: 641.365



UNIVERSIDADE FEDERAL DE MINAS GERAIS  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DE ALIMENTOS

**PPGCA**

## FOLHA DE APROVAÇÃO

ESTUDO DO PERFIL DE CONTAMINAÇÃO DE ALIMENTOS POR DIOXINAS  
QUANTO À POSSÍVEL ORIGEM POR MEIO DE MODELOS DE CLASSIFICAÇÃO  
EM PLS-DA

**MATHEUS MENDES MALAQUIAS FRIAÇA GLÓRIA**

Dissertação submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em CIÊNCIA DE ALIMENTOS, como requisito para obtenção do grau de Mestre em CIÊNCIA DE ALIMENTOS, área de concentração CIÊNCIA DE ALIMENTOS.

Aprovada em 30 de abril de 2020, pela banca constituída pelos membros:

Prof. Dr. Leandro Soares de Oliveira (Orientador e Presidente da Comissão)  
Escola de Engenharia - UFMG

Profa. Dra. Scheilla Vitorino Carvalho de Souza Ferreira  
Faculdade de Farmácia - UFMG

Prof. Dr. Marcelo Martins de Sena  
Instituto de Ciências Exatas - UFMG

Belo Horizonte, 30 de abril de 2020.

## AGRADECIMENTOS INSTITUCIONAIS

Ao Programa de Pós-Graduação em Ciência de Alimentos (PPGCA) da Faculdade de Farmácia (FAFAR) da Universidade Federal de Minas Gerais (UFMG).

Ao Laboratório Federal de Defesa Agropecuária de Minas Gerais (LFDA-MG) do Ministério da Agricultura, Pecuária e Abastecimento (MAPA), pela infraestrutura e treinamentos recebidos.

Ao Laboratório de Resíduos e Contaminantes Alimentares (RCA) do LFDA-MG, pelo apoio dos colaboradores e disponibilização da infraestrutura.

## AGRADECIMENTOS PESSOAIS

À minha família, em especial à minha esposa, pelo apoio, incentivo e compreensão pela ausência em certos momentos.

Ao Rafael Pissinatti, pela confiança depositada, oportunidade de aprendizagem e colaboração em todas as etapas de planejamento e discussões técnicas acerca deste projeto.

Aos amigos de graduação e pós-graduação, Henrique, Glaucimar e Ana pela ajuda, ideias e discussões técnicas.

Aos demais colegas do LFDA-MG, pela ajuda e companheirismo durante todo esse período.

## RESUMO

Eventos de contaminação por dioxinas e furanos se caracterizam por serem problemas científicos complexos que requerem conhecimentos específicos sobre processos produtivos, levantamento de prováveis fontes de contaminação, compreensão dos padrões de bioacumulação e transferência, sendo que a fonte de contaminação geralmente possui um perfil específico. Além disso, demandam uma grande quantidade de investimento, tempo e esforço laboral para a sua elucidação. A necessidade de se mitigar, no menor tempo, um evento cria a necessidade do uso de ferramentas que possam classificar o perfil de contaminação encontrado em relação a um grupo de perfis com a origem conhecida. Perfis de várias origens de contaminação foram compilados da literatura existente, sendo classificados em doze grupos distintos. A análise discriminante por mínimos quadrados parciais foi utilizada com o propósito de se construir um modelo que permitisse classificar perfis de amostras de dioxinas e furanos quanto à provável origem de contaminação. O modelo construído foi avaliado quanto as figuras de mérito sensibilidade, especificidade, taxa de falso positivo e taxa de falso negativo. Os valores para a sensibilidade, do modelo, variaram de 73,7 a 100 %. Valores para a especificidade variaram de 76,9 a 100%. Taxas de falso positivo e falso negativo foram de até 26,3 e 23,1 %. Gráficos de vetores de regressão indicaram quais variáveis são mais significativas para a classificação de um determinado perfil de contaminação quanto à um determinado grupo. Perfis de amostras das matrizes gordura de aves, ovos e ração, fornecidas pelo Laboratório de Dioxinas e PCB's do Laboratório Federal de Defesa Agropecuária de Minas Gerais (MAPA), foram classificadas, sendo que o resultado obtido foi condizente como a classificação esperada, apesar de taxas de bioacumulação e transferência não terem sido consideradas. O modelo construído pode ser utilizado para nortear o processo de investigação, sugerindo a provável origem desta, simplificando e trazendo agilidade ao processo de monitoramento destes contaminantes alimentares no Brasil.

**Palavras-chave:** dioxinas; furanos; PLS-DA.

## ABSTRACT

Dioxin and furan contamination events are characterized by being complex scientific problems that require specific knowledge about production processes, survey of likely sources of contamination, understanding of bioaccumulation and transfer patterns, once the source of contamination has generally a specific fingerprint. Moreover, the contamination events demand a large amount of investment, time and labor effort to elucidate them. The need to mitigate, in the shortest time, an event creates the need to use tools that can classify the contamination profile, found in relation to a group of profiles which have a known origin. Profiles from various sources of contamination were compiled from the existing literature, being classified into twelve different groups. The discriminant analysis by partial least squares was used in order to build a model that would allow classifying profiles of samples of dioxins and furans as to the probable source of contamination. The developed model was evaluated for the following figures of merit: sensitivity, specificity, false positive rate and false negative rate. The model's sensitivity values ranged from 73.7 to 100%. Values for specificity ranged from 76.9 to 100%. False positive and false negative rates were up to 26.3 and 23.1%. Regression vector plots indicated which variables are the most significant for the classification of a given contamination profile for a given group. Profiles of poultry fat, eggs and feed samples, supplied by the Dioxins and PCB's Laboratory from the Federal Laboratory of Animal and Plant Health and Inspection, were classified, and the result obtained was consistent with the expected classification, although bioaccumulation and transfer rates are not considered. The developed model can be used to guide the investigation process, suggesting its probable origin, simplifying and speeding up the process of monitoring these food contaminants in Brazil.

**Keywords:** dioxins; furans; PLS-DA.

## Lista de Figuras

Figura 1: Estrutura básica para dibenzo-p-dioxinas (à esquerda) e dibenzo-furanos(à direita).....	21
Figura 2: Estrutura química da 2,3,7,8-tetraclorodibenzo-p-dioxina.....	23
Figura 3: Perfil de congêneres médio de emissões de crematórios, fábricas de cimento, sínteses de cloroanil, queimas de lixo eletrônico, fundições primárias de ferro, efluentes químicos, emissões de veículos com motores a diesel, fundição secundária de alumínio e resíduos de branqueamento de papel (Adaptado de ZHU et al., 2008).Possíveis origens de contaminação por PCDD/PCDF.....	26
Figura 4: Perfil de congêneres médio de emissões de incineradores, de incineradores de lixo hospitalar, emissões de veículos que utilizam gasolina sem aditivos e gasolina aditivada com compostos de chumbo, emissões de metalurgia secundária de cobre e de chumbo, emissões de caldeiras industriais e da contaminação do pentaclorofenol (Adaptado de ZHU et al., 2008).....	27
Figura 5: Possíveis origens de contaminação por PCDD/PCDF. (Adaptado de WEBER et al., 2018).....	29
Figura 6: Representação gráfica da análise de componentes principais (adaptado de ALMEIDA 2015).....	33
Figura 7: Figura 7: Decomposição das matrizes <b>X</b> e <b>Y</b> para elaboração do modelo PLS (adaptado de ALMEIDA, 2015).....	35
Figura 8: Dispersão das amostras do grupo MIN em duas componentes principais. Entre parênteses, % explicada por cada componente principal. Dados em verde, azul, vermelho e marrom representam caulinita, bentonita, montmorilonita e argila respectivamente.....	51
Figura 9: Dispersão do grupo MTL em 3 CP. Dados em azul, verde, vermelho e marrom correspondem à metalurgia de cobre, magnésio, ferro e alumínio respectivamente. ....	52
Figura 10: Dispersão do grupo PTL em 2 CP. Dados em vermelhos corresponde à processos de PVC e dados em verde corresponde à processos de produção de EDC. ....	53

Figura 11: Dispersão dos perfis de contaminação do grupo CIM no espaço criado pelas duas primeiras componentes principais e a correlação de cada variável com cada componente. Perfis em vermelho e variáveis em azul.....	54
Figura 12: Dispersão dos perfis de contaminação do grupo INC no espaço criado pelas duas primeiras componentes principais e a correlação de cada variável com cada componente. Perfis em vermelho e variáveis em azul.....	55
Figura 13: Dispersão dos perfis de contaminação do grupo PCB no espaço criado pelas duas primeiras componentes principais e a correlação de cada variável com cada componente. Perfis em vermelho e variáveis em azul.....	56
Figura 14: Dispersão dos perfis de contaminação do grupo PGM no espaço criado pelas duas primeiras componentes principais e a correlação de cada variável com cada componente. Perfis em vermelho e variáveis em azul.....	57
Figura 15: Dispersão dos perfis de contaminação do grupo AGR no espaço criado pelas duas primeiras componentes principais.....	58
Figura 16: Dispersão dos perfis de contaminação do grupo CMB no espaço criado pelas duas primeiras componentes principais.....	59
Figura 17: Dispersão dos perfis de contaminação do grupo CMB no espaço criado pelas duas primeiras componentes principais. Informação sobre o teor de cloreto nos combustíveis.....	59
Figura 18: Distribuição das amostras no plano criado pelas duas primeiras componentes principais. Triângulo vermelhos são referentes ao grupo AGR, quadrados azuis são referentes ao grupo PGM, losangos rosas são referentes ao grupo MIN, círculos verde escuro são referentes ao grupo CMB, losangos verde claro são referentes ao grupo MTL, estrelas marrons são referentes ao grupo PTL, triângulos laranjas são referentes ao grupo PCB, triângulos pretos são referentes ao grupo INC e triângulos roxos são referentes ao grupo CIM.....	60
Figura 19: Hierarquia de classificação adotada para a classificação do perfil de contaminantes quanto à sua provável origem (modelo III).....	61

Figura 20: Gráfico de classificação para a etapa de classificação I. À esquerda conjunto de treinamento e à direita conjunto teste. Conjunto formado pelos grupos AGR-A, AGR-B, PGM e MIN representados por triângulos vermelhos. Conjunto formado pelos grupos MTL-A, MTL-B, AGR-C, PCB, INC, CIM, CMB, PTL representados por círculos azuis.....65

Figura 21: Gráfico de vetores de regressão referentes à classificação para a etapa de classificação I. Conjunto formado pelos grupos AGR-A, AGR-B, PGM e MIN representados como classe “positiva”. Conjunto formado pelos grupos MTL-A, MTL-B, AGR-C, PCB, INC, CIM, CMB, PTL representados como classe “negativa”.....66

## Lista de tabelas

Tabela 1: 17 congêneres monitorados nas análises do PNCRC Animal: nome e denominação comum.....	22
Tabela 2. Valores de TEFs (Fatores de Equivalência Tóxica) estabelecidos pela Organização Mundial da Saúde (OMS) para dibenzo-p-dioxinas (PCDDs), dibenzofuranos (PCDFs) (Fonte: Adaptado de OMS, 2006).....	23
Tabela 3: Ano, local, matriz afetada, origem da contaminação e referência de alguns eventos de contaminação por dioxinas.....	28
Tabela 4: Tabela 4: Limites de referência adotados para as matrizes analisadas no Plano Nacional de Controle de Resíduos Contaminantes quanto à contaminação por PCDD, PCDF (Fonte: Adaptado de MAPA, 2016).....	30
Tabela 5: Trabalhos recentes na área de alimentos com o uso de ferramentas quimiométricas.....	32
Tabela 6: Tabela com as variáveis utilizadas e a sua respectiva descrição.....	41
Tabela 7: Tabela de correlação do código adotado e o processo industrial.....	43
Tabela 8: Codificação utilizada, matriz, origem da contaminação e referência bibliográfica de cada amostra a ser classificada.....	43
Tabela 9: Localidade referente à cada amostra e a matriz analisada.....	49
Tabela 10: Número de VLs, variâncias explicadas e erros de classificação de validação cruzada (ECVC) referente aos modelos em PLS-Da para as etapas de classificação.....	62
Tabela 11: Figuras de mérito estimadas para etapa do modelo de classificação – Conjunto de treinamento.....	63
Tabela 12: Figuras de mérito estimadas para etapa do modelo de classificação – Conjunto de teste.....	64
Tabela 13: Tabela com o código da amostra, a classificação esperada e a classificação final obtida (etapa III).....	66
Tabela 14: Tabela com o código, local, matriz e classificação final respectivo das amostras da tabela 9.....	70

## Lista de Abreviaturas e Siglas

1,2,3,4,6,7,8-HpCDD	1,2,3,4,6,7,8-Heptaclorodibenzo-p-dioxina
1,2,3,4,6,7,8-HpCDF	1,2,3,4,6,7,8-Heptaclorodibenzofurano
1,2,3,4,7,8,9-HpCDF	1,2,3,4,7,8,9-Heptaclorodibenzofurano
1,2,3,4,7,8-HxCDD	1,2,3,4,7,8-Hexaclorodibenzo-p-dioxina
1,2,3,4,7,8-HxCDF	1,2,3,4,7,8-Hexaclorodibenzofurano
1,2,3,6,7,8-HxCDD	1,2,3,6,7,8-Hexaclorodibenzo-p-dioxina
1,2,3,6,7,8-HxCDF	1,2,3,6,7,8-Hexaclorodibenzofurano
1,2,3,7,8,9-HxCDD	1,2,3,7,8,9-Hexaclorodibenzo-p-dioxina
1,2,3,7,8,9-HxCDF	1,2,3,7,8,9-Hexaclorodibenzofurano
1,2,3,7,8-PeCDD	1,2,3,7,8-Pentaclorodibenzo-p-dioxina
1,2,3,7,8-PeCDF	1,2,3,7,8-Pentaclorodibenzofurano
2,3,4,6,7,8-HxCDF	2,3,4,6,7,8-Hexaclorodibenzofurano
2,3,4,7,8-PeCDF	2,3,4,7,8-Pentaclorodibenzofurano
2,3,7,8-TCDD	2,3,7,8-Tetraclorodibenzo-p-dioxina
2,3,7,8-TCDF	2,3,7,8-Tetraclorodibenzofurano
CP	Componente principal
ECVC	Erro de classificação de validação cruzada
HCA	Análise de agrupamentos hierárquicos
IARC	Agência Internacional de Pesquisa em Câncer
LFDA-MG	Laboratório Federal de Defesa Agropecuária em Minas Gerais
LDP	Laboratório de Dioxinas e PCB's
MAPA	Ministério da Agricultura, Pecuária e Abastecimento
OCDD	Octaclorodibenzo-p-dioxina
OCDF	Octaclorodibenzofurano
OMS	Organização Mundial de Saúde

PCA	Análise de componentes principais
PCB-dl	Bifenilas policloradas semelhantes a dioxinas
PCDD	Dibenzo-p-dioxinas policloradas
PCDF	Dibenzo-furanos policlorados
PCDT	Dibenzo-tiofeno policlorados
PLS	Mínimos Quadrados parciais
PLS-DA	Análise discriminante por mínimos quadrados parciais
PNCRC	Programa Nacional de Controle de Resíduos e Contaminantes em Produtos de Origem Animal
TEFs	Fator de Equivalência tóxica ( <i>Toxic equivalency factors</i> )
TEQ	Equivalência tóxica Total ( <i>Total toxic equivalency</i> )
UBA	<i>Umweltbundesamt</i> – Agência Ambiental Alemã
PNUMA	Programa das Nações Unidas para o Meio Ambiente

## SUMÁRIO

<b>1. INTRODUÇÃO</b> .....	<b>16</b>
<b>2. OBJETIVOS</b> .....	<b>20</b>
<b>3. REVISÃO DE LITERATURA</b> .....	<b>21</b>
3.1 Dioxinas e Furanos .....	21
3.1.1 Características .....	21
3.1.2 Rotas de exposição de dioxinas aos alimentos.....	27
3.2 Quimiometria .....	31
3.2.1 Análise de componentes principais (PCA) .....	33
3.2.2 Análise Discriminante por Mínimos Quadrados Parciais (PLS-DA).....	34
<b>4. MATERIAIS E MÉTODOS</b> .....	<b>39</b>
4.1 Metodologia Analítica .....	39
4.2 Amostras (Perfis de Contaminação).....	40
4.2.1 Estrutura da matriz de dados .....	40
<b>5. RESULTADOS</b> .....	<b>51</b>
5.1 Análise exploratória dos dados compilados da literatura .....	51
5.2 Classificação .....	60
<b>6. DISCUSSÃO</b> .....	<b>72</b>
<b>7. CONCLUSÃO</b> .....	<b>75</b>
<b>8. REFERÊNCIAS BIBLIOGRÁFICAS</b> .....	<b>76</b>
<b>APÊNDICE A</b> .....	<b>84</b>
<b>APÊNDICE B</b> .....	<b>92</b>
<b>APÊNDICE C</b> .....	<b>95</b>
<b>APÊNDICE D</b> .....	<b>99</b>
<b>APÊNDICE E</b> .....	<b>107</b>

APÊNDICE F.....	115
-----------------	-----

## 1. Introdução

A revolução industrial, iniciada no século XVIII trouxe um enorme desenvolvimento científico e tecnológico. No entanto, junto com esse intenso desenvolvimento científico e tecnológico, vieram sérios problemas de poluição do ar, água, solos e contaminação de alimentos (ASSUNÇÃO; PESQUERO, 1999).

Nos últimos anos, a segurança alimentar tornou-se uma questão notória e prioritária para órgãos de regulamentação, indústrias, comércio exterior e consumidores. Análises de resíduos de contaminantes em alimentos, que podem ser divididos nos grupos de poluentes ambientais persistentes e substâncias naturalmente tóxicas como metais pesados e micotoxinas, intensificaram-se consideravelmente (SUN et al., 2011).

O grupo de contaminantes conhecido como “dioxinas” (dibenzo-p-dioxina policloradas (PCDD, do inglês polychlorinated dibenzodioxins) e dibenzofuranos policlorados (PCDF, do inglês polychlorinated dibenzofurans)) constitui uma classe de hidrocarbonetos aromáticos halogenados que são substâncias extremamente tóxicas e persistentes no meio ambiente. Os efeitos adversos desses contaminantes sobre a saúde humana incluem anormalidades cutâneas, imunotoxicidade, desordens no sistema reprodutivo, desregulação endócrina e carcinogenicidade (BERG *et al.*, 1998; ASSUNÇÃO; PESQUERO, 1999)

Esses contaminantes, de formação principalmente antropogênica, são subprodutos indesejados em vários processos químicos, como na produção de clorofenóis. Além disso, podem ser formados durante a queima de resíduos contendo cloro (KANAN *et al.*, 2018). Eles se apresentam na forma de misturas de centenas de compostos ou congêneres, com limites máximos regulamentados para gêneros alimentícios e para gêneros destinados à alimentação animal extremamente baixos, da ordem de pg/g de gordura ou amostra (MAPA, 2018).

A análise de dioxinas é dispendiosa, pois demanda a identificação e quantificação de resíduos de inúmeras compostos reconhecidamente de maior efeito adverso. Geralmente, são monitorados pelo Ministério da Agricultura Pecuária e Abastecimento (MAPA), no Plano Nacional de Controle de Resíduos Contaminantes (PNCRC), sete congêneres de dioxinas e dez congêneres de furanos, sabidamente os mais tóxicos (MAPA, 2020).

As dioxinas apresentam propriedades lipossolúveis, tendendo a se bioacumular na gordura corporal, tanto em animais quanto em seres humanos, sendo assim transmitidas por meio da cadeia alimentar.

Para reduzir as quantidades absorvidas pelos seres humanos, é necessário diminuir os níveis das substâncias na cadeia alimentar, sendo que o consumo de alimentos contaminados constitui a principal forma de exposição dos seres humanos. O Ministério da Agricultura, Pecuária e Abastecimento (MAPA) tem como objetivo a redução deste contaminante em toda a cadeia produtiva, visto que estudos sobre eventos de contaminação demonstram que a alimentação animal pode contribuir significativamente para a contaminação de gêneros alimentícios. Cerca de 90% da exposição média de dioxinas à seres humanos se deve à ingestão de alimentos de origem animal (MALISCH; KOTZ, 2014).

A principal forma de controle de contaminações desta natureza é o Plano Nacional de Controle de Resíduos Contaminantes em Produtos de Origem Animal (PNCRC Animal), que tem foco na segurança alimentar e na garantia da inocuidade dos alimentos comercializados em níveis nacional e internacional. Além de controle de dioxinas, também estão incluídos neste plano outros resíduos e contaminantes como metais pesados, drogas veterinárias, agroquímicos e micotoxinas (MAPA, 2020).

A busca de mitigação da contaminação com dioxinas em diferentes matrizes tem levado diferentes órgãos de pesquisa no mundo a realizar diversos estudos de identificação de potenciais fontes de contaminação em fábricas, locais de acondicionamento e até mesmo formas de transporte dos produtos agropecuários. Nestes estudos, foram identificadas inúmeras fontes de contaminação (TOOLKIT, 2013; PNUMA 2008).

A detecção da fonte “real” de um evento de contaminação na cadeia alimentar é um problema científico complexo e requer conhecimento específico sobre processos de produção, levantamento das possíveis fontes de contaminação e entendimento dos padrões de bioacumulação. Padrões de bioacumulação podem mudar devido à cinética e metabolismo em animais (MALISCH, 2017). A estabilidade de certos congêneres em animais também pode oferecer informações importantes em relação

ao início e à duração do incidente de contaminação (HOOGENBOOM *et al.*, 2015). Estudos indicam que os perfis de congêneres de dioxinas em alimentos de localidades distintas são diferentes, sendo que isso pode estar relacionado a diferentes fontes de contaminação (WANG *et al.*, 2017).

Uma vez que a análise de dioxinas apresenta um alto custo, a identificação da fonte de contaminação de gêneros alimentícios é extremamente demorada e requer um alto investimento. Um processo de investigação adequado poderia levar semanas, até meses, com possibilidade de não se obter uma conclusão satisfatória ao final.

Um modelo de classificação que permita nortear um processo de investigação, correlacionando o perfil de contaminação da amostra com o perfil de congêneres típico de alguma fonte seria extremamente interessante do ponto de vista financeiro, laboral e de saúde pública, já que a causa do problema pode ser descoberta e mitigada de forma mais ágil. Alguns estudos tentaram correlacionar o perfil típico de algumas fontes de contaminação, no entanto, se restringiram a um número limitado de fontes de emissão e/ou um número pequeno de amostras (perfis) e às análises apenas exploratórias. Um recente relatório técnico da agência ambiental alemã (ALEMANHA, 2015; FIEDLER *et al.*, 2018) é o maior estudo feito até o momento, em relação ao número de fontes de contaminação (14 classificadas como primárias e 10 classificadas como secundárias) e de amostras (280 amostras), correlacionando-as utilizando a técnica de análise de agrupamentos hierárquicos (HCA). Resultados preliminares deste relatório técnico foram publicados por Malisch (2017). Contudo, não existe nenhum estudo de classificação de amostras brasileiras correlacionando-as com a origem da contaminação.

Outra opção para a identificação da origem de contaminação poderia ser a investigação de outros indicadores. Análises de outras dioxinas, furanos e policlorados dibenzo tiofenos (PCDT) poderiam oferecer um indicativo da possível origem de contaminação, no entanto, essas opções iriam consistir em aumentar um problema inicial (aumento de análises e métodos validados), sem a garantia de resolução do mesmo (RAPPE, 1994; BUEKENS *et al.*, 2018).

Uma vez que se possui um problema complexo e se deseja obter uma informação sobre os dados existentes de uma amostra em relação aos dados típicos de uma determinada origem, sem a necessidade de novas análises químicas, análises

com ferramentas quimiométricas são as mais indicadas para se chegar a um resultado conclusivo. Ferramentas, como análise de componentes principais (PCA) e análise de agrupamentos hierárquicos (HCA), podem ser utilizadas de forma exploratória para se ter um entendimento de como é a dispersão dos dados no espaço amostral e a ferramenta análise discriminante por mínimos quadrados parciais (PLS-DA) pode ser utilizada com o intuito de classificação de uma amostra em relação à determinada (BRERETON, 2009).

Considerando, assim, as vantagens do emprego de métodos exploratórios e classificatórios para a determinação da origem de contaminação, assim como a não existência de estudos sobre as origens de contaminação de amostras brasileiras, o presente estudo objetivou aplicar as ferramentas quimiométricas (PCA e PLS-DA) para identificar as possíveis origens de contaminação de dioxinas e furanos em alimentos com foco nos produtos destinados à alimentação animal, frango e ovos.

## **2. Objetivos**

Classificar amostras brasileiras de frango e ovos, previamente analisadas no âmbito do PNCRC Animal para o teor de dioxinas e furanos, quanto à possível origem de contaminação por meio de modelos de classificação em PLS-DA.

### 3. Revisão de Literatura

#### 3.1 Dioxinas e Furanos

##### 3.1.1 Características

Dibenzo-p-dioxinas policloradas (PCDD – *polychlorinated-p-dibenzodioxin*) e dibenzo-furanos policlorados (PCDF – *polychlorinated-dibenzofurans*), conhecidos como dioxinas e furanos respectivamente, ou coletivamente denominados apenas pelo termo “dioxinas”, formam um grupo de substâncias tóxicas largamente reconhecidas como contaminantes ocupacionais e alimentares (DOPICO *et al.*, 2015; WEBER *et al.*, 2018), cujos efeitos adversos sobre a saúde humana incluem anormalidades cutâneas, imunotoxicidade, desordens no sistema reprodutivo, teratogenicidade, desregulação endócrina e carcinogenicidade (BERG *et al.*, 1998, ASSUNÇÃO; PESQUERO, 1999).

Este grupo de substâncias são hidrocarbonetos aromáticos clorados, formados por uma estrutura de três anéis, sendo dois anéis aromáticos interconectados por um anel oxigenado (fig. 1).

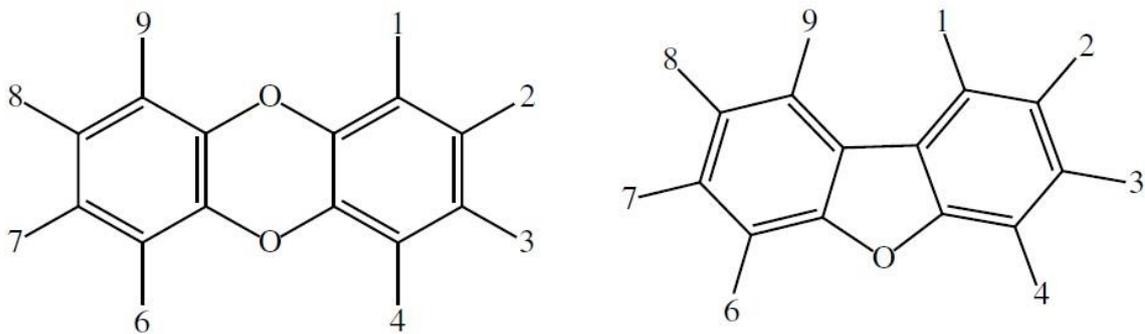


Figura 1: Estrutura básica para dibenzo-p-dioxinas (à esquerda) e dibenzo-furanos (à direita)

Cada posição numerada (fig. 1) pode ser substituída por um átomo de cloro ou de hidrogênio. Dependendo da quantidade de cloros, que pode variar de 1 a 8, e da posição à qual estão ligados aos anéis aromáticos, teoricamente, tem-se 75 congêneres de PCDD e 135 congêneres de PCDF possíveis (DOPICO *et al.*, 2015). Isômeros com no mínimo quatro substituições de cloro nas posições 2, 3, 7, 8

constituem um subconjunto de interesse especial devido à similaridade entre as suas características de toxicidade, estabilidade e bioacumulação. Esse subconjunto é composto por 7 PCDD e 10 PCDF que são frequentemente monitoradas, pois são sabidamente os mais tóxicos (RAPPE, 1994) (tabela 1).

Tabela 1: 17 congêneres monitorados nas análises do PNCRC Animal: nome e denominação comum

Composto	Denominação comum
<b>PCDDs</b>	
2,3,7,8-tetraclorodibenzo-p-dioxina	2,3,7,8-TCDD
1,2,3,7,8-pentaclorodibenzo-p-dioxina	1,2,3,7,8-PeCDD
1,2,3,4,7,8-hexaclorodibenzo-p-dioxina	1,2,3,4,7,8-HxCDD
1,2,3,6,7,8-hexaclorodibenzo-p-dioxina	1,2,3,6,7,8-HxCDD
1,2,3,7,8,9-hexaclorodibenzo-p-dioxina	1,2,3,7,8,9- HxCDD
1,2,3,4,6,7,8-heptaclorodibenzo-p-dioxina	1,2,3,4,6,7,8-HpCDD
octaclorodibenzo-p-dioxina	OCDD
<b>PCDFs</b>	
2,3,7,8-tetraclorodibenzofurano	2,3,7,8-TCDF
1,2,3,7,8-pentaclorodibenzofurano	1,2,3,7,8-PeCDF
2,3,4,7,8-pentaclorodibenzofurano	2,3,4,7,8-PeCDF
1,2,3,4,7,8-hexaclorodibenzofurano	1,2,3,4,7,8-HxCDF
1,2,3,6,7,8-hexaclorodibenzofurano	1,2,3,6,7,8-HxCDF
2,3,4,6,7,8-hexaclorodibenzofurano	2,3,4,6,7,8-HxCDF
1,2,3,7,8,9-hexaclorodibenzofurano	1,2,3,7,8,9-HxCDF
1,2,3,4,6,7,8-heptaclorodibenzofurano	1,2,3,4,6,7,8-HpCDF
1,2,3,4,7,8,9-heptaclorodibenzofurano	1,2,3,4,7,8,9HpCDF
octaclorodibenzofurano	OCDF

Fonte: Adaptado de OMS, 2006

Dentro desse subconjunto de 17 congêneres, o de maior destaque é o congêneres que possui átomos de cloro nas posições 2, 3, 7, 8 (2,3,7,8-tetracloro-para-dioxina (2,3,7,8-TCDD)), (figura 2). A 2,3,7,8-TCDD é classificada pela Agência Internacional de Pesquisa em Câncer (IARC, do inglês *International Agency for Research on Cancer*) no grupo 1, como carcinogênico à seres humanos (quando há evidências suficientes). Outro congêneres também classificado no grupo 1 é o 2,3,4,7,8-pentaclorodibenzofurano (2,3,4,7,8 PeCDF). Todos os outros congêneres 2,3,7,8 substituídos, no momento, são classificados no grupo 3, quando as evidências não são adequadas para se concluir sobre a carcinogenicidade ou não foram classificados (IARC, 2012).

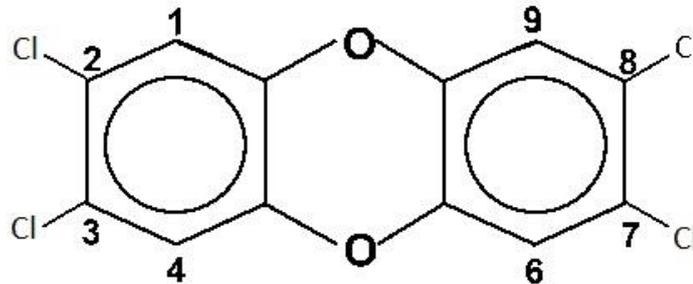


Figura 2: Estrutura química da 2,3,7,8-tetraclorodibenzo-p-dioxina

A ocorrência de dioxinas e furanos nunca se dá por apenas um congêneres exclusivo e sim por uma mistura complexa com diversos congêneres e outros compostos. Diante dessa situação, surge a necessidade de se avaliar a toxicidade geral em relação às dioxinas e furanos. Devido à essa necessidade de se avaliar a toxicidade desses 17 congêneres foi introduzido o conceito do fator de equivalência tóxica (TEF). O TEF apresenta um método útil de se estimar a toxicidade geral de uma mistura em um único valor chamado toxicidade equivalente (TEQ), conforme representado na equação 1 (RAPPE, 1993).

$$TEQ = \sum (\text{concentração congêneres} \times \text{TEF}) \quad (1)$$

Vários valores de TEF para dioxinas e furanos foram sugeridos pelos órgãos competentes, sendo que o utilizado atualmente é o proposto pela Organização Mundial de Saúde (OMS) (tabela 2). Uma vez que o 2,3,7,8-TCDD é considerado o congêneres mais tóxico, sempre a ele é atribuído o valor de TEF igual a 1 e aos outros congêneres uma fração desse valor (BERG et al, 2006).

Tabela 2. Valores de TEFs (Fatores de Equivalência Tóxica) estabelecidos pela Organização Mundial da Saúde (OMS) para dibenzo-p-dioxinas (PCDDs), dibenzofuranos (PCDFs) (Fonte: Adaptado de OMS, 2006).

Composto	TEF
PCDDs	
2,3,7,8-TCDD	1
1,2,3,7,8-PeCDD	1
1,2,3,4,7,8-HxCDD	0,1
1,2,3,6,7,8-HxCDD	0,1
1,2,3,7,8,9-HxCDD	0,1
1,2,3,4,6,7,8-HpCDD	0,01

OCDD	0,0003
PCDFs	TEF
2,3,7,8-TCDF	0,1
1,2,3,7,8-PeCDF	0,03
2,3,4,7,8-PeCDF	0,3
1,2,3,4,7,8-HxCDF	0,1
1,2,3,6,7,8-HxCDF	0,1
1,2,3,7,8,9-HxCDF	0,1
2,3,4,6,7,8-HxCDF	0,1
1,2,3,4,6,7,8-HpCDF	0,01
1,2,3,4,7,8,9-HpCDF	0,01
OCDF	0,0003

Dioxinas e furanos são de formação principalmente antropogênica, entretanto nunca foram utilizados como produtos comerciais, tampouco foram fabricados de maneira intencional senão para serem utilizados em laboratórios. Em relação ao seu processo de formação, as rotas podem ser divididas em duas categorias. Eles podem ser subprodutos de processos térmicos, resultados de uma combustão incompleta, ou subprodutos em alguns processos industriais como por exemplo a síntese de alguns agrotóxicos ou pigmentos (TOOLKIT, 2013).

Processos de combustão incompleta, em que se tem a probabilidade de formação de dioxinas e furanos, se caracterizam por ocorrerem em temperaturas que variam entre 200 °C e 900 °C, e serem pobres em teor de oxigênio, além da evidente presença de cloro (de FRE *et al.*, 1989). Dois mecanismos de formação são sugeridos para a formação de PCDD e PCDF em reações de combustão:

- o mecanismo via síntese “*de novo*” ocorre através da queima da matéria carbonácea, com oxidação e cloração simultâneas. As substâncias que participam dessa reação não estão correlacionadas entre si (moléculas simples contendo carbono, oxigênio, cloro);

- o mecanismo via síntese de precursores se caracteriza pela decomposição térmica e rearranjo molecular dos compostos aromáticos (TOOLKIT, 2013).

Ambos os mecanismos são extremamente afetados pela presença de centros metálicos (MOKHTAR *et al.*, 2014).

Processos industriais de síntese, em que há a probabilidade de formação destes contaminantes se caracterizam por temperaturas superiores à 150 °C, condições

alcalinas, presença de catalizadores metálicos e o uso de radiação ultravioleta ou outros iniciadores radicalares (TOOLKIT, 2013).

Segundo o guia para a padronização e quantificação das emissões de dioxinas e furanos (TOOLKIT, 2013), do Programa das Nações Unidas para o Meio Ambiente (PNUMA – Programa das Nações Unidas para o Meio Ambiente), as fontes de emissão de dioxinas e furanos podem ser divididas nas seguintes categorias:

a) *incineração de lixo*

Destaque para as emissões dos incineradores de lixo municipais, de lixo médico, de resíduos industriais, de resíduos perigosos, de biomassa e de carcaças de animais;

b) *metalurgia primária e secundária de metais ferrosos e não ferrosos*

Notoriedade para as emissões da metalurgia primária e secundária para a produção de ferro, aço, cobre, alumínio, chumbo, zinco, magnésio, latão, bronze, outros metais não ferrosos e coque;

c) *geração de calor e energia (termoelétricas)*

Ênfase para as emissões provenientes de termoelétricas a base de combustível fóssil, termoelétricas a base de combustível renovável e aquecimento doméstico;

d) *produção de produtos minerais*

Destaque para as emissões da produção de cimento, cal, tijolo, vidro, cerâmica, misturas asfálticas e processamento de óleo de xisto;

e) *transporte*

Destaque para as emissões de motores 2 tempos, motores 4 tempos e motores a diesel;

f) *processos de queimas sem controle*

Destaque para as emissões de queimas “abertas” e sem controle de biomassa e lixo;

g) *produção e uso de químicos e bens de consumo*

Notoriedade para as emissões da produção de polpa e papel, produtos químicos clorados orgânicos (agrotóxicos), produção de pigmentos, produção de bifenilas policloradas entre outros;

h) *“miscelânea”*

Ênfase em processos de secagem de biomassa, crematórios e processos de defumação;

i) *descarte e aterro sanitário*

Ênfase em descarte de resíduos potencialmente contaminados e descartados em aterros sanitários despreparados para esta finalidade;

j) *sítios contaminados*

Destaque para locais de aplicação de agrotóxicos contaminados, locais de produção de produtos químicos potencialmente contaminados entre outros.

As figuras 3 e 4 ilustram o perfil médio de congêneres das emissões de algumas fontes de contaminação.

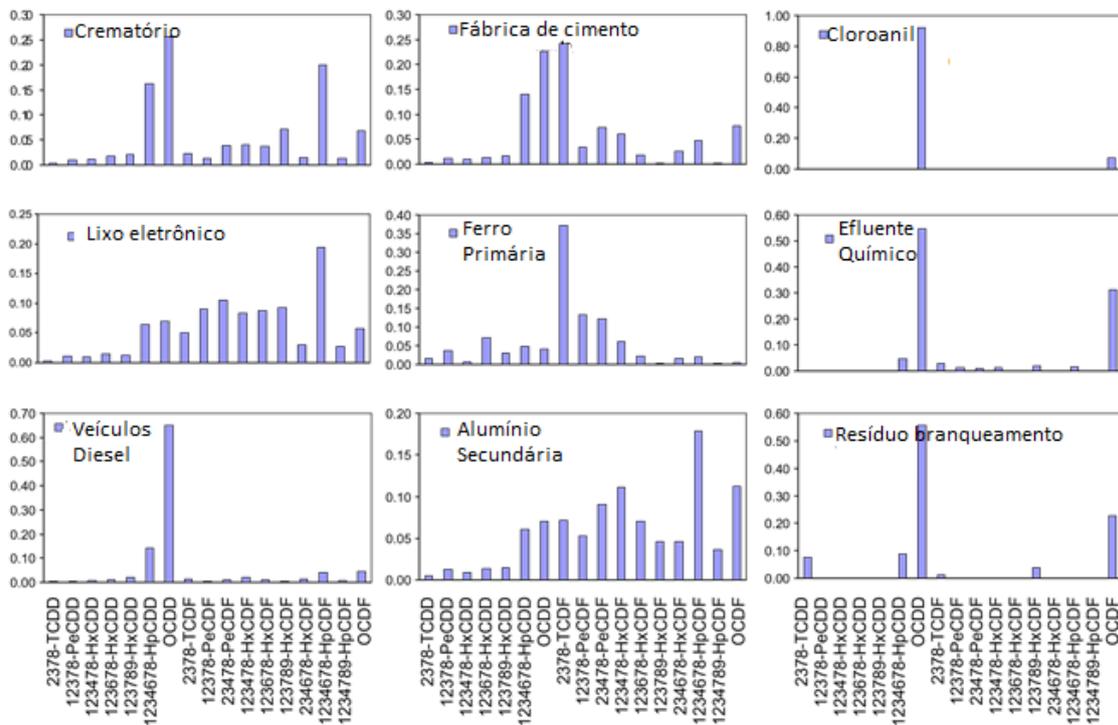


Figura 3: Perfil de congêneres médio de emissões de crematórios, fábricas de cimento, sínteses de cloroanil, queimas de lixo eletrônico, fundições primárias de ferro, efluentes químicos, emissões de veículos com motores a diesel, fundição secundária de alumínio e resíduos de branqueamento de papel (Adaptado de ZHU et al., 2008).

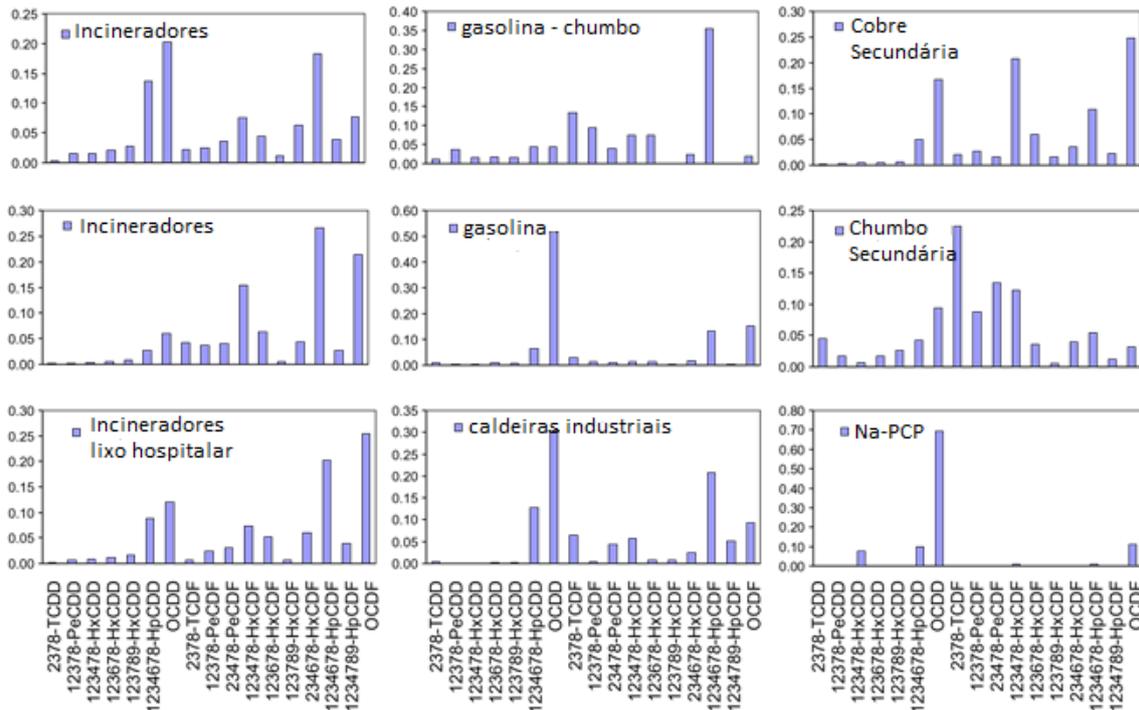


Figura 4: Perfil de congêneres médio de emissões de incineradores, de incineradores de lixo hospitalar, emissões de veículos que utilizam gasolina sem aditivos e gasolina aditivada com compostos de chumbo, emissões de metalurgia secundária de cobre e de chumbo, emissões de caldeiras industriais e da contaminação do pentaclorofenol (Adaptado de ZHU et al., 2008).

### 3.1.2 Rotas de exposição de dioxinas aos alimentos

Incidentes de contaminação de alimentos com dioxinas e furanos são uma ameaça à saúde humana e à economia de um país, principalmente às cadeias em que processos agropecuários estão envolvidos. Medidas de controle, que podem variar de bloqueio temporário de instalações ao recall de produtos no mercado, podendo ter um severo impacto financeiro, além de contribuírem negativamente para a imagem de um produtor ou país no comércio exterior (ALCOSER *et al.*, 2011). A tabela 3 mostra alguns dos eventos de contaminação que tiveram destaque na literatura disponível e tiveram a origem da contaminação identificada. Estimativas, do evento de contaminação ocorrido na Bélgica, em 1999, devido a reutilização de gordura contaminada na produção de ração para frangos e porcos, indicam um prejuízo total de aproximadamente 4 bilhões de euros (MALISCH, 2017).

Tabela 3: Ano, local, matriz afetada, origem da contaminação e referência de alguns eventos de contaminação por dioxinas.

Ano	Local	Matriz afetada	Origem contaminação	Referência
1994	EUA	frango	argila	HOLCOMB et al., 1999
1998	Brasil	manteiga/leite	polpa cítrica	MALISCH, 2000
1999	Bélgica	frango	gordura	COVACI et al., 2002
2007	Índia	goma guar	pentaclorofenol	MALISCH, 2017
2008	Irlanda	porco	combustível	TLUSTOS, 2009
2008	Chile	porco	óxido de zinco	KIM <i>et al.</i> , 2011
2010	Alemanha	ovo/frango	ácido graxo	WEBER et al., 2011

Níveis de dioxinas e furanos em alimentos podem ser correlacionados com a contaminação local ou a certos processos produtivos, contudo padrões de bioacumulação de PCDD e PCDF podem depender do congênera, espécie e tecido (WEBER *et al.*, 2018).

A principal fonte de contaminação de dioxinas e furanos em alimentos consiste na ingestão de gêneros destinados a alimentação animal contaminados pelo animal gerador do alimento. Eventos de contaminação em produtos destinados à alimentação animal e derivados podem ser pertinentes ao uso de aditivos minerais, farinhas diversas (ossos, penas, sangue, peixe), processos térmicos de secagem direta e uso de materiais contaminados. O reaproveitamento de resíduos orgânicos de origem vegetal e animal, estratégia usada como alternativa por fabricantes de rações e derivados, para manter a competitividade no mercado pode ser uma possível origem de contaminação. Exemplos de uso de resíduos contaminados com dioxinas e furanos que originaram grandes eventos de contaminação são os referentes à Bélgica, em 1999 (COVACI et al., 2002), em que gordura contaminada foi reutilizada para produzir ração para frangos e porcos, além do evento do Chile, em 2008 em que óxido de zinco produzido a partir de sucata (latão) foi utilizado como aditivo mineral em ração para porcos (KIM *et al.*, 2011).

Devido às suas propriedades lipofílicas e estabilidade química, que propiciam o seu acúmulo em níveis superiores da cadeia alimentar, a maior rota de exposição de seres humanos a dioxinas e furanos é o consumo de alimentos contaminados com alto teor de gordura, como por exemplo laticínios, ovos e carnes. Estimativas indicam que cerca de 90% da exposição de seres humanos, a esse contaminante, seja devido ao consumo de alimentos contaminados (MALISCH, 2000).

A avaliação de eventos de contaminação em frangos caipiras e ovos, esses gêneros alimentícios são particularmente propensos às condições ambientais, principalmente devido à ingestão de solo durante a alimentação do animal. Estimativas apontam para uma ingestão média entre 11 até 30 gramas de solo por dia. Contaminações do solo que superem 0,4 ng TEQ/Kg de massa seca (aproximadamente 50% do nível máximo da União Europeia para alimentos) teria potencial para exceder o limite de PCDD e PCDF em ovos (WEBER *et al.*, 2018). Outros dois fatores que devem ser avaliados são a contaminação da maravalha utilizada e a contaminação da ração e derivados. A figura 5 indica as possibilidades de contaminação de frangos e ovos.

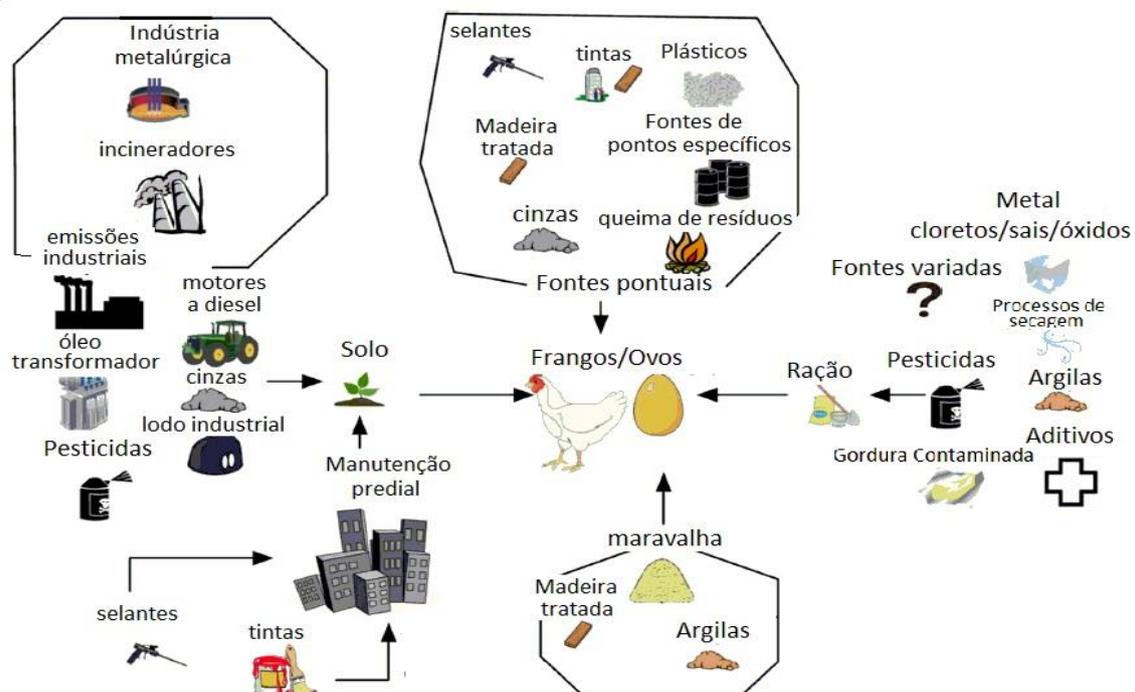


Figura 5: Possíveis origens de contaminação por PCDD/PCDF. (Adaptado de WEBER *et al.*, 2018)

Para proteger a cadeia de suprimentos de alimentos de uma possível contaminação e, por consequência, reduzir a exposição da população a esses contaminantes, várias medidas devem ser adotadas, sendo que o objetivo principal de todas é evitar a entrada desses contaminantes na cadeia alimentar. Em caso de um evento de contaminação, que propicie a entrada desse contaminante na cadeia alimentar, a origem da contaminação, sempre que possível, deve ser identificada, mitigada o quanto antes e, quando possível, erradicada. Processos que possuem a potencialidade de produzir a Contaminação devem ser evitados, de modo a proteger a cadeia alimentar. (PNUMA, 2018)

A estratégia das autoridades brasileiras em segurança alimentar para a garantia da inocuidade de alimentos ofertada ao consumo, quanto à presença de resíduos e contaminantes alimentares é o Plano Nacional de Controle de Resíduos e Contaminantes em Produtos de Origem Animal (PNCRC Animal), do Ministério da Agricultura, Pecuária e Abastecimento (MAPA) (MAPA, 1999). O PNCRC Animal visa, entre outros, monitorar o teor de dioxinas, furanos nas principais matrizes de alimentos de origem animal, produzidos no Brasil quanto à contaminação. Os limites de referência adotados para as matrizes analisadas no PNCRC Animal, quanto à análise de dioxinas e furanos estão listados na tabela 4.

Tabela 4: Limites de referência adotados para as matrizes analisadas no Plano Nacional de Controle de Resíduos Contaminantes quanto à contaminação por PCDD, PCDF (Fonte: Adaptado de MAPA, 2016).

Limites de Referência	
Matriz	Somatório de dioxinas e furanos (pg PCDD/PCDF TEQ- OMS 2005/g de gordura ou g de amostra)
Gordura de Bovinos	2,5
Gordura de Suínos	1,0

Matriz	Somatório de dioxinas e furanos (pg PCDD/PCDF TEQ- OMS 2005/g de gordura ou g de amostra)
Gordura de Aves	1,75
Ovos	2,5
Leite	2,5
Pescado*	3,5

\*Pescado de Captura e Cultivo

Vários estudos sobre os padrões de origem de contaminação, nos últimos anos, foram realizados, no entanto a maioria deles concentram as análises de padrões na área ocupacional/ambiental (WENNING *et al.*, 1993; HAGENMAIER *et al.*, 1994; RAPPE, 1994; RUIZ *et al.*, 2002; ZHU *et al.*, 2008; LIU *et al.*, 2015). Alguns autores demonstraram a aplicação de ferramentas quimiométricas exploratórias (PCA e HCA) para a identificação da origem do evento de contaminação de alimentos e de gêneros destinados à alimentação animal (ALCOCK *et al.*, 2002; HOOGENBOOM *et al.*, 2016; MALISH *et al.*, 2017). Entretanto esses estudos não aprofundam nas conclusões que podem ser retiradas dos modelos gerados. Informações como qual congênere ou conjunto de congêneres são importantes para a identificação de uma determinada fonte de contaminação não são amplamente debatidas, por exemplo.

### 3.2 Quimiometria

A procura por regularidades e padrões, em várias áreas como química, biologia, geologia, foi objetivo e continua norteando vários trabalhos. Nesse contexto ferramentas de estatística multivariada vem ganhando destaque nos últimos anos em que se verifica um crescente número de artigos que citam o uso de distintas ferramentas. O uso de ferramentas de estatística multivariada, que permitem a análises de várias variáveis simultaneamente, possibilitando a extração de

informações de um conjunto de dados, na Química, é denominado como quimiometria (MINGOTI, 2005).

Os métodos quimiométricos utilizados para reconhecer as semelhanças e as diferenças em um conjunto com vários tipos de amostras, agrupando-as e classificando-as são divididos basicamente em dois grupos: métodos supervisionados e métodos não supervisionados.

Nos métodos supervisionados, existe uma informação sobre cada amostra analisada pertencer ou não pertencer a uma classe pré-estabelecida. Essa informação é utilizada na construção dos modelos. O objetivo básico dos métodos supervisionados é identificar em qual grupo uma amostra desconhecida seria classificado (FERREIRA, 2015).

Ao contrário dos métodos supervisionados, a informação sobre as classes (ou rótulos) dos dados não são utilizados ou não está disponível. O objetivo é encontrar agrupamentos naturais para amostra analisada.

O uso de ferramentas quimiométrica requer uma organização dos dados em tabelas ou matrizes, geralmente linhas representam as amostras e as colunas representam as variáveis (ALMEIDA, 2015).

Na Tabela 5, são listados alguns trabalhos recentes na área de alimentos e que fizeram uso de ferramentas quimiométricas para atingir o objetivo desejado.

Tabela 5: Trabalhos recentes na área de alimentos com o uso de ferramentas quimiométricas

MATRIZ	OBJETIVO	REFERÊNCIA
Ácidos graxo - peixe	Estimar o teor de dioxinas	Liu et al (2018)
Mel	Discriminar origem geográfica	Karabagias et al (2017)
Ameixa	Discriminar variedades	Jakubíková et al (2018)
Carne/peixe	Monitoramento da qualidade e autenticidade	Hassoun et al (2019)
Whisky	Identificação de Fraudes	Martins et al (2017)

MATRIZ	OBJETIVO	REFERÊNCIA
Farelo de Arroz	Discriminação entre farelo de arroz branco e farelo de arroz vermelho	Sabir et al (2017)

### 3.2.1 Análise de componentes principais (PCA)

A análise de componentes principais é um método para projeção dos dados multivariados em um espaço de dimensão menor, reduzindo, portanto, a dimensionalidade do espaço original do conjunto de dados, sem que as relações entre as amostras sejam afetadas (FERREIRA, 2015). O objetivo principal é o de explicar a estrutura de variância e covariância de um vetor aleatório, composto de  $p$ -variáveis aleatórias, através da construção de combinações lineares das variáveis originais. Estas combinações lineares são chamadas de componentes principais (CP) e são ortogonais entre si, portanto não são correlacionadas, ou seja, a informação contida em uma componente principal não se encontra em outra (MINGOTI, 2005).

Do ponto de vista matemático, a PCA pode ser descrito como a decomposição da matriz de dados originais  $\mathbf{X}$  ( $m \times n$ ) ( $m$  amostras e  $n$  variáveis), em duas matrizes, uma de escores (“scores”)  $\mathbf{T}$ , com  $m$  linhas e  $a$  colunas (números de componentes principais, novas coordenadas dentro do novo sistema de eixos) e uma matriz ortonormal de pesos  $\mathbf{P}$  (“loadings”) com  $a$  linhas e  $n$  colunas (informação da contribuição de cada variável original na formação dos novos eixos) (ALMEIDA 2015), de modo que, como descrito na equação 2:

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} \quad (2),$$

O diagrama ilustra a decomposição da matriz  $\mathbf{X}$  em  $\mathbf{T}$ ,  $\mathbf{P}^T$  e  $\mathbf{E}$ . A matriz  $\mathbf{X}$  é representada por um retângulo com  $n$  linhas e  $m$  colunas. À sua direita, um sinal de igualdade precede a matriz  $\mathbf{T}$  (retângulo com  $n$  linhas e  $a$  colunas). Segue-se a matriz  $\mathbf{P}^T$  (retângulo com  $a$  linhas e  $m$  colunas), precedida por um símbolo de multiplicação implícito. Um sinal de adição (+) precede a matriz  $\mathbf{E}$  (retângulo com  $n$  linhas e  $m$  colunas).

Figura 6: Representação gráfica da análise de componentes principais (adaptado de ALMEIDA 2015).

Onde **E** é uma matriz ( $m \times n$ ) que contém a informação não modelada, constituída principalmente de variações aleatórias, tais como erros experimentais. Pode ser observado assim a separação da estrutura inerente dos dados do “ruído” experimental (FERREIRA, 2015).

Considerando uma análise exploratória, a informação descrita na matriz dos escores fornece a relação das CPs correlacionadas às amostras enquanto a informação descrita na matriz de pesos fornece a mesma relação correlacionada às variáveis. Os gráficos de escores e pesos fornecem informações sobre o padrão de associação dos dados, permitindo observar agrupamentos de objetos ou variáveis e a caracterização de “outliers” (amostras mal ajustadas ao modelo) (SENA et al., 2000; BARROS NETO *et al.*, 2006).

### **3.2.2 Análise Discriminante por Mínimos Quadrados Parciais (PLS-DA)**

O PLS-DA é um método discriminante que adapta a técnica regressão por Mínimos Quadrados Parciais (PLS do inglês partial least squares) a uma tarefa de classificação supervisionada, estabelecendo uma relação entre uma matriz de variáveis independentes (**X**) e uma matriz de variáveis dependentes (**Y** ou **y**). A matriz de variáveis dependentes contém uma variável binária/categórica que indica a classe a que cada amostra pertence, sendo que 1 indica “pertencente” e 0 “não pertencente” (SANTANA et al., 2020)

O PLS é um método de regressão multivariada que faz uso da PCA para a redução da dimensionalidade do conjunto de dados para uma correlação linear multivariada entre a matriz de dados originais, **X** (resposta analítica), e uma propriedade de interesse, matriz **Y**.

Um fato importante a se destacar é que apesar do PLS fazer uso do PCA para reduzir a dimensionalidade dos dados e assim evitar problemas como a colinearidade entre as linhas, retirar informação irrelevante como ruído e informação aleatória e trabalhar com dados com um número elevado de variáveis, os escores e pesos obtidos em cada método não são idênticos. Para ser ter a maior explicação da variância em

**X** e a maior correlação com **Y** existe uma perda da ortogonalidade das componentes principais para otimizar a previsão das variáveis dependentes. A combinação de variáveis que irão gerar o novo espaço em que as amostras são projetadas deixa de ser chamada de componentes principais e passa a ser chamada de variáveis latentes (FERREIRA et al., 1999; SENA et al., 2000; BRERETON, 2009)

No PLS, as matrizes **X** e **Y** são decompostas simultaneamente por meio das equações descritas em (3) e (4).

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} \quad (3)$$

$$\mathbf{Y} = \mathbf{UQ}^T + \mathbf{F} \quad (4)$$

**T** e **U** são as matrizes dos escores, **P** e **Q** são as matrizes dos pesos (loadings); **E** e **F** são os resíduos de **X** e **Y**, respectivamente (SOUZA et al, 2013). A figura 7 mostra graficamente a decomposição das matrizes **X** e **Y**

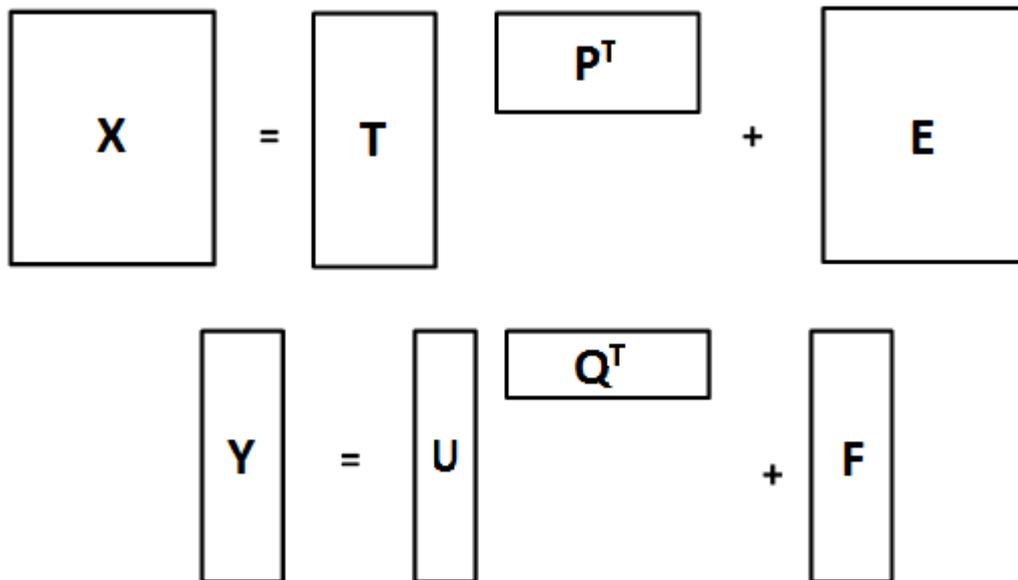


Figura 7: Decomposição das matrizes **X** e **Y** para elaboração do modelo PLS (adaptado de ALMEIDA, 2015).

A equação (5) correlaciona as matrizes **X** e **Y** de forma linear

$$u_h = b_h \cdot t_h \quad (5)$$

para “h” variáveis latentes. Os valores de  $b_h$  são agrupados na matriz diagonal  $\mathbf{B}$ , que contém os coeficientes de regressão entre as matrizes de escores  $\mathbf{T}$  de  $\mathbf{X}$  e  $\mathbf{U}$  de  $\mathbf{Y}$ . A melhor relação linear possível entre os escores das duas matrizes é alcançada a partir de pequenas rotações das variáveis latentes das matrizes  $\mathbf{X}$  e  $\mathbf{Y}$ . A matriz  $\mathbf{Y}$  pode ser calculada de  $u_h$ , como descrito na equação 6:

$$\mathbf{Y} = \mathbf{TBQ}^T + \mathbf{F} \quad (6)$$

Os valores preditos para novas amostras  $\hat{\mathbf{Y}}$  podem ser então calculados de acordo com a equação, com base nos seus escores  $\mathbf{T}^*$ , substituídos na equação anterior.

$$\hat{\mathbf{Y}} = \mathbf{T}^* \mathbf{BQ} \quad (7)$$

Uma etapa importante para a construção de um modelo PLS-DA é a determinação do número adequado de variáveis latentes. Um número demasiado de variáveis latentes pode indicar um sobreajuste do modelo, em que há a inclusão de informação irrelevante (aleatória) ao modelo. O modelo provavelmente irá funcionar bem em prever o comportamento das amostras contidas na matriz de dados original, no entanto existirá a grande probabilidade em falhar na previsão de novas amostras. Ao contrário, o subajuste é a escolha diminuta de variáveis latentes, podendo deixar informação relevante de fora da construção do modelo. A escolha do número de variáveis latentes é comumente feita através do procedimento de validação cruzada, onde um determinado subconjunto de amostras, do conjunto originais de dados, é, sistematicamente ou aleatoriamente, retirado do modelo e previstas no modelo com as amostras resultantes. Para calcular a eficiência do procedimento com diferentes números de variáveis latentes, no modelo em desenvolvimento, pode ser utilizado como critério o número de variáveis latentes que retornam o menor valor de erro de classificação de validação cruzada (ECVC) (FERREIRA, et al. 1999).

Existem duas variantes do método PLS-DA, conhecidos com PLS1-DA e PLS2-DA. No método PLS1-DA, cada classe em  $\mathbf{y}$  é modelada individualmente. No modelo PLS2-DA, caso haja mais de duas classes na matriz  $\mathbf{Y}$ , todas elas são modeladas conjuntas, sendo obtido valores únicos para escores e variáveis latentes. Nessa

dissertação será utilizado apenas a variante PLS1-DA, portanto o indicador “1” será omitido (BRERETON, 2009).

O valor obtido,  $y$  previsto, pelo modelo PLS-DA será um número não exatamente 0 ou 1 (conforme a construção da matriz  $Y$ ). Dessa forma é necessário estabelecer valores limites entre as classes, esses valores são estimados empregando o Teorema de Bayes. Esse valor é estabelecido de forma a minimizar o número de resultados de falsos positivos e falsos negativos.

A avaliação do desempenho do modelo de classificação será realizada através das figuras de mérito taxa de falso positivo, taxa de falso negativo, sensibilidade e especificidade (GODIM et al., 2011).

A taxa de falso positivo é a probabilidade de uma amostra não pertencente à classe ser classificada como pertencente à classe. A razão de falso positivo é calculada pela equação 8:

$$\text{taxa de falso positivo} = \frac{fp}{vn+fp} \times 100 \quad (8) ,$$

em que  $fp$  é o número de amostras falso positivas e  $vn$  é o número de amostras verdadeiras negativas.

De maneira análoga, a taxa de falso negativo é a probabilidade de uma amostra pertencente à classe ser classificada como não pertencente à classe. A razão de falso positivo é calculada pela equação 9:

$$\text{taxa de falso negativo} = \frac{fn}{vp+fn} \times 100 \quad (9) ,$$

em que  $fn$  é o número de amostras falso negativas e  $vp$  é o número de amostras verdadeira positivas. Equações similares podem ser usadas para calcular sensibilidade e especificidade. A sensibilidade é a habilidade de o modelo classificar corretamente as amostras positivas e pode ser calculada de acordo com a equação abaixo:

$$\text{Sensibilidade} = \frac{vp}{vp+fn} \quad (10) .$$

A especificidade é a capacidade do modelo em identificar corretamente as amostras negativas, e é calculada de acordo com a equação 11:

$$\text{Especificidade} = \frac{vn}{vp+fp} \quad (11).$$

## 4. Materiais e Métodos

### 4.1 Metodologia Analítica

Protocolos de análises de dioxinas geralmente são baseados no método EPA 1613, revisão B, que trata da determinação de dioxinas e furanos em matrizes como água, solo, sedimentos e tecidos por cromatografia de alta resolução acoplada com espectrometria de massas de alta resolução. Procedimentos específicos podem possuir pequenas variações, dependendo do laboratório em que são realizados e com qual matriz se está trabalhando, no entanto, a análise consiste basicamente em etapas de extração de gordura, uma vez que os analitos são lipofílicos, e etapas de limpeza/purificação (clean-up) para a remoção de interferentes e fracionamento. Etapas de concentração dos extratos, por meio de evaporação do solvente, são frequentes e tem como objetivo aumentar a concentração dos analitos que geralmente estão na ordem de pg/g. Volumes dos extratos finais podem variar entre 10 e 20  $\mu\text{L}$ . Procedimentos para extração da fase lipídica geralmente fazem uso de solventes como tolueno, diclorometano e hexano. Para o caso de análises de emissão, são utilizados cartuchos de adsorção adequados, sendo que o solvente é aplicado no cartucho para a extração dos analitos adsorvidos. As etapas de clean-up do extrato podem ser duas ou três, dependendo da complexidade da matriz e da presença de interferentes.

No início de toda análise existe a adição de padrão interno, isotopicamente marcado, às amostras antes das etapas de limpeza. A adição de padrão interno isotopicamente marcado com carbono de massa 13 possui duas finalidades. Adicionado ao início da etapa analítica permite a compensação de perdas e erros ao longo de todo o processo analítico. Dessa forma, se ocorre perda do analito alvo, a perda do padrão interno será da mesma ordem de magnitude, já que ambos foram submetidos ao mesmo tratamento. Como a quantidade do padrão interno adicionada a amostra é conhecida, o percentual de recuperação pode ser calculado e utilizado como uma medição indireta da recuperação do analito alvo. A segunda finalidade do padrão interno é a diluição isotópica que consiste na modificação da composição isotópica natural do analito alvo, presente na amostra, pela adição de uma quantidade

conhecida de um análogo isotopicamente marcado. A confirmação dos analitos se dá pelo tempo de retenção na análise cromatográfica e pela razão isotópica específica.

A análise cromatográfica é realizada frequentemente em um cromatógrafo gasoso e geralmente são utilizadas colunas capilares de polaridade baixa com dimensões de 60 m de comprimento por 25 mm de espessura e filme interno de 0,25 µm. Espectrômetro de massas de alta resolução são utilizados com uma resolução mínima de 10000 ppm a 10% no vale. São operados no modo impacto de elétrons (EI), com ionização positiva, com energia de 35 eV.

Perfis fornecidos pelo LDP, foram obtidos por meio de métodos acreditados segundo requisitos da norma ISO 17025. A extração da fase lipídica, ocorreu de acordo com qual matriz se está trabalhando. Amostras de ovos e gêneros destinados à alimentação animal utilizam extração líquida pressurizada. Amostras de gordura de aves (frango) são derretidas em forno micro-ondas. Três etapas de clean-up foram realizadas e o extrato final é de 20 µL. A detecção foi realizada em um equipamento de cromatografia gasosa acoplada a um espectrômetro de massas de alta resolução com analisador de setor magnético.

## **4.2 Amostras (Perfis de Contaminação)**

### **4.2.1 Estrutura da matriz de dados**

Para a elaboração do modelo de classificação dos perfis de contaminação por dioxinas e furanos quanto à possível origem, foram compilados na literatura existente 557 perfis de contaminação com os valores das concentrações dos dezessete congêneres monitorados. Valores menores que o limite de quantificação (< LQ) foram tomados como zero, para evitar que características específicas do método de análise pudessem interferir no modelo. Devido ao fato de que algumas referências bibliográficas reportam os resultados de alguns congêneres como a soma destes, para efeitos de padronização, os resultados compilados foram reunidos de acordo com o número de cloros na molécula do contaminante. Para manter a proporção entre os perfis compilados e evitar que resultados de uma mesma contaminação fossem avaliados como contaminações distintas, devido ao grau de exposição à uma

determinada fonte, todos os perfis foram normalizados fazendo com que o maior resultado de concentração tivesse o valor de 100 e os outros resultados uma proporção relativa desse valor. Além disso, foi determinada a proporção entre os resultados de dioxinas e furanos.

Em relação à toxicidade equivalente, uma vez que existem algumas diferenças entre os valores de TEF em relação à congêneres com o mesmo número de átomos de cloro, foram determinadas mais duas variáveis: TETRAS-TEQ e PENTAS-TEQ, que correspondem ao somatório TEQ em relação às moléculas com quatro átomos de cloro e às moléculas com cinco átomos de cloro respectivamente. Os somatórios em relação às moléculas com seis, sete e oito átomos de cloros não foram utilizados pois não existe diferença dos valores de TEF entre moléculas com o mesmo número de cloro. O mesmo procedimento, para normalização dos valores em percentuais, foi realizado com os valores em TEF.

De forma complementar foram determinadas mais 3 variáveis, sendo elas o somatório TEQ devido às dioxinas, o somatório TEQ devido aos furanos e o somatório TEQ total. Para o cálculo dos somatórios, no caso de perfis de emissão, foi considerado que 1 Nm<sup>3</sup> ou 1 m<sup>3</sup> de qualquer emissão teria aproximadamente a massa de 1 quilograma. A tabela inicial de dados consistia inicialmente de 557 perfis (linhas) e 12 variáveis (colunas). Na Tabela 6, são listadas as variáveis obtidas a partir dos resultados compilados e as respectivas descrições.

Tabela 6: Tabela com as variáveis utilizadas e a sua respectiva descrição

VARIÁVEL	DESCRIÇÃO
TETRAS %	Somatório das concentrações de 2378 TCDD e 2378 TCDF – Normalizado
PENTAS %	Somatório das concentrações de 12378 PeCDF, 23478 PeCDF e 12378 PeCDD – Normalizado
HEXAS %	Somatório das concentrações de 123478 HxCDF, 123678 HxCDF, 234678 HxCDF, 123789 HxCDF, 123478 HxCDD, 123678 HxCDD, 123789 HxCDD – Normalizado
HEPTAS %	Somatório das concentrações de 1234678 HPCDF, 1234789 HpCDF e 1234678 HxCDD – Normalizado

VARIÁVEL	DESCRIÇÃO
OCTAS %	Somatório das concentrações de OCDF e OCDD – Normalizado
DIOXINAS %	Somatório das concentrações das dioxinas – Normalizado
FURANOS %	Somatório das concentrações dos furanos – Normalizado
TETRAS TEQ	Somatório do produto das concentrações de 2378 TCDD e 2378 TCDF multiplicados pelos respectivos valores de TEF – Normalizado
PENTAS TEQ	Somatório do produto das concentrações de 12378 PeCDD, 12378 PeCDF e 23478 PeCDF multiplicados pelos respectivos valores de TEF - Normalizado
DIOXINAS TEQ	Somatório dos TEQ devido as dioxinas – Normalizado
FURANOS TEQ	Somatório dos TEQ devido aos furanos – Normalizado
TEQ TOTAL	Somatório TEQ devido às dioxinas e furanos

Os 557 perfis obtidos foram filtrados quanto à dois critérios:

- perfis com valor de TEQ TOTAL inferior a 1 foram descartados, pois dificilmente irão consistir na origem de um evento de contaminação. As fontes de emissão de dioxinas e furanos com valor de TEQ TOTAL inferior a 1 provavelmente irão compor o valor de contaminação natural, ou “background”; e
- perfis com no máximo dois valores positivos dentre os dezessete congêneres foram descartados pois o perfil de contaminação provavelmente ficou omitido pelo limite de quantificação da análise.

Os resultados compilados foram classificados em 9 grupos, de acordo com possíveis similaridades entre os processos industriais a que estão submetidos. Na Tabela 7, apresentam-se as classes e os códigos adotados, em que os dados compilados na literatura foram classificados. Após a aplicação dos dois critérios acima, a tabela de dados originais será constituída de 293 perfis (linhas) e 12 variáveis (colunas) classificados em 9 grupos. A tabela com todos os perfis compilados, a respectiva codificação, a classe em que foi classificado, o somatório TEQ TOTAL e a respectiva referência bibliográfica pode ser encontrada no Apêndice 1.

Tabela 7: Tabela de correlação do código adotado e o processo industrial.

CÓDIGO DA CLASSE	DESCRIÇÃO
AGR	Perfis de contaminação em agrotóxicos
PCB	Perfis de contaminação em PCBs comerciais
PGM	Perfis de contaminação em corantes e pigmentos
CIM	Perfis de emissão e contaminação em cimenteiras
MTL	Perfis de emissão e contaminação em processos metalúrgicos de fundição de metais primários e secundários
MIN	Perfis de contaminação de minerais obtidos pela indústria extrativista
CMB	Perfis de emissão de processos de combustão de termoelétricas e motores
INC	Perfis de emissão de incineradores
PLT	Perfis de contaminação e emissão de processos de produção e emissões de queimas de plásticos

Resultados de análises, em que possui o interesse de se classificar, quanto a possível origem de contaminação, foram fornecidos pelo Laboratório de Dioxinas e Furanos do Laboratório Federal de Defesa Agropecuária. Esses dados consistem em resultados de análises de amostras de gêneros alimentícios destinados à alimentação animal, gorduras de aves e ovos (97 amostras). Além disso, foram fornecidas pelo laboratório algumas informações relevantes quanto à localidade de alguns conjuntos de amostras que, a princípio, deveriam constituir a mesma origem de contaminação. Tais informações estão compiladas nas Tabelas 8 e 9. Da literatura, também foram compilados 41 perfis de contaminação com a origem conhecida. As informações relevantes acerca desses perfis se encontram compiladas também na Tabela 8. Esses perfis de contaminação, com origens conhecidas, irão constituir um conjunto de “validação externa” em que se poderá avaliar o desempenho do modelo de classificação quanto a amostras reais. Os perfis das amostras da Tabela 8 também foram normalizados e os valores utilizados para a classificação estão de acordo com a Tabela 6.

Tabela 8: Codificação utilizada, matriz, origem da contaminação e referência bibliográfica de cada amostra a ser classificada.

CÓDIGO AMOSTRA	MATRIZ	PROVÁVEL ORIGEM DE CONTAMINAÇÃO	TEQ TOTAL	REFERÊNCIA BIBLIOGRÁFICA
FF1	polpa cítrica	plástico	5,5	MALISCH, 2000
FF2	polpa cítrica	plástico	7,1	MALISCH, 2000
FF3	polpa cítrica	plástico	5,7	MALISCH, 2000

CÓDIGO AMOSTRA	MATRIZ	PROVÁVEL ORIGEM DE CONTAMINAÇÃO	TEQ TOTAL	REFERÊNCIA BIBLIOGRÁFICA
FF4	polpa cítrica	plástico	9,1	MALISCH, 2000
FF5	polpa cítrica	plástico	5,8	MALISCH, 2000
FF6	polpa cítrica	plástico	7,5	MALISCH, 2000
FF7	polpa cítrica	plástico	7,2	MALISCH, 2000
FF8	polpa cítrica	plástico	8,7	MALISCH, 2000
FF9	polpa cítrica	plástico	5,4	MALISCH, 2000
FF10	polpa cítrica	plástico	5,9	MALISCH, 2000
FF11	polpa cítrica	plástico	14,1	MALISCH, 2000
FF12	polpa cítrica	plástico	4,6	MALISCH, 2000
FF13	polpa cítrica	plástico	5,2	MALISCH, 2000
FF14	polpa cítrica	plástico	7,9	MALISCH, 2000
FF15	gelatina - gordura reciclada	crise dioxinas - Bélgica (PCB)	49,0	HOOGENBOOM et al., 2007
FF16	gelatina - gordura reciclada	crise dioxinas - Bélgica (PCB)	439,8	HOOGENBOOM et al., 2007
FF17	gelatina - gordura reciclada	crise dioxinas - Bélgica (PCB)	7,4	HOOGENBOOM et al., 2007
FF18	ração	crise dioxinas - Bélgica (PCB)	35,4	HOOGENBOOM et al., 2004.
FF19	Ração	óxido de zinco - Chile	13,6	KIM et al., 2011
FF20	premix – ZnO	óxido de zinco - Chile	800,2	KIM et al., 2011
FF21	premix – ZnO	óxido de zinco - Chile	6.673	KIM et al., 2011
FF22	ZnO	óxido de zinco - Chile	17.147	KIM et al., 2011
FF23	ração	crise dioxinas - Bélgica (PCB)	7.566	COVACI et al., 2002
FF24	Ovo	fundição secundária de alumínio	221,4	SQUADRONE et al., 2015
FF25	Ovo	fundição secundária de alumínio	173,3	SQUADRONE et al., 2015
FF26	Ovo	fundição secundária de alumínio	126,4	SQUADRONE et al., 2015
FF27	Ovo	fundição secundária de alumínio	104,0	SQUADRONE et al., 2015
FF28	Biodiesel	clorofenol	123,5	WEBER et al., 2011
FF29	gordura animal	desconhecido	0,7	ELJARRAT et al., 2002
FF30	gordura animal	desconhecido	0,6	ELJARRAT et al., 2002
FF31	gordura animal	desconhecido	0,5	ELJARRAT et al., 2002
FF32	óleo de peixe	desconhecido	1,9	ELJARRAT et al., 2002
FF33	óleo de peixe	desconhecido	2,4	ELJARRAT et al., 2002
FF34	óleo de peixe	desconhecido	2,3	ELJARRAT et al., 2002

CÓDIGO AMOSTRA	MATRIZ	PROVÁVEL ORIGEM DE CONTAMINAÇÃO	TEQ TOTAL	REFERÊNCIA BIBLIOGRÁFICA
FF35	farinha de peixe	desconhecido	2,8	ELJARRAT et al., 2002
FF36	farinha de peixe	desconhecido	2,4	ELJARRAT et al., 2002
FF37	farinha de peixe	desconhecido	1,3	ELJARRAT et al., 2002
FF38	farinha de peixe	desconhecido	2,2	ELJARRAT et al., 2002
FF39	farinha de peixe	desconhecido	7,7	ELJARRAT et al., 2002
FF40	farinha de carne e ossos	desconhecido	2,7	ELJARRAT et al., 2002
FF41	farinha de carne e ossos	desconhecido	7,0	ELJARRAT et al., 2002
LDP1	Aves	desconhecido	1,2	AMOSTRAS LDP - LFDA
LDP2	Ração	desconhecido	1,3	AMOSTRAS LDP - LFDA
LDP3	Aves	desconhecido	3,0	AMOSTRAS LDP - LFDA
LDP4	Aves	desconhecido	3,0	AMOSTRAS LDP - LFDA
LDP5	Ração	desconhecido	1,6	AMOSTRAS LDP - LFDA
LDP6	Ração	desconhecido	19,1	AMOSTRAS LDP - LFDA
LDP7	Aves	desconhecido	1,1	AMOSTRAS LDP - LFDA
LDP8	Aves	desconhecido	1,3	AMOSTRAS LDP - LFDA
LDP9	Aves	desconhecido	2,8	AMOSTRAS LDP - LFDA
LDP10	Aves	desconhecido	1,9	AMOSTRAS LDP - LFDA
LDP11	Aves	desconhecido	1,4	AMOSTRAS LDP - LFDA
LDP12	Aves	desconhecido	1,5	AMOSTRAS LDP - LFDA
LDP13	Aves	desconhecido	1,4	AMOSTRAS LDP - LFDA
LDP14	Aves	desconhecido	2,1	AMOSTRAS LDP - LFDA
LDP15	Aves	desconhecido	1,6	AMOSTRAS LDP - LFDA
LDP16	Aves	desconhecido	3,5	AMOSTRAS LDP - LFDA
LDP17	Aves	desconhecido	2,9	AMOSTRAS LDP - LFDA
LDP18	Aves	desconhecido	1,2	AMOSTRAS LDP - LFDA
LDP19	Aves	desconhecido	1,5	AMOSTRAS LDP - LFDA

CÓDIGO AMOSTRA	MATRIZ	PROVÁVEL ORIGEM DE CONTAMINAÇÃO	TEQ TOTAL	REFERÊNCIA BIBLIOGRÁFICA
LDP20	Aves	desconhecido	1,7	AMOSTRAS LDP - LFDA
LDP21	Aves	desconhecido	1,5	AMOSTRAS LDP - LFDA
LDP22	Aves	desconhecido	1,3	AMOSTRAS LDP - LFDA
LDP23	Aves	desconhecido	1,9	AMOSTRAS LDP - LFDA
LDP24	Aves	desconhecido	1,5	AMOSTRAS LDP - LFDA
LDP25	Aves	desconhecido	1,7	AMOSTRAS LDP - LFDA
LDP26	Ração	desconhecido	2,6	AMOSTRAS LDP - LFDA
LDP27	Ração	desconhecido	1,2	AMOSTRAS LDP - LFDA
LDP28	Ração	desconhecido	22,9	AMOSTRAS LDP - LFDA
LDP29	Ração	desconhecido	7,0	AMOSTRAS LDP - LFDA
LDP30	Ração	desconhecido	5,4	AMOSTRAS LDP - LFDA
LDP31	Ração	desconhecido	5,4	AMOSTRAS LDP - LFDA
LDP32	Aves	desconhecido	1,3	AMOSTRAS LDP - LFDA
LDP33	Ovo	desconhecido	4,5	AMOSTRAS LDP - LFDA
LDP34	Aves	desconhecido	1,6	AMOSTRAS LDP - LFDA
LDP35	Aves	desconhecido	2,1	AMOSTRAS LDP - LFDA
LDP36	Aves	desconhecido	1,8	AMOSTRAS LDP - LFDA
LDP37	Aves	desconhecido	1,1	AMOSTRAS LDP - LFDA
LDP39	Aves	desconhecido	1,4	AMOSTRAS LDP - LFDA
LDP40	Ração	desconhecido	3,1	AMOSTRAS LDP - LFDA
LDP41	Ração	desconhecido	1,4	AMOSTRAS LDP - LFDA
LDP42	Ovo	desconhecido	4,5	AMOSTRAS LDP - LFDA
LDP43	Ovo	desconhecido	4,2	AMOSTRAS LDP - LFDA
LDP44	Ovo	desconhecido	4,2	AMOSTRAS LDP - LFDA
LDP45	Ovo	desconhecido	4,7	AMOSTRAS LDP - LFDA

CÓDIGO AMOSTRA	MATRIZ	PROVÁVEL ORIGEM DE CONTAMINAÇÃO	TEQ TOTAL	REFERÊNCIA BIBLIOGRÁFICA
LDP46	Ovo	desconhecido	4,2	AMOSTRAS LDP - LFDA
LDP47	Ração	desconhecido	9,1	AMOSTRAS LDP - LFDA
LDP48	Aves	desconhecido	1,5	AMOSTRAS LDP - LFDA
LDP49	Ovo	desconhecido	5,4	AMOSTRAS LDP - LFDA
LDP50	Ovo	desconhecido	5,8	AMOSTRAS LDP - LFDA
LDP51	Ovo	desconhecido	5,6	AMOSTRAS LDP - LFDA
LDP52	Ovo	desconhecido	5,4	AMOSTRAS LDP - LFDA
LDP53	Ovo	desconhecido	5,2	AMOSTRAS LDP - LFDA
LDP54	Ração	desconhecido	6,4	AMOSTRAS LDP - LFDA
LDP55	Ração	desconhecido	4,5	AMOSTRAS LDP - LFDA
LDP56	Aves	desconhecido	7,1	AMOSTRAS LDP - LFDA
LDP57	Aves	desconhecido	7,5	AMOSTRAS LDP - LFDA
LDP58	Aves	desconhecido	7,0	AMOSTRAS LDP - LFDA
LDP59	Aves	desconhecido	6,6	AMOSTRAS LDP - LFDA
LDP60	Aves	desconhecido	7,0	AMOSTRAS LDP - LFDA
LDP61	Ração	desconhecido	5,7	AMOSTRAS LDP - LFDA
LDP62	Ração	desconhecido	3,7	AMOSTRAS LDP - LFDA
LDP63	Ração	desconhecido	1,1	AMOSTRAS LDP - LFDA
LDP64	Ração	desconhecido	4,6	AMOSTRAS LDP - LFDA
LDP65	Ração	desconhecido	1,2	AMOSTRAS LDP - LFDA
LDP66	Aves	desconhecido	1,7	AMOSTRAS LDP - LFDA
LDP67	Ração	desconhecido	3,6	AMOSTRAS LDP - LFDA
LDP68	Aves	desconhecido	1,7	AMOSTRAS LDP - LFDA
LDP69	Ração	desconhecido	9,2	AMOSTRAS LDP - LFDA
LDP70	Ração	desconhecido	5,9	AMOSTRAS LDP - LFDA

CÓDIGO AMOSTRA	MATRIZ	PROVÁVEL ORIGEM DE CONTAMINAÇÃO	TEQ TOTAL	REFERÊNCIA BIBLIOGRÁFICA
LDP71	Ração	desconhecido	1,3	AMOSTRAS LDP - LFDA
LDP72	Ração	desconhecido	1,1	AMOSTRAS LDP - LFDA
LDP73	Ração	desconhecido	1,1	AMOSTRAS LDP - LFDA
LDP74	Ração	desconhecido	12,1	AMOSTRAS LDP - LFDA
LDP75	Ração	desconhecido	9,8	AMOSTRAS LDP - LFDA
LDP76	Ração	desconhecido	11,8	AMOSTRAS LDP - LFDA
LDP77	Ração	desconhecido	8,0	AMOSTRAS LDP - LFDA
LDP78	Ração	desconhecido	9,2	AMOSTRAS LDP - LFDA
LDP79	Ração	desconhecido	1,5	AMOSTRAS LDP - LFDA
LDP80	Ração	desconhecido	6,4	AMOSTRAS LDP - LFDA
LDP81	Ração	desconhecido	1,2	AMOSTRAS LDP - LFDA
LDP82	Aves	desconhecido	1,2	AMOSTRAS LDP - LFDA
LDP83	Aves	desconhecido	1,0	AMOSTRAS LDP - LFDA
LDP84	Ovo	desconhecido	2,1	AMOSTRAS LDP - LFDA
LDP85	Ovo	desconhecido	6,1	AMOSTRAS LDP - LFDA
LDP86	Aves	desconhecido	1,9	AMOSTRAS LDP - LFDA
LDP87	Aves	desconhecido	1,4	AMOSTRAS LDP - LFDA
LDP88	Aves	desconhecido	1,0	AMOSTRAS LDP - LFDA
LDP89	Aves	desconhecido	1,3	AMOSTRAS LDP - LFDA
LDP90	Ração	desconhecido	5,8	AMOSTRAS LDP - LFDA
LDP91	Ração	desconhecido	4,3	AMOSTRAS LDP - LFDA
LDP92	Ração	desconhecido	5,6	AMOSTRAS LDP - LFDA
LDP93	Ração	desconhecido	3,3	AMOSTRAS LDP - LFDA
LDP94	Ração	desconhecido	5,3	AMOSTRAS LDP - LFDA
LDP95	Ração	desconhecido	1,1	AMOSTRAS LDP - LFDA

CÓDIGO AMOSTRA	MATRIZ	PROVÁVEL ORIGEM DE CONTAMINAÇÃO	TEQ TOTAL	REFERÊNCIA BIBLIOGRÁFICA
LDP96	Ração	desconhecido	1,2	AMOSTRAS LDP - LFDA
LDP97	Ração	desconhecido	12,3	AMOSTRAS LDP - LFDA

Tabela 9: Localidade referente à cada amostra e a matriz analisada

LOCAL	CÓDIGO AMOSTRA	MATRIZ
A	LDP3	aves
A	LDP4	aves
A	LDP5	ração
A	LDP26	ração
A	LDP29	ração
B	LDP68	aves
B	LDP75	ração
C	LDP56	aves
C	LDP57	aves
C	LDP58	aves
C	LDP59	aves
C	LDP60	aves
D	LDP86	aves
D	LDP87	aves
D	LDP88	aves
D	LDP89	aves
D	LDP90	ração
D	LDP91	ração
D	LDP92	ração
D	LDP93	ração
D	LDP94	ração
E	LDP42	ovo
E	LDP43	ovo
E	LDP44	ovo
E	LDP45	ovo
E	LDP46	ovo
E	LDP49	ovo
E	LDP50	ovo
E	LDP51	ovo
E	LDP52	ovo
E	LDP53	ovo
F	LDP7	aves
F	LDP8	aves
F	LDP10	aves
F	LDP11	aves

LOCAL	CÓDIGO AMOSTRA	MATRIZ
F	LDP12	aves
F	LDP13	aves
F	LDP14	aves
F	LDP15	aves
F	LDP17	aves
F	LDP18	aves
F	LDP19	aves
F	LDP20	aves
F	LDP21	aves
F	LDP22	aves
F	LDP23	aves
F	LDP24	aves
F	LDP25	aves
F	LDP28	ração
G	LDP9	aves
G	LDP16	aves
H	LDP40	ração
H	LDP61	ração
H	LDP62	ração
I	LDP47	ração
I	LDP54	ração
I	LDP55	ração
J	LDP30	ração
J	LDP31	ração

## 5. Resultados

### 5.1 Análise exploratória dos dados compilados da literatura

A desigualdade entre perfis de produção/emissão de contaminantes, dentro de cada grupo, pode ser devido a inúmeros fatores, sendo que, no caso desse modelo, a principal fonte de variabilidade entre os resultados se deve provavelmente às características intrínsecas de cada processo, dentro de cada grupo. Pequenas variações de um processo, como mudanças de fornecedores de insumos ou parâmetros de otimização de produção provavelmente devem produzir pequenas variações dentro do perfil de congêneres. Variações maiores podem significar a combinação de diferentes fontes de emissão, sendo que provavelmente serão observadas como amostras anômalas dentro do grupo. Para avaliar a dispersão dos dados, para cada grupo foi feita uma PCA exploratória. Os

- Análise exploratória do grupo MIN

Da Figura 8, observa-se uma distinção clara, na primeira componente, entre o perfil da caulinita e os perfis referentes à contaminação dos demais minerais, sendo que as variáveis TETRA %, PENTAS %, HEXAS %, HEPTAS %, TETRAS TEQ, PENTAS TEQ e TEQ TOTAL estão correlacionadas com valores positivos da primeira componente principal. Não foi possível, nessa análise exploratória, a distinção dos perfis da montmorilonita, bentonita e argila.

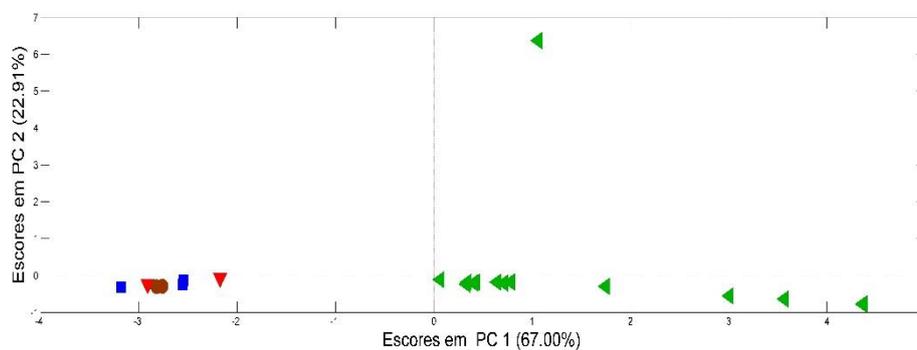


Figura 8: Dispersão das amostras do grupo MIN em duas componentes principais. Entre parênteses, % explicada por cada componente principal. Dados em verde, azul, vermelho e marrom representam caulinita, bentonita, montmorilonita e argila respectivamente.

- Análise exploratória do grupo MTL

Da Figura 9, observa-se uma formação de dois subgrupos, em que se pode notar que os perfis de contaminação oriundos da metalurgia de cobre e magnésio estão correlacionados com valores negativos para a primeira componente principal, sendo as variáveis mais correlacionadas com esses valores negativos OCTAS %, DIOXINAS %, FURANOS – TEQ e TEQ TOTAL. Perfis de contaminação oriundos da metalurgia de ferro estão correlacionados com valores positivos para a primeira componente principal, sendo que as variáveis mais correlacionadas com esses valores positivos são TETRAS %, PENTAS %, HEXAS % e HEPTAS.

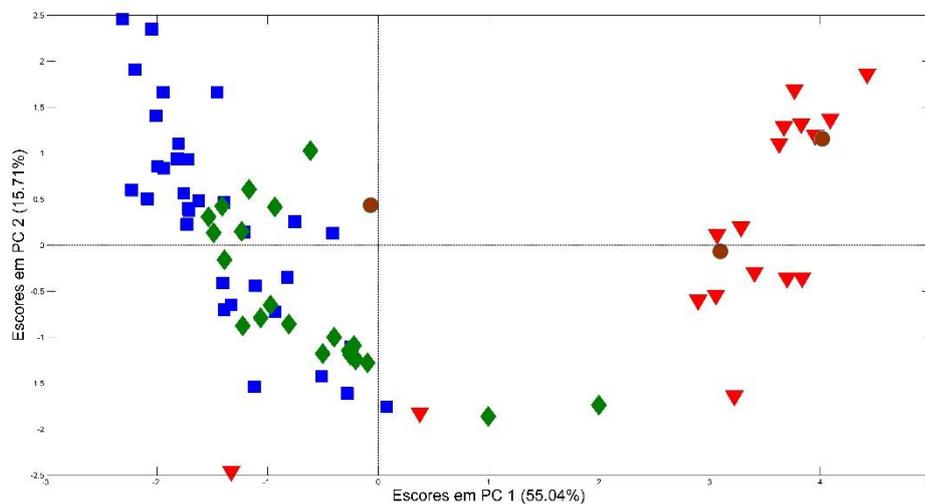


Figura 9: Dispersão do grupo MTL em 3 CP. Dados em azul, verde, vermelho e marrom correspondem à metalurgia de cobre, magnésio, ferro e alumínio respectivamente.

- Análise exploratória do grupo PLT

O conjunto PLT consiste em perfis de degradação de PVC e perfis de efluentes de processos de produção de dicloreto de etileno (EDC) e PVC. Não foi observada

uma distinção clara entre as amostras (Figura 10).

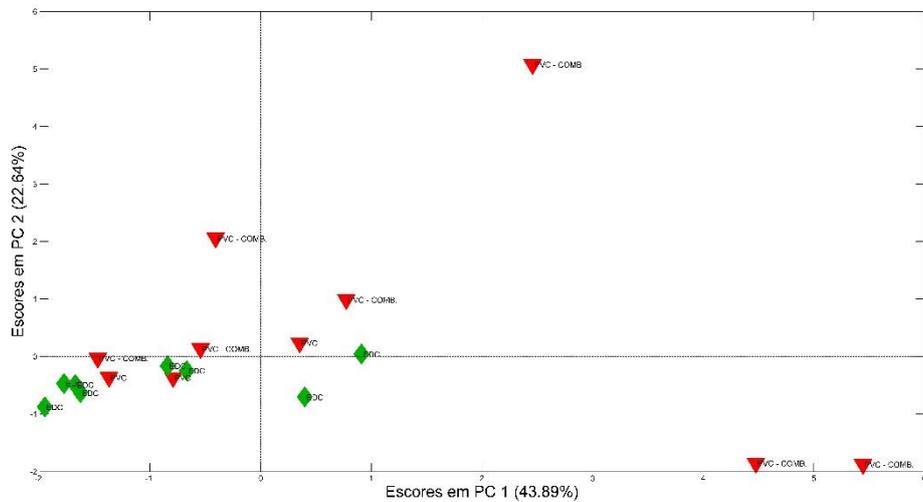


Figura 10: Dispersão do grupo PTL em 2 CP. Dados em vermelhos corresponde à processos de PVC e dados em verde corresponde à processos de produção de EDC.

#### - Análise exploratória do grupo CIM

A análise exploratória dos dados do grupo CIM (Figura 11), com duas componentes principais, explica 76 % da informação. O conjunto CIM consiste de dados de apenas uma referência bibliográfica, apresentando assim pouca variabilidade entre processos de fábricas em localidades diferentes. Observa-se que, nesse grupo, as variáveis TETRAS %, PENTAS %, HEXA %, HEPTAS %, FURANOS % e TEQ TOTAL estão correlacionadas com valores positivos da primeira componente principal, enquanto as variáveis OCTAS % e DIOXINAS % estão correlacionadas com valores negativos dessa mesma componente. As variáveis TETRA TEQ e PENTA TEQ estão correlacionadas com valores positivos na segunda componente principal ao passo que a variável DIOXINA – TEQ está correlacionada com valores negativos nessa componente.



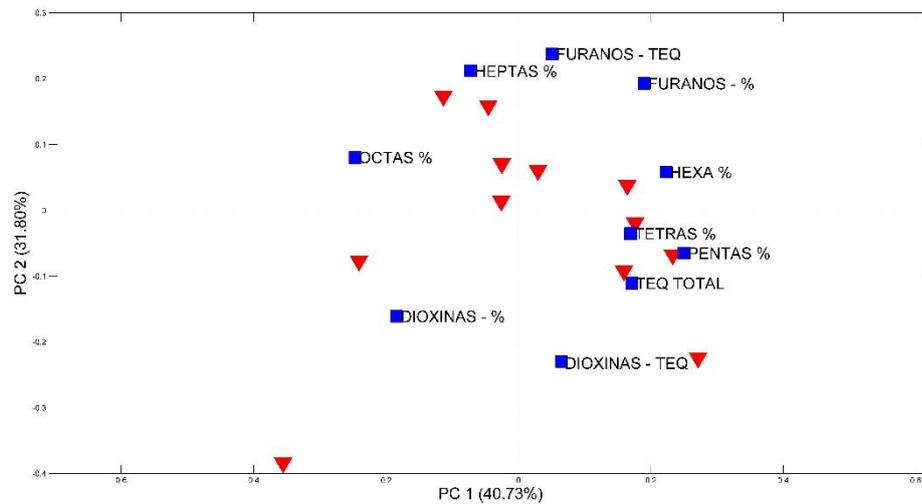


Figura 12: Dispersão dos perfis de contaminação do grupo INC no espaço criado pelas duas primeiras componentes principais e a correlação de cada variável com cada componente. Perfis em vermelho e variáveis em azul

#### - Análise exploratória do grupo PCB

A análise exploratória dos dados do grupo PCB, com duas componentes principais, explica 80 % da informação (Figura 13). É possível verificar uma separação entre os fabricantes do PCB, sendo que os perfis referentes aos PCB's da marca KC se apresentam com valores próximos a zero em relação à primeira componente e valores positivos em relação a segunda componente. Perfis referentes à marca Aroclor se apresentam dispersos em relação à primeira componente de valores negativos em relação à segunda componente. Em relação às variáveis, pode-se observar que, nesse grupo, as variáveis TETRAS %, PENTAS %, TETRAS-TEQ, PENTAS-TEQ, FURANOS-TEQ, DIOXINAS-TEQ e TEQ TOTAL estão correlacionadas com valores positivos da primeira componente principal, enquanto as variáveis HEXAS %, HEPTAS %, OCTAS %, DIOXINAS % e DIOXINAS-TEQ estão correlacionadas com valores negativos dessa mesma componente. As variáveis PENTAS %, HEXA %, HEPTAS %, PENTAS-TEQ e FURANOS-TEQ estão correlacionadas com valores positivos na segunda componente principal enquanto as variáveis TETRAS %, OCTAS %, DIOXINAS %, TETRAS-TEQ, DIOXINA – TEQ e TEQ TOTAL estão correlacionadas com valores negativos nessa.

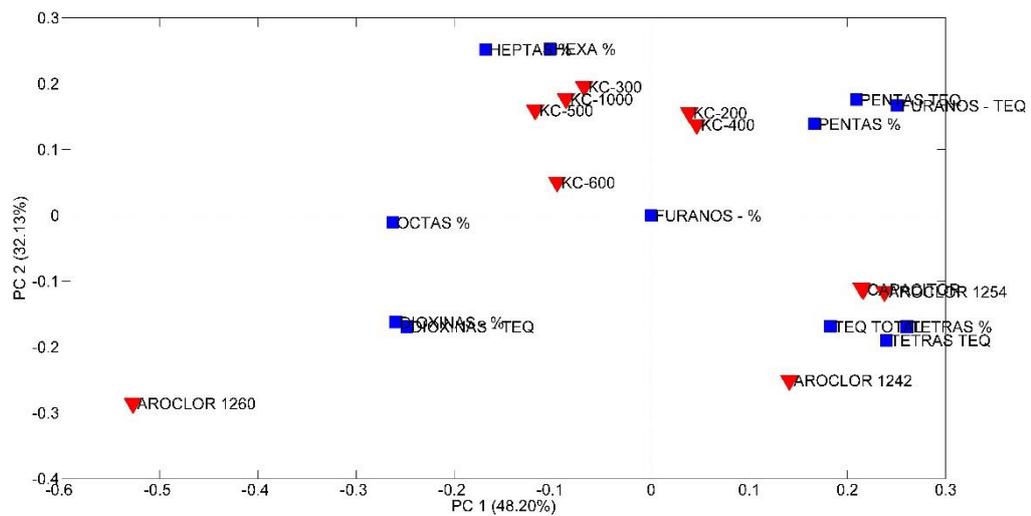


Figura 13: Dispersão dos perfis de contaminação do grupo PCB no espaço criado pelas duas primeiras componentes principais e a correlação de cada variável com cada componente. Perfis em vermelho e variáveis em azul

#### - Análise exploratória do grupo PGM

A análise exploratória dos dados do grupo PGM, com duas componentes principais, explica 60 % da informação (Figura 14). Observa-se que, nesse grupo, as variáveis OCTAS %, DIOXINAS % e DIOXINAS-TEQ estão correlacionadas com valores positivos da primeira componente (PC1) e valores negativos da segunda componente (PC2). As variáveis PENTAS %, HEXA %, TETRAS-TEQ e PENTAS-TEQ estão correlacionadas com valores positivos da PC1 e valores positivos da PC2 e as variáveis HEPTAS %, FURANOS % e TEQ TOTAL estão correlacionadas com valores positivos de PC2 e valores negativos de PC1.

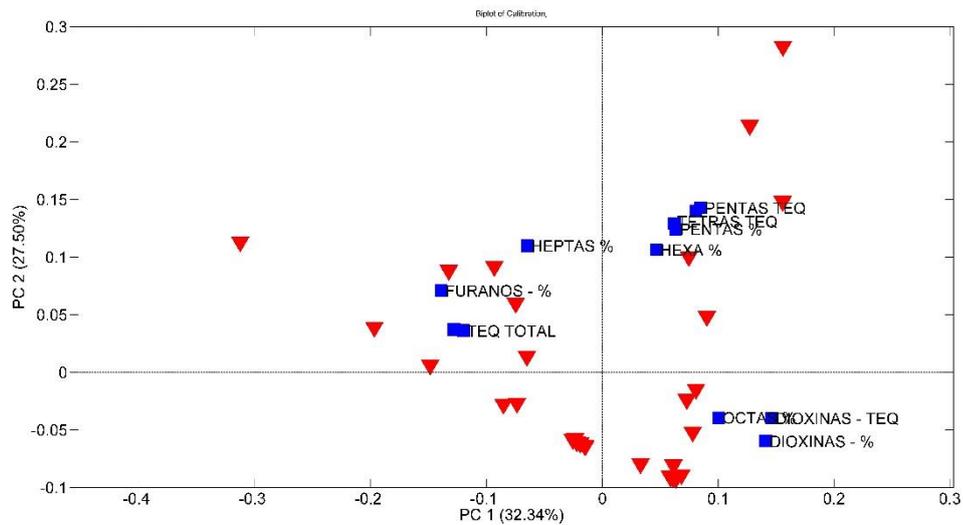


Figura 14: Dispersão dos perfis de contaminação do grupo PGM no espaço criado pelas duas primeiras componentes principais e a correlação de cada variável com cada componente. Perfis em vermelho e variáveis em azul

#### - Análise exploratória do grupo AGR

A análise exploratória dos dados do grupo AGR, com cinco componentes principais, explica 81,47 % da informação. Na Figura 15, é apresentada a dispersão dos dados no espaço criado pelas duas primeiras componentes principais. Verifica-se, a princípio, a separação das amostras em três subgrupos distintos. Em relação às variáveis, pode-se observar que FURANOS - % e FURANOS TEQ estão correlacionadas com valores positivos de PC1 e PC2. HEPTAS-%, OCTAS-% e TEQ TOTAL estão correlacionadas com valores positivos de PC2 e valores negativos de PC1. DIOXINAS-%, TETRAS-TEQ e DIOXINAS-TEQ estão associadas à valores negativos de PC1 e PC2 e as variáveis TETRA-%, PENTAS-%, HEXAS-% e PENTAS-TEQ estão associadas à valores positivos de PC1 e negativos de PC2. Observa-se que perfis referentes ao agrotóxico 2,4-D se encontram nos grupos representados por triângulos vermelhos e no grupo representado por círculos verdes, não sendo possível

assim, a princípio, uma correlação entre o subgrupo e o tipo do agrotóxico.

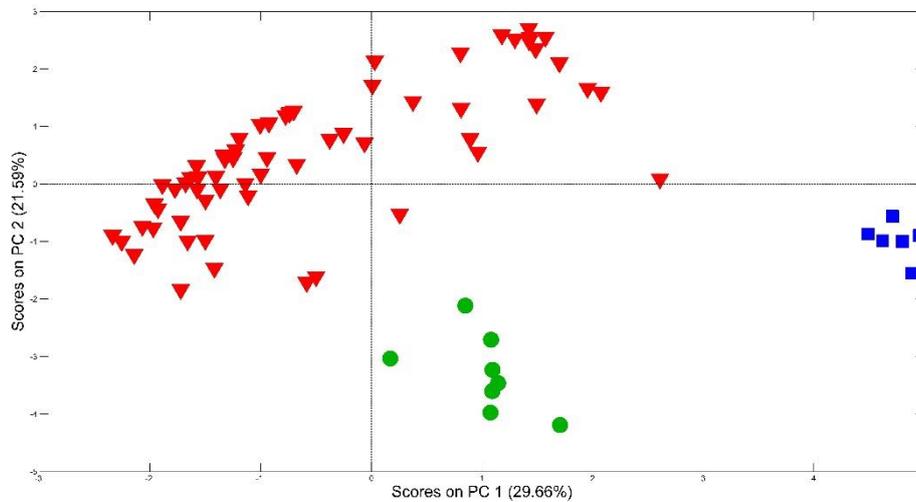


Figura 15: Dispersão dos perfis de contaminação do grupo AGR no espaço criado pelas duas primeiras componentes principais.

#### - Análise exploratória do grupo CMB

A análise exploratória dos dados do grupo CMB, com três componentes principais, explica 72 % da informação. A análise exploratória (Figura 16) sugere algumas separações entre as origens dos perfis de contaminação, no entanto, a variabilidade é muito baixa. Em algumas referências havia a informação sobre o teor de cloreto (Cl) em ppm, variando de 300 a 2000 ppm (Figura 17). Não houve também uma separação clara entre os perfis de contaminação gerados por processos com maior teor de cloreto daqueles gerados por processos com menor teor de cloreto. Essa observação sugere que a quantidade de cloreto, no caso das amostras deste grupo, não é importante para perfil de congêneres gerado, sendo aparentemente intrínseca ao processo. No entanto, a variabilidade é pequena, necessitando de mais dados para confirmar essa afirmação. Em relação às variáveis, TETRAS-%, PENTAS-%, HEPTAS-%, TETRAS-TEQ, PENTAS-TEQ e TEQ-TOTAL estão correlacionadas com valores positivos de PC1 e PC2. OCTAS-% e DIOXINAS-% estão correlacionadas com valores negativos de PC1 e valores positivos de PC2 e HEXAS-%, FURANOS-% e FURANOS-TEQ estão correlacionadas com valores positivos de PC1 e valores negativos de PC-2.

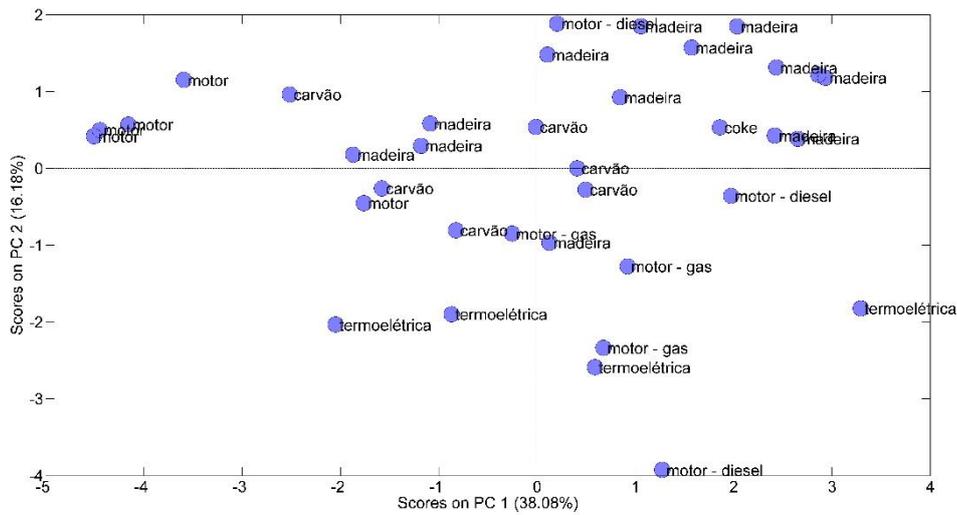


Figura 16: Dispersão dos perfis de contaminação do grupo CMB no espaço criado pelas duas primeiras componentes principais.

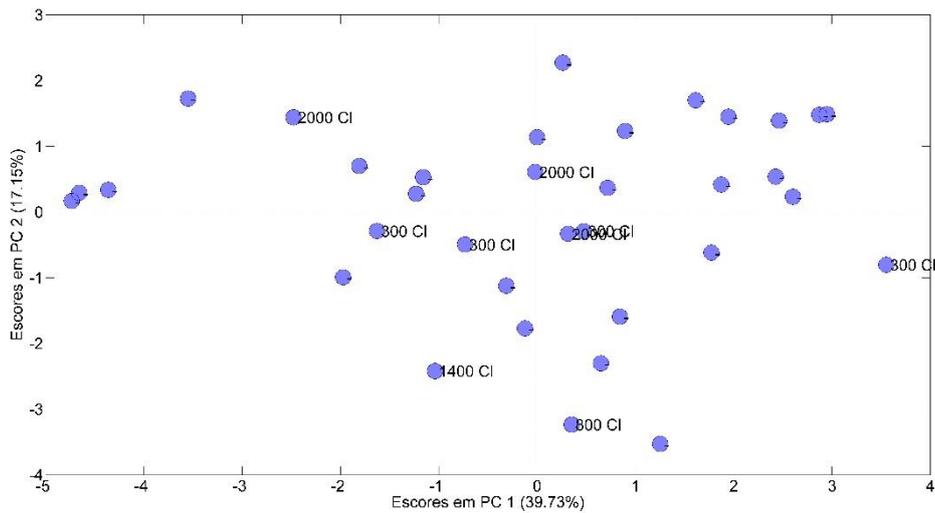


Figura 17: Dispersão dos perfis de contaminação do grupo CMB no espaço criado pelas duas primeiras componentes principais. Informação sobre o teor de cloreto nos combustíveis.

## 5.2 Classificação

Como não foi observada, na análise exploratória, uma separação clara com todos os grupos reunidos (Figura 18), que permitisse uma classificação direta, a análise de classificação foi realizada de forma a criar subgrupos que permitissem a classificação subsequente das amostras com menores valores de erro quanto às taxas de sensibilidade, especificidade, falso positivo e falso negativo. Nesse modelo de classificação, foi utilizado o método de PLS1-DA, em que uma classe de amostras é considerada “positiva” e o restante das amostras é considerada “negativo”. Foram propostos três arranjos para a classificação, que podem ser observados no Apêndice 2. Na Figura 19, é apresentado o arranjo de classificação utilizado para classificar as amostras de aves, ovos e gêneros de alimentação animal quanto à possível origem de contaminação, que apresentou melhor desempenho.

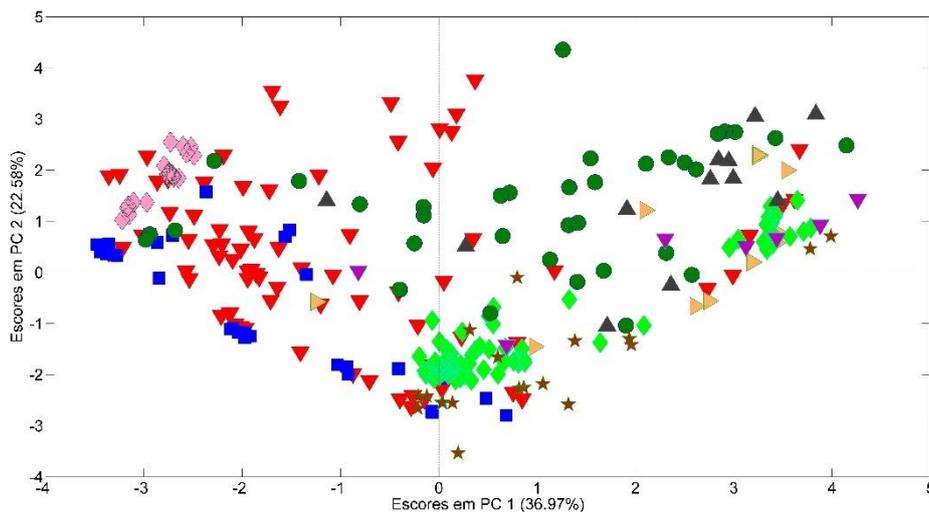


Figura 18: Distribuição das amostras no plano criado pelas duas primeiras componentes principais. Triângulo vermelhos são referentes ao grupo AGR, quadrados azuis são referentes ao grupo PGM, losangos rosas são referentes ao grupo MIN, círculos verde-escuros são referentes ao grupo CMB, losangos verde-claros são referentes ao grupo MTL, estrelas marrons são referentes ao grupo PTL, triângulos laranjas são referentes ao grupo PCB, triângulos pretos são referentes ao grupo INC e triângulos roxos são referentes ao grupo CIM

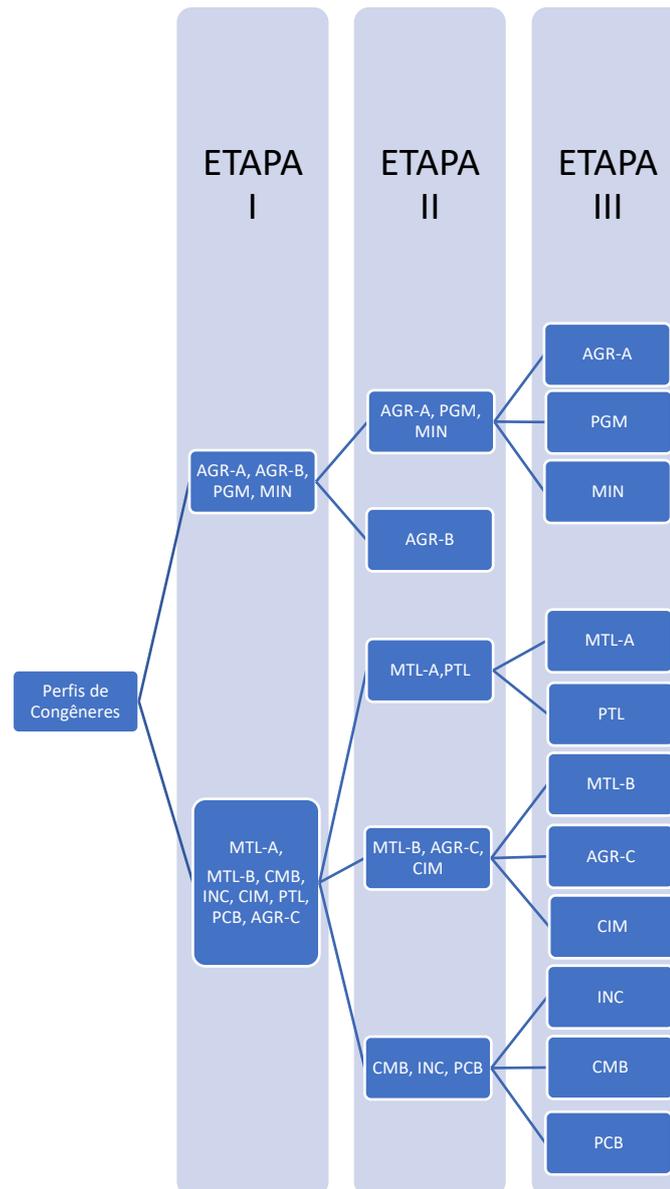


Figura 18: Hierarquia de classificação adotada para a classificação do perfil de contaminantes quanto à sua provável origem (modelo III).

Uma vez que os grupos AGR e MTL apresentavam separações internas que estavam reduzindo o desempenho da classificação em relação às figuras de mérito, no arranjo de classificação utilizado esses grupos foram desmembrados em AGR-A, AGR-B, AGR-C, para o caso dos perfis de contaminação associados à agrotóxicos, e em MTL-A e MTL-B para o caso dos perfis de contaminação associados à indústria metalúrgica. Nessa hierarquia de classificação, foram utilizadas três etapas de classificação, sendo que a primeira etapa (I) separa as amostras em dois grandes conjuntos, um contendo os grupos AGR-A, AGR-B, MIN e PGM e o outro contendo os grupos MTL-A, MTL-B, PCB, INC, PTL, CMB e CIM. Na segunda etapa (II), os perfis são reorganizados em 5

conjuntos que apresentam maiores similaridades dentre deles. Na terceira etapa (III), cada grupo é separado individualmente.

Para a criação dos modelos de classificação, cada grupo foi separado em um conjunto de treinamento (dois terços dos números de perfis) e um conjunto de teste (um terço do número de perfis), utilizando o algoritmo de Kennard-Stone. Para selecionar o melhor número de variáveis latentes, em cada etapa, foi realizada uma validação cruzada usando o critério de subconjuntos aleatórios. O número de VL é escolhido de acordo com aquele que retorna o menor valor para a razão entre amostras erroneamente classificadas e o número total de amostras no conjunto de treinamento (ECVC). Na Tabela 10, é mostrado o número de variáveis latentes utilizadas em cada classificação e a variância explicada em X e Y, ambos em % e os valores de ECVC. Na Tabela 11, são mostrados os valores das figuras de mérito obtidos em cada classificação no conjunto de treinamento e no conjunto teste (Tabela 12). Valores para os outros arranjos de classificação se encontram no Apêndice 3.

Os valores de variância em X variam de 32,16 a 89,90 e os valores de variância em Y variam de 37,98 a 93,58 (Tabela 10). Em todos os modelos, os dados foram auto escalados, sendo os valores de cada variável subtraídos da média e divididos pelo desvio padrão respectivo da variável correspondente.

Tabela 10: Número de VLs, variâncias explicadas e erros de classificação de validação cruzada (ECVC) referente aos modelos em PLS-Da para as etapas de classificação

ETAPA	GRUPOS SEPARADOS	VARIÁVEIS LATENTES	Variância Explicada X (%)	Variância Explicada Y (%)	ECVC
I	AGR-A, AGR-B, PGM, MIN	8	91,37	67,62	0,12
I	MTL-A, MTL-B, CMB, INC, CIM, PTL, PCB, AGR-C	8	91,37	67,62	0,12
II	AGR-B	3	57.95	93.58	0.007
II	AGR-A, PGM, MIN	3	57.95	93.58	0.007
II	MTL-A, PTL	1	32.16	79.29	0.04

ETAPA	GRUPOS SEPARADOS	VARIÁVEIS LATENTES	Variância Explicada X (%)	Variância Explicada Y (%)	ECVC
II	MTL-B, AGR-C, CIM	4	71.5	57.73	0.12
II	CMB, INC, PCB	4	73.83	63.68	0.17
III	AGR-A	3	44.15	51.18	0.28
III	PGM	4	63.38	37.98	0.28
III	MIN	2	39,02	72,43	0.04
III	PTL	2	45.82	54.54	0.2
III	MTL-A	2	45.82	54.54	0.2
III	MTL-B	2	37.5	67.25	0.24
III	AGR-C	2	52.11	90.26	0.13
III	INC	4	62.51	52.82	0.41
III	CIM	2	47.49	54.26	0.13
III	PCB	3	61.9	74.2	0.18
III	CMB	4	62.51	52.82	0.41

Tabela 11: Figuras de mérito estimadas para etapa do modelo de classificação – Conjunto de treinamento

ETAPA	GRUPOS SEPARADOS	SENSIBILIDADE DE	ESPECIFICIDADE	FALSO POSITIVO	FALSO NEGATIVO
I	AGR-A, AGR-B, PGM, MIN	83.8	93.1	16.2	6.9
I	MTL-A, MTL-B, CMB, INC, CIM, PTL, PCB, AGR-C	93.1	83.8	6.9	16.2
II	AGR-B	98.6	100	1.4	0
II	AGR-A, PGM, MIN	100	98.6	0	1.4
II	MTL-A, PTL	98.1	94.9	1.9	5.1
II	MTL-B, AGR-C, CIM	90.5	92.3	9.5	7.7
II	CMB, INC, PCB	86.1	91.7	13.9	8.3
III	AGR-A	73.7	88.2	26.3	11.8
III	PGM	89.5	79.2	10.5	20.8
III	MIN	100	100	0	0
III	PTL	97.4	76.9	2.6	23.1
III	MTL-A	97.4	76.9	2.6	23.1
III	MTL-B	91.7	88.9	8.3	11.1

ETAPA	GRUPOS SEPARADOS	SENSIBILIDADE DE	ESPECIFICIDADE	FALSO POSITIVO	FALSO NEGATIVO
III	AGR-C	100	100	0	0
III	INC	88.9	95.7	11.1	4.3
III	CIM	80	93.8	20	6.2
III	PCB	100	100	0	0
III	CMB	88.9	95.7	11.1	4.3

Tabela 12: Figuras de mérito estimadas para etapa do modelo de classificação – Conjunto de teste

ETAPA	GRUPOS SEPARADOS	SENSIBILIDADE DE	ESPECIFICIDADE DE	FALSO POSITIVO	FALSO NEGATIVO
I	AGR-A, AGR-B, PGM, MIN MTL-A, MTL-B,	78	94.5	22	5.5
I	CMB, INC, CIM, PTL, PCB, AGR-C	94.5	78	5.5	22
II	AGR-B	100	100	0	0
II	AGR-A, PGM, MIN	100	100	0	0
II	MTL-A, PTL	84	96.7	16	3.3
II	MTL-B, AGR-C, CIM	100	84.8	0	15.2
II	CMB, INC, PCB	58.8	97.1	41.2	2.9
III	AGR-A	57.1	88.2	42.9	11.8
III	PGM	80	71.4	20	28.6
III	MIN	85,7	96,9	14,3	3.1
III	PTL	88.9	100	11.1	0
III	MTL-A	88.9	100	11.1	0
III	MTL-B	100	75	0	25
III	AGR-C	100	100	0	0
III	INC	100	83.3	0	16.7
III	CIM	50	100	50	0
III	PCB	25	100	75	0
III	CMB	100	83.3	0	16.7

Os gráficos de classificação do conjunto de treinamento e do conjunto teste, referentes à classificação da etapa I, podem ser observados na Figura 18. No Apêndice 4, podem ser observados os gráficos de classificação de todas as etapas de classificação. Os valores de classificação para cada classe/modelo foram estimados pelo próprio software (Matlab 7.10 – PLS – toolbox) pela teoria de Bayes, para minimizar os erros de classificação.

A interpretação dos modelos pode ser feita a partir da observação dos coeficientes de regressão que podem ser observados no Apêndice 5. Na Figura 20, são apresentados os vetores de regressão da etapa de classificação I. Os resultados das amostras, de origem de contaminação desconhecidas, assim como as amostras compiladas na literatura com a origem conhecida, foram processadas no modelo obtido. Na Tabela 13, reúnem-se os resultados obtidos pelo modelo de classificação acerca sobre a possível origem de contaminação e, na Tabela 14, reúnem-se os resultados de classificação para uma mesma localidade das amostras citadas na Tabela 9.

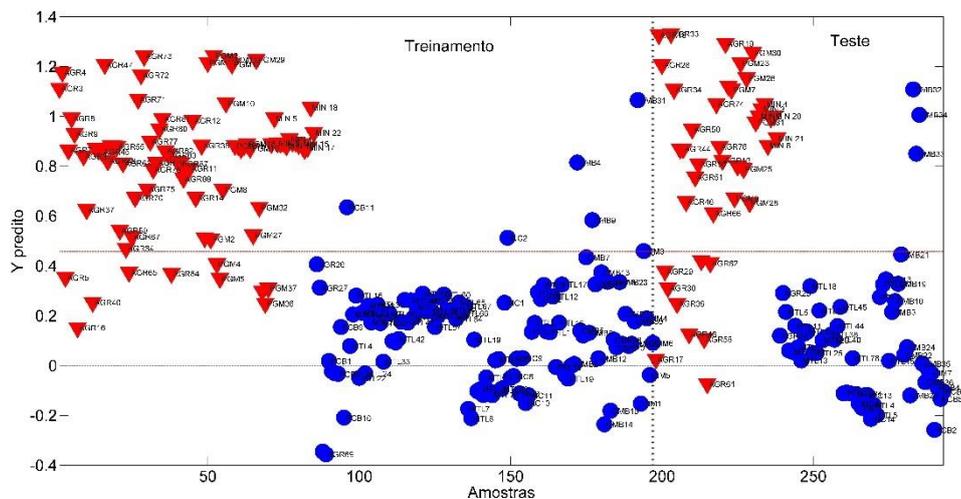


Figura 20: Gráfico de classificação para a etapa de classificação I. À esquerda conjunto de treinamento e à direita conjunto teste. Conjunto formado pelos grupos AGR-A, AGR-B, PGM e MIN representados por triângulos vermelhos. Conjunto formado pelos grupos MTL-A, MTL-B, AGR-C, PCB, INC, CIM, CMB, PTL representados por círculos azuis.

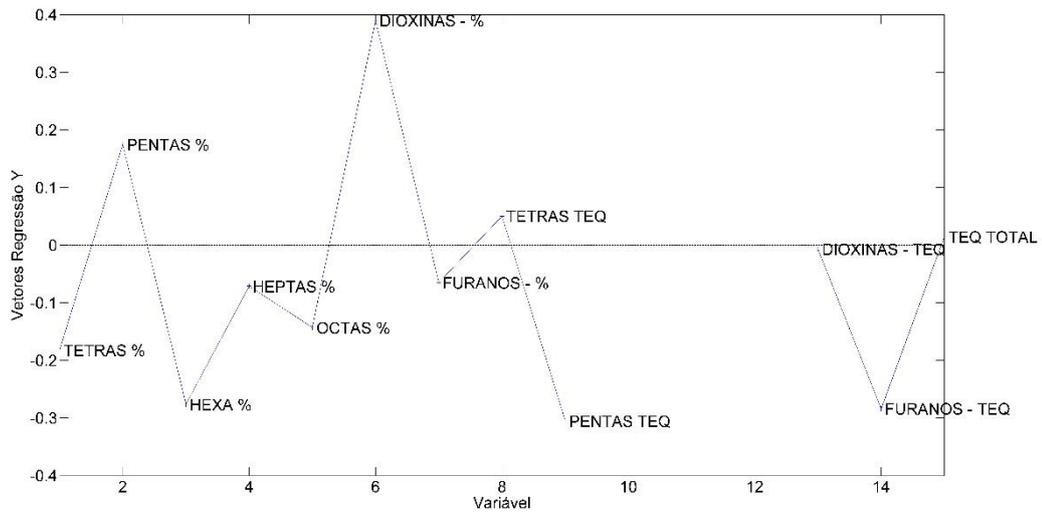


Figura 21: Gráfico de vetores de regressão referentes à classificação para a etapa de classificação I. Conjunto formado pelos grupos AGR-A, AGR-B, PGM e MIN representados como classe “positiva”. Conjunto formado pelos grupos MTL-A, MTL-B, AGR-C, PCB, INC, CIM, CMB, PTL representados como classe “negativa”.

Tabela 13: Tabela com o código da amostra, a classificação esperada e a classificação final (etapa III)

CÓDIGO AMOSTRA	MATRIZ	CLASSIFICAÇÃO ESPERADA	CLASSIFICAÇÃO FINAL
FF1	polpa cítrica	PLT	INC
FF2	polpa cítrica	PLT	INC
FF3	polpa cítrica	PLT	INC
FF4	polpa cítrica	PLT	INC
FF5	polpa cítrica	PLT	INC
FF6	polpa cítrica	PLT	INC
FF7	polpa cítrica	PLT	INC
FF8	polpa cítrica	PLT	INC
FF9	polpa cítrica	PLT	INC
FF10	polpa cítrica	PLT	INC
FF11	polpa cítrica	PLT	PTL
FF12	polpa cítrica	PLT	INC
FF13	polpa cítrica	PLT	INC
FF14	polpa cítrica	PLT	INC
FF15	gelatina - gordura reciclada	PCB	AGR-B
FF16	gelatina - gordura reciclada	PCB	AGR-B
FF17	gelatina - gordura reciclada	PCB	AGR-B

CÓDIGO AMOSTRA	MATRIZ	CLASSIFICAÇÃO ESPERADA	CLASSIFICAÇÃO FINAL
FF18	ração	PCB	CMB
FF19	ração	MTL	CIM
FF20	premix - ZnO	MTL	CIM
FF21	premix - ZnO	MTL	CIM
FF22	ZnO	MTL	MTL-A
FF23	ração	PCB	AGR-A
FF24	ovo	MTL	MIN
FF25	ovo	MTL	AGR-A, PGM, MIN
FF26	ovo	MTL	MIN
FF27	ovo	MTL	AGR-A, PGM, MIN
FF28	biodiesel	AGR-A	AGR-B
FF29	gordura animal	N.A.	MIN
FF30	gordura animal	N.A.	MIN
FF31	gordura animal	N.A.	MIN
FF32	óleo de peixe	N.A.	CMB
FF33	óleo de peixe	N.A.	MIN
FF34	óleo de peixe	N.A.	AGR-A
FF35	farinha de peixe	N.A.	MIN
FF36	farinha de peixe	N.A.	MIN
FF37	farinha de peixe	N.A.	MIN
FF38	farinha de peixe	N.A.	MIN
FF39	farinha de peixe	N.A.	CMB
FF40	farinha de carne e ossos	N.A.	MIN
FF41	farinha de carne e ossos	N.A.	AGR-A, PGM, MIN
LDP1	aves	N.A.	INC
LDP2	ração	N.A.	CMB
LDP3	aves	N.A.	CMB
LDP4	aves	N.A.	CMB
LDP5	ração	N.A.	INC
LDP6	ração	N.A.	INC
LDP7	aves	N.A.	CMB
LDP8	aves	N.A.	CMB
LDP9	aves	N.A.	CMB
LDP10	aves	N.A.	CMB
LDP11	aves	N.A.	CMB
LDP12	aves	N.A.	CMB
LDP13	aves	N.A.	CMB
LDP14	aves	N.A.	CMB
LDP15	aves	N.A.	CMB
LDP16	aves	N.A.	CMB
LDP17	aves	N.A.	CMB
LDP18	aves	N.A.	CMB
LDP19	aves	N.A.	CMB

CÓDIGO AMOSTRA	MATRIZ	CLASSIFICAÇÃO ESPERADA	CLASSIFICAÇÃO FINAL
LDP20	aves	N.A.	CMB
LDP21	aves	N.A.	CMB
LDP22	aves	N.A.	CMB
LDP23	aves	N.A.	CMB
LDP24	aves	N.A.	CMB
LDP25	aves	N.A.	CMB
LDP26	ração	N.A.	INC
LDP27	ração	N.A.	INC
LDP28	ração	N.A.	INC
LDP29	ração	N.A.	CMB
LDP30	ração	N.A.	INC
LDP31	ração	N.A.	CMB
LDP32	aves	N.A.	CMB
LDP33	ovo	N.A.	CMB
LDP34	aves	N.A.	CMB
LDP35	aves	N.A.	CMB
LDP36	aves	N.A.	CMB
LDP37	aves	N.A.	CMB
LDP39	aves	N.A.	CMB, INC, PCB
LDP40	ração	N.A.	INC
LDP41	ração	N.A.	AGR-B
LDP42	ovo	N.A.	CMB
LDP43	ovo	N.A.	CMB
LDP44	ovo	N.A.	CMB, MTL-B, AGR-C
LDP45	ovo	N.A.	CMB
LDP46	ovo	N.A.	CMB
LDP47	ração	N.A.	GRUPO 2
LDP48	aves	N.A.	CMB
LDP49	ovo	N.A.	CMB
LDP50	ovo	N.A.	CMB
LDP51	ovo	N.A.	CMB
LDP52	ovo	N.A.	CMB
LDP53	ovo	N.A.	CMB
LDP54	ração	N.A.	CMB
LDP55	ração	N.A.	CMB
LDP56	aves	N.A.	CMB
LDP57	aves	N.A.	CMB
LDP58	aves	N.A.	CMB
LDP59	aves	N.A.	CMB
LDP60	aves	N.A.	CMB
LDP61	ração	N.A.	INC
LDP62	ração	N.A.	INC
LDP63	ração	N.A.	CMB

CÓDIGO AMOSTRA	MATRIZ	CLASSIFICAÇÃO ESPERADA	CLASSIFICAÇÃO FINAL
LDP64	ração	N.A.	CMB
LDP65	ração	N.A.	CMB
LDP66	aves	N.A.	CMB
LDP67	ração	N.A.	AGR-B
LDP68	aves	N.A.	CMB
LDP69	ração	N.A.	CMB
LDP70	ração	N.A.	CMB
LDP71	ração	N.A.	INC
LDP72	ração	N.A.	CMB
LDP73	ração	N.A.	CMB
LDP74	ração	N.A.	CMB
LDP75	ração	N.A.	CMB, MTL-B
LDP76	ração	N.A.	CMB
LDP77	ração	N.A.	CMB
LDP78	ração	N.A.	CMB
LDP79	ração	N.A.	CMB
LDP80	ração	N.A.	INC
LDP81	ração	N.A.	CMB
LDP82	aves	N.A.	CMB
LDP83	aves	N.A.	CMB
LDP84	ovo	N.A.	AGR-B
LDP85	ovo	N.A.	CMB
LDP86	aves	N.A.	CMB
LDP87	aves	N.A.	CMB, MTL-B
LDP88	aves	N.A.	CMB, MTL-B
LDP89	aves	N.A.	CMB, MTL-B
LDP90	ração	N.A.	INC
LDP91	ração	N.A.	INC
LDP92	ração	N.A.	INC
LDP93	ração	N.A.	INC
LDP94	ração	N.A.	INC
LDP95	ração	N.A.	AGR-B
LDP96	ração	N.A.	CMB
LDP97	ração	N.A.	CMB

N.A.: não se aplica

Tabela 14: Tabela com o código, local, matriz e classificação final respectivo das amostras da Tabela 9.

LOCAL	CÓDIGO AMOSTRA	MATRIZ	CLASSIFICAÇÃO FINAL
A	LDP03	aves	CMB
A	LDP04	aves	CMB
A	LDP05	ração	INC
A	LDP26	ração	INC
A	LDP29	ração	CMB
B	LDP68	aves	CMB
B	LDP75	ração	CMB, MTL-B
C	LDP56	aves	CMB
C	LDP57	aves	CMB
C	LDP58	aves	CMB
C	LDP59	aves	CMB
C	LDP60	aves	CMB
D	LDP86	aves	CMB
D	LDP87	aves	CMB, MTL-B
D	LDP88	aves	CMB, MTL-B
D	LDP89	aves	CMB, MTL-B
D	LDP90	ração	INC
D	LDP91	ração	INC
D	LDP92	ração	INC
D	LDP93	ração	INC
D	LDP94	ração	INC
E	LDP42	ovo	CMB
E	LDP43	ovo	CMB
E	LDP44	ovo	CMB, MTL-B, AGR-C
E	LDP45	ovo	CMB
E	LDP46	ovo	CMB
E	LDP49	ovo	CMB
E	LDP50	ovo	CMB
E	LDP51	ovo	CMB
E	LDP52	ovo	CMB
E	LDP53	ovo	CMB
F	LDP07	aves	CMB
F	LDP08	aves	CMB
F	LDP10	aves	CMB
F	LDP11	aves	CMB
F	LDP12	aves	CMB
F	LDP13	aves	CMB
F	LDP14	aves	CMB
F	LDP15	aves	CMB
F	LDP17	aves	CMB

LOCAL	CÓDIGO AMOSTRA	MATRIZ	CLASSIFICAÇÃO FINAL
F	LDP18	aves	CMB
F	LDP19	aves	CMB
F	LDP20	aves	CMB
F	LDP21	aves	CMB
F	LDP22	aves	CMB
F	LDP23	aves	CMB
F	LDP24	aves	CMB
F	LDP25	aves	CMB
F	LDP28	ração	INC
G	LDP09	aves	CMB
G	LDP16	aves	CMB
H	LDP40	ração	INC
H	LDP61	ração	INC
H	LDP62	ração	INC
I	LDP47	ração	GRUPO 2
I	LDP54	ração	CMB
I	LDP55	ração	CMB
J	LDP30	ração	INC
J	LDP31	ração	CMB

## 6. Discussão

Por meio da análise exploratória, foi possível observar agrupamentos internos e associações entre os perfis e as variáveis utilizadas para a construção do modelo de classificação. No grupo MIN, observa-se uma distinção clara entre os perfis referentes à caulinita e os restantes (Figura 6). No grupo MTL, observa-se também uma distinção clara entre os perfis referentes à metalurgia de cobre e magnésio e os perfis referentes à metalurgia de ferro (Figura 7). No grupo AGR, é possível notar a separação dos perfis em três grupos (Figura 13). No entanto, não existe uma separação entre os tipos de agrotóxicos, sendo que o motivo dessa formação possa ser alguma variável que não foi observada no levantamento dos perfis nas referências bibliográficas.

Em relação ao modelo de classificação, os arranjos foram sugeridos de forma a reduzir a influência do desbalanceamento entre as classes de perfis. Apesar da possibilidade de viés na classificação, o modelo III foi o utilizado. Pode-se observar, em algumas etapas de classificação, uma certa discrepância entre os valores das figuras de mérito obtidas no conjunto de treinamento e no conjunto teste (Tabelas 11 e 12). Tal discrepância sugere um sobreajuste do modelo, situação na qual o modelo prevê bem as amostras internas (treinamento), no entanto, prevendo mal amostras externas (teste). Uma solução para a melhoria do desempenho do modelo de classificação e redução do viés, causado pelo desbalanceamento entre as classes de perfis, seria o aumento da variabilidade dos dados, que em alguns grupos foi bastante reduzida, como no caso de perfis referentes ao grupo INC.

Apesar de modelo de classificação apresentar um desempenho razoável, considerando as figuras de mérito, deve-se ter em mente que, quanto menor o valor do  $\Sigma$  TEQ, maior será a influência da contaminação nativa basal (background), podendo assim induzir a uma classificação equivocada. Logo, o modelo construído terá aplicabilidade para amostras contaminadas. No entanto, não foi estimado qual seria o valor mínimo de  $\Sigma$  TEQ para a adequada classificação. Frequentemente, em um evento de contaminação real, tem-se uma situação complexa, em que várias fontes podem estar contribuindo para o perfil de contaminação. Avaliações dos gráficos dos vetores de regressão (Figura 19) podem fornecer informações valiosas no contexto de qual o conjunto de variáveis que são mais significativas para a

confirmação de que um determinado perfil pertence à uma determinada classe. A avaliação dos gráficos de classificação (Figura 18) pode também fornecer informações interessantes quanto a que tipo de perfis constituem os falsos positivos e falsos negativos. Na etapa I de classificação, verifica-se que as variáveis PENTAS %, DIOXINAS % e TETRAS TEQ apresentam maior relevância para a separação do conjunto formado pelos grupos AGR-A, AGR-B, PGM e MIN, enquanto as variáveis HEXAS %, HEPTAS %, OCTAS % , FURANOS %, PENTAS-TEQ, FURANOS-TEQ são mais relevantes para a classificação do conjunto formado pelos grupos MTL-A, MTL-B, AGR-C, PCB, CMB, CIM, INC e PTL (Figura 19). Essas informações podem sugerir que os processos do primeiro grupo tendem a gerar um percentual maior de dioxinas, ao contrário do segundo grupo em que se sugere que os processos tendem a gerar um maior percentual de furanos. Observa-se também que os processos do grupo 2 tendem a ter um maior percentual de HEXAS, HEPTAS e OCTAS. Falsos negativos, considerando o grupo 1 como positivo e o grupo 2 como negativo (Figura 18) são devidos principalmente ao grupo AGR. Tal situação pode ser devido à extensa variabilidade dos processos de produção dos agrotóxicos. Falsos positivos são devidos ao grupo CMB.

Em relação à classificação dos perfis de contaminação da Tabela 13, pode-se observar uma aparente falta de coerência entre a classificação esperada (reportada na respectiva referência bibliográfica) e a classificação final, obtida pelo modelo de classificação construído (modelo III). Tal falta de coerência sugere que o modelo ainda necessita ser melhorado quanto à variabilidade dos perfis. O aumento da variabilidade dos perfis pode expor formações de agrupamentos internos em um grupo específico, sugerindo uma alteração no arranjo de classificação. Outro fator que não foi considerado na construção do modelo de classificação foram os fatores de transferência e de bioacumulação, quando as matrizes “aves” e “ovos” são avaliadas. Esses fatores podem atuar com intensidades diferentes nas variáveis (absorção e acumulação maior ou menor), promovendo uma classificação que poderia não corresponder à realidade. Uma alternativa poderia ser a seleção de variáveis que fossem impactadas com a mesma intensidade quanto aos fatores de transferência e bioacumulação. Os resultados da Tabela 14 podem ser utilizados como base para a seleção dessas variáveis, uma vez que perfis que pertencem a um mesmo evento de

contaminação, na mesma localidade deveriam apresentar a mesma “classificação final”. Apesar das divergências em alguns resultados da Tabela 14, a classificação final para os perfis fornecidos pelo Laboratório de Dioxinas e PCB não foi incoerente, uma vez que se esperava uma maior predominância de perfis referentes à processos de combustão de biomassa.

## 7. Conclusão

O modelo elaborado pode classificar, com razoável exatidão (considerando os valores das figuras de mérito), os perfis de contaminação fornecidos pelo LDP.

O modelo pode ser melhorado com um aumento da variabilidade dos perfis dentro de um grupo, tornando-o mais robusto, reduzindo o viés de classificação e podendo revelar também agrupamentos internos que poderiam alterar o arranjo de classificação.

A adição de dados referentes às bifenilas policloradas semelhantes às dioxinas (PCB's-dl), que são analisadas conjuntamente com dioxinas e furanos, podem melhorar o poder de previsão do modelo, inserindo um número maior de informação que poderia diferenciar o perfil de contaminação de cada grupo.

O resultado de classificação, obtido pelo modelo, pode ser utilizado como uma sugestão para nortear o processo de investigação, na busca de mitigar o evento de contaminação no menor tempo possível.

## 8. Referências bibliográficas

1. ALCOCK, R. E.; SWEETMAN, A. J.; ANDERSON, D. R.; FISHER, R.; JENNINGS, R. A.; JONES, K. C. Using PCDD/F congener patterns to determine the source of elevated TEQ concentrations in cows milk: a case study. **Chemosphere**, v. 46, p.383-391, 2002.
2. ALCOSER, V. H. L.; VELTHUIS, A. G. J.; HOOGENBOOM, L. A. P.; FELS-KLERX, H. J. van Der. Financial Impact of a Dioxin Incident in the Dutch Dairy Chain. **Journal of Food Protection**, v. 74, n. 6, p.967-979, 1 jun. 2011. International Association for Food Protection.
3. ALEMANHA, ROLAND WEBER. **Analyse und trendabschätzung der belastung der umwelt und von lebensmitteln mit ausgewählten POPs und erweiterung des datenbestandes der POP-dioxin-datenbank des bundes und der länder mit dem ziel pfadbezogener ursachenaufklärung**. P&d Report (UBA), 2015. 528 p. Disponível em: <[https://www.umweltbundesamt.de/sites/default/files/medien/378/publikationen/doku\\_114\\_2015\\_analyse\\_un](https://www.umweltbundesamt.de/sites/default/files/medien/378/publikationen/doku_114_2015_analyse_un)>. Acesso em: 28 jan. 2019.
4. ALMEIDA, M. R. de. **Espectroscopia raman e quimiometria como ferramentas analíticas para química forense**. 2015. 177 f. Tese (Doutorado) - Curso de Química, Universidade Estadual de Campinas, Campinas, 2015.
5. ASSUNÇÃO, J. V. de; PESQUERO, C. R.. Dioxinas e furanos: origens e riscos. **Revista de Saúde Pública**, São Paulo, v. 33, n. 5, p.523-530, out. 1999.
6. BARROS NETO, B. de; SCARMINIO, I. S.; BRUNS, R. E. 25 anos de quimiometria no Brasil. **Química Nova**, v. 29, n. 6, p.1401-1406, 2006.
7. BERG, M van Den et al. Toxic equivalency factors (TEFs) for PCBs, PCDDs, PCDFs for humans and wildlife. **Environmental Health Perspectives**, v. 106, n. 12, p.775-792, 1998. Environmental Health Perspectives.
8. BLACK, R.R.; MEYER, C.P. (mick); YATES, A.; VAN ZWEITEN, L.; MUELLER, J.F. Formation of artefacts while sampling emissions of PCDD/PCDF from open burning of biomass. **Chemosphere**, v. 88, n. 3, p.352-357, 2012.
9. BRERETON, R. G. **Chemometrics: for pattern recongnition**. United Kingdon: John Wiley, 2009. 513 p.
10. BUEKENS, A.; ZHANG, M.; LI, X.. Statistical analysis as a tool for discriminating dioxin formation pathways. **Journal of Material Cycles And Waste Management**, v. 20, n. 3, p.1516-1529, 2018.
11. CARROL JUNIOR, W. F.; BERGER, T. C.; BORRELLI, F. E.; LEWIS, J. W.; LEDVINA, J.; MCCREEDY, R. L.; SMITH, T. P.; TUHOVAK, D. R.; WESTON, A. F. Characterization of emissions of dioxins and furans from ethylene dichloride (EDC), vinyl chloride monomer (VCM) and polyvinyl chloride (PVC) facilities in the United States. III. oxychlorination catalyst. **Organohalogen Compounds**, v. 41, p.133-136, 1999.
12. COVACI, A.; RYAN, J. J.; SCHEPENS, P. Patterns of PCBs and PCDD/PCDFs in chicken and pork fat following a Belgian food contamination incident. **Chemosphere**, v. 47, p.207-217, 2002.
13. DOPICO, M.; GÓMEZ, A. Review of the current state and main sources of dioxins around the world. **Journal Of The Air & Waste Management Association**, v. 65, n. 9, p.1033-1049, 11 jun. 2015.

14. ELJARRAT, E.; CAIXACH, J.; RIVERA, J.. Determination of PCDDs and PCDFs in different animal feed ingredients. **Chemosphere**, v. 46, p.1403-1407, 2002.
15. ESTADOS UNIDOS DA AMÉRICA. **An Inventory of Sources and Environmental Releases of Dioxin-Like Compounds in the United States for the Years 1987, 1995, and 2000**. Washington: U.S. Environmental Protection Agency, 2006. 677 p.
16. FERRARIO, J. B.; BYRNE, C. J.; CLEVERLY, D. H.. 2,3,7,8-Dibenzo-p-dioxins in mined clay products from the United States: evidence for possible natural origin. **Environmental Science Technology**, v. 34, p.4524-4532, 2000.
17. FERREIRA, M.M.C.; ANTUNES, A.M.; MELGO, M.S.; VOLPE, P.L.O.. Quimiometria I: calibração multivariada, um tutorial. **Química Nova**, v. 22, n.5, 1999
18. FERREIRA, M. M. C. **Quimiometria: conceitos, métodos e aplicações**. Campinas: Editora da Unicamp, 2015. 486 p.
19. FIEDLER, H.; MALISCH, R.; SCHACHTELE, A., HOOGENBOOM, R.; van LEEUWEN, S.; STEPHANOWITZ, R.; KNETSCH, G.. Pattern database for identification of sources and transfers of polychlorinated dibenzo-p-dioxins, dibenzofurans and biphenyls. **Organohalogen Compounds**, v. 80, p.714-718, 2018.
20. FONT, R.; GÁLVEZ, A.; MOLTÓ, J.; FULLANA, A.; ARACIL, I. Formation of polychlorinated compounds in the combustion of PVC with iron nanoparticles. **Chemosphere**, v. 78, n. 2, p.152-159, 2010.
21. FRE, R. de; RYMEN, T.. PCDD and PCDF formation from hydrocarbon combustion in the presence of hydrogen chloride. **Chemosphere**, v. 19, p.331-336, 1989.
22. GODIM, C.S., JUNQUEIRA, R.G., SOUZA, S.V.C. Tendências em validação de métodos de ensaios qualitativos. *Rev. Inst Adolfo Lutz*. 70(4): 433-47, 2011.
23. GÖTZ, R.; BERGEMANN, M.; STACHEL, B.; UMLAUF, G.. Dioxin in the river Elbe. **Chemosphere**, v. 183, p.229-241, set. 2017.
24. HADDAD, A.; MOQBEL, S. Modeling of dioxin levels in pine needles exposed to solid waste open combustion emissions. **Waste Management**, v. 79, p.510-515, set. 2018.
25. HAGENMAIER, H.; LINDIG, C.; SHE, J.. Correlation of environmental occurrence of polychlorinated dibenzo-p-dioxins and dibenzofurans with possible sources. **Chemosphere**, v. 29, n. 9-11, p.2163-2174, 1994.
26. HAGENMAIER, H.; BRUNNER, H.. Isomerspecific analysis of pentachlorophenol and sodium pentachlorophenate for 2,3,7,8-substituted pcdd and pcdf at sub-ppb levels. **Chemosphere**, v. 16, p.1759-1764, 1987.
27. HASSOUN, A.; SAHAR, A.; LAKHAL, L.; AÏT-KADDOUR, A. Fluorescence spectroscopy as a rapid and non-destructive method for monitoring quality and authenticity of fish and meat products: Impact of different preservation conditions. **Lwt**, v. 103, p.279-292, abr. 2019.
28. HOLCOMB, J.; FERRARIO, J.; BYRNE, C.. Analysis of mined clay products for pcdds/pcdfs by high-resolution mass spectrometry (hrgc/hrms) and quadrupole ion storage mass spectrometry/mass spectrometry (qisms/ms). **Organohalogen Compounds**, v. 40, p.137-141, 1999.

29. HOLT, E.; WEBER, R.; STEVENSON, G.; GAUS, C.. Formation of dioxins during exposure of pesticide formulations to sunlight. **Chemosphere**, [s.l.], v. 88, n. 3, p.364-370, 2012.
30. HOLT, E.; WEBER, R.; STEVENSON, G.; GAUS, C.. Polychlorinated dibenzo-p-dioxins and dibenzofurans (PCDD/Fs) impurities in pesticides: a neglected source of contemporary relevance. **Environmental Science Technology**, v. 44, p.5409-5415, 2010.
31. HOOGENBOOM, L.A.P.; KAN, C.A.; BOVEE, T.H.F.; WEG, G. van der; ONSTENK, C.; TRAAG, W.A.. Residues of dioxins and PCBs in fat of growing pigs and broilers fed contaminated feed. **Chemosphere**, v. 57, p.35-42, 2004.
32. HOOGENBOOM, L.A.P.; VAN EIJKEREN, J.C.H.; ZEILMAKER, M.J.; MENGELERS, M.I.B.; HERBES, R.; IMMERZEEL, J.; TRAAG, W.A.. A novel source for dioxins present in recycled fat from gelatin production. **Chemosphere**, v. 68, n. 5, p.814-823, jun. 2007.
33. HOOGENBOOM, R.; MALISCH, R.; DAM, G. T.; VAN LEEUWEN, S.; HOVE, H.; FERNANDES, A.; SCHÄTELE, A.; ROSE, M.. Congener patterns of polychlorinated dibenzo-p-dioxins and dibenzofurans as a useful aid to source identification during food contamination incidents. **Organohalogen Compounds**, v. 78, p.819-823, 2016.
34. HOOGENBOOM, R.; TRAAG, W.; FERNANDES, A.; ROSE, M.. European developments following incidents with dioxins and PCBs in the food and feed chain. **Food Control**, v. 50, p.670-683, abr. 2015.
35. HOOGENBOOM, R.; ZEILMAKER, M.; VAN EIJKEREN, J.; KAN, K.; MENGELERS, M.; LUYKX, D.; TRAAG, W.. Kaolinic clay derived PCDD/Fs in the feed chain from a sorting process for potatoes. **Chemosphere**, v. 78, n. 2, p.99-105, 2010.
36. HORSTMANN, M.; MCLACHLAN, M.S.. Results of an initial survey of polychlorinated dibenzo-p-dioxins (PCDD) and dibenzofurans (PCDF) in textiles. **Chemosphere**, v. 31, n. 2, p.2579-2589, 1995.
37. HU, J.; ZHENG, M.; NIE, Z.; LIU, W.; LIU, G.; ZHANG, B.; XIAO, K. Polychlorinated dibenzo-p-dioxin and dibenzofuran and polychlorinated biphenyl emissions from different smelting stages in secondary copper metallurgy. **Chemosphere**, v. 90, n. 1, p.89-94, 2013.
38. HUANG, J.; GAO, J.; YU, G.; YAMAZAKI, N.; DENG, S.; WANG, B.; WEBER, R. Unintentional formed PCDDs, PCDFs, and DL-PCBs as impurities in chinese pentachloronitrobenzene products. **Environmental Science And Pollution Research**, v. 22, n. 19, p.14462-14470, 30 ago. 2014.
39. HUANG, J.; MATSUMURA, T.; YU, G.; DENG, S.; YAMAUCHI, M.; YAMAZAKI, N.; WEBER, R.. Determination of PCBs, PCDDs and PCDFs in insulating oil samples from stored chinese electrical capacitors by HRGC/HRMS. **Chemosphere**, v. 85, n. 2, p.239-246, set. 2011.
40. IARC (International Agency for Research on Cancer). Chemical agents and related occupations volume 100 F A review of human carcinogens. IARC monographs on the evaluation of carcinogenic risks to humans; v. 100F. Lyon: International Agency for Research on Cancer. World Health Organization, p. 599, 2012.
41. JAKUBÍKOVÁ, M.; SÁDECKÁ, J.; KLEINOVÁ, A.. On the use of the fluorescence, ultraviolet-visible and near infrared spectroscopy with

- chemometrics for the discrimination between plum brandies of different varietal origins. **Food Chemistry**, v. 239, p.889-897, 2018.
42. JIANG, K.; LI, L.; CHEN, Y.; JIN, J.. Determination of PCDD/Fs and dioxin-like PCBs in chinese commercial PCBs and emissions from a testing PCB incinerator. **Chemosphere**, v. 34, p.941-950, 1997.
43. KANAN, S.; SAMARA, F.. Dioxins and furans: A review from chemical and environmental perspectives. **Trends In Environmental Analytical Chemistry**, v. 17, p.1-13, 2018.
44. KARABAGIAS, I. K.; LOUPPIS, A. P.; KARABOURNIOTI, S.; KONTAKOS, S.; PAPASTEPHANOU, C.; KONTOMINAS, M. G.. Characterization and geographical discrimination of commercial citrus spp. honeys produced in different Mediterranean countries based on minerals, volatile compounds and physicochemical parameters, using chemometrics. **Food Chemistry**, [s.l.], v. 217, p.445-455, 2017.
45. KIM, M.; KIM, D.; CHOI, S.; GUERRERO, P.; NORAMBUENA, J.; CHUNG, G.. Formation of polychlorinated dibenzo-p-dioxins/dibenzofurans (PCDD/Fs) from a refinery process for zinc oxide used in feed additives: a source of dioxin contamination in chilean pork. **Chemosphere**, v. 82, n. 9, p.1225-1229, 2011.
46. LIU, G.; JIANG, X.; WANG, M.; DONG, S.; ZHENG, M.. Comparison of PCDD/F levels and profiles in fly ash samples from multiple industrial thermal sources. **Chemosphere**, v. 133, p.68-74, ago. 2015.
47. LIU, G.; ZHENG, M.; BA, T.; LIU, W.; GUO, L.. A preliminary investigation on emission of polychlorinated dibenzo-p-dioxins/dibenzofurans and dioxin-like polychlorinated biphenyls from coke plants in China. **Chemosphere**, v. 75, n. 5, p.692-695, 2009.
48. LIU, W.; LIU, C.; YU, J.; ZHANG, Y.; LI, J.; CHEN, Y.; ZHENG, L.. Discrimination of geographical origin of extra virgin olive oils using terahertz spectroscopy combined with chemometrics. **Food Chemistry**, v. 251, p.86-92, jun. 2018.
49. LIU, W.; LI, H.; TAO, F.; LI, S.; TIAN, Z.; XIE, H.. Formation and contamination of PCDD/Fs, PCBs, PeCBz, HxCBz and polychlorophenols in the production of 2,4-D products. **Chemosphere**, v. 92, n. 3, p.304-308, 2013.
50. LIU, W.; TAO, F.; ZHANG, W.; LI, S.; ZHENG, M.. Contamination and emission factors of PCDD/Fs, unintentional PCBs, HxCBz, PeCBz and polychlorophenols in chloranil in China. **Chemosphere**, v. 86, n. 3, p.248-251, 2012.
51. LIU, W.; ZHENG, M.; WANG, D.; XING, Y.; ZHAO, X.; MA, X.; QIAN, Y.. Formation of PCDD/Fs and PCBs in the process of production of 1,4-dichlorobenzene. **Chemosphere**, v. 57, n. 10, p.1317-1323, 2004.
52. MALISCH, R. et al. Interactive data base of PCDD/F and PCB congener patterns to aid identification of contamination sources in feed and food. **Organohalogen Compounds**, v. 79, p.545-547, 2017.
53. MALISCH, R.. Incidents with dioxins and PCBs in food and feed-investigative work, risk management and economic consequences. **Journal Of Environmental Protection**, v. 08, n. 06, p.744-785, 2017.
54. MALISCH, R.. Increase of the PCDD/F: contamination of milk, butter and meat samples by use of contaminated citrus pulp. **Chemosphere**, v. 40, p.1041-1053, 2000.

55. MALISCH, R.; KOTZ, A.. Dioxins and PCBs in feed and food — review from european perspective. **Science Of The Total Environment**, v. 491-492, p.2-10, set. 2014.
56. MARTINS, A. R.; TALHAVINI, M.; VIEIRA, M. L.; ZACCA, J. J.; BRAGA, J. W. B.. Discrimination of whisky brands and counterfeit identification by UV–Vis spectroscopy and multivariate data analysis. **Food Chemistry**, v. 229, p.142-151, ago. 2017.
57. MASUNAGA, S.; TAKASUGA, T.; NAKANISHI, J.. Dioxin and dioxin-like PCB impurities in some Japanese agrochemical formulations. **Chemosphere**, v. 44, p.873-885, 2001.
58. MINGOTI, S. A.. **Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada**. Belo Horizonte: Editora UFMG, 2005. 300 p.
59. MINISTÉRIO DA AGRICULTURA, PECUÁRIA E ABASTECIMENTO. Instrução normativa nº 42, de 20 de dezembro de 1999. Diário Oficial da União de 22 de dez. 1999, Brasília, Seção 1, p.213. Disponível em: <<http://www.agricultura.gov.br/assuntos/inspecao/produtos-animal/plano-de-nacional-de-controle-de-residuos-e-contaminantes>> Acesso em: 28 maio, 2019.
60. MINISTÉRIO DA AGRICULTURA, PECUÁRIA E ABASTECIMENTO. Instrução normativa nº 9, de 12 de maio de 2016. Diário Oficial da União de 17 de maio. 2016, Brasília, Seção 1. Disponível em: <<http://www.agricultura.gov.br/assuntos/insumos-agropecuarios/insumos-pecuarios/alimentacao-animal/arquivos-alimentacao-animal/legislacao/instrucao-normativa-no-9-de-12-de-maio-de-2016.pdf>>. Acesso em: 18 outubro, 2019.
61. MIYABARA, Y.; HASHIMOTO, S.; SAGAI, M.; MORITA, M.. PCDDs and PCDFs in vehicle exhaust particles in Japan. **Chemosphere**, v. 39, n. 1, p.143-150, 1999.
62. MOKHTAR, M. M.; TAIB, R. M.; HASSIM, M. H.. Measurement of PCDD/Fs emissions from a coal-fired power plant in Malaysia and establishment of emission factors. **Atmospheric Pollution Research**, v. 5, n. 3, p.388-397, 2014.
63. NI, Y.; ZHANG, Z.; ZHANG, Q.; CHEN, J.; WU, Y.; LIANG, X.. Distribution patterns of PCDD/Fs in chlorinated chemicals. **Chemosphere**, v. 60, p.779-784, 2005.
64. NIE, Z.; LIU, G.; LIU, W.; ZHANG, B.; ZHENG, M.. Characterization and quantification of unintentional POP emissions from primary and secondary copper metallurgical processes in China. **Atmospheric Environment**, v. 57, p.109-115, set. 2012.
65. OMS- Organização Mundial da Saúde. The International Programme on Chemical Safety. 7 julho 2006. Disponível em <[https://www.who.int/ipcs/assessment/tef\\_values.pdf](https://www.who.int/ipcs/assessment/tef_values.pdf)>. Acesso em 10 maio. 2018.
66. PEREIRA, H. V.; AMADOR, V. S.; SENA, M. M.; AUGUSTI, R.; PICCIN, E.. Paper spray mass spectrometry and PLS-DA improved by variable selection for the forensic discrimination of beers. **Analytica Chimica Acta**, v. 940, p.104-112, out. 2016.

67. Programa das Nações Unidas para o meio ambiente. Diretrizes sobre melhores técnicas disponíveis e orientação provisional sobre melhores práticas ambientais conforme o artigo 5 e anexo C do convênio de Estocolmo sobre contaminantes orgânicos persistentes, 2008. Disponível em: <<http://www.pops.int/Implementation/BATandBEP/BATBEPGuidelinesArticle5/tabid/187/Default.aspx>> Acesso em: 25 abr. 2019.
68. RAPPE, C. Dioxin, patterns and source identification. **Journal Of Analytical Chemistry**, v. 348, p.63-75, 1994.
69. REMMERS, J.; DUPUY, A.; MCDANIEL, D.; HARLESS, R.; STEELE, D.. Polychlorinated dibenzo-p-dioxin and dibenzofuran contamination in chloranil and carbazole violet. **Chemosphere**, v. 25, n. 7, p.1505-1508, 1992.
70. REY, M.D.; FONT, R.; ARACIL, I.. PCDD/F emissions from light-duty diesel vehicles operated under highway conditions and a diesel-engine based power generator. **Journal of Hazardous Materials**, v. 278, p.116-123, ago. 2014.
71. REZAEI, E.; FARAHANI, A.; BUEKENS, A.; CHEN, T.; LU, S. Y.; HABIBINEJAD, M.; DAMERCHELI, F.; MOGHADAM, S.H.; GANDOMKAR, M.; BAHMANI, A.. Dioxins and furans releases in iranian mineral industries. **Chemosphere**, v. 91, n. 6, p.838-843, 2013.
72. RIGGS, K. B.; BROWN, T. D.; SCHROCK, M. E.. PCDD/PCDF emissions from coal-fired power plants. **Organohalogen Compounds**, v. 24, p.51-54, 1995.
73. RUIZ, M. L.; SANZ, P.; FABRELLAS, B.. Principal component analysis as a tool to elucidate the origin of PCDD/Fs releases to the atmosphere. **Organohalogen Compounds**, v. 56, p.261-264, 2002.
74. SABIR, A.; RAFI, M.; DARUSMAN, L. K.. Discrimination of red and white rice bran from Indonesia using HPLC fingerprint analysis combined with chemometrics. **Food Chemistry**, v. 221, p.1717-1722, abr. 2017.
75. SAKAI, S.; HAYAKAWA, K.; TAKATSUKI, H.; KAWAKAMI, I.. Dioxin-like PCBs released from waste incineration and their deposition flux. **Environmental Science Technology**, v. 35, p.3601-3607, 2001.
76. SANTANA, F. B.; SOUZA, A.M.; ALMEIDA, M.R.; BREITKREITZ, M.C.; FILGUEIRAS, P.R.; SENA, M.M.; POPPI, R.J. Experimento didático de quimiometria por espectroscopia no infravermelho médio combinado com análise discriminante por mínimos quadrados parciais um tutorial, parte V. **Química Nova**, v.43, 371-381, 2020.
77. SALTHAMMER, T.; KLIPP, H.; PEEK, R-d.; MARUTZKY, R.. Formation of polychlorinated dibenzo-p-dioxins (PCDD) and polychlorinated dibenzofurans (PCDF) during the combustion of impregnated wood. **Chemosphere**, v. 30, p.2051-2060, 1995.
78. SANTL, H.; GRUBER, L.; STÖHRER, E.. Some new sources of polychlorinated dibenzodioxins (PCDDs) and dibenzofurans (PCDFs) in waste papers and recycled pulps. **Chemosphere**, v. 29, n. 9-11, p.1995-2003, 1994.
79. SCHECTER, A.; PÄPKE, O.; LIS, A.; BALL, M.. Chlorinated dioxin and dibenzofuran content in 2,4-D amine salt form UFA, Russia. **Organohalogen Compounds**, v. 11, p.325-328, 1993.
80. SENA, M. M de; POPPI, R. J.; FRIGHETTO, R. T. S.; VALARINI, P. J.. Avaliação do uso de métodos quimiométricos em análise de solos. **Química Nova**, v. 23, n. 4, p.547-556, 2000.

81. SOUZA, A. M.; BREITKREITZ, M. C.; FILGUEIRAS, P. R.; ROHWEDDER, J. J. R.; POPPI, R. J.. Experimento didático de quimiometria para calibração multivariada na determinação de paracetamol em comprimidos comerciais utilizando espectroscopia no infravermelho próximo: um tutorial, parte ii. **Química Nova**, v. 36, n. 7, p.1057-1065, 2013.
82. SQUADRONE, S.; BRIZIO, P.; NESPOLI, R.; STELLA, C.; ABETE, M.c.. Human dietary exposure and levels of polychlorinated dibenzo-p-dioxins (PCDDs), polychlorinated dibenzofurans (PCDFs), dioxin-like polychlorinated biphenyls (DL-PCBs) and non-dioxin-like polychlorinated biphenyls (NDL-PCBs) in free-range eggs close to a secondary aluminum smelter, northern Italy. **Environmental Pollution**, v. 206, p.429-436, 2015.
83. STRINGER, R.I.; COSTNER, P.; JOHNSTON, P.A.. PVC manufacture as a source of PCDD/Fs. **Organohalogen Compounds**, v. 24, p.119-123, 1995.
84. SUN, H.; GE, X.; LV, Y.; WANG, A.. Application of accelerated solvent extraction in the analysis of organic contaminants, bioactive and nutritional compounds in food and feed. **Journal Of Chromatography A**, v. 1237, p.1-23, 2012.
85. TAKASUGA, T.; KUMAR, K. S.; NOMA, Y.; SAKAI, S.. Chemical characterization of polychlorinated biphenyls, -dibenzo-p-dioxins, and -dibenzofurans in technical Kanechlor PCB formulations in Japan. **Archives Of Environmental Contamination And Toxicology**, v. 49, n. 3, p.385-395, 2005.
86. THUB, U.; POPP, P.. Domestic Lignite Combustion as source of polychlorodibenzodioxins and -furans (PCDD/F). **Chemosphere**, v. 31, p.2591-2604, 1995.
87. TLUSTOS, C.. The dioxin contamination incident in Ireland 2008. **Organohalogen Compounds**, v. 71, p.1155-1159, 2009.
88. TONDEUR, Y.; NESTRICK, T.; SILVA, H. A.; VINING, B.; HART, J.. Analytical procedures for the determination of polychlorinated-p-dioxins, polychlorinated dibenzofurans, and hexachlorobenzene in pentachlorophenol. **Chemosphere**, v. 80, n. 2, p.157-164, jun. 2010.
89. TOOLKIT for identification and quantification of releases of dioxins, furans and other unintentional POPs under article 5 of the Stockholm convention on persistent organic pollutants, Stockholm Convention on persistent organic pollutants (POPs). Programa das Nações Unidas Para O Meio Ambiente: 2013. Disponível em: <<https://toolkit.pops.int/Publish/Downloads/UNEP-POPS-TOOLKIT-2012-En.pdf>> Acesso em: 25 abr. 2019.
90. VEJERANO, E. P.; HOLDER, A. L.; MARR, L. C.. Emissions of polycyclic aromatic hydrocarbons, polychlorinated dibenzo-p-Dioxins, and dibenzofurans from incineration of nanomaterials. **Environmental Science & Technology**, v. 47, n. 9, p.4866-4874, 10 abr. 2013.
91. WANG, L.; DING, G.; ZHOU, Z.; LIU, X.; WANG, Y.; XIE, H. Q.; XU, T.; WANG, P.; ZHAO, B. Patterns and dietary intake of polychlorinated dibenzo- p -dioxins and polychlorinated dibenzofurans in food products in China. **Journal Of Environmental Sciences**, v. 51, p.165-172, 2017.
92. WEBER, R.; A WATSON,. Assessment of the PCDD/PCDF fingerprint of the dioxin food scandal from bio-diesel in Germany and possible PCDD/F sources. **Organohalogen Compounds**, v. 73, p.400-403, 2011.

93. WEBER, R.; HEROLD, C.; HOLLERT, H.; KAMPHUES, J.; BLEPP, M.; BALLSCHMITER, K.. Reviewing the relevance of dioxin and PCB sources for food from animal origin and the need for their inventory, control and management. **Environmental Sciences Europe**, v. 30, n. 1, p.1-42, 2018.
94. WEBER, R.; VARBELOW, H. G.. Dioxin/POPs legacy of pesticide production in Hamburg: part 1—securing of the production area. **Environmental Science and Pollution Research**, v. 20, n. 4, p.1918-1924, 2012.
95. WENNING, R.; PAUSTENBACH, D.; JOHNSON, G.; EHRLICH, R.; HARRIS, M.; BEDBURY, H.. Chemometric analysis of potential sources of polychlorinated dibenzo-p-dioxins and dibenzofurans in surficial sediments from Newark bay, New Jersey. **Chemosphere**, v. 27, n. 1-3, p.55-64, 1993.
96. WILLIAMS, D. T.; LEBEL, G. L.; BENOIT, F. M.. Polychlorodibenzodioxins and polychlorodibenzofurans in dioxazine dyes and pigments. **Chemosphere**, v. 24, n. 2, p.169-180, 1992.
97. YASUHARA, A.; KATAMI, T.; OKUDA, T.; OHNO, N.; SHIBAMOTO, T.. Formation of dioxins during the combustion of newspapers in the presence of sodium chloride and poly(vinyl chloride). **Environmental Science Technology**, v. 35, p.1373-1378, 2001.
98. ZHANG, M.; BUEKENS, A.; OLIE, K.; LI, X.. PCDD/F-isomers signature - effect of metal chlorides and oxides. **Chemosphere**, v. 184, p.559-568, out. 2017.
99. ZHANG, R.; LUO, Y.; YIN, R.. Experimental study on dioxin formation in an MSW gasification-combustion process: an attempt for the simultaneous control of dioxins and nitrogen oxides. **Waste Management**, v. 82, p.292-301, 2018.
100. ZHAO, Z.; NI, M.; LI, X.; CHEN, T.; BUEKENS, A.; YAN, J.. PCDD/F formation during thermal desorption of chlorobenzene contaminated soil. **Environmental Science And Pollution Research**, v. 24, n. 29, p.23321-23330, 24 ago. 2017.
101. ZHU, J.; HIRAI, Y.; YU, G.; SAKAI, S.. Levels of polychlorinated dibenzo-p-dioxins and dibenzofurans in China and chemometric analysis of potential emission sources. **Chemosphere**, v. 70, n. 4, p.703-711, 2008.

**Apêndice A - Tabela com a codificação dos perfis compilados, código da classe, somatório TEQ TOTAL (ng/Kg) e a respectiva referência bibliográfica.**

CÓDIGO AMOSTRA	CÓDIGO CLASSIFICAÇÃO	TEQ TOTAL	REFERÊNCIA
AGR - 3	AGR	461.046	HAGENMAIER et al., 1987
AGR - 4	AGR	60.011	HAGENMAIER et al., 1987
AGR - 5	AGR	120	SCHECTER et al., 1993
AGR - 6	AGR	2.510.047	MASUNAGA et al., 2001
AGR - 7	AGR	38.604	MASUNAGA et al., 2001
AGR - 8	AGR	75.027	MASUNAGA et al., 2001
AGR - 9	AGR	35.807	MASUNAGA et al., 2001
AGR - 10	AGR	900.500	MASUNAGA et al., 2001
AGR - 11	AGR	411.651	MASUNAGA et al., 2001
AGR - 12	AGR	11.717	MASUNAGA et al., 2001
AGR - 13	AGR	746	MASUNAGA et al., 2001
AGR - 14	AGR	522	MASUNAGA et al., 2001
AGR - 15	AGR	188	MASUNAGA et al., 2001
AGR - 16	AGR	17,2	MASUNAGA et al., 2001
AGR - 17	AGR	8,9	MASUNAGA et al., 2001
AGR - 18	AGR	2,2	MASUNAGA et al., 2001
AGR - 19	AGR	13,2	MASUNAGA et al., 2001
AGR - 20	AGR	2.656.844	SANTL et al., 1994
AGR - 21	AGR	1.641.901	DHARMARATHNE et al., 2019
AGR - 22	AGR	6.242.300	DHARMARATHNE et al., 2019
AGR - 23	AGR	5.049.899	DHARMARATHNE et al., 2019
AGR - 24	AGR	499	LIU et al., 2004
AGR - 25	AGR	1.470	LIU et al., 2004
AGR - 26	AGR	2.795	LIU et al., 2004
AGR - 27	AGR	33,9	LIU et al., 2004
AGR - 28	AGR	16,4	HOLT et al., 2010
AGR - 29	AGR	10,8	HOLT et al., 2010
AGR - 30	AGR	3,2	HOLT et al., 2010
AGR - 31	AGR	2,7	HOLT et al., 2010
AGR - 32	AGR	1,9	HOLT et al., 2010
AGR - 33	AGR	3,6	HOLT et al., 2010
AGR - 34	AGR	85,9	HOLT et al., 2010
AGR - 35	AGR	72,9	HOLT et al., 2010
AGR - 36	AGR	81,6	HOLT et al., 2010
AGR - 37	AGR	2,0	HOLT et al., 2010
AGR - 38	AGR	5,1	HOLT et al., 2010
AGR - 39	AGR	56,8	HOLT et al., 2010
AGR - 40	AGR	53,8	HOLT et al., 2010
AGR - 41	AGR	1.706	HOLT et al., 2010

CÓDIGO AMOSTRA	CÓDIGO CLASSIFICAÇÃO	TEQ TOTAL	REFERÊNCIA
AGR - 42	AGR	1.816	HOLT et al., 2010
AGR - 43	AGR	2.940	HOLT et al., 2010
AGR - 44	AGR	2.831	HOLT et al., 2010
AGR - 45	AGR	4.205	HOLT et al., 2010
AGR - 46	AGR	5,3	HOLT et al., 2010
AGR - 47	AGR	9,4	HOLT et al., 2010
AGR - 48	AGR	7,8	HOLT et al., 2010
AGR - 49	AGR	13.627	HOLT et al., 2010
AGR - 50	AGR	2,3	ZHAO et al., 2017
AGR - 51	AGR	3,0	ZHAO et al., 2017
AGR - 52	AGR	156.212	WEBER et al., 2013
AGR - 53	AGR	392.360	WEBER et al., 2013
AGR - 54	AGR	230.403.600	WEBER et al., 2013
AGR - 55	AGR	272.484	WEBER et al., 2013
AGR - 56	AGR	73,3	HOTL et al., 2012
AGR - 57	AGR	13,4	LIU et al., 2013
AGR - 58	AGR	268	LIU et al., 2013
AGR - 59	AGR	293	LIU et al., 2013
AGR - 60	AGR	891	LIU et al., 2013
AGR - 61	AGR	695	LIU et al., 2013
AGR - 62	AGR	262	HUANG et al., 2014
AGR - 63	AGR	931	HUANG et al., 2014
AGR - 64	AGR	161	HUANG et al., 2014
AGR - 65	AGR	657	HUANG et al., 2014
AGR - 66	AGR	387	HUANG et al., 2014
AGR - 67	AGR	453.780	TONDEUR et al., 2010
AGR - 68	AGR	11.250	GÖTZ et al., 2017
AGR - 69	AGR	24.326	GÖTZ et al., 2017
AGR - 70	AGR	405	GÖTZ et al., 2017
AGR - 71	AGR	572	GÖTZ et al., 2017
AGR - 72	AGR	1.868	GÖTZ et al., 2017
AGR - 73	AGR	10.605	GÖTZ et al., 2017
AGR - 74	AGR	2.909	GÖTZ et al., 2017
AGR - 75	AGR	685.055	GÖTZ et al., 2017
AGR - 76	AGR	272.484	GÖTZ et al., 2017
AGR - 77	AGR	308.024	GÖTZ et al., 2017
AGR - 78	AGR	229.535	GÖTZ et al., 2017
AGR - 79	AGR	122.080	GÖTZ et al., 2017
AGR - 80	AGR	25.039	GÖTZ et al., 2017
AGR - 81	AGR	304.828	GÖTZ et al., 2017
AGR - 82	AGR	223.760	GÖTZ et al., 2017
AGR - 83	AGR	126.822	GÖTZ et al., 2017

CÓDIGO AMOSTRA	CÓDIGO CLASSIFICAÇÃO	TEQ TOTAL	REFERÊNCIA
AGR - 84	AGR	77.198	GÖTZ et al., 2017
AGR - 85	AGR	205.149	GÖTZ et al., 2017
AGR - 86	AGR	250.611	GÖTZ et al., 2017
AGR - 87	AGR	307.725	GÖTZ et al., 2017
AGR - 88	AGR	308.722	GÖTZ et al., 2017
AGR - 89	AGR	369.092	GÖTZ et al., 2017
PCB - 1	PCB	47,2	TAKASUGA et al., 2005
PCB - 2	PCB	80,1	TAKASUGA et al., 2005
PCB - 3	PCB	236	TAKASUGA et al., 2005
PCB - 4	PCB	187	TAKASUGA et al., 2005
PCB - 5	PCB	176	TAKASUGA et al., 2005
PCB - 6	PCB	162	TAKASUGA et al., 2005
PCB - 7	PCB	69.349	HUANG et al., 2011
PCB - 8	PCB	70.578	HUANG et al., 2011
PCB - 9	PCB	186.846	JIANG et al., 1997
PCB - 10	PCB	314.807	JIANG et al., 1997
PCB - 11	PCB	81,8	COVACI et al., 2002
PGM - 1	PGM	82.930	REMMERS et al., 1992
PGM - 2	PGM	2.608.300	REMMERS et al., 1992
PGM - 3	PGM	247.260	REMMERS et al., 1992
PGM - 4	PGM	2.938.800	REMMERS et al., 1992
PGM - 5	PGM	168.460	REMMERS et al., 1992
PGM - 6	PGM	88,4	SANTL et al., 1994
PGM - 7	PGM	60,7	SANTL et al., 1994
PGM - 8	PGM	41,6	SANTL et al., 1994
PGM - 9	PGM	17,7	SANTL et al., 1994
PGM - 10	PGM	19.535	WILLIAMS et al., 1992
PGM - 11	PGM	9.551	WILLIAMS et al., 1992
PGM - 12	PGM	6.352	WILLIAMS et al., 1992
PGM - 13	PGM	100	WILLIAMS et al., 1992
PGM - 14	PGM	17.871	WILLIAMS et al., 1992
PGM - 15	PGM	1.189	WILLIAMS et al., 1992
PGM - 16	PGM	12.501	WILLIAMS et al., 1992
PGM - 17	PGM	9.186	WILLIAMS et al., 1992
PGM - 18	PGM	2.362	WILLIAMS et al., 1992
PGM - 19	PGM	2.114	WILLIAMS et al., 1992
PGM - 20	PGM	148	LIU et al., 2012
PGM - 21	PGM	22.251	LIU et al., 2012
PGM - 22	PGM	1.292.172	LIU et al., 2012
PGM - 23	PGM	81,8	LIU et al., 2012
PGM - 24	PGM	62,0	LIU et al., 2012
PGM - 25	PGM	1,6	HORSTMANN et al., 1995

CÓDIGO AMOSTRA	CÓDIGO CLASSIFICAÇÃO	TEQ TOTAL	REFERÊNCIA
PGM - 26	PGM	1,5	HORSTMANN et al., 1995
PGM - 27	PGM	6,3	HORSTMANN et al., 1995
PGM - 28	PGM	1,4	HORSTMANN et al., 1995
PGM - 29	PGM	1,1	HORSTMANN et al., 1995
PGM - 30	PGM	17,0	HORSTMANN et al., 1995
PGM - 31	PGM	1,4	HORSTMANN et al., 1995
PGM - 32	PGM	179	HORSTMANN et al., 1995
PGM - 33	PGM	205	HORSTMANN et al., 1995
PGM - 34	PGM	60,3	NI et al., 2005
PGM - 35	PGM	653	NI et al., 2005
PGM - 36	PGM	120	NI et al., 2005
PGM - 37	PGM	11,8	NI et al., 2005
CIM - 1	CIM	100	REZAEI et al., 2013
CIM - 2	CIM	102	REZAEI et al., 2013
CIM - 3	CIM	45,3	REZAEI et al., 2013
CIM - 4	CIM	135	REZAEI et al., 2013
CIM - 5	CIM	99,2	REZAEI et al., 2013
CIM - 6	CIM	100	REZAEI et al., 2013
CIM - 7	CIM	99,0	REZAEI et al., 2013
MTL - 1	MTL	3,9	U.S.EPA, 2006
MTL - 2	MTL	7,4	U.S.EPA, 2006
MTL - 3	MTL	400	U.S.EPA, 2006
MTL - 4	MTL	14.010	U.S.EPA, 2006
MTL - 5	MTL	3,1	U.S.EPA, 2006
MTL - 6	MTL	2,9	U.S.EPA, 2006
MTL - 7	MTL	11,1	U.S.EPA, 2006
MTL - 8	MTL	45,4	U.S.EPA, 2006
MTL - 9	MTL	1,1	HU et al., 2013
MTL - 10	MTL	4,0	NIE et al., 2012
MTL - 11	MTL	2,6	NIE et al., 2012
MTL - 12	MTL	152	NIE et al., 2012
MTL - 13	MTL	1.377	NIE et al., 2012
MTL - 14	MTL	2,2	NIE et al., 2012
MTL - 15	MTL	3.723	NIE et al., 2012
MTL - 16	MTL	36.740	NIE et al., 2012
MTL - 17	MTL	32,8	CHEN et al., 2016
MTL - 18	MTL	101	REZAEI et al., 2013
MTL - 19	MTL	108	REZAEI et al., 2013
MTL - 20	MTL	128	REZAEI et al., 2013
MTL - 21	MTL	102	REZAEI et al., 2013
MTL - 22	MTL	17.685	GÖTZ et al., 2017
MTL - 23	MTL	5.680	GÖTZ et al., 2017

CÓDIGO AMOSTRA	CÓDIGO CLASSIFICAÇÃO	TEQ TOTAL	REFERÊNCIA
MTL - 24	MTL	10.488	GÖTZ et al., 2017
MTL - 25	MTL	10.596	GÖTZ et al., 2017
MTL - 26	MTL	1.509	GÖTZ et al., 2017
MTL - 27	MTL	17.561	GÖTZ et al., 2017
MTL - 28	MTL	12.782	GÖTZ et al., 2017
MTL - 29	MTL	7.121	GÖTZ et al., 2017
MTL - 30	MTL	16.049	GÖTZ et al., 2017
MTL - 31	MTL	4.011	GÖTZ et al., 2017
MTL - 32	MTL	4.442	GÖTZ et al., 2017
MTL - 33	MTL	12.429	GÖTZ et al., 2017
MTL - 34	MTL	6.810	GÖTZ et al., 2017
MTL - 35	MTL	6.137	GÖTZ et al., 2017
MTL - 36	MTL	1.986	GÖTZ et al., 2017
MTL - 37	MTL	2.675	GÖTZ et al., 2017
MTL - 38	MTL	2.134	GÖTZ et al., 2017
MTL - 39	MTL	3.680	GÖTZ et al., 2017
MTL - 40	MTL	2.389	GÖTZ et al., 2017
MTL - 41	MTL	2.895	GÖTZ et al., 2017
MTL - 42	MTL	2.149	GÖTZ et al., 2017
MTL - 43	MTL	9.518	GÖTZ et al., 2017
MTL - 44	MTL	62,5	GÖTZ et al., 2017
MTL - 45	MTL	264	GÖTZ et al., 2017
MTL - 46	MTL	70.686	GÖTZ et al., 2017
MTL - 47	MTL	34.650	GÖTZ et al., 2017
MTL - 48	MTL	166.319	GÖTZ et al., 2017
MTL - 49	MTL	95.523	GÖTZ et al., 2017
MTL - 50	MTL	74.545	GÖTZ et al., 2017
MTL - 51	MTL	83.824	GÖTZ et al., 2017
MTL - 52	MTL	64.916	GÖTZ et al., 2017
MTL - 53	MTL	69.701	GÖTZ et al., 2017
MTL - 54	MTL	44.003	GÖTZ et al., 2017
MTL - 55	MTL	45.370	GÖTZ et al., 2017
MTL - 56	MTL	61.989	GÖTZ et al., 2017
MTL - 57	MTL	14.385	GÖTZ et al., 2017
MTL - 58	MTL	34.211	GÖTZ et al., 2017
MTL - 59	MTL	46.266	GÖTZ et al., 2017
MTL - 60	MTL	36.131	GÖTZ et al., 2017
MTL - 61	MTL	49.633	GÖTZ et al., 2017
MTL - 62	MTL	7.227	GÖTZ et al., 2017
MTL - 63	MTL	20.256	GÖTZ et al., 2017
MTL - 64	MTL	51.175	GÖTZ et al., 2017
MTL - 65	MTL	33.286	GÖTZ et al., 2017

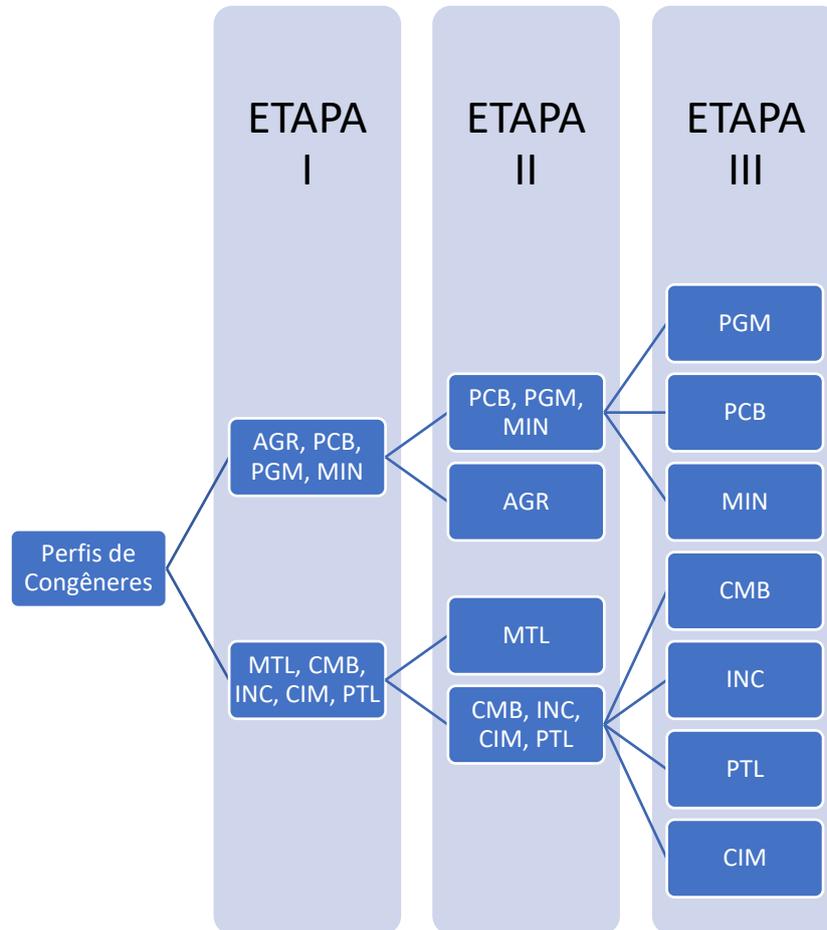
CÓDIGO AMOSTRA	CÓDIGO CLASSIFICAÇÃO	TEQ TOTAL	REFERÊNCIA
MTL - 66	MTL	61.086	GÖTZ et al., 2017
MTL - 67	MTL	44.936	GÖTZ et al., 2017
MTL - 68	MTL	46,7	GÖTZ et al., 2017
MTL - 69	MTL	31,0	GÖTZ et al., 2017
MTL - 70	MTL	3,0	GÖTZ et al., 2017
MTL - 71	MTL	27,7	GÖTZ et al., 2017
MTL - 72	MTL	2,1	GÖTZ et al., 2017
MTL - 73	MTL	1,1	GÖTZ et al., 2017
MTL - 74	MTL	10,6	GÖTZ et al., 2017
MTL - 75	MTL	8,5	GÖTZ et al., 2017
MTL - 76	MTL	8,4	GÖTZ et al., 2017
MTL - 77	MTL	9,3	GÖTZ et al., 2017
MTL - 78	MTL	2,8	GÖTZ et al., 2017
MTL - 79	MTL	1,5	GÖTZ et al., 2017
MTL - 80	MTL	1,0	GÖTZ et al., 2017
MTL - 81	MTL	1,2	GÖTZ et al., 2017
MTL - 82	MTL	1,6	GÖTZ et al., 2017
MIN - 1	MIN	1,9	FERRARIO et al., 2000
MIN - 2	MIN	6,8	FERRARIO et al., 2000
MIN - 3	MIN	13,2	FERRARIO et al., 2000
MIN - 4	MIN	5,6	FERRARIO et al., 2000
MIN - 5	MIN	29,9	FERRARIO et al., 2000
MIN - 6	MIN	1.625	HOOGENBOOM et al., 2010
MIN - 7	MIN	1.358	HOOGENBOOM et al., 2010
MIN - 8	MIN	249	HOOGENBOOM et al., 2010
MIN - 9	MIN	206	HOOGENBOOM et al., 2010
MIN - 10	MIN	292	HOOGENBOOM et al., 2010
MIN - 11	MIN	967	HOOGENBOOM et al., 2010
MIN - 12	MIN	238	HOOGENBOOM et al., 2010
MIN - 13	MIN	169	HOOGENBOOM et al., 2010
MIN - 14	MIN	306	HOOGENBOOM et al., 2010
MIN - 15	MIN	259	HOOGENBOOM et al., 2010
MIN - 16	MIN	32,0	HOOGENBOOM et al., 2010
MIN - 17	MIN	1.804	HOOGENBOOM et al., 2010
MIN - 18	MIN	13,4	HOLCOMB et al., 1999
MIN - 19	MIN	29,9	HOLCOMB et al., 1999
MIN - 20	MIN	6,8	HOLCOMB et al., 1999
MIN - 21	MIN	235	ELJARRAT et al., 2002
MIN - 22	MIN	464	ELJARRAT et al., 2002
CMB - 1	CMB	3,0	MIYABARA et al., 1999
CMB - 2	CMB	4,1	MIYABARA et al., 1999
CMB - 3	CMB	2,6	MIYABARA et al., 1999

CÓDIGO AMOSTRA	CÓDIGO CLASSIFICAÇÃO	TEQ TOTAL	REFERÊNCIA
CMB - 4	CMB	15,5	MIYABARA et al., 1999
CMB - 5	CMB	6,8	MIYABARA et al., 1999
CMB - 6	CMB	15,9	MIYABARA et al., 1999
CMB - 7	CMB	233	MIYABARA et al., 1999
CMB - 8	CMB	82,4	THUB et al., 1995
CMB - 9	CMB	131	THUB et al., 1995
CMB - 10	CMB	100	THUB et al., 1995
CMB - 11	CMB	11,7	THUB et al., 1995
CMB - 12	CMB	21,2	THUB et al., 1995
CMB - 13	CMB	10,4	THUB et al., 1995
CMB - 14	CMB	1,2	RIGGS et al., 1995
CMB - 15	CMB	5,1	RIGGS et al., 1995
CMB - 16	CMB	5,5	RIGGS et al., 1995
CMB - 17	CMB	1,3	RIGGS et al., 1995
CMB - 18	CMB	2,0	SALTHAMMER et al., 1995
CMB - 19	CMB	1,4	SALTHAMMER et al., 1995
CMB - 20	CMB	1,6	SALTHAMMER et al., 1995
CMB - 21	CMB	1,4	SALTHAMMER et al., 1995
CMB - 22	CMB	1,8	SALTHAMMER et al., 1995
CMB - 23	CMB	1,5	SALTHAMMER et al., 1995
CMB - 24	CMB	1.693	SALTHAMMER et al., 1995
CMB - 25	CMB	1.508	SALTHAMMER et al., 1995
CMB - 26	CMB	1.534	SALTHAMMER et al., 1995
CMB - 27	CMB	1.079	SALTHAMMER et al., 1995
CMB - 28	CMB	1.706	SALTHAMMER et al., 1995
CMB - 29	CMB	1.577	SALTHAMMER et al., 1995
CMB - 30	CMB	1.420	SALTHAMMER et al., 1995
CMB - 31	CMB	5,5	REY et al., 2014
CMB - 32	CMB	8,4	REY et al., 2014
CMB - 33	CMB	4,0	REY et al., 2014
CMB - 34	CMB	1,8	REY et al., 2014
CMB - 35	CMB	15,1	BLACK et al., 2012
CMB - 36	CMB	1,4	LIU et al., 2009
INC - 1	INC	390	SAKAI et al., 2001
INC - 2	INC	10,5	SAKAI et al., 2001
INC - 3	INC	8,2	YASUHARA et al., 2001
INC - 4	INC	11,2	YASUHARA et al., 2001
INC - 5	INC	26,6	YASUHARA et al., 2001
INC - 6	INC	2.691	YASUHARA et al., 2001
INC - 7	INC	2.794	YASUHARA et al., 2001
INC - 8	INC	4.323	YASUHARA et al., 2001

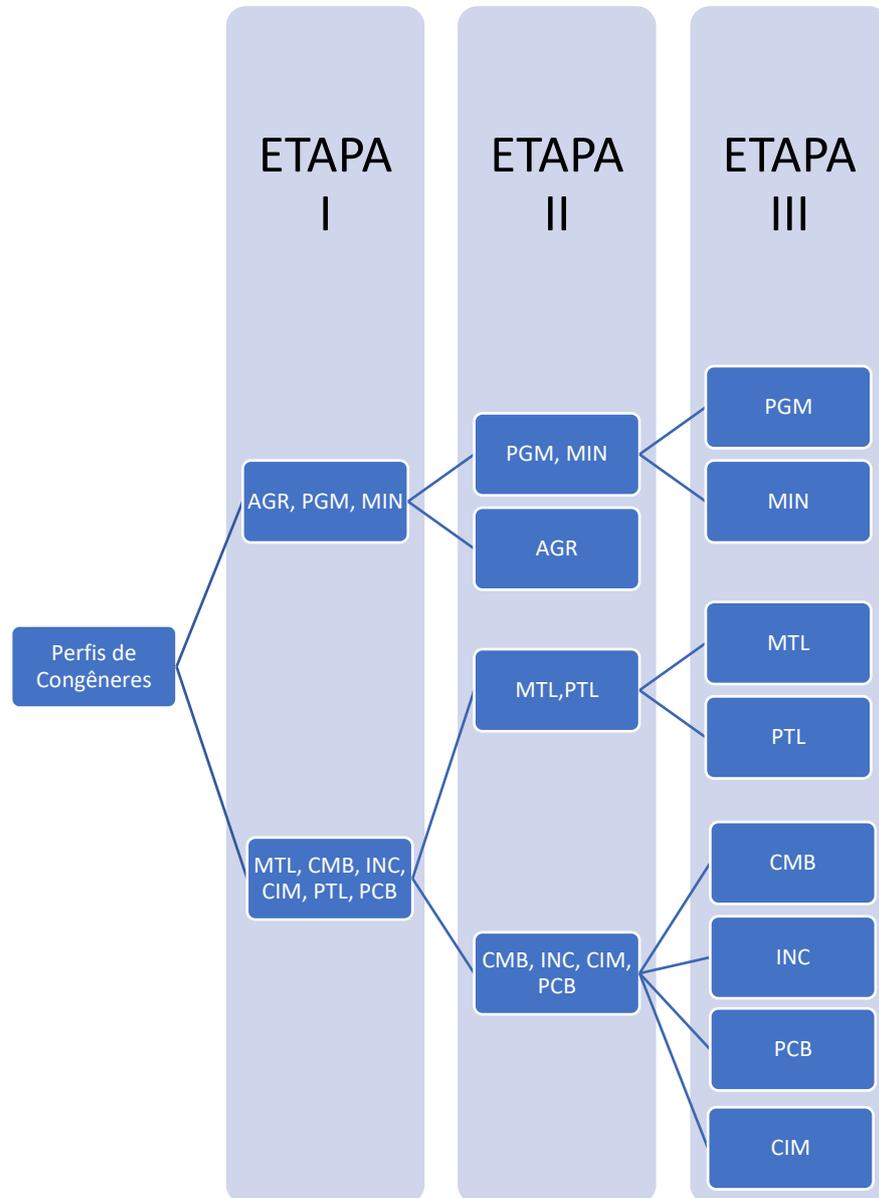
CÓDIGO AMOSTRA	CÓDIGO CLASSIFICAÇÃO	TEQ TOTAL	REFERÊNCIA
INC - 9	INC	14,5	YASUHARA et al., 2001
INC - 10	INC	1,4	ZHANG et al., 2018
INC - 11	INC	1,0	ZHANG et al., 2018
INC - 12	INC	1,0	ZHANG et al., 2018
INC - 13	INC	1,0	ZHANG et al., 2018
INC - 14	INC	2,5	ZHANG et al., 2018
PLT - 1	PLT	19.255	STINGER et al., 1995
PLT - 2	PLT	819.9218	STINGER et al., 1995
PLT - 3	PLT	2.648	STINGER et al., 1995
PLT - 4	PLT	451	VEJERANO et al., 2013
PLT - 5	PLT	426	VEJERANO et al., 2013
PLT - 6	PLT	1.473	CARROL Jr et al., 1999
PLT - 7	PLT	13.622	CARROL Jr et al., 1999
PLT - 8	PLT	13.396	CARROL Jr et al., 1999
PLT - 9	PLT	12.825	CARROL Jr et al., 1999
PLT - 10	PLT	142.522	CARROL Jr et al., 1999
PLT - 11	PLT	138.171	CARROL Jr et al., 1999
PLT - 12	PLT	92,1	CARROL Jr et al., 1999
PLT - 13	PLT	418	CARROL Jr et al., 1999
PLT - 14	PLT	21.479	CARROL Jr et al., 1999
PLT - 15	PLT	136	FONT et al., 2010
PLT - 16	PLT	200.366	FONT et al., 2010
PLT - 17	PLT	7.369.007	FONT et al., 2010
PLT - 18	PLT	6.587	FONT et al., 2010
PLT - 19	PLT	3.961	FONT et al., 2010

## Apêndice B - Hierarquia dos modelos de classificação sugeridos

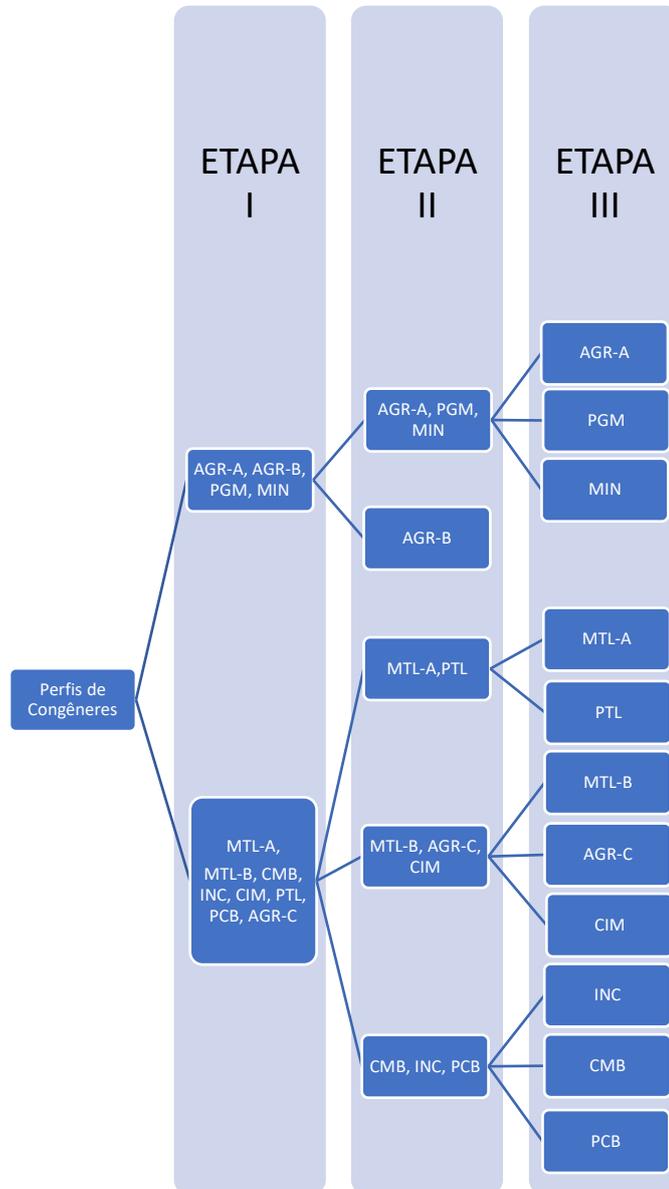
### Hierarquia de classificação do Modelo I



## Hierarquia de classificação do Modelo II



### Hierarquia de classificação do Modelo III



## Apêndice C

Tabela 16: Número de VLs, variâncias explicadas e erros de classificação de validação cruzada (ECVC) referente aos modelos em PLS-Da para as etapas de classificação (modelo I)

ETAPA	GRUPOS SEPARADOS	VARIÁVEIS LATENTES	Variância Explicada X (%)	Variância Explicada Y (%)	ECVC
I	AGR, PCB, PGM, MIN	7	87.94	57.75	0.11
I	MTL, CMB, INC, CIM, PTL	7	87.94	57.75	0.11
II	AGR	4	66.61	51.80	0.22
II	PCB, PGM, MIN	4	66.61	51.80	0.22
II	MTL	7	86.34	53.51	0.19
II	CMB, INC, CIM, PTL	7	86.34	53.51	0.19
III	PGM	4	82.78	84.80	0.13
III	MIN	5	90.48	88.51	0.04
III	CMB	6	75.07	59.35	0.33
III	INC	6	81.03	35.95	0.44
III	CIM	3	59.19	47.40	0.35
III	PTL	1	33.84	55.53	0.11

Tabela 17: Figuras de mérito estimadas para etapa do modelo de classificação – Conjunto de treinamento (modelo I)

ETAPA	GRUPOS SEPARADOS	SENSIBILIDADE	ESPECIFICIDADE	FALSO POSITIVO	FALSO NEGATIVO
I	AGR, PCB, PGM, MIN	83.3	95.0	16.7	5.0
I	MTL, CMB, INC, CIM, PTL	95.0	83.3	5.0	16.7
II	AGR	79.2	86.0	20.8	14.0
II	PCB, PGM, MIN	86.0	79.2	14.0	20.8
II	MTL	90.0	80.0	10.0	20.0
II	CMB, INC, CIM, PTL	80.0	90.0	20.0	10.0
III	PGM	100.0	100.0	0.0	0.0
III	MIN	100.0	96.4	0.0	3.6
III	CMB	95.7	81.5	4.3	18.5

ETAPA	GRUPOS SEPARADOS	SENSIBILIDADE	ESPECIFICIDADE	FALSO POSITIVO	FALSO NEGATIVO
III	INC	88.9	82.9	11.1	17.1
III	CIM	100.0	93.3	0.0	6.7
III	PTL	92.3	89.2	7.7	10.8

Tabela 18: Figuras de mérito estimadas para etapa do modelo de classificação – Conjunto de teste (modelo I)

ETAPA	GRUPOS SEPARADOS	SENSIBILIDADE	ESPECIFICIDADE	FALSO POSITIVO	FALSO NEGATIVO
I	AGR, PCB, PGM, MIN	72.3	94.0	27.7	6.0
I	MTL, CMB, INC, CIM, PTL	94.0	72.3	6.0	27.7
II	AGR	57.7	61.9	42.3	38.1
II	PCB, PGM, MIN	61.9	57.7	38.1	42.3
II	MTL	40.0	84.0	60.0	16.0
II	CMB, INC, CIM, PTL	84.0	40.0	16.0	60.0
III	PGM	90.0	100.0	10.0	0.0
III	MIN	100.0	85.7	0.0	14.3
III	CMB	91.7	100.0	8.3	0.0
III	INC	100.0	75.0	0.0	25.0
III	CIM	100.0	87.0	0.0	13.0
III	PTL	66.7	89.5	33.3	10.5

Tabela 19: Número de VLs, variâncias explicadas e erros de classificação de validação cruzada (ECVC) referente aos modelos em PLS-Da para as etapas de classificação (modelo II)

ETAPA	GRUPOS SEPARADOS	VARIÁVEIS LATENTES	Variância Explicada X (%)	Variância Explicada Y (%)	ECVC
I	AGR, PGM, MIN	6	82,09	60,27	0.1
I	MTL, CMB, INC, CIM, PTL, PCB	6	82,09	60,27	0.1
II	AGR	5	74,24	58,84	0,15
II	PGM, MIN	5	74,24	58,84	0,15
II	MTL, PTL	1	35,38	71,78	0,08

ETAPA	GRUPOS SEPARADOS	VARIÁVEIS LATENTES	Variância Explicada X (%)	Variância Explicada Y (%)	ECVC
II	CMB, INC, CIM, PCB	1	35,38	71,78	0,08
III	PGM	4	82,02	93,63	0
III	MIN	4	93,63	82,02	0
III	PTL	1	20,97	41,70	0,16
III	MTL	1	20,97	41,70	0,16
III	INC	3	55,01	38,2	0,23
III	CIM	6	80,48	46,13	0,04
III	PCB	3	60,23	68,08	0,03
III	CMB	5	75,88	49,79	0,16

Tabela 10 Figuras de mérito estimadas para etapa do modelo de classificação – Conjunto de treinamento (modelo II)

ETAPA	GRUPOS SEPARADOS	SENSIBILIDADE	ESPECIFICIDADE	FALSO POSITIVO	FALSO NEGATIVO
I	AGR, PGM, MIN	86,5	94,4	13,5	5,6
I	MTL, CMB, INC, CIM, PTL, PCB	94,4	86,5	5,6	13,5
II	AGR	81,1	88,9	18,9	11,1
II	PGM, MIN	88,9	81,1	11,1	18,9
II	MTL, PTL	95,2	88,6	4,8	11,4
II	CMB, INC, CIM, PCB	88,6	95,2	11,4	4,8
III	PGM	100	100	0	0
III	MIN	100	100	0	0
III	PTL	76,9	92,0	23,1	8
III	MTL	92,0	76,9	8	23,1
III	INC	77,8	77,1	22,2	22,9
III	CIM	100	92,3	0	7,7
III	PCB	100	94,6	0	5,4
III	CMB	91,3	76,2	8,7	23,8

Tabela 21: Figuras de mérito estimadas para etapa do modelo de classificação – Conjunto de treinamento (modelo II)

ETAPA	GRUPOS SEPARADOS	SENSIBILIDADE	ESPECIFICIDADE	FALSO POSITIVO	FALSO NEGATIVO
I	AGR, PGM, MIN	76,7	92,6	23,3	7,4
I	MTL, CMB, INC, CIM, PTL, PCB	92,6	76,7	7,4	23,3
II	AGR	50,0	76,5	50,0	23,5
II	PGM, MIN	76,5	50,0	23,5	50,0
II	MTL, PTL	41,9	82,6	58,1	17,4
II	CMB, INC, CIM, PCB	82,6	41,9	17,4	58,1
III	PGM	90,0	100	10,0	0
III	MIN	100	90,0	0	10,0
III	PTL	66,7	100	33,3	0
III	MTL	100	66,7	0	33,3
III	INC	83,3	66,7	16,7	33,3
III	CIM	100	71,4	0	28,6
III	PCB	20,0	100	80,0	0
III	CMB	92,3	100	7,7	0

## Apêndice D

Gráficos de classificação de cada arranjo representado na figura 17.

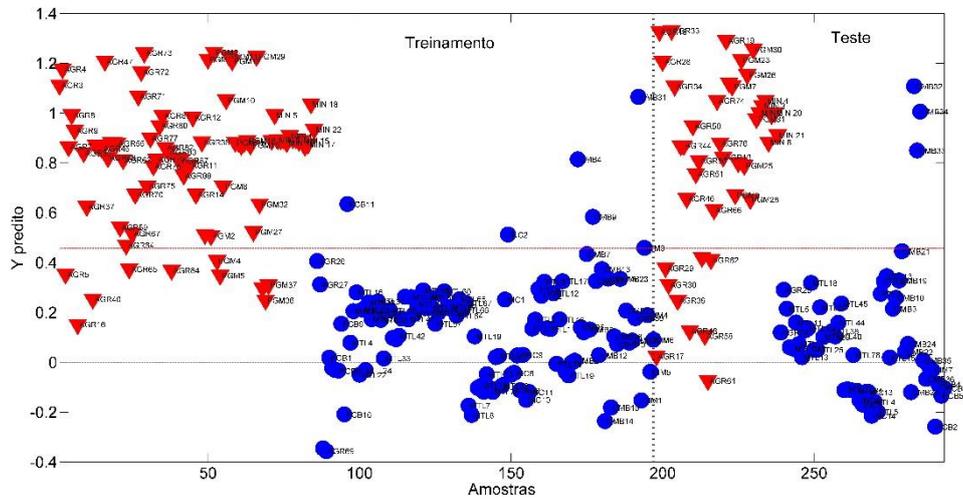


Figura 23: Gráfico de classificação para a etapa de classificação I. À esquerda conjunto de treinamento e à direita conjunto teste. Conjunto formado pelos grupos AGR-A, AGR-B, PGM e MIN representados por triângulos vermelhos. Conjunto formado pelos grupos MTL-A, MTL-B, AGR-C, PCB, INC, CIM, CMB, PTL representados por círculos azuis.

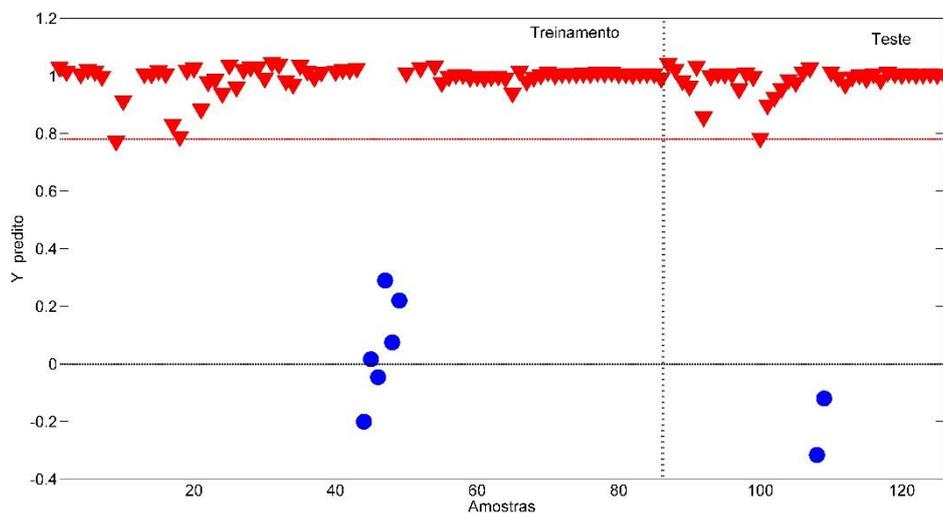


Figura 24: Gráfico de classificação para a etapa de classificação II. À esquerda conjunto de treinamento e à direita conjunto teste. Conjunto formado pelos grupos AGR-A, PGM e MIN representados por triângulos vermelhos. Conjunto formado pelos grupos AGR-B representado por círculos azuis.

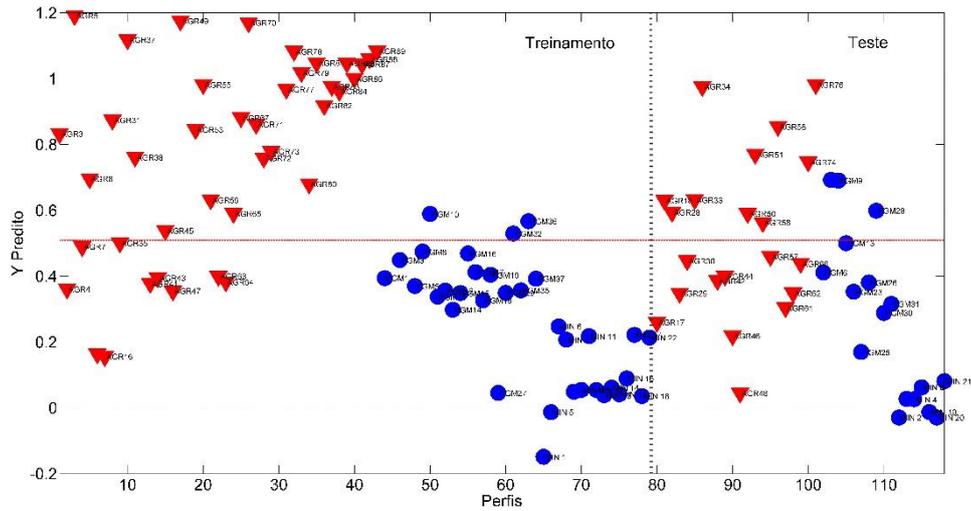


Figura 25: Gráfico de classificação para a etapa de classificação III. À esquerda conjunto de treinamento e à direita conjunto teste. Conjunto formado pelo grupo AGR-A representado por triângulos vermelhos. Conjunto formado pelos grupos PGM e MIN representados por círculos azuis.

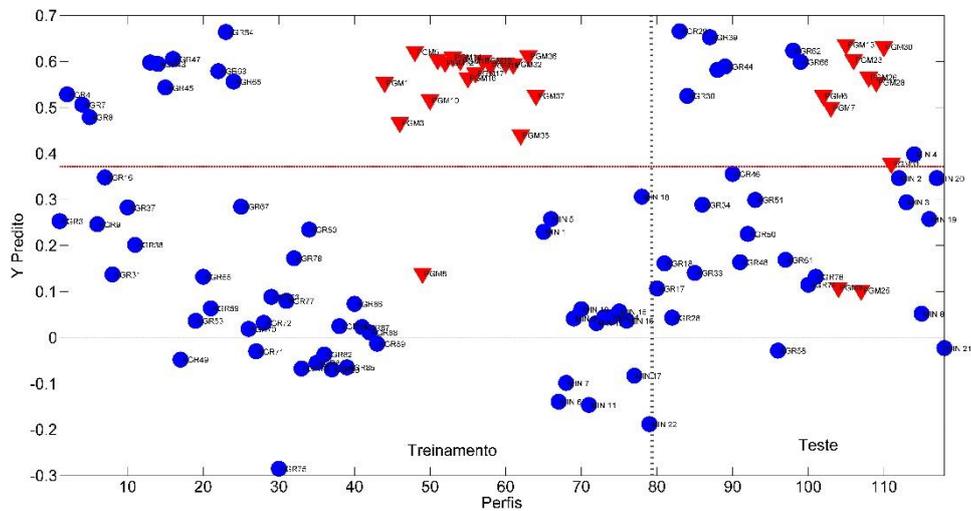


Figura 26: Gráfico de classificação para a etapa de classificação III. À esquerda conjunto de treinamento e à direita conjunto teste. Conjunto formado pelo grupo PGM representado por triângulos vermelhos. Conjunto formado pelos grupos AGR-A e MIN representados por círculos azuis.

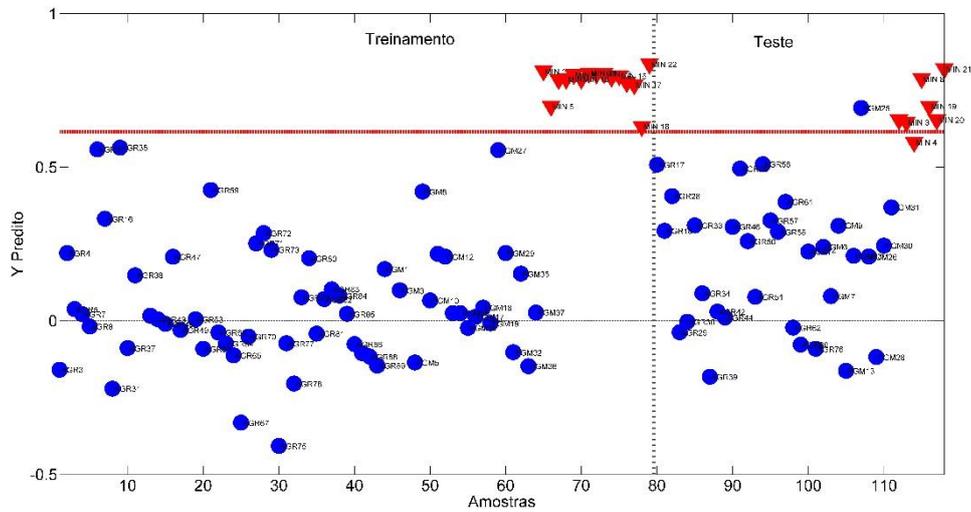


Figura 27: Gráfico de classificação para a etapa de classificação III. À esquerda conjunto de treinamento e à direita conjunto teste. Conjunto formado pelo grupo MIN representado por triângulos vermelhos. Conjunto formado pelos grupos AGR-A e PGM representados por círculos azuis.

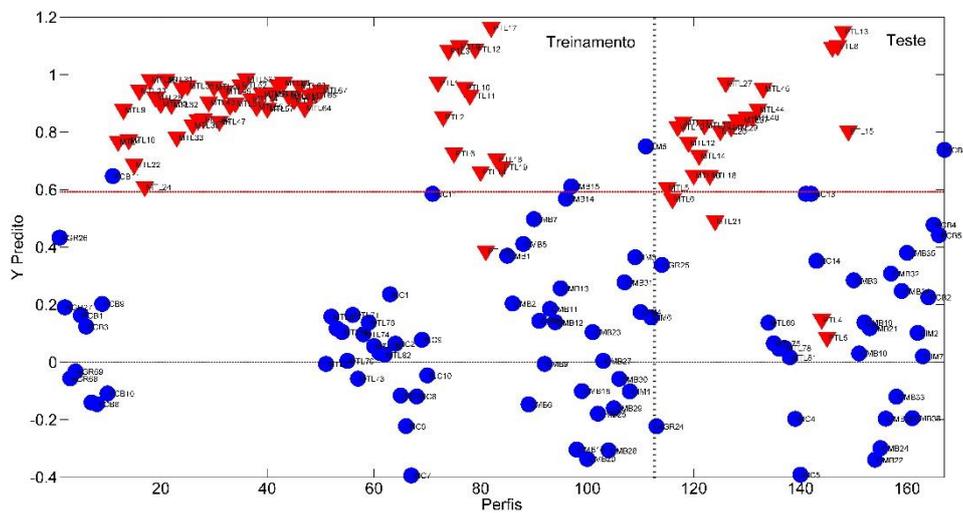


Figura 28: Gráfico de classificação para a etapa de classificação II. À esquerda conjunto de treinamento e à direita conjunto teste. Conjunto formado pelos grupos MTL-A e PTL representados por triângulos vermelhos. Conjunto formado pelos grupos MTL-B, AGR-C, CIM, CMB, INC e PCB representados por círculos azuis.

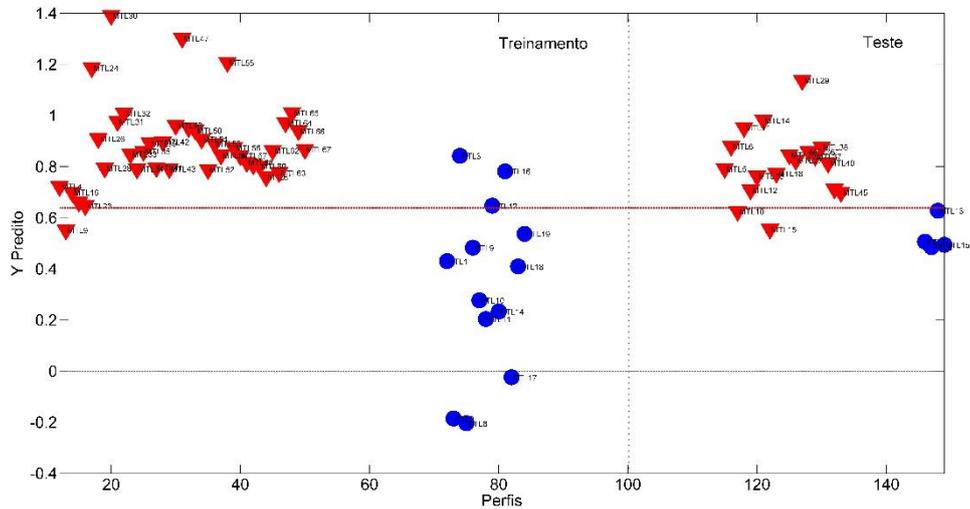


Figura 29: Gráfico de classificação para a etapa de classificação II. À esquerda conjunto de treinamento e à direita conjunto teste. Conjunto formado pelo grupo MTL-A representado por triângulos vermelhos. Conjunto formado pelo grupo PTL representado por círculos azuis.

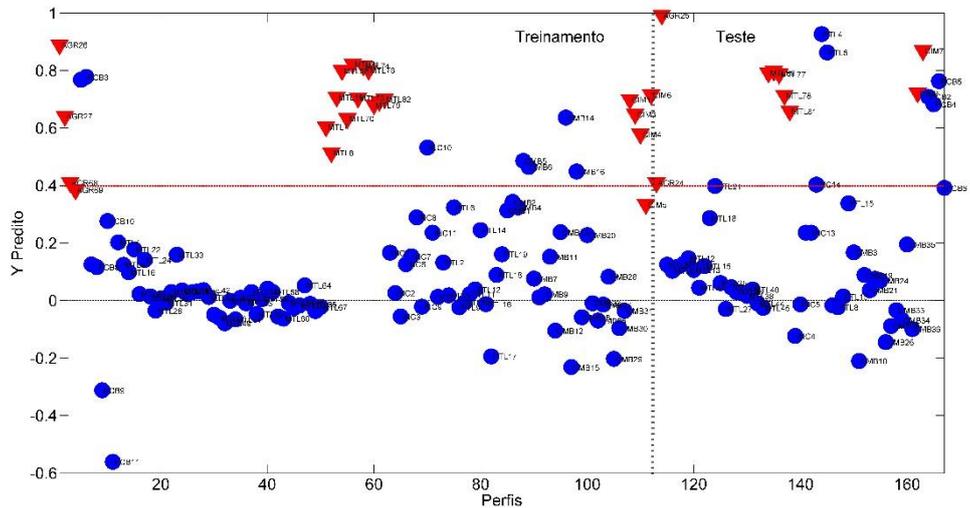


Figura 30: Gráfico de classificação para a etapa de classificação II. À esquerda conjunto de treinamento e à direita conjunto teste. Conjunto formado pelos grupos MTL-B, AGR-C, CIM representados por triângulos vermelhos. Conjunto formado pelos grupos MTL-A, PTL, CMB, INC, PCB representados por círculos azuis.

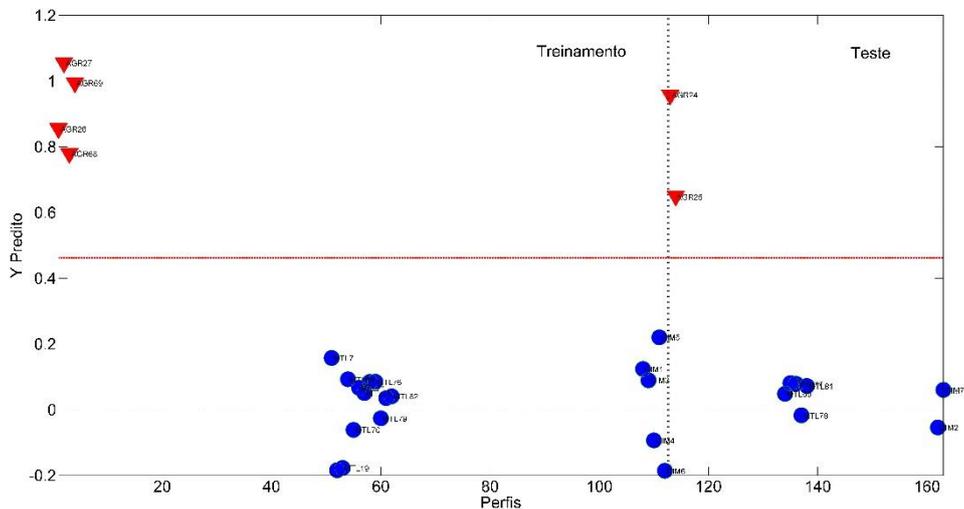


Figura 31: Gráfico de classificação para a etapa de classificação III. À esquerda conjunto de treinamento e à direita conjunto teste. Conjunto formado pelo grupo AGR-C representado por triângulos vermelhos. Conjunto formado pelos grupos MTL-B e CIM representados por círculos azuis.

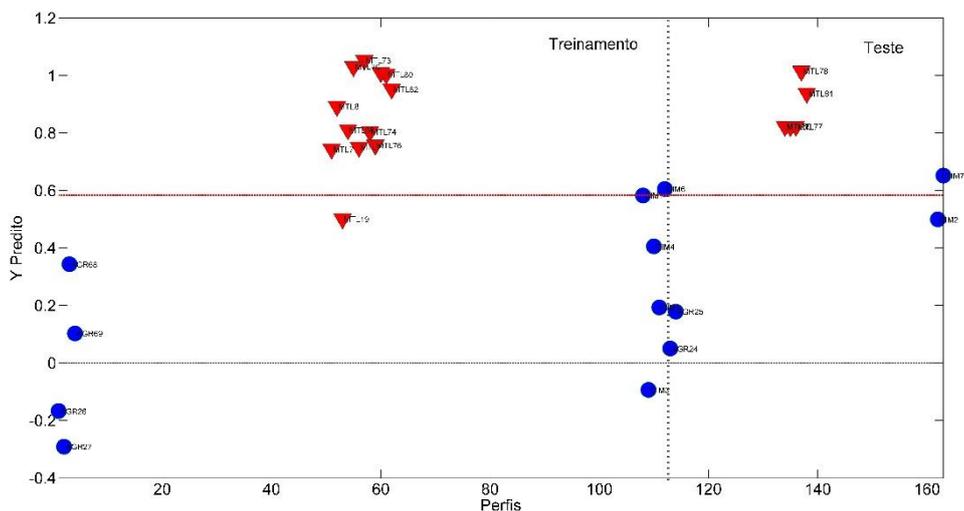


Figura 31: Gráfico de classificação para a etapa de classificação III. À esquerda conjunto de treinamento e à direita conjunto teste. Conjunto formado pelo grupo MTL-B representado por triângulos vermelhos. Conjunto formado pelos grupos AGR-C e CIM representados por círculos azuis.

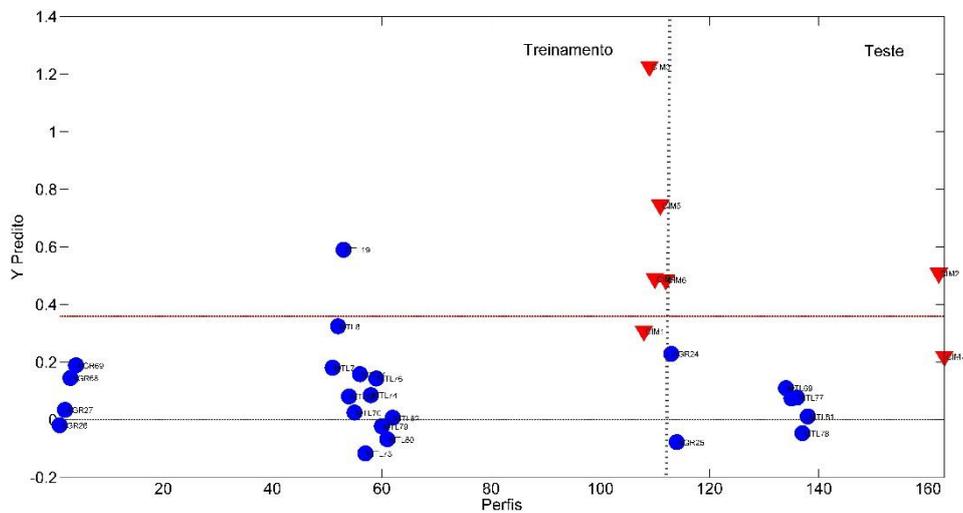


Figura 32: Gráfico de classificação para a etapa de classificação III. À esquerda conjunto de treinamento e à direita conjunto teste. Conjunto formado pelo grupo CIM representado por triângulos vermelhos. Conjunto formado pelos grupos MTL-B e AGR-C representados por círculos azuis.

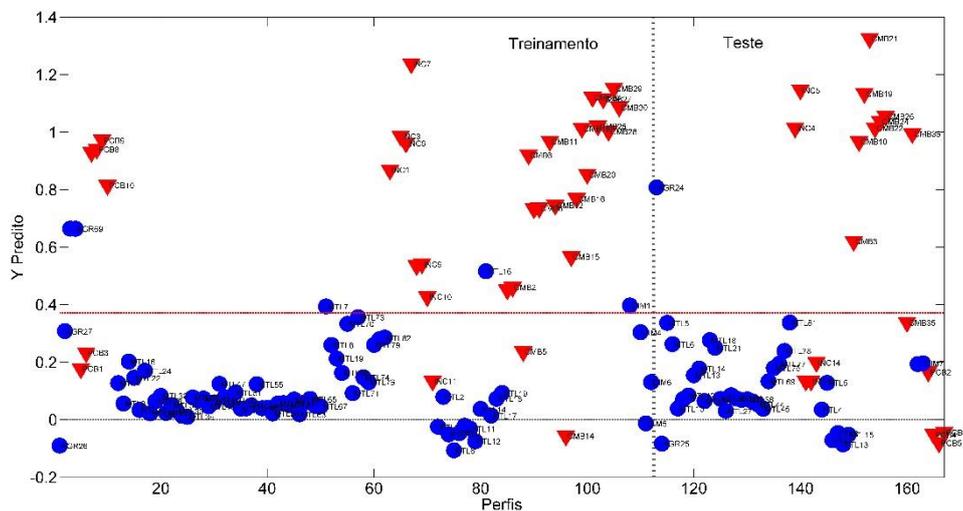


Figura 33: Gráfico de classificação para a etapa de classificação II. À esquerda conjunto de treinamento e à direita conjunto teste. Conjunto formado pelos grupos CMB, INC e PCB representados por triângulos vermelhos. Conjunto formado pelos grupos MTL-A, MTL-B, PTL, AGR-C, CIM representados por círculos azuis.

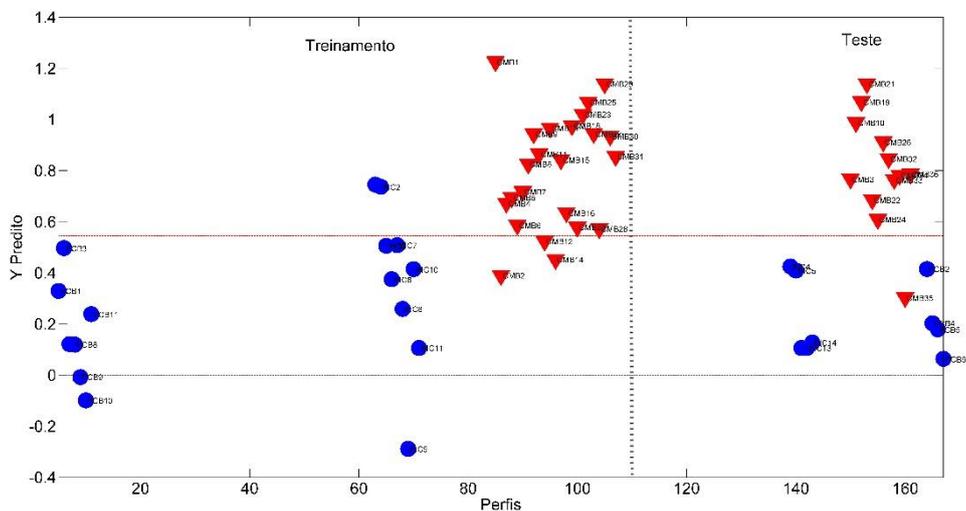


Figura 34: Gráfico de classificação para a etapa de classificação III. À esquerda conjunto de treinamento e à direita conjunto teste. Conjunto formado pelo grupo CMB representado por triângulos vermelhos. Conjunto formado pelos grupos INC e PCB representados por círculos azuis.

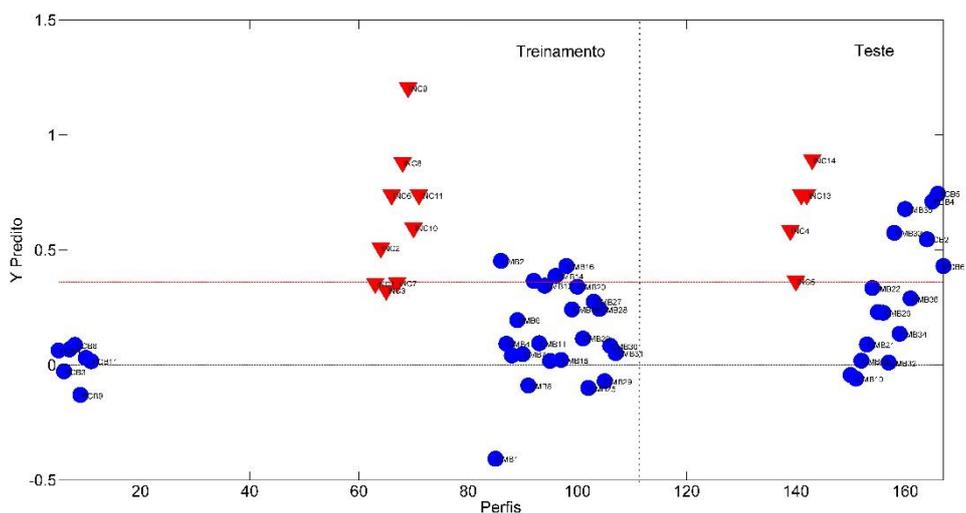


Figura 35: Gráfico de classificação para a etapa de classificação III. À esquerda conjunto de treinamento e à direita conjunto teste. Conjunto formado pelo grupo INC representado por triângulos vermelhos. Conjunto formado pelos grupos CMB e PCB representados por círculos azuis.

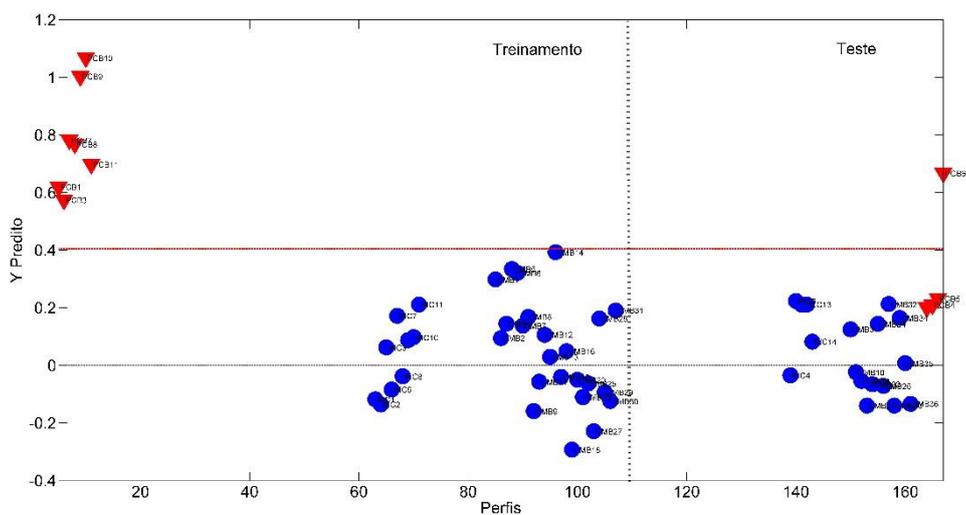


Figura 36: Gráfico de classificação para a etapa de classificação III. À esquerda conjunto de treinamento e à direita conjunto teste. Conjunto formado pelo grupo PCB representado por triângulos vermelhos. Conjunto formado pelos grupos CMB e INC representados por círculos azuis.

## Apêndice E

Gráficos dos vetores de regressão de cada etapa de classificação representado na figura 17.

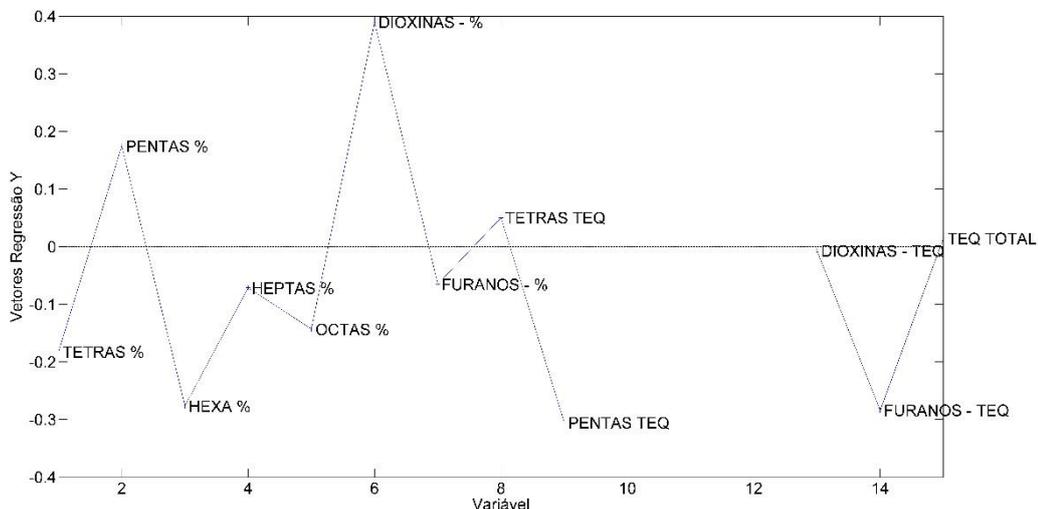


Figura 37: Gráfico de vetores de regressão referentes à classificação para a etapa de classificação I. Conjunto formado pelos grupos AGR-A, AGR-B, PGM e MIN representados como classe “positiva”. Conjunto formado pelos grupos MTL-A, MTL-B, AGR-C, PCB, INC, CIM, CMB, PTL representados como classe “negativa”.

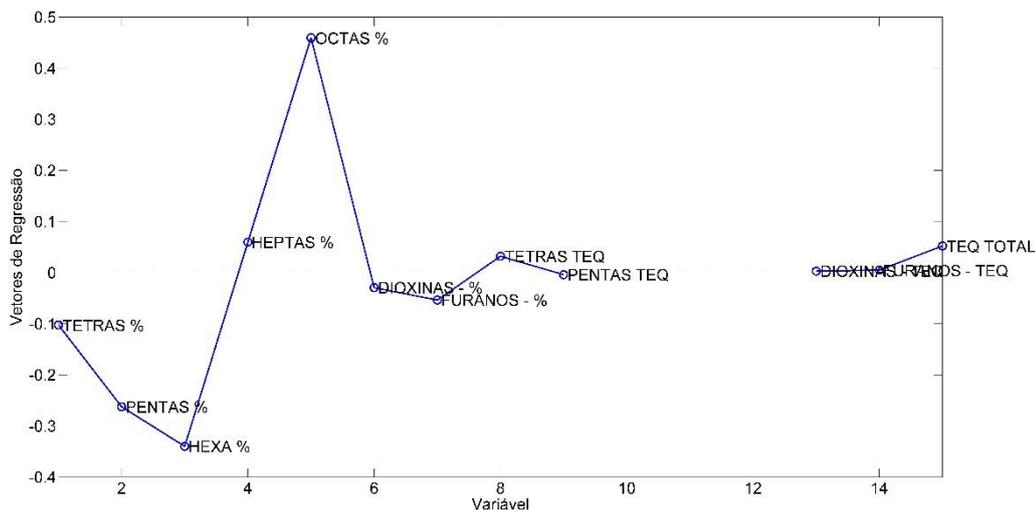


Figura 38: Gráfico de vetores de regressão referentes à classificação para a etapa de classificação II. Conjunto formado pelos grupos AGR-A, PGM e MIN representados como classe “positiva”. Conjunto formado pelos grupos AGR-B representado como classe “negativa”.

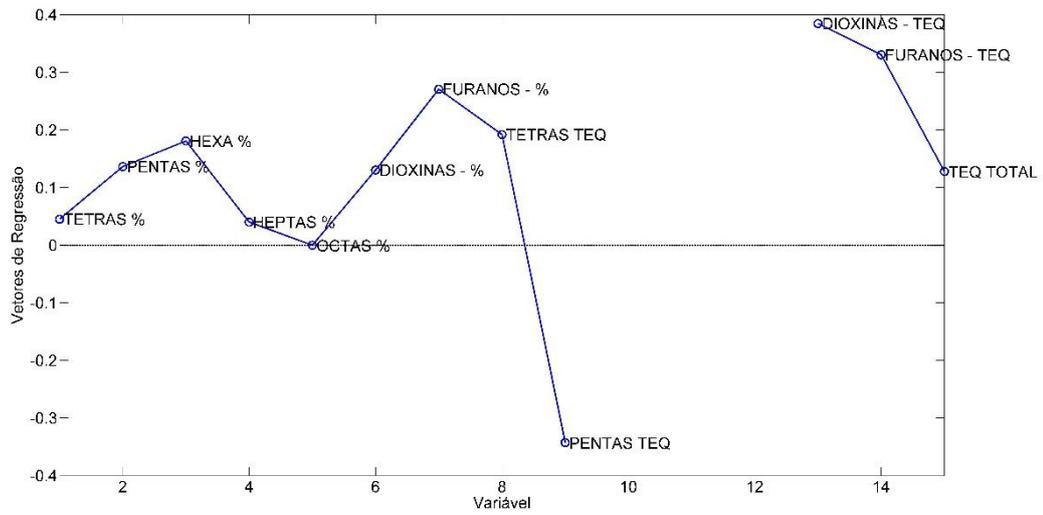


Figura 39: Gráfico de vetores de regressão referentes à classificação para a etapa de classificação III. Conjunto formado pelo grupo AGR-A representado como classe “positiva”. Conjunto formado pelos grupos PGM e MIN representados como classe “negativa”.

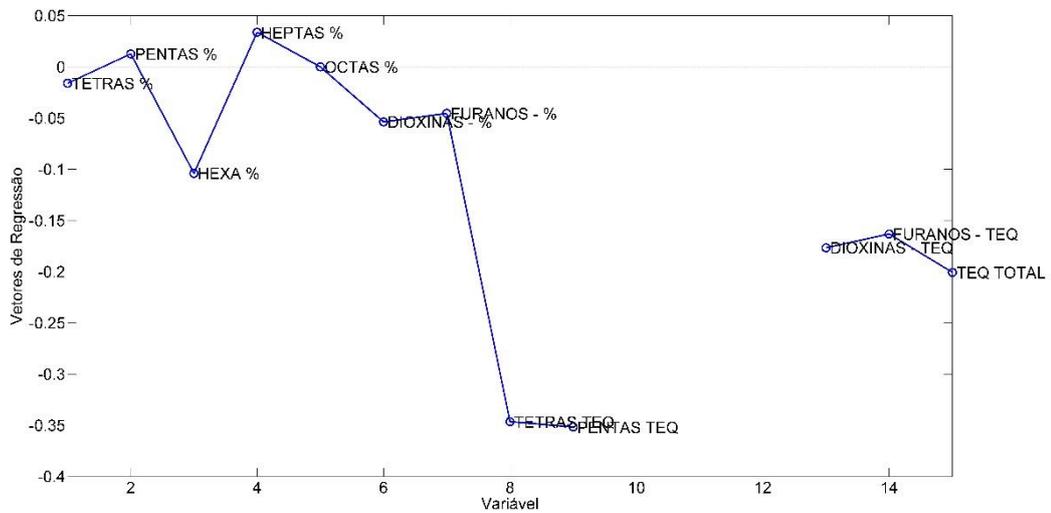


Figura 40: Gráfico de vetores de regressão referentes à classificação para a etapa de classificação III. Conjunto formado pelo grupo PGM representado como classe “positiva”. Conjunto formado pelos grupos AGR-A e MIN representados como classe “negativa”.

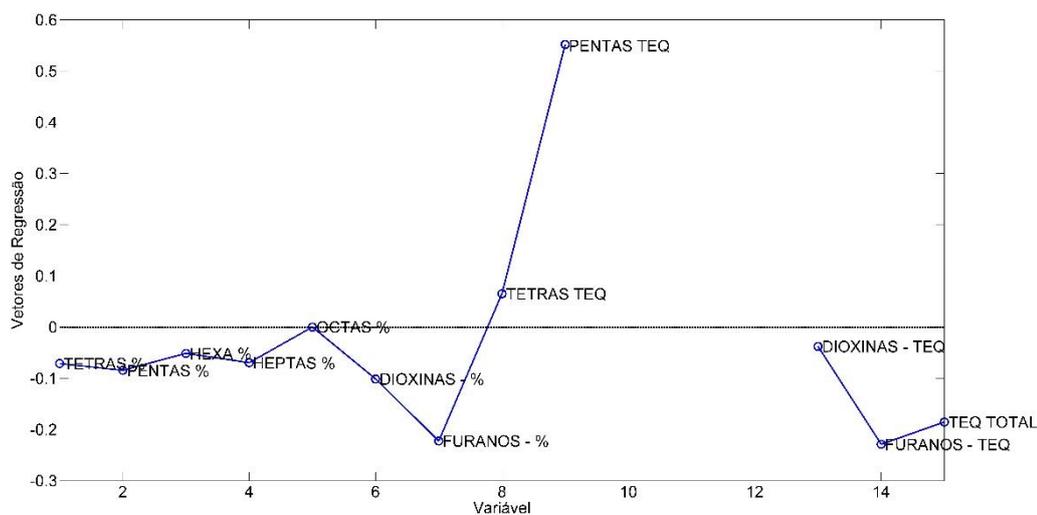


Figura 41: Gráfico de vetores de regressão referentes à classificação para a etapa de classificação III. Conjunto formado pelo grupo MIN representado como classe “positiva”. Conjunto formado pelos grupos AGR-A e PGM representados como classe “negativa”.

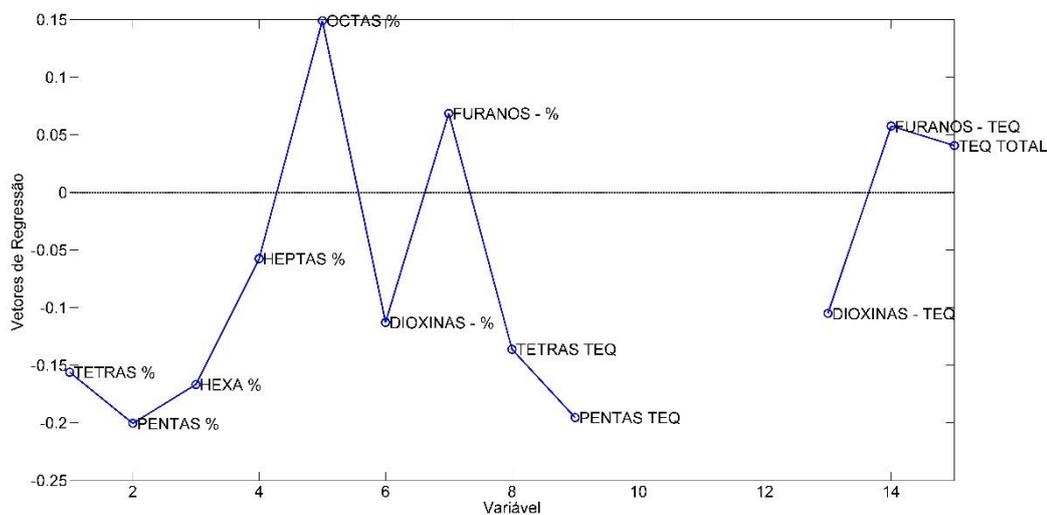


Figura 42: Gráfico de vetores de regressão referentes à classificação para a etapa de classificação II. Conjunto formado pelos grupos MTL-A e PTL representados como classe “positiva”. Conjunto formado pelos grupos MTL-B, AGR-C, PCB, INC, CIM, CMB e representados como classe “negativa”.

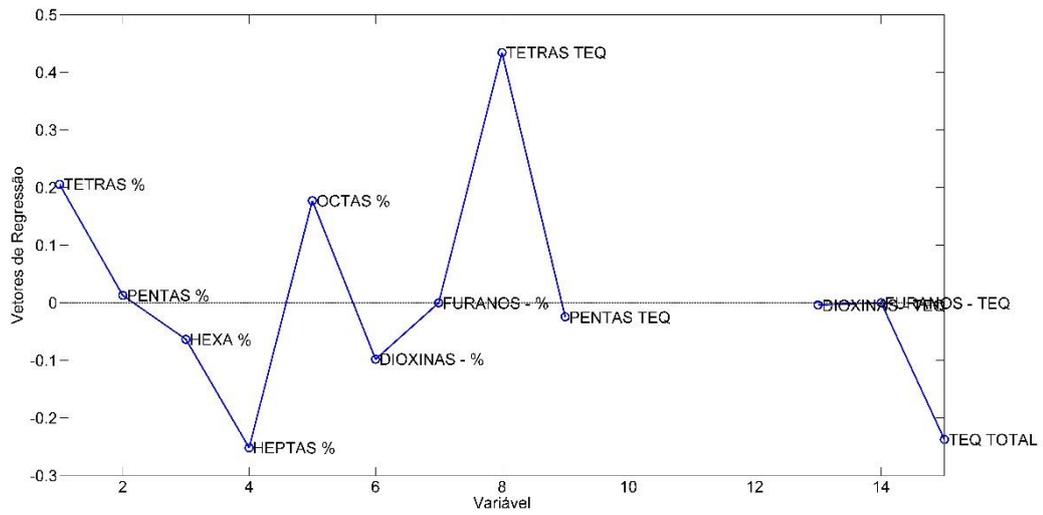


Figura 43: Gráfico de vetores de regressão referentes à classificação para a etapa de classificação III. Conjunto formado pelo grupo MTL-A representado como classe “positiva”. Conjunto formado pelo grupo PTL representado como classe “negativa”.

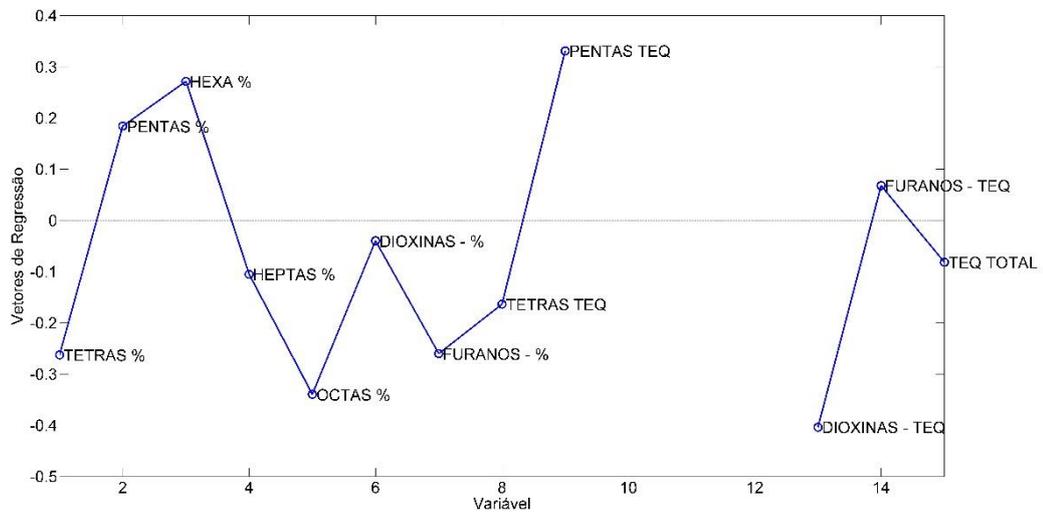


Figura 44: Gráfico de vetores de regressão referentes à classificação para a etapa de classificação II. Conjunto formado pelos grupos MTL-B, AGR-C e CIM representados como classe “positiva”. Conjunto formado pelos grupos MTL-A, PTL, CMB, INC e PCB representados como classe “negativa”.

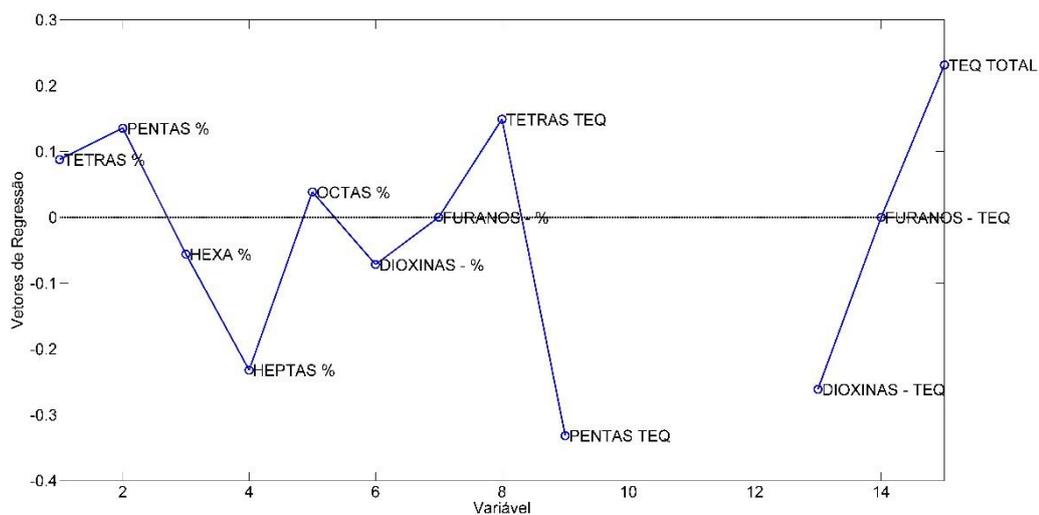


Figura 45: Gráfico de vetores de regressão referentes à classificação para a etapa de classificação III. Conjunto formado pelo grupo AGR-C representado como classe “positiva”. Conjunto formado pelos grupos MTL-B e PCB representados como classe “negativa”.

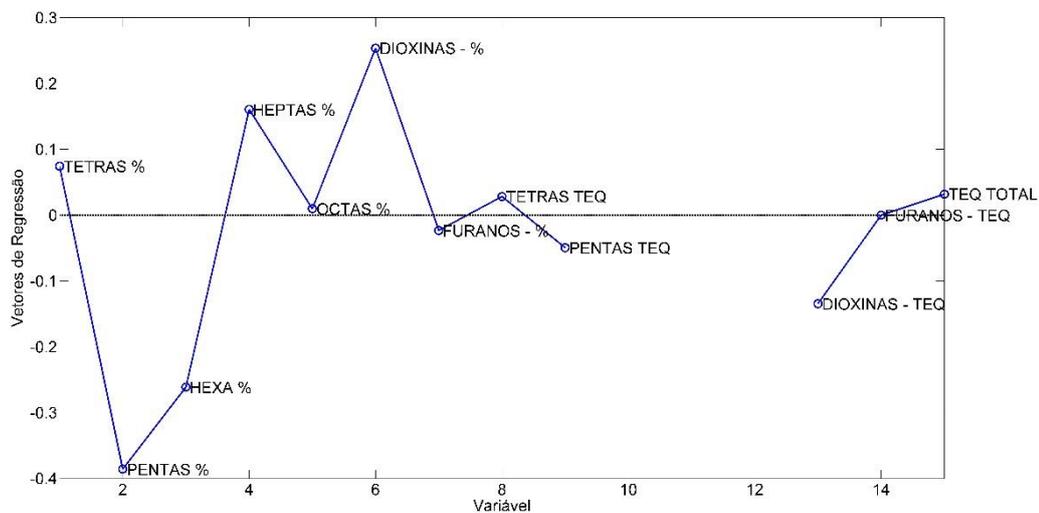


Figura 46: Gráfico de vetores de regressão referentes à classificação para a etapa de classificação II. Conjunto formado pelo grupo CIM representado como classe “positiva”. Conjunto formado pelos grupos MTL-B e CIM representados como classe “negativa”.

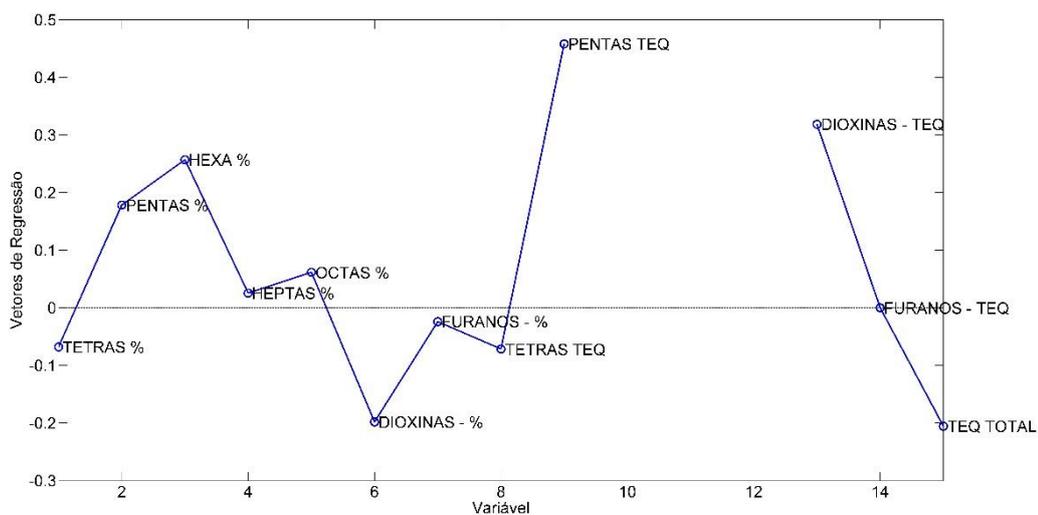


Figura 47: Gráfico de vetores de regressão referentes à classificação para a etapa de classificação III. Conjunto formado pelo grupo MTL-B representado como classe “positiva”. Conjunto formado pelos grupos AGR-C e CIM representados como classe “negativa”.

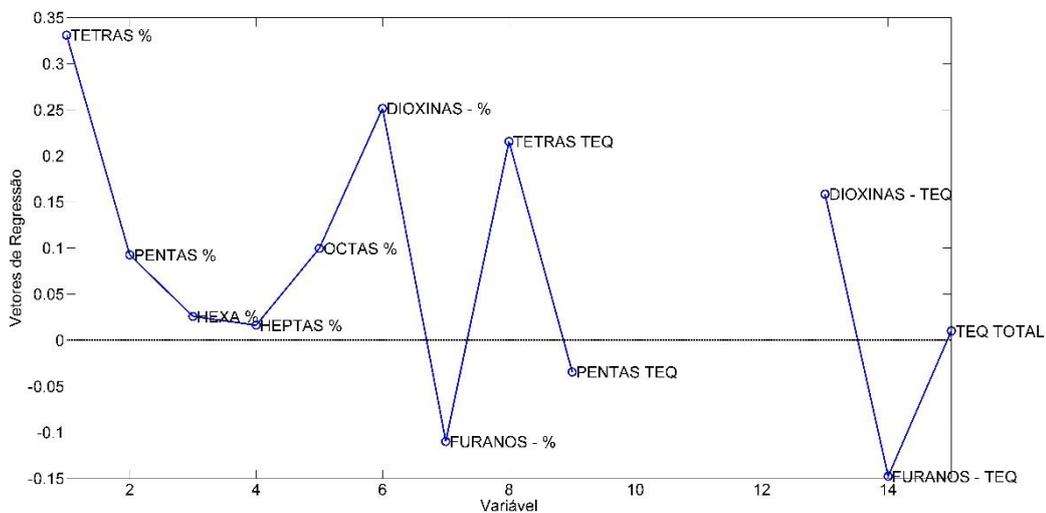


Figura 48: Gráfico de vetores de regressão referentes à classificação para a etapa de classificação II. Conjunto formado pelos grupos CMB, INC e PCB representados como classe “positiva”. Conjunto formado pelos grupos MTL-A, PTL, MTL-B, AGR-C e CIM representados como classe “negativa”.

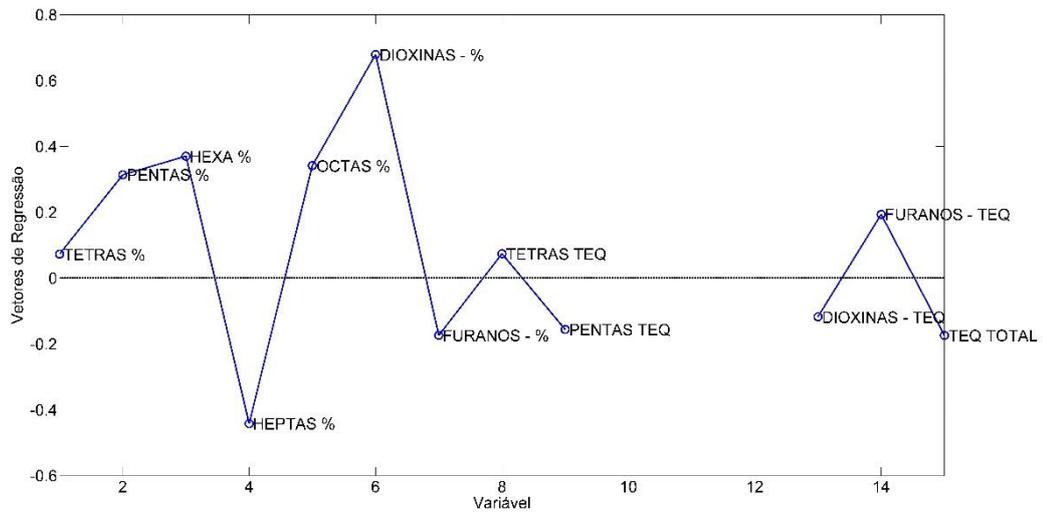


Figura 49: Gráfico de vetores de regressão referentes à classificação para a etapa de classificação II. Conjunto formado pelo grupo CMB representado como classe “positiva”. Conjunto formado pelos grupos INC e PCB representados como classe “negativa”.

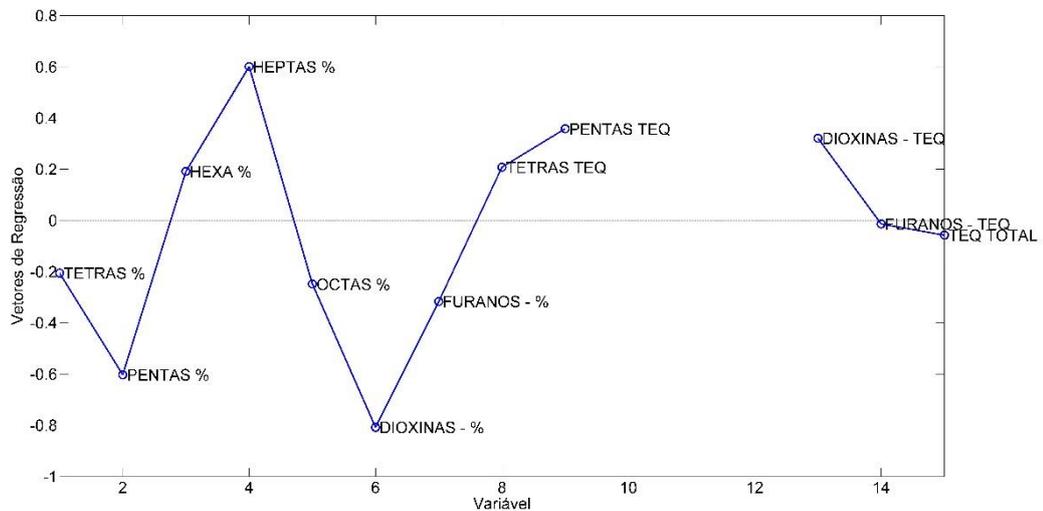


Figura 50: Gráfico de vetores de regressão referentes à classificação para a etapa de classificação II. Conjunto formado pelo grupo INC representados como classe “positiva”. Conjunto formado pelos grupos CMB e PCB representados como classe “negativa”.

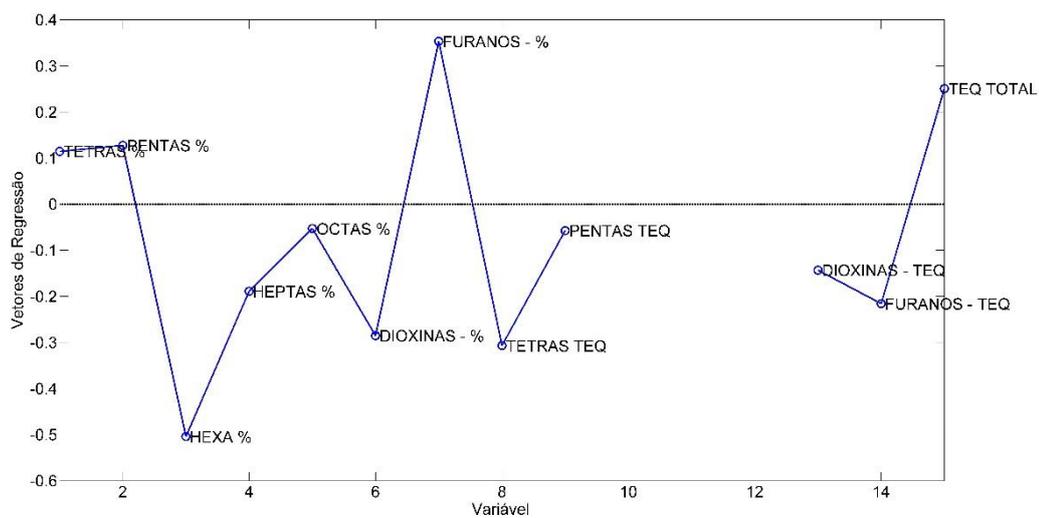


Figura 51: Gráfico de vetores de regressão referentes à classificação para a etapa de classificação II. Conjunto formado pelo grupo PCB representado como classe “positiva”. Conjunto formado pelos grupos CMB e INC representados como classe “negativa”.

## Apêndice F

Gráficos dos VIP escores de cada etapa de classificação representado na figura 17.

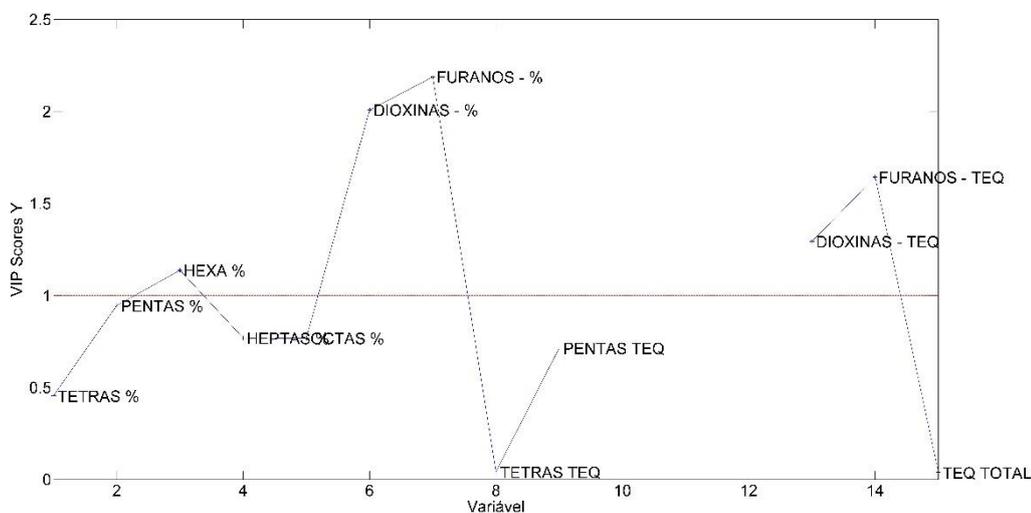


Figura 37: Gráfico dos VIP escores referentes à classificação para a etapa de classificação I. Conjunto formado pelos grupos AGR-A, AGR-B, PGM e MIN representados como classe “positiva”. Conjunto formado pelos grupos MTL-A, MTL-B, AGR-C, PCB, INC, CIM, CMB, PTL representados como classe “negativa”.

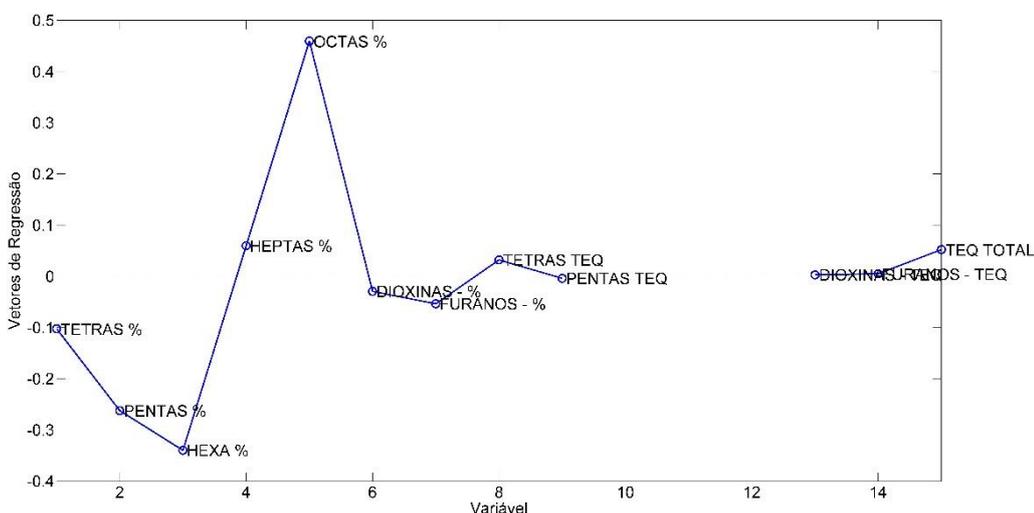


Figura 38: Gráfico dos VIP escores referentes à classificação para a etapa de classificação II. Conjunto formado pelos grupos AGR-A, PGM e MIN representados como classe “positiva”. Conjunto formado pelos grupos AGR-B representado como classe “negativa”.

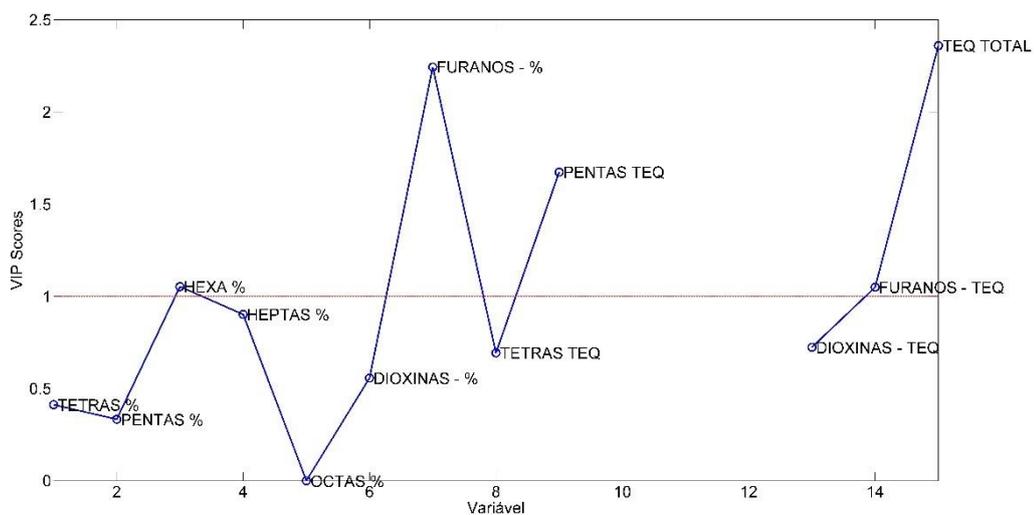


Figura 39: Gráfico dos VIP scores referentes à classificação para a etapa de classificação III. Conjunto formado pelo grupo AGR-A representado como classe “positiva”. Conjunto formado pelos grupos PGM e MIN representados como classe “negativa”.

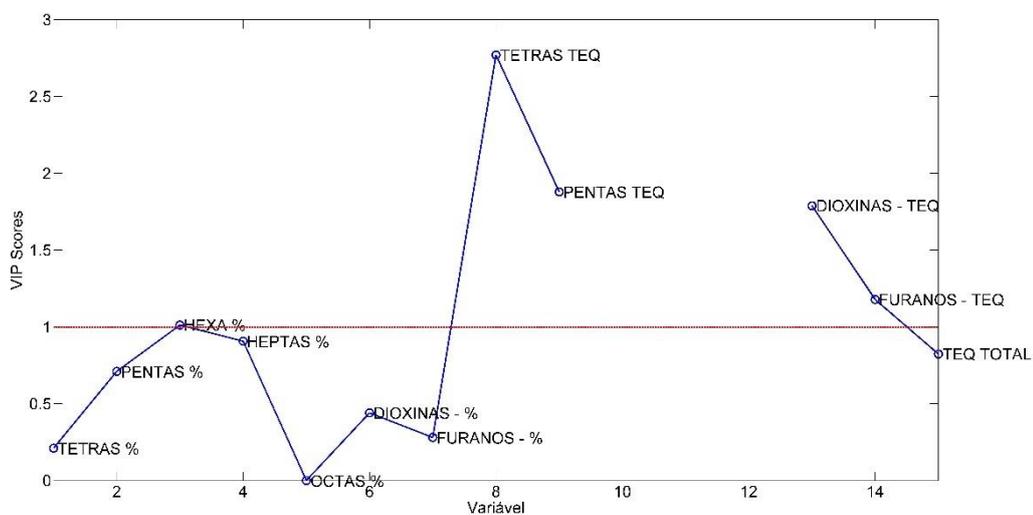


Figura 40: Gráfico dos VIP scores referentes à classificação para a etapa de classificação III. Conjunto formado pelo grupo PGM representado como classe “positiva”. Conjunto formado pelos grupos AGR-A e MIN representados como classe “negativa”.

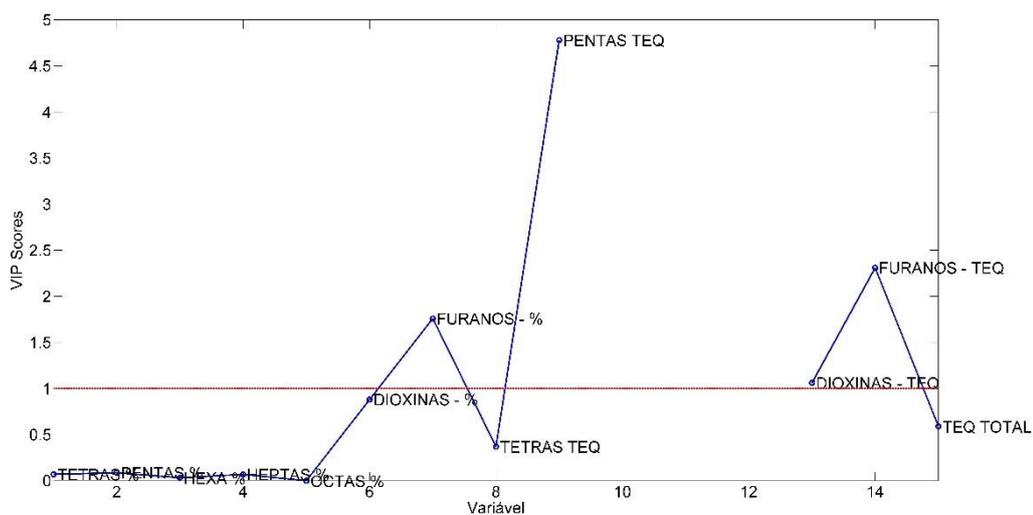


Figura 41: Gráfico dos VIP scores referentes à classificação para a etapa de classificação III. Conjunto formado pelo grupo MIN representado como classe “positiva”. Conjunto formado pelos grupos AGR-A e PGM representados como classe “negativa”.

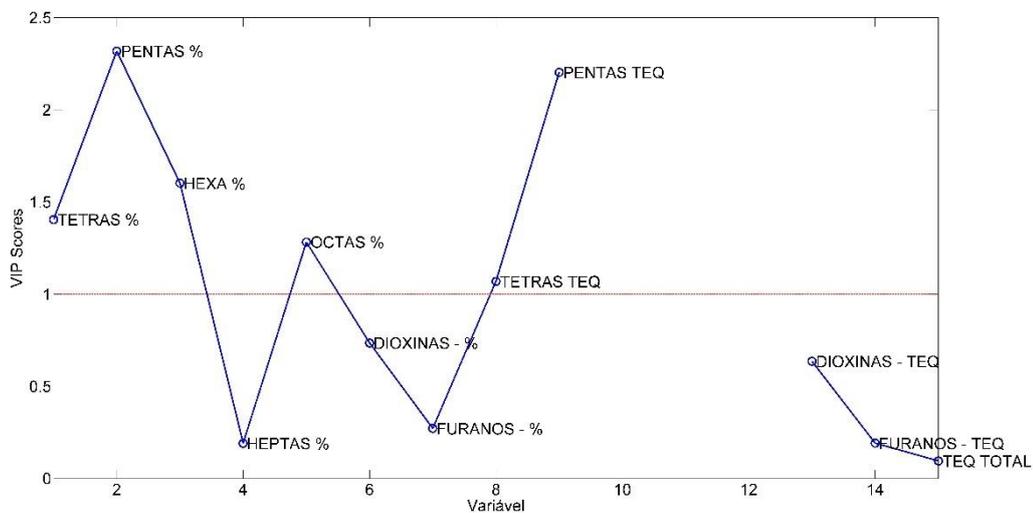


Figura 42: Gráfico dos VIP scores referentes à classificação para a etapa de classificação II. Conjunto formado pelos grupos MTL-A e PTL representados como classe “positiva”. Conjunto formado pelos grupos MTL-B, AGR-C, PCB, INC, CIM, CMB e representados como classe “negativa”.

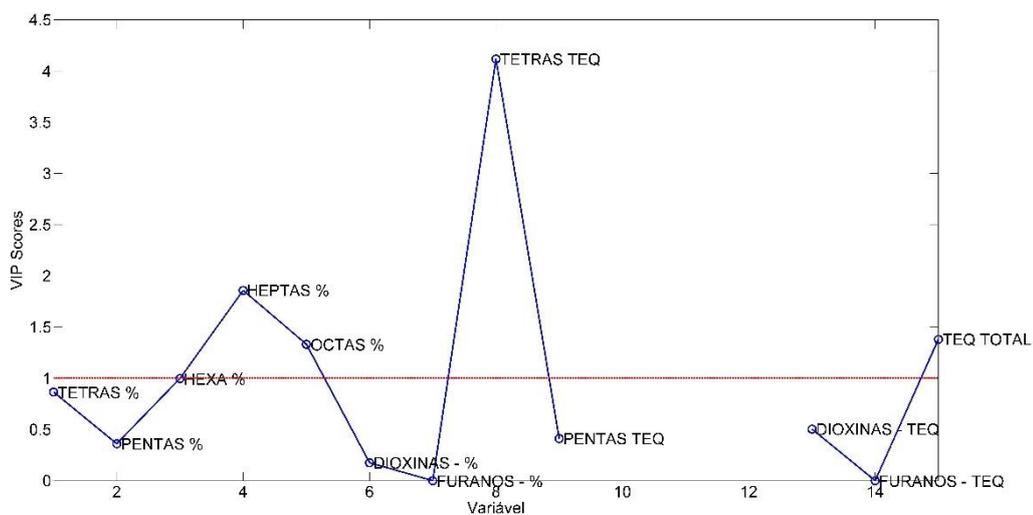


Figura 43: Gráfico dos VIP scores referentes à classificação para a etapa de classificação III. Conjunto formado pelo grupo MTL-A representado como classe “positiva”. Conjunto formado pelo grupo PTL representado como classe “negativa”.

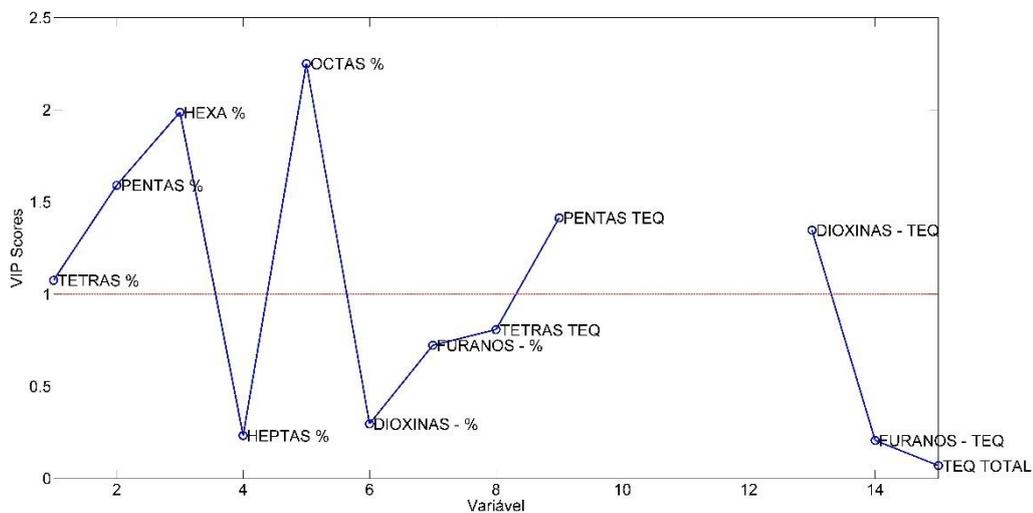


Figura 44: Gráfico dos VIP scores referentes à classificação para a etapa de classificação II. Conjunto formado pelos grupos MTL-B, AGR-C e CIM representados como classe “positiva”. Conjunto formado pelos grupos MTL-A, PTL, CMB, INC e PCB representados como classe “negativa”.

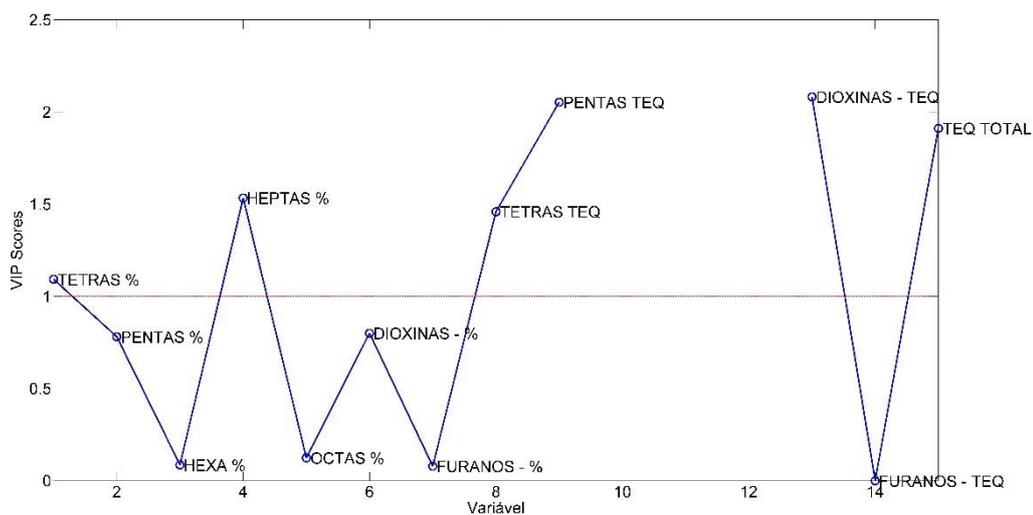


Figura 45: Gráfico dos VIP scores referentes à classificação para a etapa de classificação III. Conjunto formado pelo grupo AGR-C representado como classe “positiva”. Conjunto formado pelos grupos MTL-B e PCB representados como classe “negativa”.

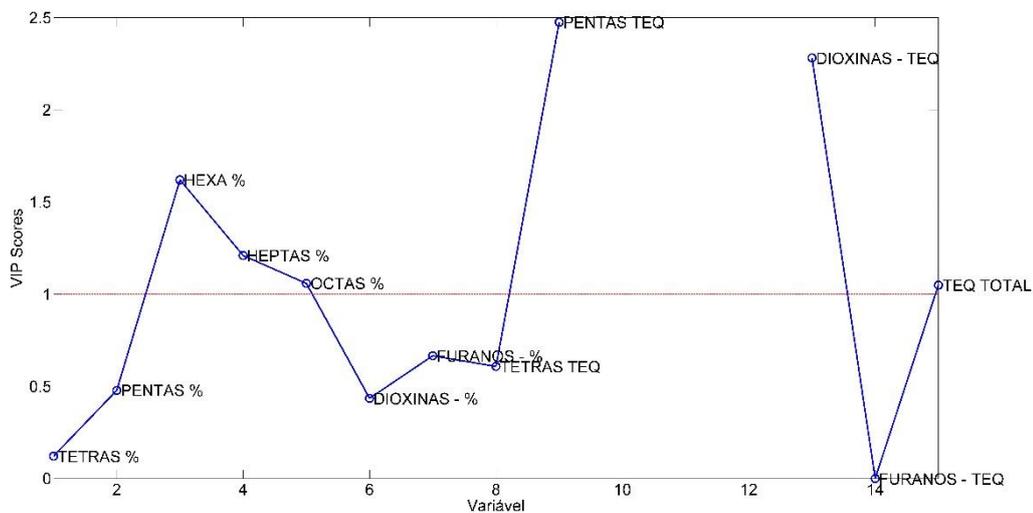


Figura 46: Gráfico dos VIP scores referentes à classificação para a etapa de classificação II. Conjunto formado pelo grupo CIM representado como classe “positiva”. Conjunto formado pelos grupos MTL-B e CIM representados como classe “negativa”.

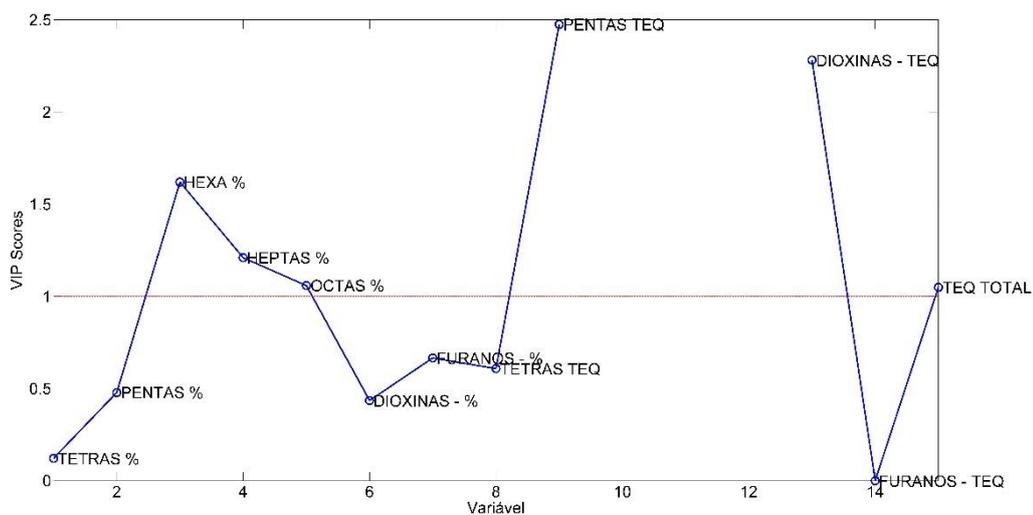


Figura 47: Gráfico dos VIP scores referentes à classificação para a etapa de classificação III. Conjunto formado pelo grupo MTL-B representado como classe “positiva”. Conjunto formado pelos grupos AGR-C e CIM representados como classe “negativa”.

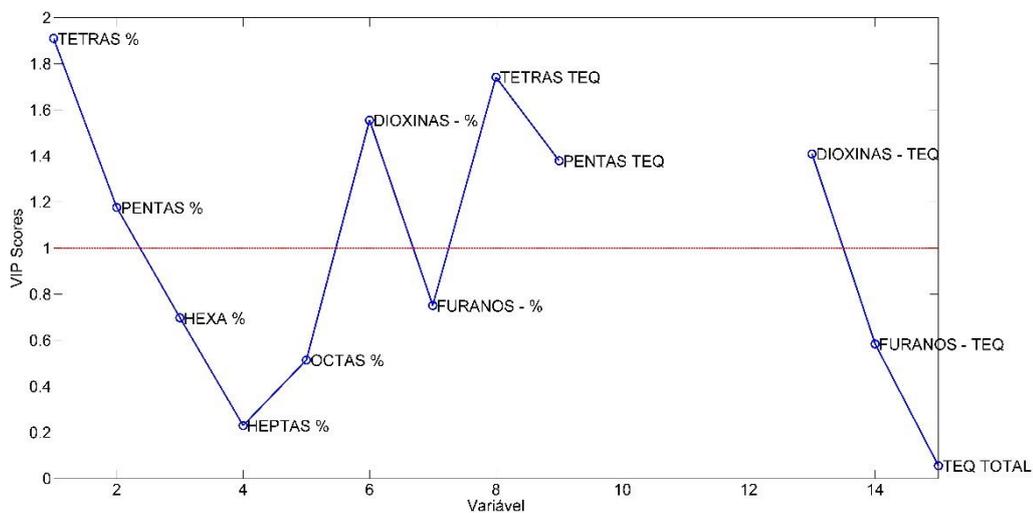


Figura 48: Gráfico dos VIP scores referentes à classificação para a etapa de classificação II. Conjunto formado pelos grupos CMB, INC e PCB representados como classe “positiva”. Conjunto formado pelos grupos MTL-A, PTL, MTL-B, AGR-C e CIM representados como classe “negativa”.

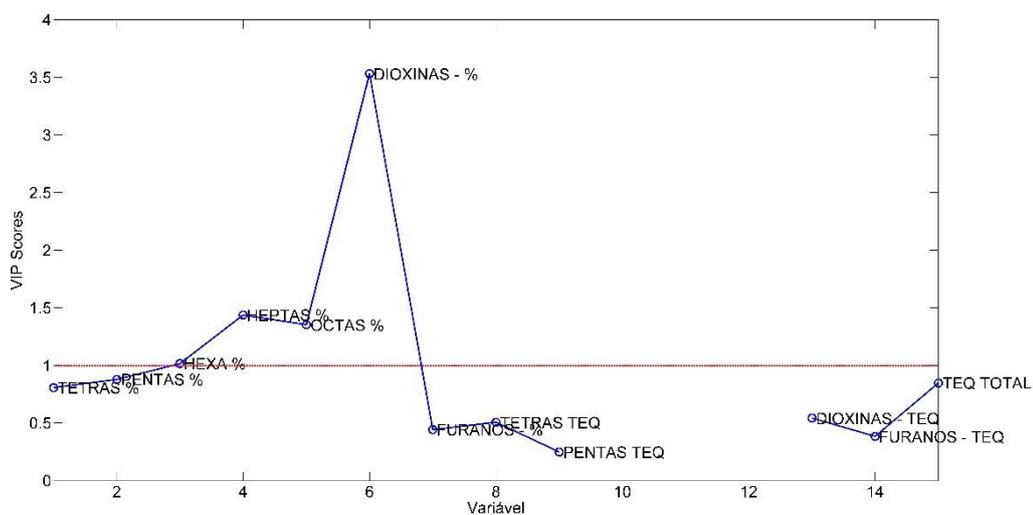


Figura 49: Gráfico dos VIP scores referentes à classificação para a etapa de classificação II. Conjunto formado pelo grupo CMB representado como classe “positiva”. Conjunto formado pelos grupos INC e PCB representados como classe “negativa”.

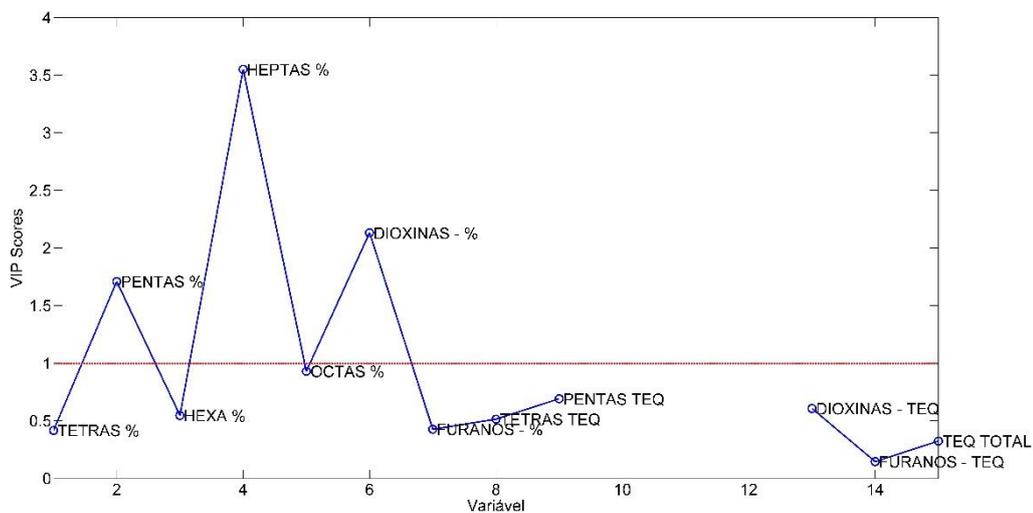


Figura 50: Gráfico dos VIP scores referentes à classificação para a etapa de classificação II. Conjunto formado pelo grupo INC representados como classe “positiva”. Conjunto formado pelos grupos CMB e PCB representados como classe “negativa”.

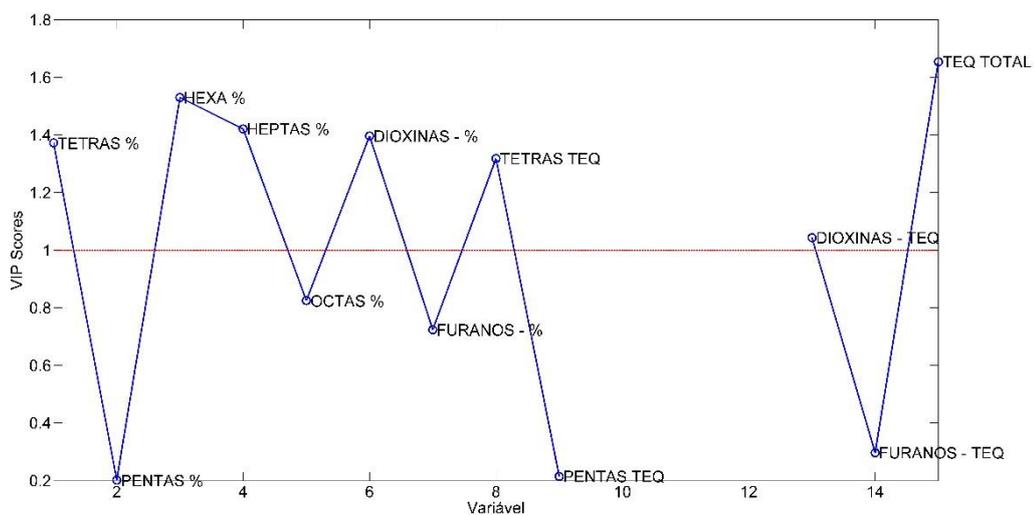


Figura 51: Gráfico dos VIP escores referentes à classificação para a etapa de classificação II. Conjunto formado pelo grupo PCB representado como classe “positiva”. Conjunto formado pelos grupos CMB e INC representados como classe “negativa”.