



BRUNO FILIPE CASTANHEIRA DOS SANTOS

Licenciado em Ciências da Engenharia Física

NOVO ALGORITMO DE INTEGRAÇÃO DO SINAL CROMATOGRÁFICO

MESTRADO EM ENGENHARIA FÍSICA

Universidade NOVA de Lisboa
Setembro, 2022

NOVO ALGORITMO DE INTEGRAÇÃO DO SINAL CROMATOGRÁFICO

BRUNO FILIPE CASTANHEIRA DOS SANTOS

Licenciado em Ciências da Engenharia Física

Coorientadores: Orlando Manuel Neves Duarte Teodoro

Professor Catedrático, Universidade Nova de Lisboa

Miguel Freire de Albuquerque Ferreira Cabral

Diretor de I&D, Amorim Cork

Júri

Presidente: Hugo Filipe Silveira Gamboa

Professor Associado com Agregação, FCT-NOVA

Arguente: André João Maurício Leitão do Valle Wemans

Professor Auxiliar, FCT-NOVA

Vogal: Miguel Freire de Albuquerque Ferreira Cabral

Professor Auxiliar, Universidade do Porto

Novo algoritmo de integração do sinal cromatográfico

Copyright © Bruno Filipe Castanheira dos Santos, Faculdade de Ciências e Tecnologia, Universidade NOVA de Lisboa.

A Faculdade de Ciências e Tecnologia e a Universidade NOVA de Lisboa têm o direito, perpétuo e sem limites geográficos, de arquivar e publicar esta dissertação através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, e de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objetivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.

Agradecimentos

Gostaria de agradecer à Amorim Cork, em particular ao Prof. Dr. Miguel Cabral, meu orientador, não só pela oportunidade de desenvolver este trabalho como também por todas as condições e apoio que proporcionaram durante os últimos meses.

Ao Prof. Dr. Orlando Teodoro, não só gostaria de agradecer pela função de orientador, como também pelo longo trabalho de criação de um ambiente de colaboração e entreaajuda no grupo de investigação no qual este trabalho se insere, o qual serviu de forte base de apoio científico.

À equipa do NDtech, em particular ao Carlos Costa e ao Joel Ferreira, gostaria de agradecer pela infundável paciência na viabilização dos testes que se realizaram e por toda a confiança que depositaram no meu trabalho.

Ao Mota e à Deolinda, por providenciarem a minha estadia nesta zona que me era desconhecida e que agora terá sempre algo de familiar, o meu muito obrigado!

À minha família quero agradecer à minha por todo o apoio, amor e carinho que senti durante toda a minha vida e em particular nos últimos 5 anos. A partilha das felicidades e das dificuldades tornou as últimas muito mais toleráveis.

Agradeço aos meus amigos, em particular (e sem ordem de preferência) ao Santos, João, Tomás, Joana, "Sparta", Tiago, Rafael, Jessica, Beatriz, Carlota(s), Susana, Mariana e Adriana.

Por último e deveras o mais importante, agradeço à Sara, minha namorada, pela paciência interminável e nervos de aço que teve para me aturar, em particular nestes últimos seis meses e pelo apoio fulcral na realização desta tese.

*“Physics is mathematical not because we know so
much about the physical world,
but because we know so little it is only its
mathematical properties that we can discover.”
(Bertrand Russell)*

Resumo

O 2,4,6-tricloroanisol (TCA) é um contaminante ocorrente nas rolhas de cortiça que prejudica as propriedades sensoriais de vinho engarrafado, levando a perdas económicas significativas.

Como parte de um esforço tecnológico transversal à indústria na resolução deste problema, a Amorim Cork utiliza o sistema NDtech para aplicar industrialmente técnicas cromatográficas de forma a efetuar análises rolha-a-rolha da concentração de TCA libertável.

Com os algoritmos de classificação atuais, 99,96% das rolhas classificadas como TCA negativas apresentam valores abaixo de 0,5 ng/L quando analisado o *headspace* da respetiva maceração. Embora este valor seja bastante significativo, é obtido também à custa de rejeição de rolhas numa proporção superior à desejável.

Pretende-se com este trabalho o desenvolvimento de um novo algoritmo, com os objetivos de diminuir os erros de classificação das rolhas e a dependência de fatores de decisão humana.

Palavras-chave: Rolhas de cortiça, 2,4,6-tricloroanisol, Análise de sinal, NDtech

Abstract

2,4,6-Trichloroanisole (TCA) is a naturally occurring contaminant in both cork and cork stoppers which, despite being non-toxic, its presence ravages the smell and taste of wine, leading to considerable economic losses.

As part of an industrywide effort towards the solution of this problem, Amorim Cork uses NDtech to apply chromatographic techniques for individual screening of every stopper's releasable TCA quantity.

With the current classification techniques, 99,96% of all corks have non-detectable amounts of TCA (under 0,5 ng/L when undergone maceration and headspace analysis). Despite that, these results are obtained in a way that a substantial amount of stoppers is rejected in this process.

The aim of this work is to reduce both the classification error of the current algorithms, especially in the case of higher quality stoppers and human-related classification errors.

Keywords: Cork stoppers, 2, 4, 6-Trichloroanisole, Signal analysis, NDtech

Índice

Índice de Figuras	xv
Índice de Tabelas	xvii
Glossário	xix
Siglas/Acrónimos	xxi
Símbolos químicos	xxiii
1 Introdução	1
1.1 Motivação	1
1.2 NDtech	4
1.2.1 Sistemas GC/ECD	5
1.2.2 Sistema GC-ECD do NDtech (GC×GC com <i>heartcut</i>)	7
1.2.3 Obtenção de rolhas NDtech®	9
1.3 Estrutura da dissertação	12
2 Sinal obtido e algoritmos	13
2.1 Morfologia do sinal cromatográfico no NDtech	13
2.2 Método atual de integração do pico cromatográfico no NDtech	15
2.2.1 Fontes de falsos positivos	16
2.2.2 Fontes de falsos negativos	18
2.3 A matemática da análise de sinal	18
2.3.1 Otimização não-linear via método dos mínimos quadrados	19
2.4 Trabalho prévio - <i>fitplus</i>	21
2.5 Alterações ao <i>fitplus</i> - NDfit	22
2.6 Métodos de procura do pico de TCA no cromatograma	24
2.6.1 Ajuste ao pico gaussiano	24
2.6.2 Ajuste de modelo de pico cromatográfico	26

3	Testes e validação	31
3.1	Validação fora de linha de produção	31
3.1.1	Algoritmo na forma manual - métodos e objetivos	31
3.1.2	Algoritmo na forma manual - resultados e discussão	31
3.1.3	Algoritmos na forma automática - métodos e objetivos	35
3.1.4	Algoritmos na forma automática - resultados e discussão	35
3.1.5	Conclusão	41
3.2	Validação em linha de produção (sem substituição de resultados) - algoritmo na forma manual	42
3.2.1	Objetivos da validação em linha de produção	42
3.2.2	Teste de média escala - Métodos	43
3.2.3	Teste de média escala - Resultados	43
3.2.4	Teste de média escala - Conclusão	45
3.2.5	Teste de pequena escala - Métodos e objetivos	47
3.2.6	Teste de pequena escala - Resultados	48
3.2.7	Teste de pequena escala - Conclusão	49
3.3	Alterações ao algoritmo decorrentes dos testes	49
3.4	Validação em linha de produção (com substituição de resultados)	50
3.4.1	Objetivos e métodos	50
3.4.2	Resultados	51
3.5	Repetitividade experimental	54
4	Trabalho futuro e conclusão	57
4.1	Trabalho futuro	57
4.1.1	Produção e manutenção	57
4.1.2	Algoritmo e à aplicação de testes	57
4.2	Conclusões	59
	Bibliografia	61
	Apêndices	
A	Interface <i>NDfit Tester</i>	65
A.1	Funcionamento	65
A.2	Instruções	66
B	Resultados experimentais	69
B.1	Resultados do teste de pequena escala	69
B.2	Resultados da validação em linha de produção (Grupos 2 e 3)	72

Índice de Figuras

1.1	Estrutura molecular do TCA	1
1.2	Exemplos de rolhas naturais, de aglomerado com dois discos, de espumante com um ou dois discos e de aglomerado	3
1.3	TCA libertável	3
1.4	Equipamento de produção no NDtech. Imagem de [20]	4
1.5	Esquema simplificado de uma montagem GC/ECD	5
1.6	Esquema simplificado do funcionamento de um ECD	6
1.7	Representação de um cromatograma convencional e ideal (simulado)	6
1.8	Esquema simplificado da ligação das várias colunas cromatográficas a um mesmo ECD	7
1.9	Esquema de funcionamento de um ECD com gás de <i>make-up</i>	7
1.10	Exemplos de cromatogramas obtidos no NDTech	8
1.11	Funcionamento do NDtech	9
1.12	Exemplo de funcionamento do algoritmo de integração utilizado no NDTech. A preto: Limites de integração. A preto, a tracejado: fundo cromatográfico A cor-de-laranja: Cromatograma obtido	10
2.1	Secções do sinal cromatográfico	13
2.2	Representação gráfica de algumas funções utilizadas como modelos de pico cromatográfico.	14
2.3	Erro de integração do Ellution	16
2.4	Rejeição por carry-over	17
2.5	Método do Gradiente	20
2.6	Método de Gauss-Newton	21
2.7	Comparação de fundo linear e polinomial	23
2.8	Algoritmo de pesquisa por ajustes sucessivos de picos gaussianos (1)	25
2.9	Algoritmo de pesquisa por ajustes sucessivos de picos gaussianos (2)	26
2.10	Exemplo de procura e integração de um pico cromatográfico pelo algoritmo de procura de padrão	27

2.11	Algoritmo de pesquisa por comparações sucessivas com o padrão	28
3.1	Cromatogramas analisados para a área mínima por coluna no conjunto de dados	33
3.2	Cromatogramas analisados para a área máxima por coluna no conjunto de dados	34
3.3	Cromatogramas analisados para a área mínima no conjunto de dados . . .	37
3.4	Cromatogramas analisados para a área máxima no conjunto de dados . . .	38
3.5	Cromatogramas analisados para a área mínima no conjunto de dados (por coluna do equipamento) A vermelho : área integrada pelo algoritmo ; A preto, a tracejado: limites de integração e fundo cromatográfico	39
3.6	Cromatogramas analisados para a área máxima no conjunto de dados (por coluna do equipamento)	41
3.7	Funcionamento da aplicação de teste de algoritmos sem substituição de resultados	42
3.8	Rejeição discriminada pelos equipamentos e pelas suas colunas cromatográficas	44
3.9	Estimativas da proporção global de rolhas para cada um dos grupos de rolhas no NDtech	45
3.10	Erro de integração do pico de TCA pelo NDfit	47
3.11	Alterações ao funcionamento do algoritmo	50
3.12	Funcionamento da aplicação de teste de algoritmos em modo de substituição de resultados	51
3.13	Resultados de medidas sucessivas das áreas dos picos de TCA de um conjunto de 8 rolhas no NDtech	54
3.14	Resultados de medidas sucessivas das áreas dos picos de TCA de um conjunto de 8 rolhas no NDtech, com filtro de média aplicado.	55
4.1	Exemplo de paralelização do algoritmo desenvolvido neste trabalho.	58
A.1	Esquema de funcionamento da interfaceNDfittestester	65
A.2	Interface gráfica para testes dos algoritmos	66

Índice de Tabelas

1.1	Análise de um conjunto de rolhas identificadas como G2	11
1.2	Análise de um conjunto de rolhas identificadas como G3	11
3.1	Comparação entre as classificações dos diferentes algoritmos	43
3.2	Classificações das rolhas por ambos os algoritmos, em percentagem do total	44
3.3	Distribuição de classificação das rolhas por grupo (em número de rolhas) .	48
3.4	Distribuição de classificação das rolhas por grupo (em número de rolhas) .	48
3.5	Classificação das rolhas em grupos pelo algoritmo atual da AC	52
3.6	Classificação das rolhas em grupos pelo algoritmo NDfitM	52
3.7	Classificação das rolhas em grupos pelo algoritmo NDfitA	52
3.8	Precisão na identificação do G1 pelos algoritmos em análise	52
3.9	Rolhas contaminadas no G1 não detetadas pelos algoritmos em análise . .	53
A.1	Instruções da aplicação	67
B.1	Comparação entre as áreas medidas pelo algoritmo atual e pelo e a quantidade de TCA medida no laboratório para as rolhas TCA-negativas	69
B.2	Comparação entre as áreas medidas pelo algoritmo atual e pelo NDfitM e a quantidade de TCA medida no laboratório para as rolhas do G1	69
B.3	Comparação entre as áreas medidas pelo algoritmo atual e pelo NDfitM e a quantidade de TCA medida no laboratório para as rolhas G2 em ambos os algoritmos	70
B.4	Comparação entre as áreas medidas pelo algoritmo atual e pelo NDfitM e a quantidade de TCA medida no laboratório para as rolhas do G1	70
B.5	Percentagem de rolhas TCA-positivas nos G2 dos vários algoritmos	72
B.6	Percentagem de rolhas TCA-positiva e TCA-negativa nos G3 dos vários algoritmos	72

Glossário

Aceitação	Proporção de rolhas que, em produção na produção de rolhas NDtech®, são atribuídas ao Grupo 1, em percentagem.
NDfit	Denominação do algoritmo desenvolvido neste trabalho
NDfitA	Denominação do algoritmo desenvolvido neste trabalho - forma automática
NDfitM	Denominação do algoritmo desenvolvido neste trabalho - forma manual
TCA libertável	Medida deste composto que, por efeito da sua quantidade, localização e outras propriedades físicas da rolha, estabelece um equilíbrio de concentração com o vinho, independentemente da quantidade total de TCA na rolha.
TCA-negativa	Rolha que, quando macerada durante 24 horas em solução hidroalcoólica de concentração 12% V/V, provoca contaminação da solução inferior a 0,5 ng/L de 2,4,6-tricloroanisol por litro. Sinónimo de não detetável (ND).
TCA-positiva	Definida por oposição a TCA-negativa

Siglas/Acrónimos

AC	Amorim Cork
AIFL	Área Integrada com Fundo Linear
AIFP	Área Integrada com Fundo Polinomial
APCOR	Associação Portuguesa da Cortiça
ECD	Sigla da língua inglesa <i>electron capture detector</i> (detetor de captura eletrónica)
EMG	Sigla da língua inglesa <i>exponentially modified gaussian</i> (gaussiana modificada exponencialmente)
FE	Fase estática
FM	Fase móvel
FN	Falso(s) negativo(s)
FP	Falso(s) positivo(s)
G	Grupo
GC	Sigla da língua inglesa <i>gas chromatography</i> (cromatografia gasosa)
GC×GC	Cromatografia gasosa 2-D
GC/ECD	Sigla da língua inglesa <i>gas chromatography - electron capture detector</i> (cromatografia gasosa com detetor de captura eletrónica)
GC/MS	Sigla da língua inglesa <i>gas chromatography - mass spectrometry</i> (cromatografia gasosa com espetrometria de massa)
I&D	Investigação e Desenvolvimento
LM	Levenberg–Marquardt
ML	<i>Machine Learning</i>
ND	não detetável

ROSA	Sistema patenteado pela Amorim Cork, com o nome do seu acrónimo em língua inglesa <i>Rate of Optimal Steam Application</i> (taxa ótima de aplicação de vapor)
SPME	Sigla da língua inglesa <i>solid phase microextraction</i> (microextração em fase sólida)
TCA	2,4,6-tricloroanisol
TR	Tempo de retenção

Símbolos químicos

Cl Átomo ou elemento cloro

CO₂ Dióxido de carbono

N₂ Molécula de azoto ou azoto molecular

⁶³Ni Níquel 63

Introdução

1.1 Motivação

A indústria da cortiça, da qual Portugal é o maior exportador mundial, tem como maior fonte de receita a produção de rolhas de cortiça, tendo esta representado 73% dos mais de 1134 milhões de euros de receitas associadas a esta indústria no ano de 2021[2].

Este material apresenta-se como um material de excelência devido às suas propriedades mecânicas, nomeadamente a nível da sua compressibilidade, que, em conjunto com tratamentos de superfície aplicados, permitem um elevado grau de hermeticidade ao ar [3].

Infelizmente, por ser um material de origem natural sujeito à ocorrência de contaminantes provenientes do meio em que se encontra, as suas propriedades químicas nem sempre acompanham a qualidade das restantes. A existência de anisóis, entre os quais se destaca o 2,4,6-tricloroanisol (TCA) (fig. 1.1), e fenóis compromete as propriedades sensoriais - nomeadamente o cheiro e o sabor - do vinho, levando ao seu descarte por motivos coloquialmente designados como sabor ou "cheiro a rolha" e a perdas económicas associadas, tendo chegado aos vários milhares de milhões de dólares nos anos 90 [4-7].

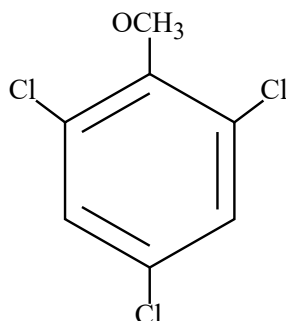


Figura 1.1: Estrutura molecular do TCA

Acresce a este problema o facto dos compostos responsáveis terem limiares de deteção pelo ser humano extremamente baixos, chegando à ordem dos nanogramas por litro em

função do tipo de vinho [5, 8], ou das décimas de nanograma por litro em água [7]. Nesta ordem de grandeza, os métodos de análise química sofrem de dificuldades de deteção e quantificação deste composto [7, 8]. Consequentemente, este problema motiva o avanço das técnicas analíticas utilizadas e das soluções de engenharia necessárias para a sua implementação.

A atenuação do problema do TCA nas rolhas de cortiça pode ser conseguida através da combinação de três tipos de abordagens: a prevenção da formação do composto ou de incorporação de rolhas contaminadas na cadeia de produção, extração deste composto ou o controlo da sua presença seguida de descarte das rolhas contaminadas. As abordagens de controlo e extração podem ser aplicadas separadamente ou em conjunto. Já as medidas preventivas são aplicadas em conjunto com as restantes. Para diminuir a carga de esforço sobre estes processos de combate ao TCA existe também um conjunto de boas práticas expressas no Código Internacional das Práticas Rolheiras [9].

Apesar das abordagens de erradicação não serem o foco deste trabalho, pela sua importância, serão seguidamente resumidas.

Os sistemas ROSA (*Rate of Optimal Steam Application*)[10] e as suas variantes (ROSA Evolution, Super ROSA [11] e Cork Nova) permitem a redução dos níveis de 2,4,6 - tri-cloroanisol nas rolhas naturais, nos discos para rolhas TwinTop® e de espumante e nos aglomerados para fabrico de rolhas técnicas através da aplicação de vapor de água e temperatura elevada (próxima dos 100 °C) de forma a conciliar a remoção de anisóis com a formação de defeitos mecânicos e/ou estéticos na rolha. Todos os tipos de rolha mencionados podem ser encontrados na fig. 1.2.

Outros métodos de remoção incluem a dessorção de TCA em ambiente de vácuo[16], possível de ser aplicada a rolhas naturais ou por extração por via do uso de CO₂ super-crítico [6], sendo o último destes apenas aplicável a granulados com vista à produção de rolhas técnicas.

De entre os processos de controlo, a Associação Portuguesa de Cortiça (APCOR) destaca a análise sensorial [17, 18], onde o olfato humano é utilizado, e sistemas de microextração em fase sólida (SPME) seguida de cromatografia em fase gasosa (GC) com detetores de espectrometria de massa (SPME-GC/MS) ou de captura eletrónica (SPME-GC/ECD).

No seguimento do esforço de modernização tecnológica transversal à indústria corticeira, a AC registou a marca NDtech®, orientada para o segmento de vinhos de qualidade superior, assegurando controlo de qualidade rolha-a-rolha, pela medição do nível de TCA libertável¹ (fig. 1.3). Esta medição é efetuada com recurso a um sistema GC/ECD industrial desenvolvido pela AC em conjunto com uma empresa inglesa. Este sistema difere de outros do mesmo tipo para utilização em laboratório pela necessidade de adequar a duração do processo à cadência de produção de uma unidade industrial. Como tal, o sinal é truncado mecanicamente, através de um sistema de válvulas de controlo do fluxo.

¹"Medida deste composto que, por efeito da sua quantidade, localização e outras propriedades físicas da rolha, estabelece um equilíbrio de concentração com o vinho, independentemente da quantidade total de TCA na rolha."[8]



Figura 1.2: Exemplos de rolhas naturais (a), de aglomerado com dois discos (b), de espumante com um ou dois discos (c) e de aglomerado (d). As primeiras são fabricadas a partir de tratamentos mecânicos da cortiça. Todas as outras são fabricadas a partir da colagem e moldação de discos e grânulos de cortiça com colas poliméricas.

A atuação destas introduz artifícios no sinal que, em conjunto com a eluição de outros compostos, dificultam a análise da localização e quantificação do analito de interesse numa dada rolha.

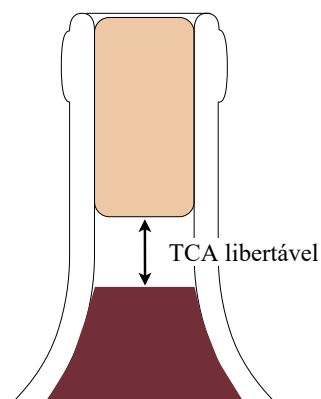


Figura 1.3: TCA libertável, o qual se encontra em equilíbrio de concentrações com o vinho

Dadas estas dificuldades, a empresa utilizava dois *softwares* distintos que analisavam cada cromatograma de forma a acordarem sobre o eventual descarte de uma dada rolha para produção de outros produtos. Mais recentemente tem vindo a ser usado apenas o programa Ellution devido à disparidade de resultados entre este e a alternativa.

Propõe-se com este trabalho a conceção de um novo algoritmo de integração do sinal cromatográfico de forma a colmatar as falhas dos modelos de decisão existentes, tendo como principais objetivos:

1. Diminuir a quantidade de rolhas marcadas incorretamente como TCA-negativas - falsos negativos (FN);
2. Diminuir a quantidade de rolhas marcadas incorretamente como TCA-positivas - falsos positivos (FP);
3. Diminuir a necessidade de interação humana nos fatores que condicionam a quantificação do TCA libertável presente numa dada rolha.

Acresce a estes, como objetivo secundário, incrementar o conhecimento sobre os motivos de falha do NDtech.

1.2 NDtech

O NDtech, cujas rolhas foram introduzidas no mercado em 2016, foi o primeiro sistema industrial de análise individual da quantidade de TCA libertável, com a garantia de uma quantidade deste inferior a 0,5 ng/L.

A introdução das rolhas NDtech® representou um passo importante para a indústria em que, à data, nenhum outro concorrente conseguia assegurar a inexistência de TCA libertável nas suas rolhas de cortiça naturais, apenas nas rolhas aglomeradas. Esse feito foi possível à custa dos vários milhares de milhões de euros investidos ao longo dos 5 anos precedentes ao início da comercialização deste produto [19].

Existem presentemente três versões dos equipamentos de produção: os da versão inicial, denominados "1.0", os de uma segunda versão - "2.0" e um conjunto de equipamentos dedicados exclusivamente à análise de rolhas de espumante. Este projeto incide maioritariamente sobre os equipamentos do NDtech 1.0, representados na fig. 1.4.



Figura 1.4: Equipamento de produção no NDtech. Imagem de [20]

1.2.1 Sistemas GC/ECD

Um sistema de GC/ECD (figura 1.5) é uma técnica particular de entre as diversas técnicas de cromatografia gasosa - GC - em que o detetor utilizado é um ECD.

A sua componente cromatográfica tem como constituintes principais um injetor, onde uma amostra volátil ou volatilizável é colocada, uma coluna e um detetor.

O processo inicia-se com a injeção do composto a analisar no injetor, onde é aquecido. Dentro da coluna, os vários componentes da amostra oscilam entre ligações à fase estática - FE - (ao material do qual é constituído a coluna) e ao gás de arrasto (fase móvel - FM). Assim, a velocidade média de cada um dos componentes na coluna é proporcional à fração do tempo que um dado composto passa na FM, em detrimento da FE.

Uma vez que a ligação dos compostos à coluna depende de processos físicos termicamente ativados, a coluna é mantida dentro de um forno, cuja temperatura pode ser constante no tempo ou alternar entre patamares e rampas de aquecimento, de forma a alterar o sinal do cromatograma obtido, com contrações e expansões locais no eixo horizontal (temporal) do cromatograma. Estas alterações de temperatura permitem dar destaque a regiões de interesse e suprimir outras onde não se espera o aparecimento de picos cromatográficos.

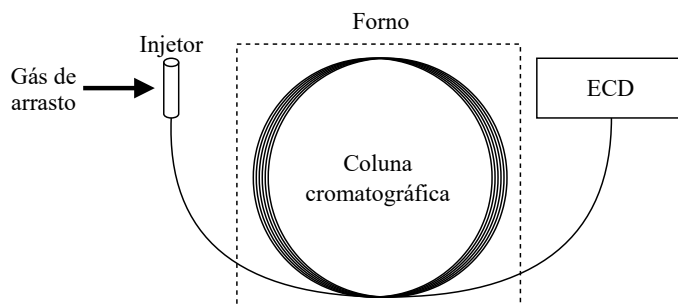


Figura 1.5: Esquema simplificado de uma montagem GC/ECD

Após a passagem pela coluna e separação dos componentes, estes atingem o detetor ao fim de um determinado tempo, denominado tempo de retenção - TR. Apesar de não identificar univocamente um composto, este é um forte indicador da sua presença. Em alternativa, poderia ser usado um espectrómetro de massa, numa montagem GC/MS para melhor distinção de compostos com TR semelhante. No entanto, os custos superiores de aquisição e manutenção deste equipamento em ambiente industrial justificam a escolha de um ECD para uso no NDtech.

Na fig. 1.6 encontra-se esquematizado um ECD, um detetor colocado no término da coluna cromatográfica cujo método de funcionamento é a medida de uma corrente que atinge um detetor de Faraday atenuada ou não pela absorção de eletrões pelo gás. Tem como constituintes principais uma rede com um emissor de radiação β^- (emissor de eletrões) de baixa energia (A). No caso do NDtech, são utilizados detetores com ^{63}Ni que, na transformação de um dos seus neutrões num próton e num par eletrão - neutrino

antieletirão confere ao primeiro energia média de $17,434 \pm 0,004$ keV e nunca superior a $66,980 \pm 0,015$ keV [21, 22]. Para além da rede, o ECD tem como componentes um detetor de Faraday, (B) e um conjunto de terminais (não representados) que aplicam um campo elétrico ao detetor ($\vec{E}$).

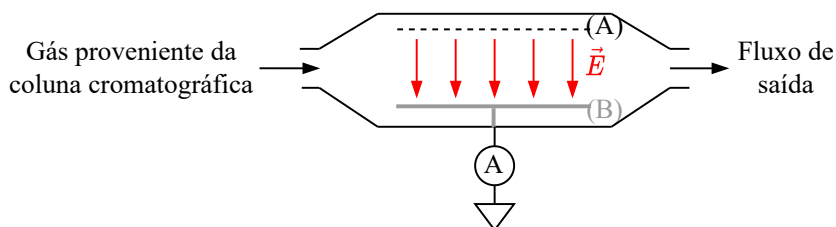


Figura 1.6: Esquema simplificado do funcionamento de um ECD

Ao ser emitida, a radiação β^- ioniza azoto - N_2 , utilizado como gás de arrasto, formando uma corrente de eletrões, posteriormente acelerados pelo campo elétrico externo na direção e sentido do detetor de Faraday.

Quando o percurso destes é interrompido pela passagem de um gás com elementos com eletronegatividade acentuada (como é o caso do TCA, em que cada molécula contém 3 átomos de cloro - Cl), uma determinada percentagem dos eletrões é absorvida pelos átomos desse gás, tornando-se mensurável uma diminuição da corrente coletada em comparação com a passagem do gás de arrasto isoladamente.

O cromatograma obtido numa montagem convencional GC/ECD é, então, uma sequência de picos estreitos (quando comparados com a totalidade do cromatograma) e, em geral, de fácil identificação e separação (fig. 1.7). Para a obtenção de um destes cromatogramas são frequentemente necessários vários minutos. Tal morfologia permite facilmente identificar os picos e quantificar a sua área, exceto em casos de coeluição. Este processo ocorre quando dois agentes químicos contidos na amostra são expelidos da coluna com pouco tempo entre eles, formando picos fracamente separados no tempo.

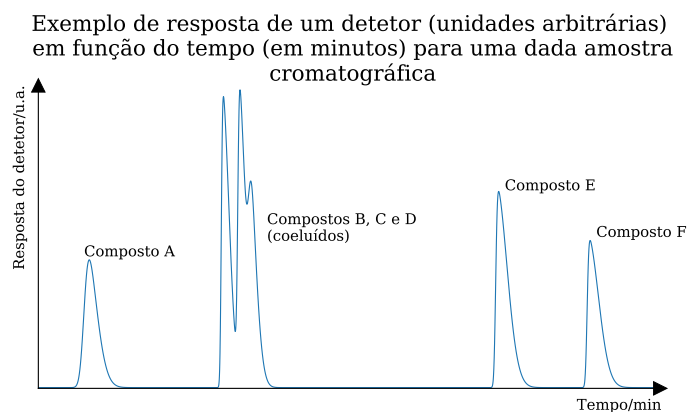


Figura 1.7: Representação de um cromatograma convencional e ideal (simulado)

1.2.2 Sistema GC-ECD do NDTech (GC×GC com *heartcut*)

Para a produção de rolhas da marca NDtech®, o processo cromatográfico deve ser acelerado de forma a adequar a velocidade da técnica à cadência de produção de uma unidade industrial.

Em primeiro lugar, o curto intervalo de tempo reservado para cada rolha no NDtech não permitiria a completa separação do pico de TCA dos outros compostos presentes na cortiça. Para tal, utiliza-se cromatografia gasosa 2-D (GC×GC), que faz uso de duas colunas cromatográficas distintas (pré-coluna e coluna), o que permite melhor separação dos compostos para um dado intervalo de tempo [23]. Para a utilização de duas colunas, é necessário o uso de válvulas que redirecionem os fluxos de gases entre as colunas e detetores, o que induz picos de pressão no sistema e, consequentemente, artefactos no sinal.

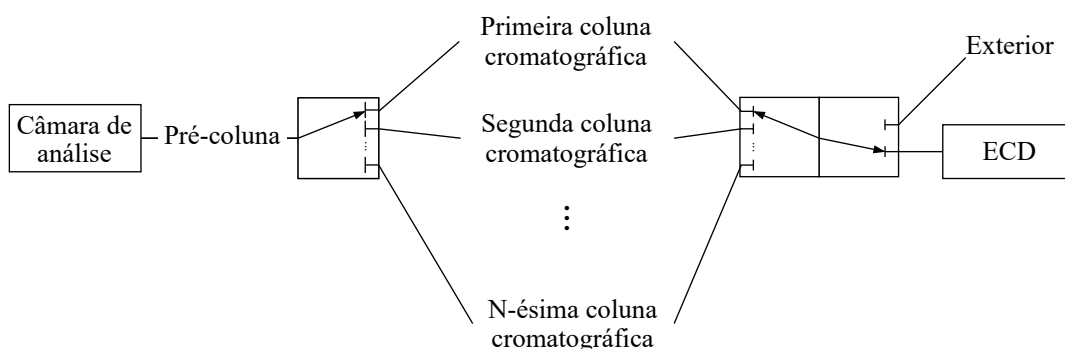


Figura 1.8: Esquema simplificado da ligação das várias colunas cromatográficas a um mesmo ECD

Este sistema permite que várias colunas partilhem um detetor, tendo cada câmara um intervalo de apenas cerca de 15 segundos alocado no detetor, tempo bastante inferior ao necessário para as técnicas cromatográficas convencionais.

Neste intervalo de tempo é necessário que a pressão se iguale entre a coluna e o sensor e que o eventual pico de eluição de TCA se manifeste.

Um sistema implementado no sentido de facilitar a extração de informação é a introdução de gás de *make-up* no detetor. Este impede o retrocesso do gás dentro da câmara do ECD, estreitando os picos dos compostos eluídos (fig. 1.9).

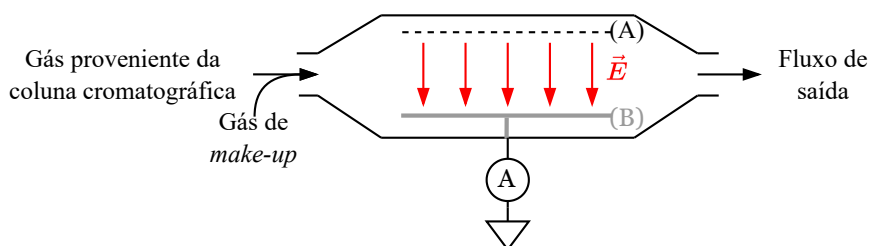


Figura 1.9: Esquema de funcionamento de um ECD com gás de *make-up*

Apesar das soluções tecnológicas implementadas, os cromatogramas obtidos continuam a demonstrar elevada variabilidade entre si, demonstrada pelos diversos perfis cromatográficos, alguns dos quais representados abaixo (fig. 1.10).

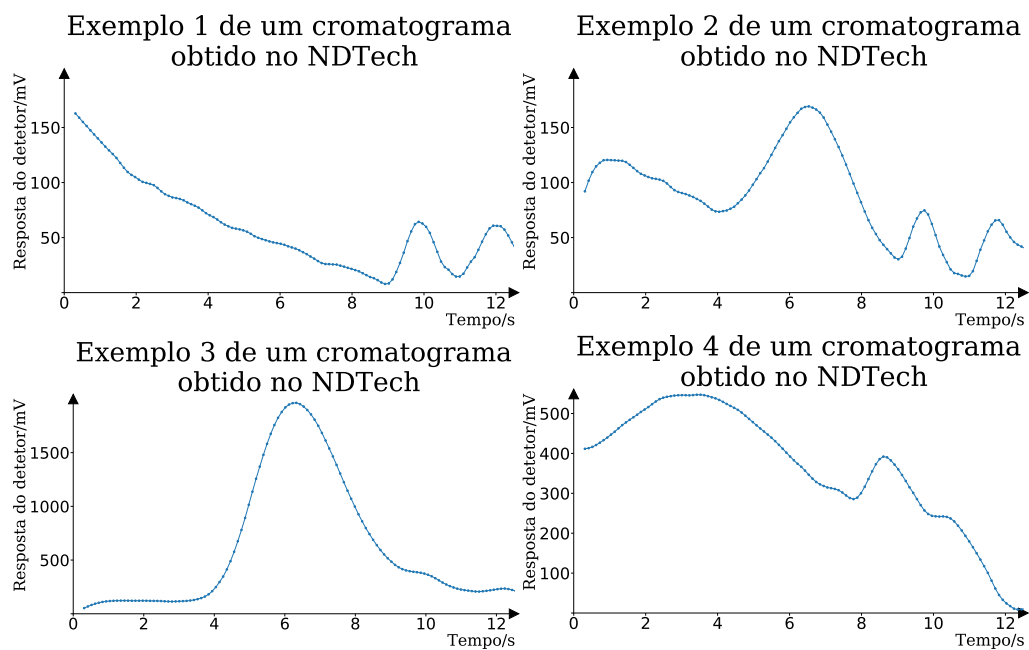


Figura 1.10: Exemplos de cromatogramas obtidos no NDTech

1.2.3 Obtenção de rolhas NDtech®

Atualmente, a análise efetuada no NDTech resulta na separação das rolhas em três grupos, denominados numericamente como G1, G2 e G3, em que o primeiro denomina a maior qualidade e uma quantidade de TCA considerada TCA-negativa ou não detetável (ND)² cuja medida deve ser inferior a 6 mV.s de área integrada. O último dos grupos denomina as maiores quantidades deste composto sendo o limiar de separação das rolhas do G2 e G3 definido, por omissão, por uma área integrada de 25 mV.s, podendo ser alteado conforme os objetivos de produção. Este processo encontra-se resumido na fig. 1.11.

Frequentemente, o G2 é sujeito aos processos de remoção de TCA mencionados anteriormente (1.1).

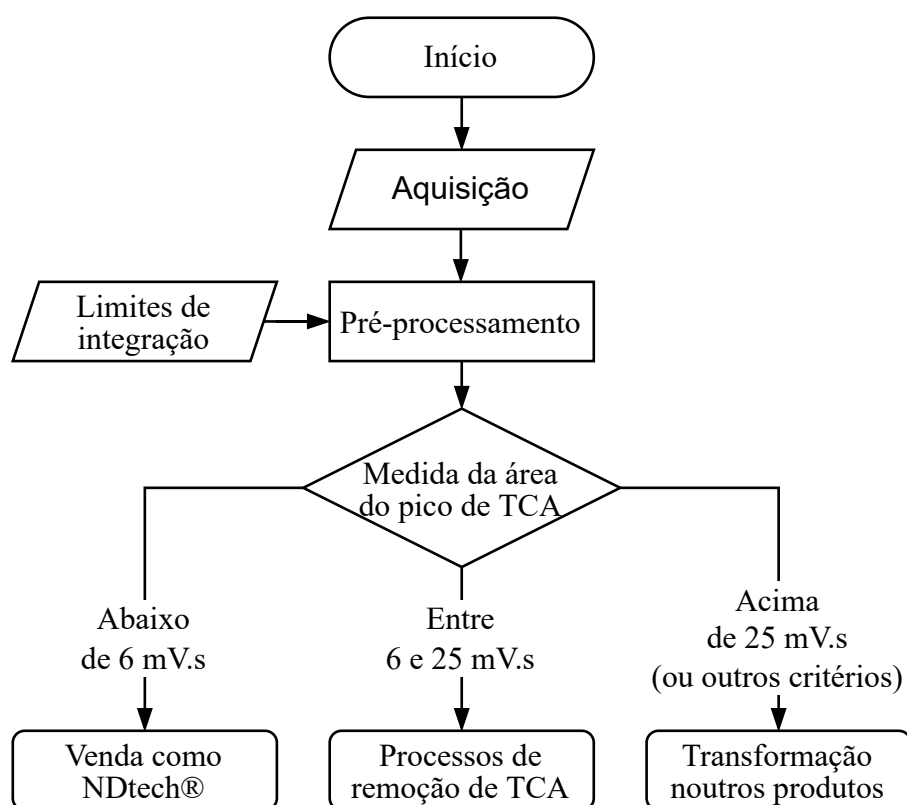


Figura 1.11: Funcionamento do NDtech

²Ver glossário

O algoritmo utilizado atualmente faz uso de dois limites de integração, ao nível do eixo horizontal, definidos por um operador. A intersecção da resposta do detetor com as retas verticais (representadas a preto, em 1.12) define univocamente um segmento de reta, sendo calculada a área limitada pela resposta do detetor (superiormente) e pelo segmento de reta (inferiormente, a tracejado). Outras particularidades sobre o funcionamento deste algoritmo serão discutidas posteriormente (secção 2.2).

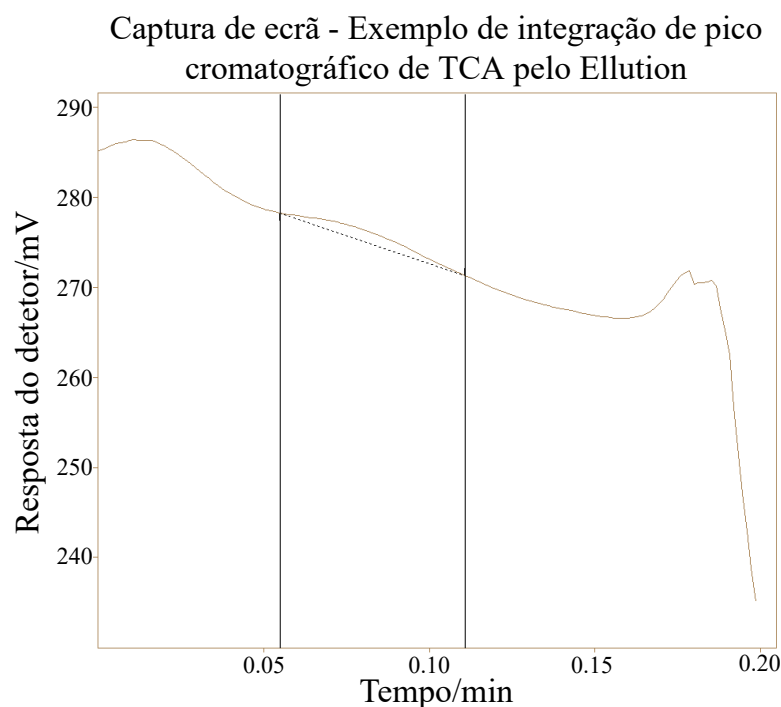


Figura 1.12: Captura de ecrã, adaptada: exemplo de funcionamento do algoritmo de integração utilizado no NDtech

A performance no NDtech é avaliada frequentemente, sendo sujeita a dois tipos de controlo:

De equipamento, que ocorre diariamente e no qual são controlados:

1. O limiar de identificação de uma amostra como TCA-positiva (G2);
Em que uma dada quantidade de solução hidroalcoólica de TCA deve ser sujeita ao mesmo processo de análise que uma rolha e deve ser atingida uma determinada área integrada do pico deste composto.
2. A retenção de compostos nas câmaras e/ou colunas do cromatógrafo entre medidas - *carry-over*;
Em que após uma medição de uma determinada quantidade de TCA, algumas ordens de grandeza superior à do controlo 1, uma medição não deve ser maior do que 50% do valor da medida que a precede.
3. Outros distúrbios ao funcionamento do sistema, como a alteração do TR do TCA.

De produto, que ocorre de 2 modos:

1. É testada uma amostra de 200 rolhas no laboratório de Investigação e Desenvolvimento (I&D) de cada lote do G1, a fim de amostrar a eventual presença de rolhas contaminadas neste grupo;
2. São testadas amostras dos grupos 2 e 3, para que se detete a presença de rolhas que deveriam ter sido classificadas como G1. Este controlo de qualidade é realizado de forma aleatória conforme a disponibilidade do laboratório de I&D

O relatório interno do NDtech de 2021 [24] dá conta de uma precisão do G1 (P_1) - calculado segundo a eq.1.1 - de 99,96% na identificação de rolhas pelos equipamentos NDtech 1.0:

$$P_1 = \frac{\text{Rolhas TCA-negativas do G1}}{\text{Total de rolhas do G1}} \times 100 (\%) \quad (1.1)$$

O mesmo documento dá conta das estatísticas para a análise de amostras pertencentes aos G2 e G3:

Tabela 1.1: Análise de um conjunto de rolhas identificadas como G2

	Rolhas G2 do NDTech 1.0	
	TCA-positivas	TCA-negativas
Quantidade	70	426
Percentagem	14,1%	85,9%
Total	496	

Tabela 1.2: Análise de um conjunto de rolhas identificadas como G3

	Rolhas G3 do NDTech 1.0	
	TCA-positivas	TCA-negativas
Quantidade	21	3
Percentagem	87,5%	12,5%
Total	24	

Quaisquer desenvolvimentos com vista a melhorar os resultados acima terão que ter cuidado especial de forma a não implicarem aumentos de rolhas FN (contaminadas com TCA e classificadas como G1).

Acrescenta-se a isto o facto de, devido à manutenção de algum TCA na coluna cromatográfica após a medida da quantidade deste composto numa rolha positiva (fenómeno de *carry-over* entre medidas), poderá ser inevitável que os G2 e G3 contenham rolhas TCA-negativas, uma vez que poderão vir no seguimento de rolhas G3 e o *carry-over* poderá ser suficiente para que uma rolha subsequente a outra seja classificada como G2 ou mesmo G3.

1.3 Estrutura da dissertação

O objetivo deste trabalho é desenvolvimento de um algoritmo que permita a melhor distinção do pico de TCA nos cromatogramas obtidos no NDtech. A dissertação está estruturada de forma cronológica. No entanto, poderão ocorrer desvios a esta ordem no sentido de viabilizar a melhor compreensão de alguns problemas e soluções encontrados.

O capítulo 2 contém uma abordagem à morfologia dos cromatogramas e aos métodos matemáticos e de análise de sinal necessários para o posterior desenvolvimento da solução que se pretende aplicada. Segue-se o cap. 3 onde, nos seus subcapítulos, serão expostos os métodos de teste e validação dos algoritmos candidatos, seguidos dos seus resultados.

No cap. 4 são expostas as conclusões e sugestões para trabalho futuro. Existem também apêndices (A e B) dos quais constam alguns detalhes sobre o funcionamento e utilização da interface utilizada para os testes cuja apresentação não faria sentido na componente principal do documento.

Sinal obtido e algoritmos

2.1 Morfologia do sinal cromatográfico no NDtech

Como introdução à morfologia geral do sinal, vale a pena recordar que este é proveniente de uma técnica cromatográfica que viu o seu processo de execução acelerado e truncado. Como tal, esperam-se artefactos decorrentes das transições entre as fases do processo de análise/fabrico, marcadas por estruturas do sinal que manifestam variações de pressão. Entre estas, espera-se uma zona com relativa estabilidade onde ocorrerá o pico de TCA. Deste modo identificam-se 3 secções - ilustradas na fig. 2.1:

Secção A ou inicial:

Onde se manifestam variações de sinal devido à troca de posição da válvula que liga a pré-coluna à coluna cromatográfica;

Secção B ou de pico:

Onde, caso exista, se apresenta o pico de TCA;

Secção C ou final:

Onde ocorre o pico causado pela ligação da coluna ao canal de saída.

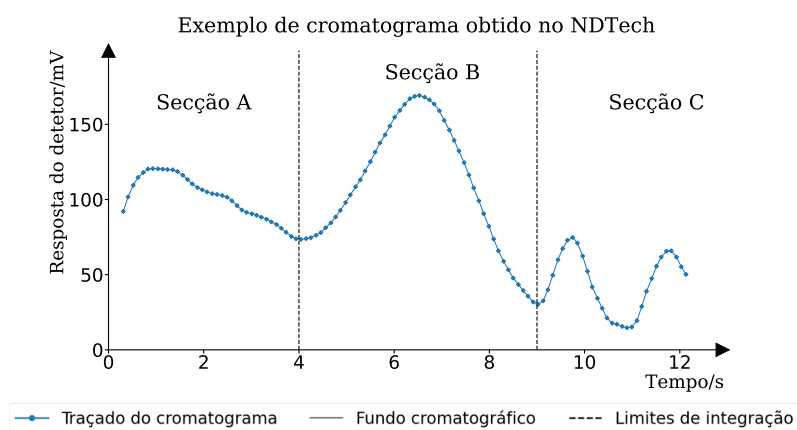


Figura 2.1: Secções do sinal cromatográfico

A separação entre as secções demonstra alguma variabilidade, especialmente no que toca à separação da secção A e B, uma vez que o TR do TCA pode ser alterado pelas condições em que é realizada a cromatografia, quer intencionalmente, quer por alterações das condições do meio como a temperatura ou a humidade.

Já o início da secção C, marcado pelo pulso proveniente da válvula que liga a coluna cromatográfica ao exterior, permite que seja mais facilmente utilizado como referência.

Apesar de Marco e Bombi indicarem a existência de "cerca de 90 fórmulas matemáticas empíricas" para a descrição do pico cromatográfico [25], é dado destaque a 5 funções, nomeadamente as funções gaussiana, gaussiana modificada exponencialmente (EMG), a distribuição de Poisson, a função log-normal e as séries de Edgeworth-Cramer e de Gram-Charlier. Na figura 2.2 são representadas as funções de densidade de probabilidade gaussiana, EMG, log-normal e função massa de probabilidade de Poisson. A Séries de Edgeworth-Cramer / Gram-Charlier não se encontram representadas uma vez que são formas de cálculo dos momentos estatísticos do pico cromatográfico.

Representação gráfica das funções em destaque.

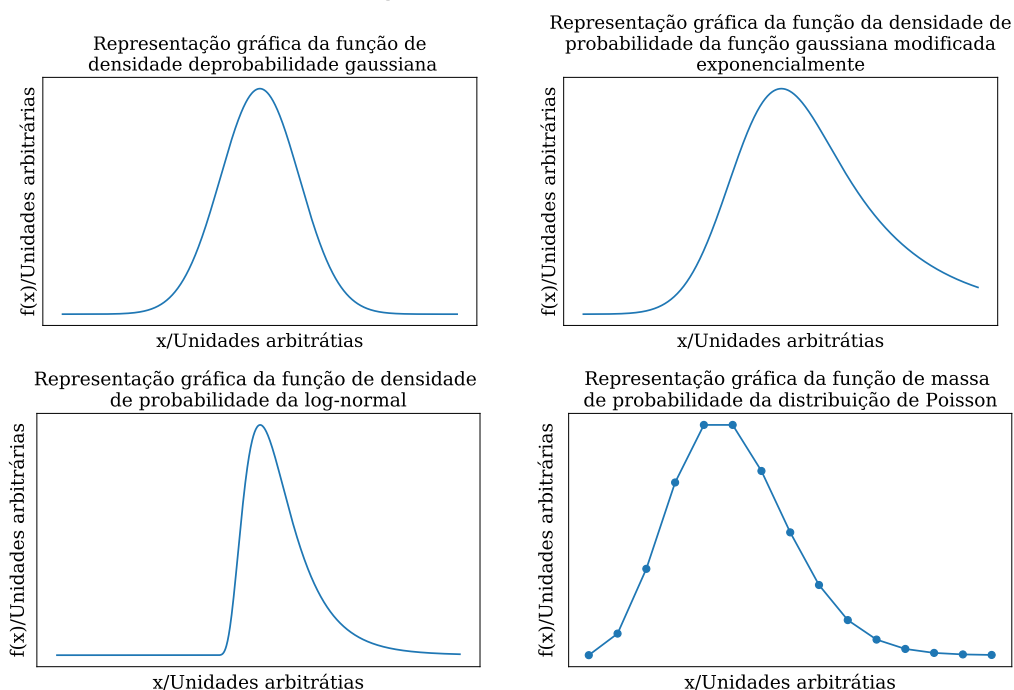


Figura 2.2: Representação gráfica de algumas funções utilizadas como modelos de pico cromatográfico.

A primeira distingue-se das seguintes pela sua simetria em relação ao máximo da distribuição, o que não é propriedade de todos os picos obtidos em experiências de GC. Todas as outras permitem ter esse facto em conta através de um ou mais parâmetros.

Vale a pena notar que a utilização destas funções é empírica uma vez que resultam de tentativas de ajuste de funções matemáticas conhecidas (em geral perturbações da distribuição gaussiana) aos picos obtidos nas técnicas analíticas e não na tentativa de

estudo dos fenómenos físicos envolvidos com vista à criação de um modelo do pico. Como tal, os diversos parâmetros dos modelos matemáticos fornecem significado físico limitado [26].

Sobre as restantes componentes do cromatograma: o fundo cromatográfico engloba todo o sinal que não provém de picos de eluição de compostos com assinatura cromatográfica e é a componente que se pretende separar dos picos cromatográficos. Ao longo deste trabalho, esta noção será generalizada afim de incluir todas as estruturas cromatográficas (incluindo picos de outros compostos) que não o pico de TCA. Assim, pretende-se que os algoritmos desenvolvidos neste trabalho efetuem a separação fundo-pico de forma a que o resultado desta separação mantenha as características morfológicas do pico separado do fundo, indicando que esta separação permite a integração completa de toda a área relativa à assinatura cromatográfica do TCA.

2.2 Método atual de integração do pico cromatográfico no NDtech

Tal como descrito em 1.2.3, atualmente a quantificação do pico de TCA dá-se pela integração da área entre o cromatograma obtido e uma reta, compreendido entre 2 limites temporais fixos de integração.

Foi também explicitada em comunicação pessoal com a empresa que produz o *software* utilizado (Ellution) que existem 3 regras em relação à integração dos picos:

1. Os limites de integração são arredondados ao ponto de dados mais próximo
2. Sob nenhuma circunstância o cromatograma pode cruzar a reta que atua como limite inferior de integração, sendo os limites de integração alterados no sentido de esta regra ser respeitada.
3. Em caso de sobreposição de dois picos, o limite de integração é traçado como uma linha vertical que separa os dois compostos, no ponto mais baixo da zona de sobreposição.

A forma exata como o programa faz cumprir a segunda não é conhecida, no entanto, a implementação deste algoritmo aparenta apresentar deficiências devido a:

Concavidades do cromatograma: a quantização do nível de TCA em zonas com concavidade acentuada, seja ela positiva ou negativa, pode levar a sub-estimas ou sobre-estimas da quantidade presente.

Alterações ao TR: podem levar a integrações incompletas do pico.

Coeluição de outros compostos: leva à sobreposição dos picos cromatográficos, dificultando a escolha de limites de integração que permitam discernir ambos os picos.

Quantidades elevadas de TCA: uma vez que uma contaminação superior leva a um alargamento do pico e consequentemente à sua integração parcial.

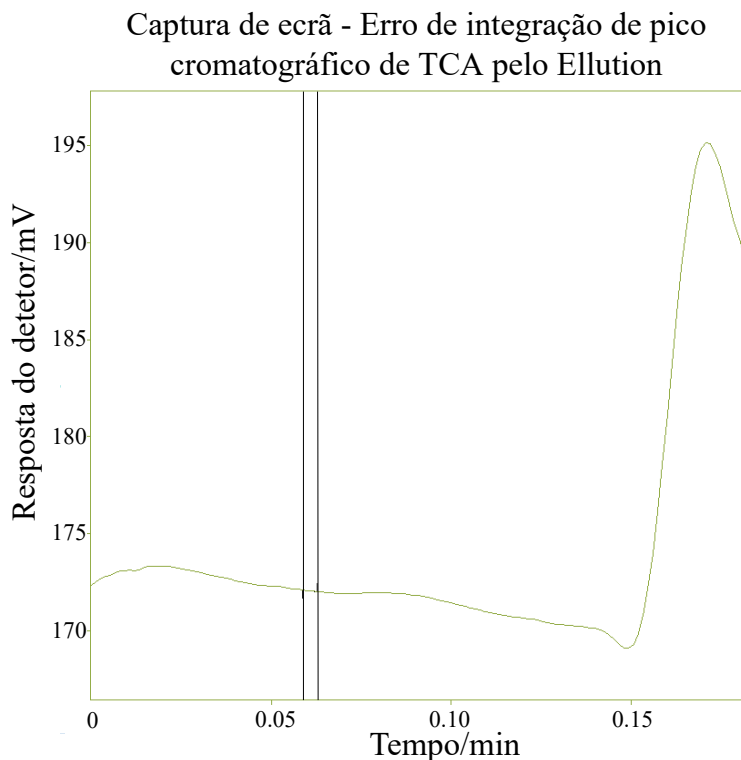


Figura 2.3: (Captura de ecrã adaptada) erro de integração do Ellution. É visível que os limites de integração (linhas verticais a preto) despreza o pico cromatográfico à direita destes.

Note-se que a última é manifestamente menos importante no propósito da marca NDtech®, uma vez que a distinção das quantidades deste composto em quaisquer rolhas com quantidade de TCA superior à que o G1 permite é irrelevante no que à produção de rolhas NDtech® diz respeito.

Os resultados são confirmados (por amostragem) no laboratório de I&D da AC através de uma análise SPME GC/ECD do macerado de 24 horas de cada rolha em solução hidroalcoólica de 12% V/V. Nesta análise, as rolhas NDtech® deverão apresentar valores de TCA < 0,5 ng/L.

2.2.1 Fontes de falsos positivos

Foram identificadas duas fontes significativas de FP: o *carry-over* e as rejeições sucessivas/recorrentes.

O fenómeno de *carry-over*, consiste na manutenção de TCA resultante da rolha anterior na câmara ou coluna cromatográfica, gerando uma falsa leitura. Este fenómeno tem o efeito de limitar superiormente as melhorias possíveis ao sistema de classificação do NDtech.

O limite de retenção de TCA que justifica a paragem de um determinado equipamento é, presentemente, a repetição de 50% do valor anterior entre medidas. Mesmo em equipamentos com quantidades de *carry-over* abaixo do limiar que justifica a sua paragem, a rejeição de rolhas subsequentes a rolhas TCA-positivas é mais frequente do que após rolhas TCA-negativas, sendo este facto confirmado pelos operadores dos equipamentos.

De forma a analisar a efeito do *carry-over* foi retirada da linha de produção uma rolha cujas medidas de área do pico cromatográfico indicam a presença de TCA e esta foi colocada na câmara 1 do equipamento, permanecendo as restantes vazias. Foi desativado o equipamento que coloca e retira as rolhas do cromatógrafo de forma a que as condições não se alterassem e foi deixado o equipamento a funcionar nestas condições.

A área do pico de TCA¹ medido na câmara 1 foi de 70 ± 4 mV.s e, mesmo com uma medida de *carry-over* de $31 \pm 4\%$, ocorreu a rejeição das 4 rolhas seguintes. Nota-se também na câmara 5 - que partilha a coluna cromatográfica com a coluna 1 - a medição de áreas do pico cromatográfico que não acompanha a tendência de diminuição verificada nas restantes câmaras.

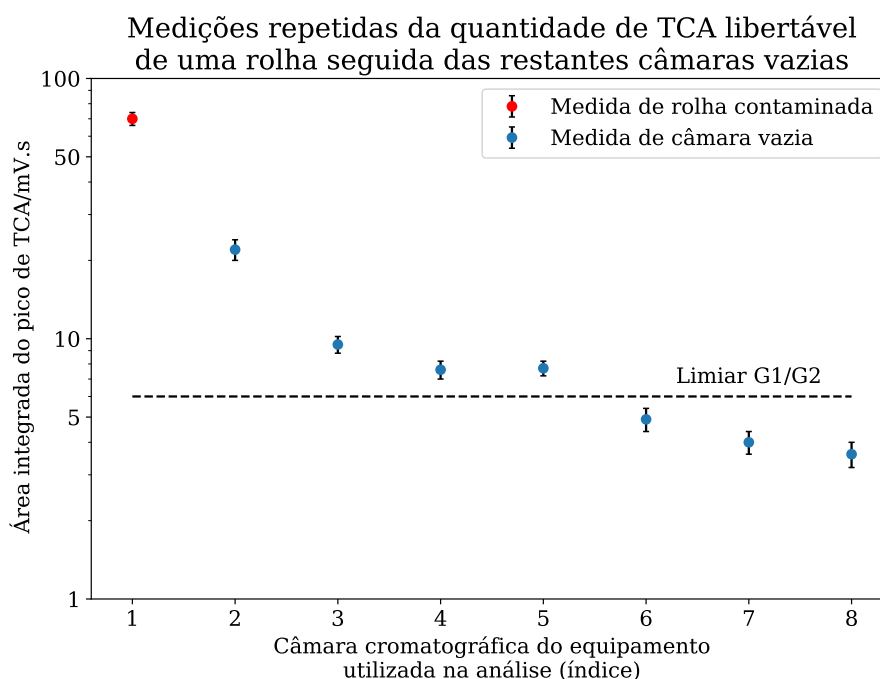


Figura 2.4: Rejeição por *carry-over*

A rejeição sucessiva manifesta-se quando a combinação de parâmetros escolhida pelo operador resulta em aceitações próximas do espectável² na maioria das colunas e numa rejeição total ou quase total das rolhas de uma das colunas cromatográficas.

Tal acontece de forma díspar com a aceitação não só das restantes colunas, como com os restantes cromatógrafos que analisam o mesmo lote. Este tipo de rejeição é resultante

¹ medida com o programa utilizado atualmente (Ellution)

²Próximas da média das aceitações de todos os equipamentos a analisar um mesmo lote de rolhas.

na inadaptabilidade do fundo linear aos vários tipos de perfil cromatográfico, nomeadamente nos casos em que o pico de TCA se encontra sobre um sinal de fundo com concavidade negativa na vizinhança do pico. Apesar de o diagnóstico deste problema poder ser efetuado por via da inspeção visual do perfil, o programa de controlo da produção utilizado pela AC não permite a separação das estatísticas de produção pelas colunas do cromatógrafo, o que não permite o cálculo dos efeitos deste problema na produção, em tempo real.

Já a rejeição recorrente ocorre quando a aceitação de um dado cromatógrafo é manifestamente inferior à aceitação dos outros que analisam a quantidade de TCA, devido à utilização de fatores de sensibilidade da medição demasiado elevados ou de intervalos de integração demasiado abrangentes.

Os resultados que permitiram a identificação destes problemas encontram-se no capítulo 3.

2.2.2 Fontes de falsos negativos

Devido ao número reduzido de rolhas FN, inferior a 0,2% das rolhas do G1 [24], é manifestamente mais difícil encontrar as fontes de FN, em comparação com as de FP. Como tal, não existe nenhuma causa que possa ser apontada como razão clara para este erro de classificação.

Uma das possíveis fontes destas classificações erróneas apresentada pela responsável do laboratório de I&D da AC é a hipótese de que ligeiras alterações ao TR do TCA causarem integrações incompletas do pico, o que levaria a que a área medida não fosse suficiente para a rejeição de uma dada rolha. Este é o caso visível na fig. 2.3. Esta hipótese poderia ser confirmada pela análise da implementação do algoritmo de medida do programa Ellution (e das regras de 2.2). Outra explicação possível é a de que a maceração de 24 horas permita uma análise em maior profundidade dos compostos de uma dada rolha, por existir eventual maior poder de extração do TCA libertável pela solução.

Para o cobrimento da totalidade de causas de erros de classificação, crê-se que seja possível que erros de execução dos *softwares* utilizados ou mau funcionamento dos equipamentos de separação das rolhas possam causar problemas tanto de FN e como de FP. Porém, na falta de qualquer tipo de investigação sobre a ocorrência destes fenómenos e da sua frequência, não é possível assegurar a existência destes fenómenos ou quantificar a sua frequência.

2.3 A matemática da análise de sinal

Pretende-se com este capítulo uma explicação breve, não exaustiva, dos fundamentos matemáticos considerados necessários para a abordagem de análise do sinal cromatográfico.

2.3.1 Otimização não-linear via método dos mínimos quadrados

A análise por otimização não-linear parte do pressuposto que, pelo menos na vizinhança do pico de TCA, uma soma de funções suficientemente próximas visual e matematicamente, quer do sinal de fundo quer do pico do analito de interesse, são suficientes para assistir ao cálculo da sua área.

Considere-se que o sinal cromatográfico obtido, $s(t)$, é uma interpolação linear do sinal amostrado numa sucessão de pontos $t \in \{t_0, t_1, \dots, t_n\}$, que pode ser replicado por uma função $f(t, a_f, b_f, c_f, \dots)$ - que funciona como sinal de fundo - e que o pico cromatográfico pode ser replicado por uma função $g(t, a_g, b_g, c_g, \dots)$, ambas com parâmetros $a_f, b_f, c_f, \dots$ e $a_g, b_g, c_g, \dots$, definidas no intervalo $t \in [t_0, t_n]$ e escolhidas empiricamente. Considera-se otimizada a combinação de parâmetros de f e de g que minimiza a soma do quadrado das diferenças entre o sinal amostrado e o valor de $(f + g)(t)$ nos pontos onde ocorre a amostragem. Sendo μ esse quadrado das diferenças tem-se:

$$\mu(\mathbf{p}) = \sum_{t=t_0}^{t_n} \left(s(t) - f(t, a_f, b_f, c_f, \dots) - g(t, a_g, b_g, c_g, \dots) \right)^2 \quad (2.1)$$

O conjunto de parâmetros que minimiza μ é uma das soluções de:

$$\nabla \mu(\mathbf{p}) = 0, \quad \mathbf{p} = [a_f, b_f, c_f, \dots, a_g, b_g, c_g, \dots] \quad (2.2)$$

em particular, é uma das soluções de 2.2 que verifica:

$$\nabla^2 \mu(\mathbf{p}) \geq 0 \quad (2.3)$$

ou seja, a solução, é o mínimo absoluto da superfície $\mu(\mathbf{p})$, porém qualquer mínimo local desta função é solução das equações 2.2 e 2.3 e esse é um fator determinante no cálculo dos parâmetros mencionados.

Tanto os desenvolvedores da biblioteca *SciPy* [27] como os homólogos da biblioteca *LMfit* (baseada na anterior) optaram por utilizar o algoritmo de Levenberg–Marquardt (LM) [28, 29].

Este baseia-se nos métodos do gradiente e de Gauss-Newton. Ambos são métodos iterativos de procura de pontos de interesse de uma função, requerendo uma tentativa inicial e o seguimento de uma regra que escolhe o ponto seguinte, efetuando um "passo".

O primeiro, procura na vizinhança de um ponto da superfície $\mu(\mathbf{p})$ a direção de maior descida (ilustrado na fig. 2.5), sendo efetuado um passo $-\lambda \nabla \mu(\mathbf{p})$, ou seja:

$$\mathbf{p}_{i+1} = \mathbf{p}_i - \lambda \nabla \mu(\mathbf{p}) \quad (2.4)$$

Otimização de uma variável x para minimização de uma função μ segundo o método do gradiente

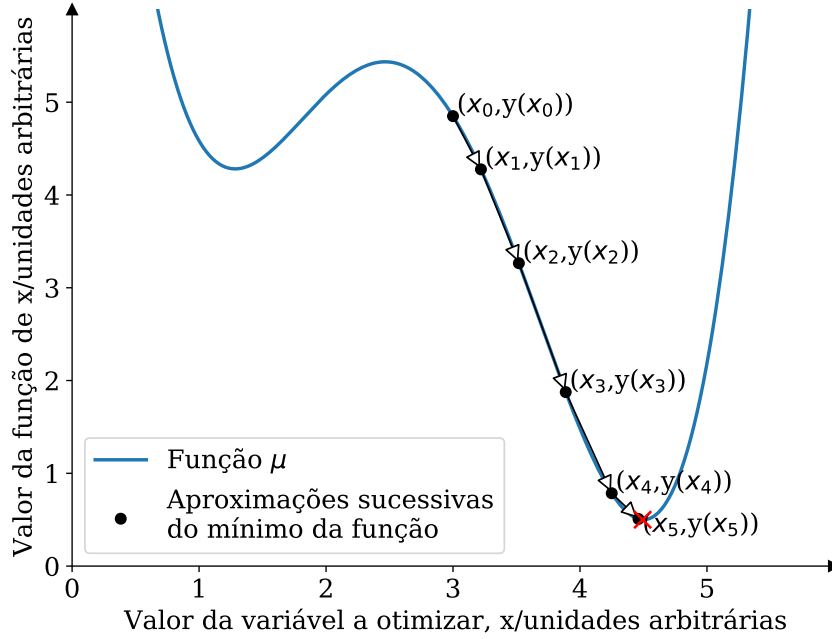


Figura 2.5: Método do Gradiente

sendo λ um escalar, em geral contido no intervalo $[0, 1]$.

A maior limitação do método do gradiente é a convergência lenta nas vizinhanças dos zeros de $\nabla\mu(\mathbf{p})$ devido à proporcionalidade entre o comprimento dos passos e o gradiente.

Já o método de Gauss-Newton define as iterações de forma diferente. A série de Taylor de primeira ordem que aproxima $\mu(\mathbf{p})$, próximo de $\mathbf{p}^*$ é:

$$\nabla\mu(\mathbf{p}) \simeq \nabla\mu(\mathbf{p}^*) + (\mathbf{p} - \mathbf{p}^*)^T \nabla^2\mu(\mathbf{p}^*) \quad (2.5)$$

o que permite resolver (numericamente) a equação:

$$\nabla\mu(\mathbf{p}) = 0 \quad (2.6)$$

seguindo passos na forma expressa matematicamente na eq. 2.7 e visualmente na fig. 2.6.

$$\mathbf{p}_{i+1} = \mathbf{p}_i - (\nabla^2\mu(\mathbf{p}_i))^{-1} \nabla\mu(\mathbf{p}_i) \quad (2.7)$$

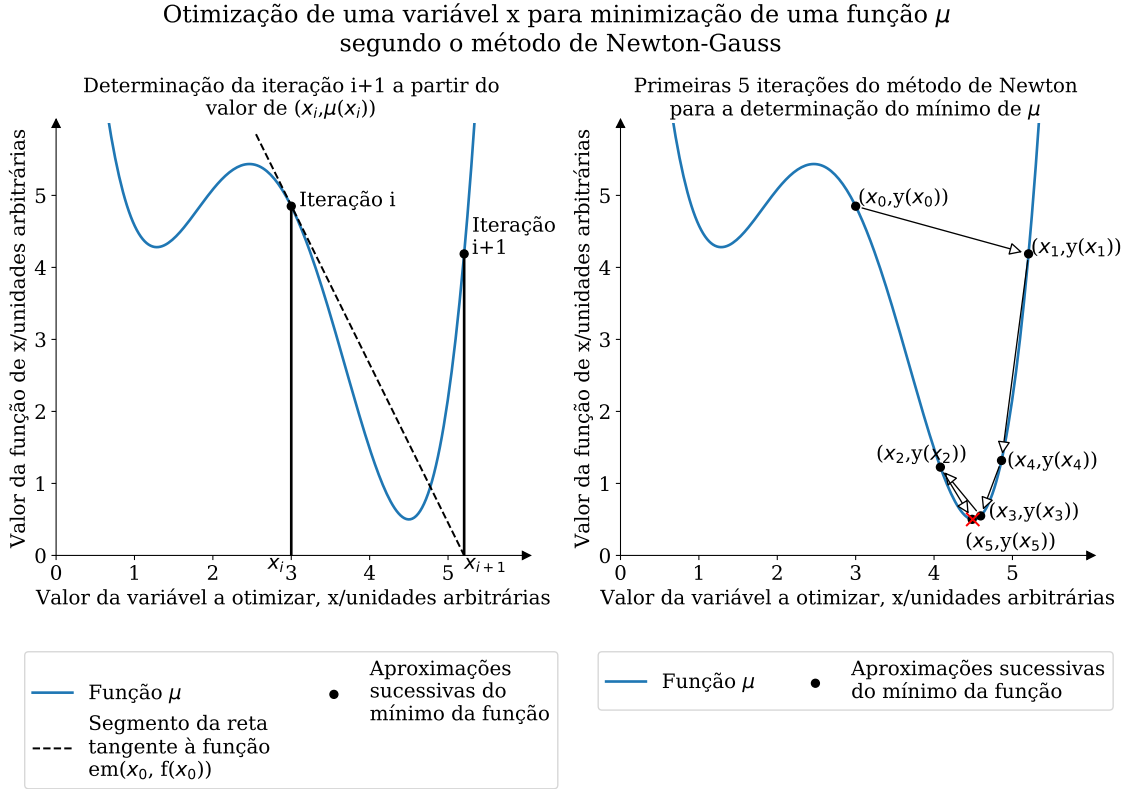


Figura 2.6: Método de Gauss-Newton

Os contributos de Levenberg [30] e de Marquardt [31] focam-se essencialmente na evidência da complementariedade destes dois métodos e na formulação de uma relação que os concilia num único algoritmo e com recurso à aproximação $\nabla^2 \mu \simeq J^T J = H$:

$$\mathbf{p}_{i+1} = \mathbf{p}_i - [H + \lambda \cdot \text{diag}(H)]^{-1} \nabla f(\mathbf{p}_i) \quad (2.8)$$

O processo de otimização é tão mais computacionalmente exigente quanto maior for o número de variáveis a otimizar. Mesmo assim, é este o método implementado por omissão em múltiplas bibliotecas, incluindo as bibliotecas *Scipy* e *LMFit* [28, 29], devido à sua velocidade de convergência em comparação com as alternativas.

2.4 Trabalho prévio - *fitplus*

O algoritmo *fitplus* resulta de colaborações anteriores entre a AC e a Universidade NOVA de Lisboa. Neste recorre-se ao uso de um fundo polinomial de 3.º grau. Os parâmetros - (a, b, c, d) - deste polinómio na forma $f(t) = a \cdot t^3 + b \cdot t^2 + c \cdot t + d$ são calculados com base nas diferenças quadradas de pontos contidos em dois intervalos selecionados pelo operador. A escolha de um polinómio do 3º grau baseia-se no facto desta função permitir mimetizar a curvatura do cromatograma sem um número excessivo de parâmetros e da crença prévia na prevalência de fundos cromatográficos com concavidade positiva, tal como o da figura 1.12.

Apesar da fração de cromatogramas deste tipo ser elevada, a a flexibilidade de utilização de um polinómio com função de fundo do 3.º grau permite o ajuste a fundos com formas variadas.

Para a definição dos intervalos onde é calculado o polinómio, ao contrário do programa Ellution, onde seriam necessários definir dois instantes t_0 e t_1 , é necessário definir um intervalo de tempo Δt de forma a criar intervalos na forma $[t_0 - \Delta t, t_0 + \Delta t]$ e $[t_1 - \Delta t, t_1 + \Delta t]$. Estes têm como propósito definir e permitir calcular o polinómio $f(t)$. Assume-se uma aproximação suficientemente boa do cromatograma em causa, caso a quantidade de TCA libertável seja nula.

Estando $f(t)$ definida no intervalo contido entre os anteriores a secção positiva da diferença $s(t) - f(t)$ é integrada, sendo o produto dessa área pelo fator de sensibilidade a medida da quantidade de TCA libertável, em mV.s.

Este método encontra-se já em uso pela equipa do NDtech, nos equipamentos mais recentes. O facto deste algoritmo ser distinto do Ellution gera necessidades de formação distintas e o aparecimento de pequenos picos de outros compostos ou estruturas provenientes do ruído do sinal que alteram as regiões $[t_0 - \Delta t, t_0 + \Delta t]$ e $[t_1 - \Delta t, t_1 + \Delta t]$ são descritos pelos operadores como elementos que dificultam a utilização e que requerem uma maior atenção nos equipamentos que utilizam este algoritmo.

2.5 Alterações ao *fitplus* - NDfit

No sentido de facilitar a utilização de um algoritmo semelhante ao *fitplus* e de viabilizar a sua aplicação nos equipamento mais antigos, torna-se necessário o retorno à seleção por parte do utilizador de um modo de escolha num único intervalo, definido por dois instantes t_0 e t_1 . Definidos estes dois valores, são necessárias as condições relativas ao polinómio descrito por f :

1. Deve passar pelo cromatograma nos limites de integração.
2. Deve ser tangente ao cromatograma nos limites de integração, sendo a inclinação do fundo cromatográfico em t_0 e t_1 calculada pela média das inclinações de uma vinhança de raio 0,7 s destes pontos.

condições que podem ser expressas matematicamente no sistema

$$\begin{cases} f(t_0) = s(t_0) \end{cases} \quad (2.9a)$$

$$\begin{cases} f(t_1) = s(t_1) \end{cases} \quad (2.9b)$$

$$\begin{cases} \frac{df}{dt}(t_0) = \frac{\Delta s}{\Delta t}(t_0) \end{cases} \quad (2.9c)$$

$$\begin{cases} \frac{df}{dt}(t_1) = \frac{\Delta s}{\Delta t}(t_1) \end{cases} \quad (2.9d)$$

que é facilmente resolúvel pela aplicação do algoritmo de LM. De forma a acelerar o processo de convergência deste algoritmo, a tentativa inicial dos valores (a, b, c, d) é o resultado da sua otimização nos cálculos anteriores para a mesma coluna, uma vez que esta mantém um perfil cromatográfico semelhante entre medidas. A utilização dos valores anteriores como solução-tentativa permite que o número de iterações seja reduzido em relação à tentativa inicial. Nesta, devido ao facto de não existir um resultado prévio, o polinómio é calculado a partir de uma tentativa inicial em que os parâmetros de f são tais que descrevem a reta que une (t_0, s_{t_0}) a (t_1, s_{t_1}) . Um exemplo de um cromatograma obtido no NDtech com os dois exemplos de integração (com fundo normal e curvilíneo) pode ser encontrado na fig. 2.7.

Exemplo da diferença entre o fundo cromatográfico utilizado após a escolha dos limites de integração por um utilizador

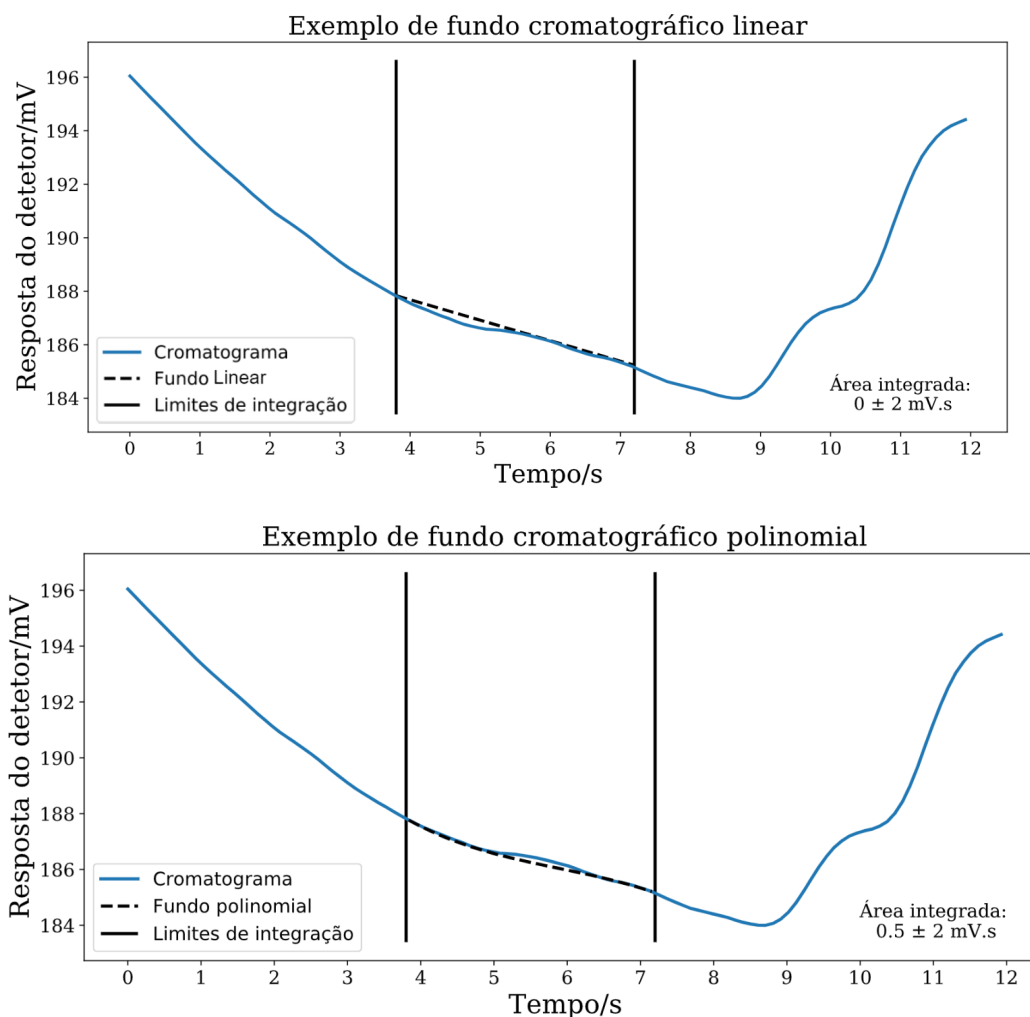


Figura 2.7: Comparação de fundo linear e polinomial

2.6 Métodos de procura do pico de TCA no cromatograma

2.6.1 Ajuste ao pico gaussiano

Neste método, é realizado iterativamente o processo descrito em 2.5. É definido um intervalo móvel $[t_i, t_i + \Delta t]$ (com Δt fixo) nos extremos do qual é calculado um polinómio $f(t)$ com as condições de 2.9.

A diferença positiva é considerada candidata a pico cromatográfico e, para esta diferença, recorre-se ao método dos mínimos quadrados para procurar a função gaussiana que melhor se ajusta à diferença positiva sinal - fundo.

A dificuldade da utilização de outras funções que mimetizam o sinal, nomeadamente as descritos em 2.1 reside na existência de mais parâmetros a otimizar, alguns dos quais correlacionados. Devido ao facto de possuir menor número de parâmetros a função gaussiana permite resolver o problema de minimização de μ de forma mais célere do que com outras funções.

O processo de cálculo do fundo polinomial e da função que minimiza a μ é repetido e, no final, a escolha do intervalo em que a área do pico é calculada recai sobre o intervalo em que μ é mínimo.

Afim de ilustrar o comportamento deste algoritmo, o seu funcionamento encontra-se exemplificado na figura 2.8 e descrito esquematicamente na figura 2.9.

Ilustração do funcionamento do algoritmo de procura de pico de TCA por comparação com picos gaussianos

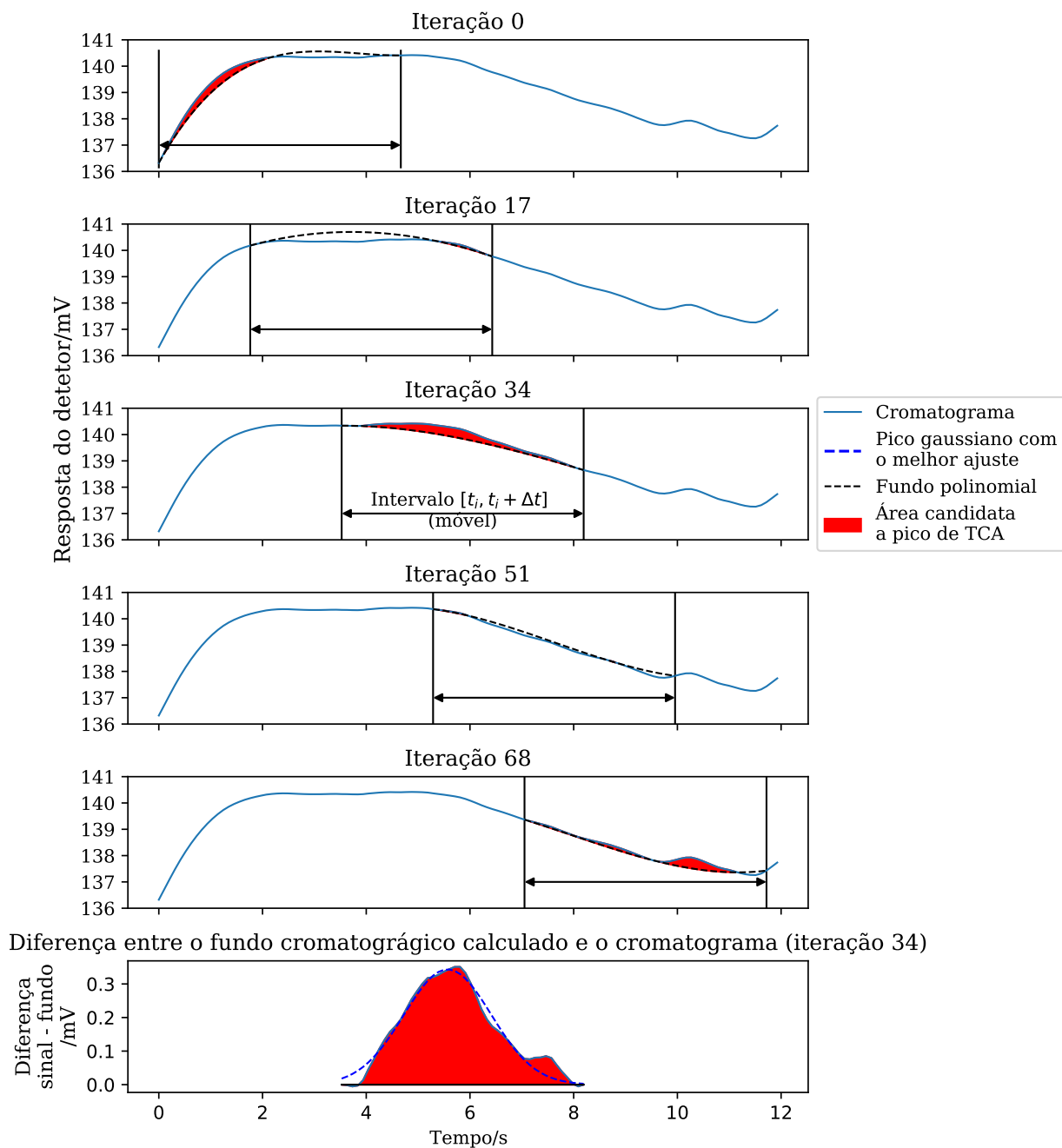


Figura 2.8: Algoritmo de pesquisa por comparações sucessivas com picos gaussianos

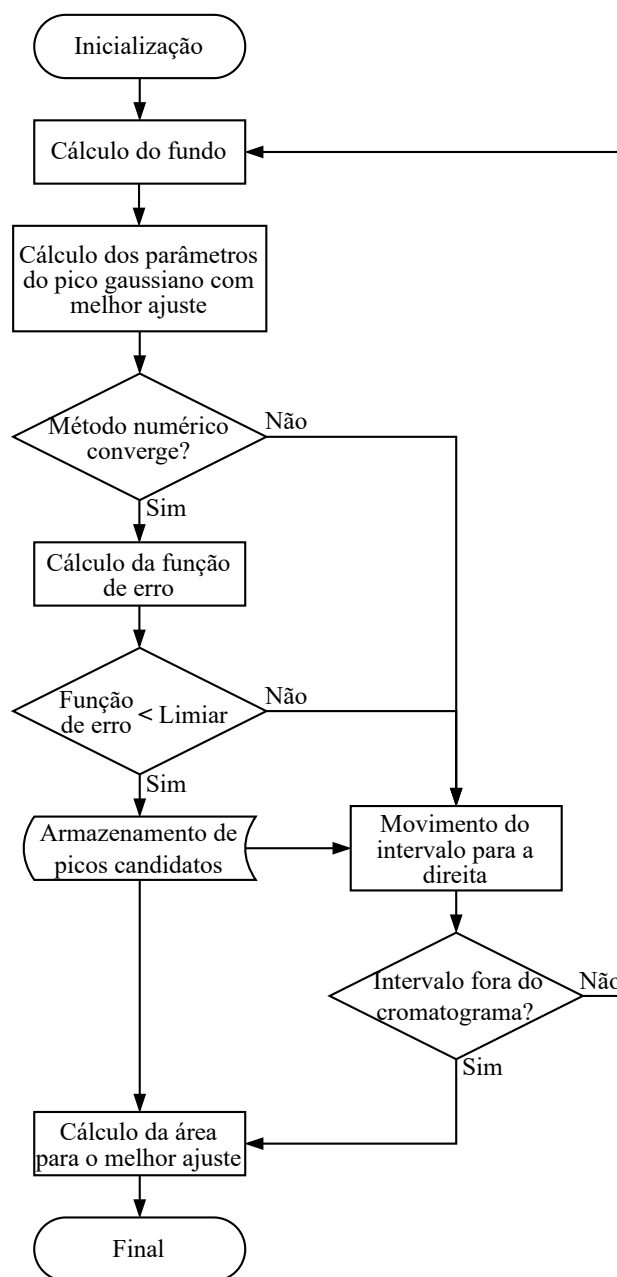


Figura 2.9: Algoritmo de pesquisa por ajustes sucessivos de fundo e pico

2.6.2 Ajuste de modelo de pico cromatográfico

A realização sucessiva do método descrito em 2.6.1 tem como problema principal o número de instruções realizadas para a detecção de pico, em particular no que toca à comparação de uma dada diferença (positiva) sinal-fundo com o pico gaussiano. A resolução do sistema de equações não-lineares várias vezes num processo iterativo leva a que a análise seja computacionalmente exigente e que o tempo necessário para os cálculos seja excessivo.

Como forma de minimizar o tempo de execução deste método, optou-se pela comparação com um padrão obtido a partir de um pico de TCA isolado de outras estruturas.

Ilustração do funcionamento do algoritmo de procura de pico de TCA por comparação com picos gaussianos

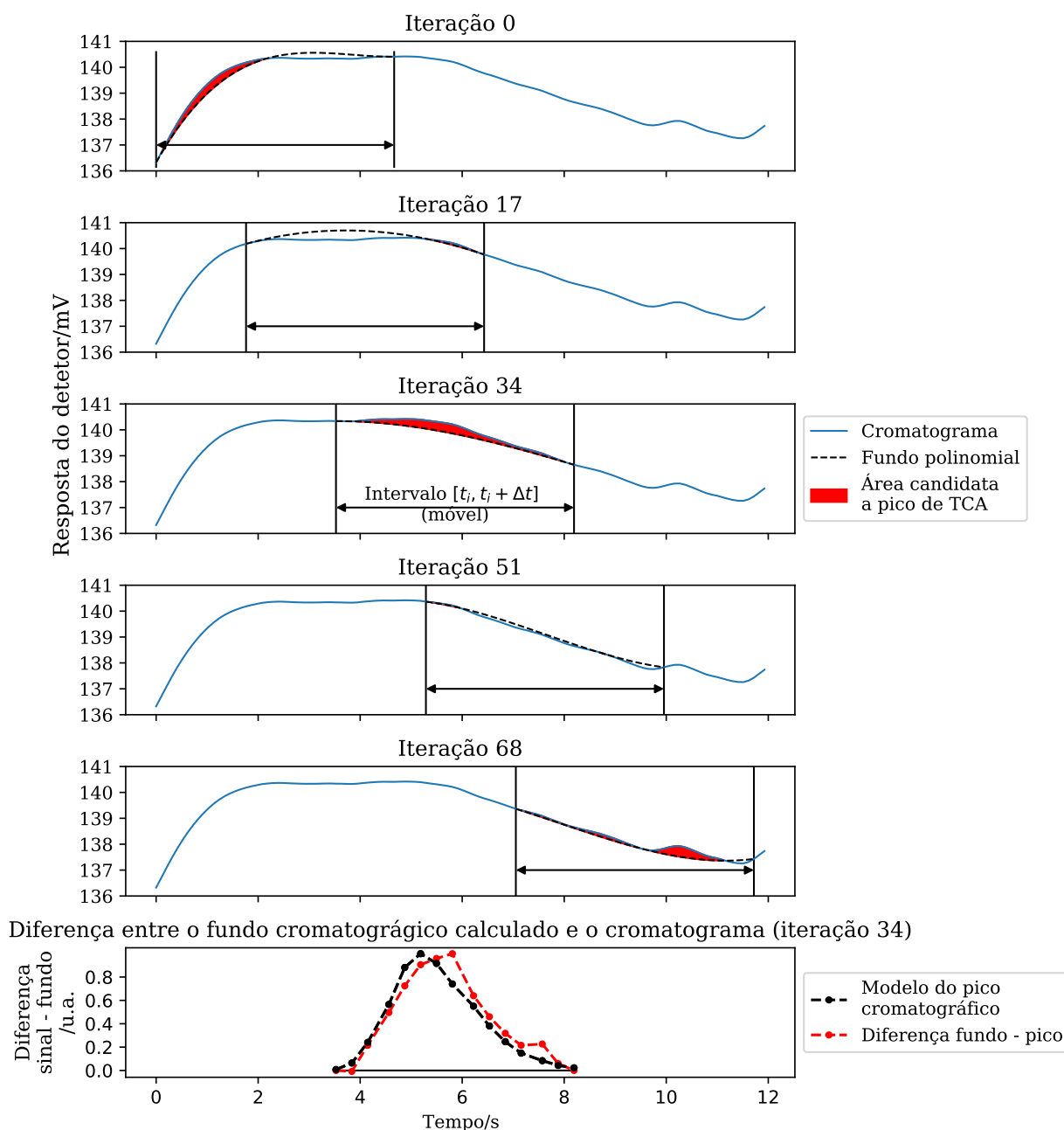


Figura 2.10: Exemplo de procura e integração de um pico cromatográfico pelo algoritmo de procura de padrão

De igual forma ao método anterior, é calculado o fundo adequado para as possíveis posições do pico cromatográfico do composto de interesse. Para as várias diferenças sinal-fundo, são escolhidos 15 pontos e a altura é reduzida a uma unidade. Assim, resta comparar estes 15 pontos com o mesmo número pontos de um padrão, sendo usada como classificação a soma das diferenças quadradas entre o pico candidato medido e o padrão - μ' , nos 15 pontos considerados.

O intervalo que minimiza μ é considerado a localização mais provável do pico, sendo nesse calculada a área integrada neste intervalo. O processo encontra-se exemplificado na fig. 2.10 e esquematizado na fig. 2.11.

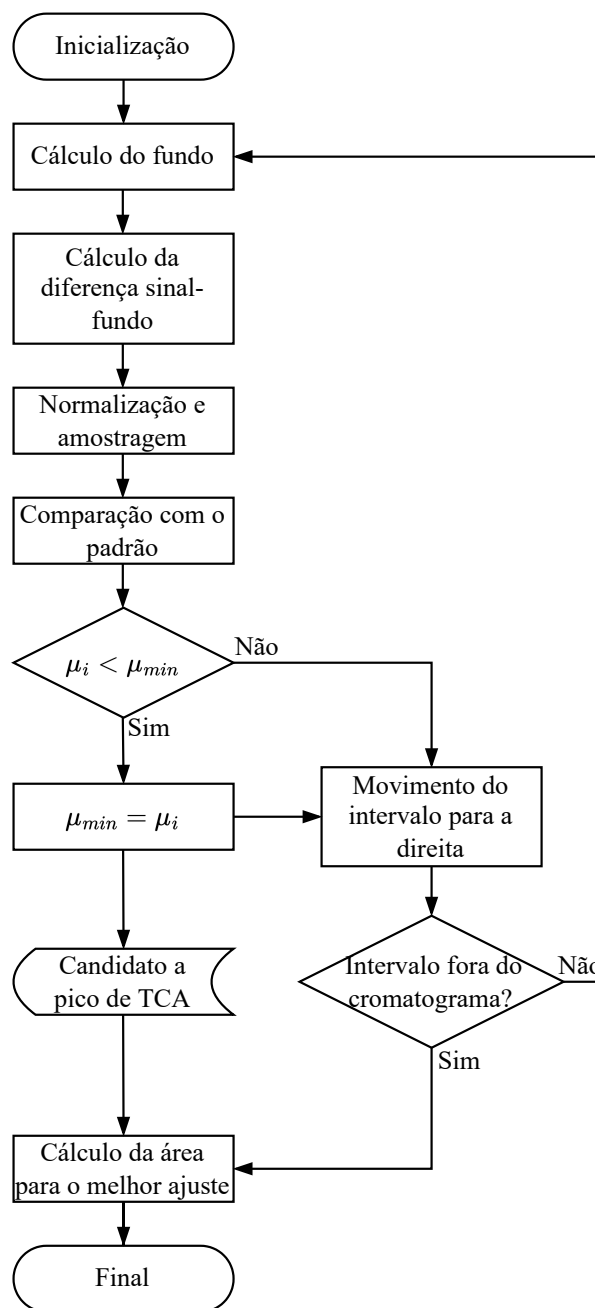


Figura 2.11: Algoritmo de pesquisa por comparações sucessivas com o padrão

As diversas tipologias de funcionamento descritas neste (2.º) capítulo, serão, em diante, denominadas como algoritmo na forma manual - como descrito em 2.5 - e na(s) forma(s) automática(s) - descritas nas subsecções 2.6.1 e 2.6.2. Posteriormente, uma destas formas obterá preferência e, a partir desse momento, as tipologias serão denominadas segundo:

Algoritmo na forma manual - NDfit Manual ou NDfitM

Algoritmo na forma automática - NDfit Auto ou NDfitA

Testes e validação

3.1 Validação fora de linha de produção

A validação fora da linha de produção consiste na análise de cromatogramas que o programa de controlo de produção permite extrair. A análise destes ficheiros tem como propósito possibilitar os passos iniciais do desenvolvimento do algoritmo, nomeadamente ao nível dos métodos de cálculo utilizados e da celeridade da execução destes. Para além disso, a ausência de intromissão no processo de fabrico das rolhas NDtech® permite assegurar a ausência de interferência quer com o processo de produção, quer com as garantias que a marca proporciona.

Assim, este tipo de validação apresenta-se como o método ideal para teste dos métodos mais fundamentais de qualquer algoritmo desenvolvido no âmbito deste trabalho.

3.1.1 Algoritmo na forma manual - métodos e objetivos

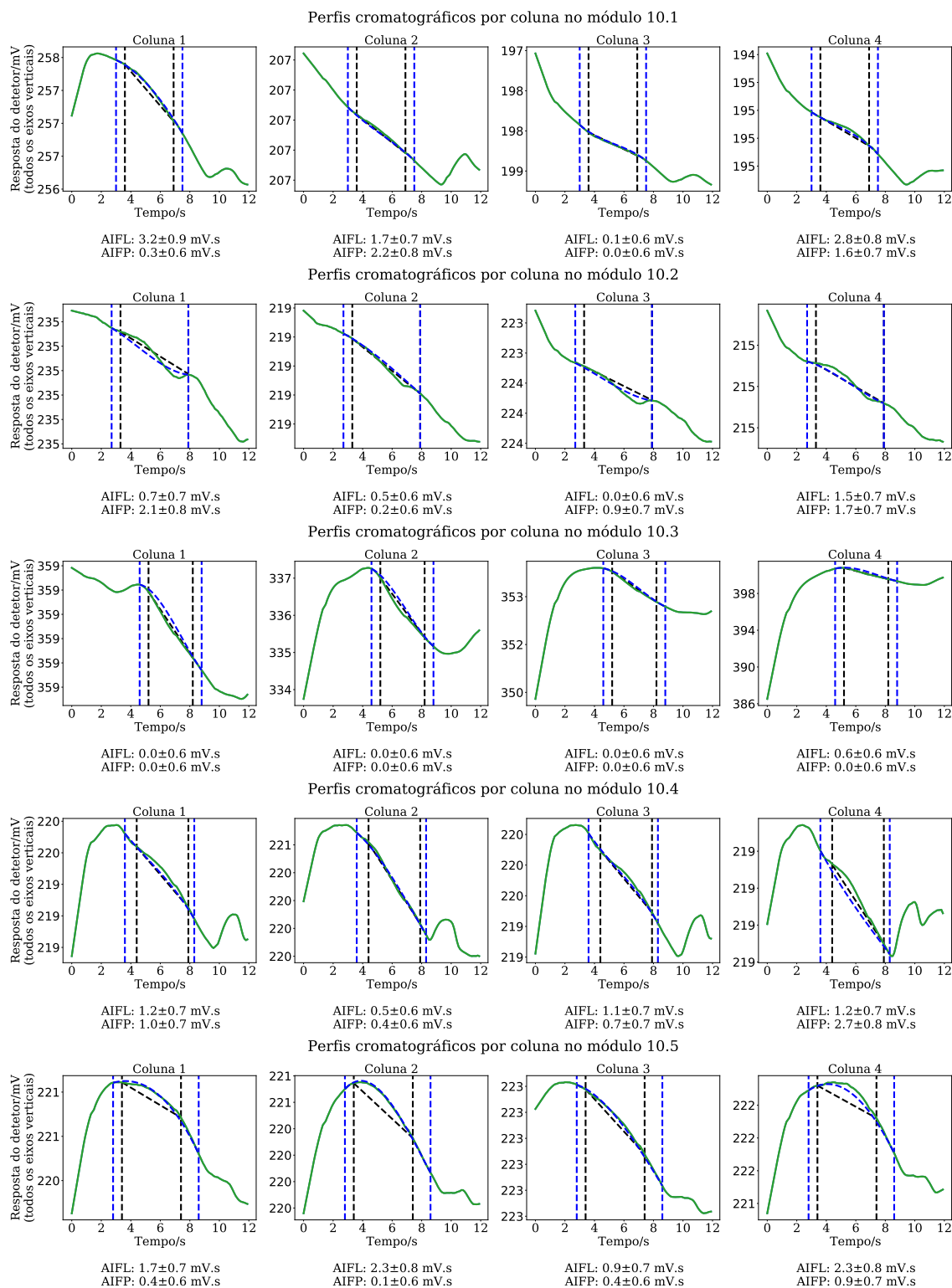
A replicação total do algoritmo atualmente utilizado não é possível uma vez que a sua implementação não é revelada em detalhe pela empresa que produz o *software*. Nesta secção (3.1), será meramente efetuada a comparação entre uma integração com fundo linear e com fundo polinomial (escolhido automaticamente ou não), para um mesmo conjunto de dados. Pretende-se nesta componente do trabalho uma validação meramente conceptual da viabilidade da utilização de um fundo curvo na integração dos picos.

De forma a possibilitar a extração e comparação destes resultados, foi criado um *script* que executa as funções de cálculo do fundo cromatográfico e da área integrada, sendo exportadas imagens onde serão visíveis os limites de integração e a região limitada pelo cromatograma e pelo fundo calculado.

3.1.2 Algoritmo na forma manual - resultados e discussão

Serão expostos nas figuras seguintes, figs. 3.1 e 3.2, para um conjunto de cromatogramas obtidos durante a produção de rolhas no dia 18 de fevereiro de 2022 os cromatogramas com menor (fig. 3.1) e maior (fig. 3.2) área integrada do pico cromatográfico. Nestas, é possível observar o fundo cromatográfico linear e os respetivos limites de integração (a

tracejado, a preto) e o fundo de integração polinomial e os respetivos limites de integração (a tracejado, a azul). É também possível observar os valores da Área Integrada com Fundo Linear (AIFL) e Área Integrada com Fundo Polinomial (AIFP) correspondentes.



3.1. VALIDAÇÃO FORA DE LINHA DE PRODUÇÃO

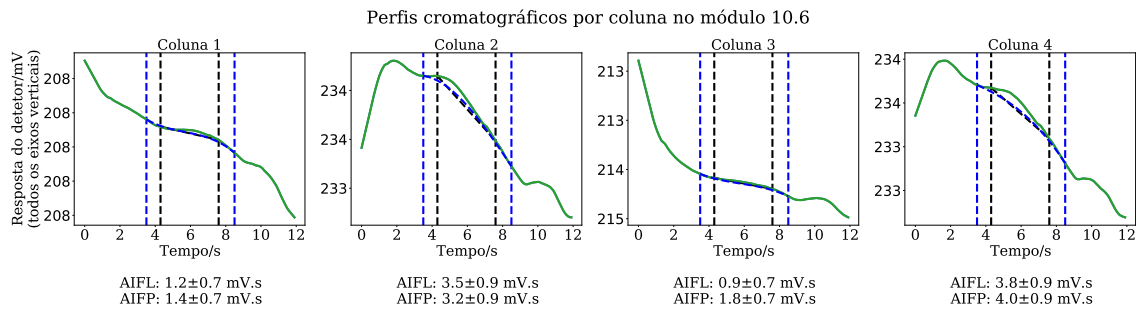
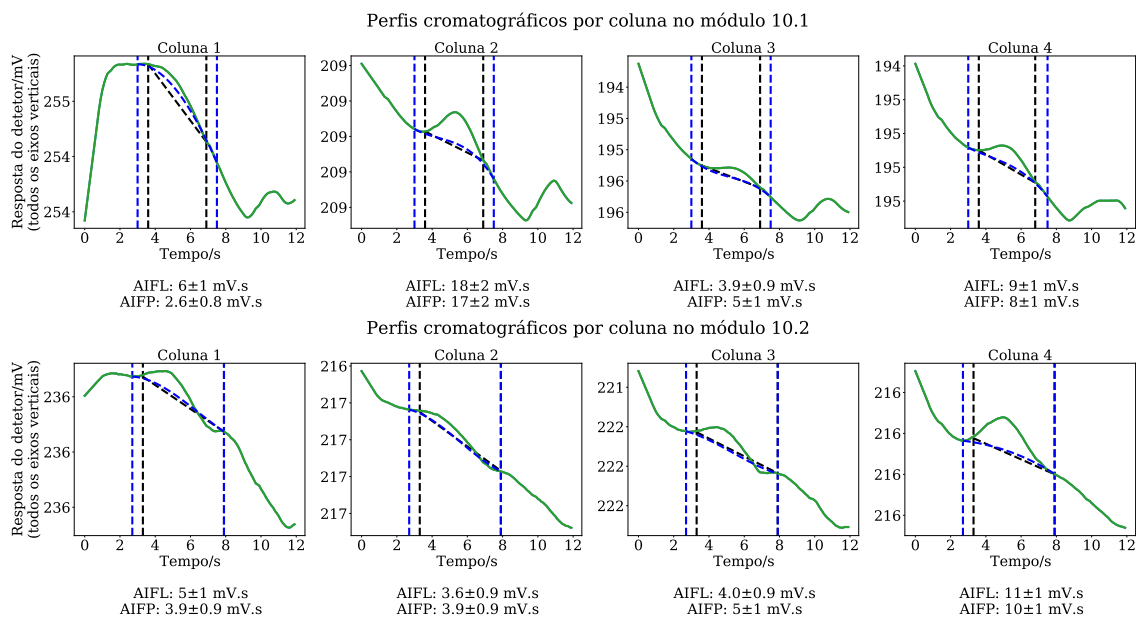


Figura 3.1: A vermelho : Traçado do cromatograma obtido
 A preto, a tracejado: Limites de integração e fundo linear
 A azul, a tracejado : Limites de integração e fundo polinomial
 Cromatogramas analisados para a área mínima por coluna no conjunto de dados

A primeira inferência da visualização destas imagens é a confirmação da elevada variabilidade dos fundos cromatográficos obtidos na produção de rolhas NDtech®, consequência de estas serem produtos naturais.

Para além disso, as figuras ilustram que, para a área mínima obtida em cada coluna de cada equipamento (acima, fig 3.1), o fundo polinomial é, regra geral, mais próximo do traçado do cromatograma do que o fundo linear.

Tal nem sempre se traduz numa área integrada menor, devido não só à escolha de limites de integração distintos como à contribuição apenas da componente positiva da diferença sinal - fundo para a AIFL e AIFP.



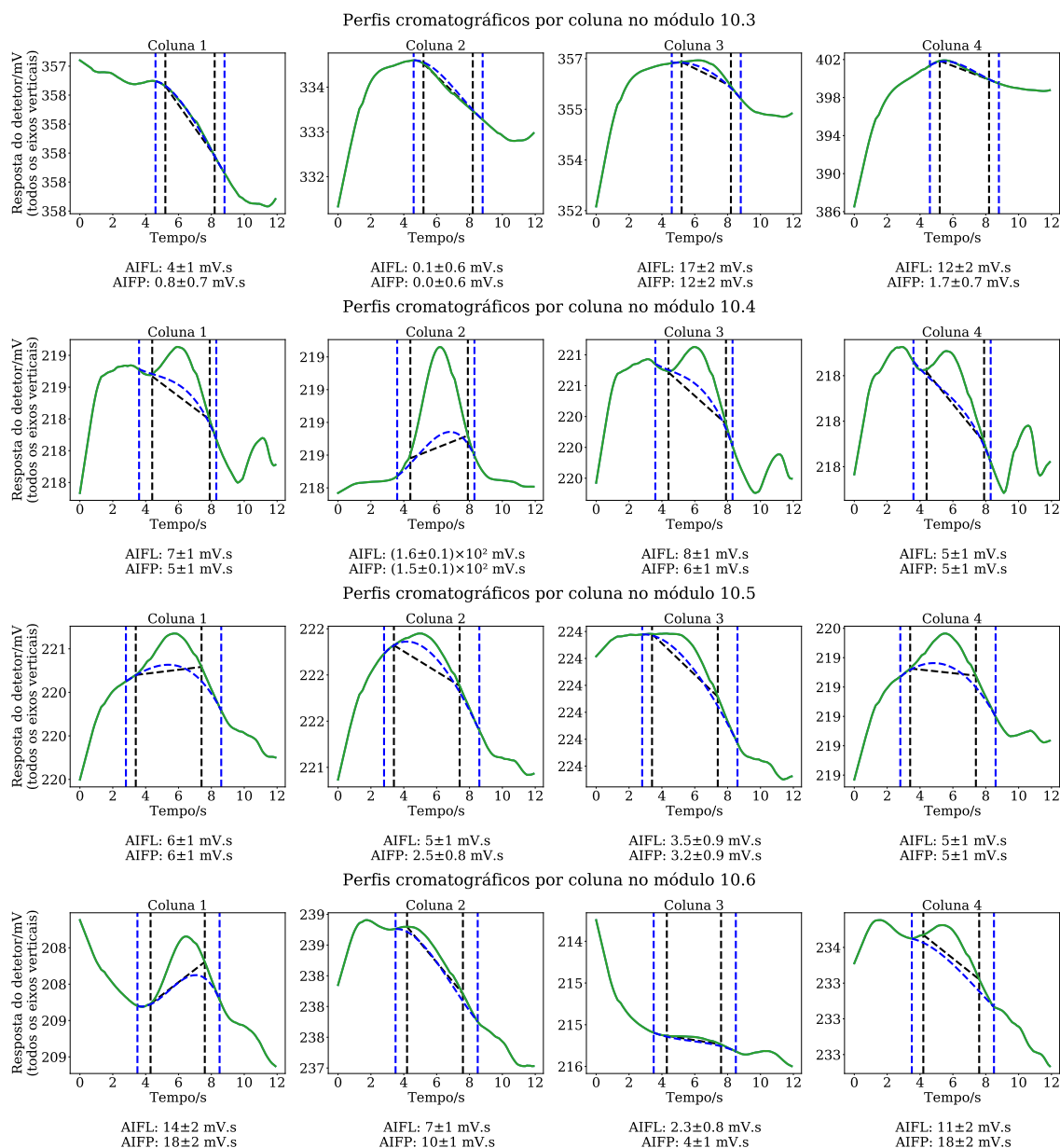


Figura 3.2: Cromatogramas analisados para a área máxima por coluna no conjunto de dados

Já para a área maior (anteriormente, fig 3.2), ocorrem valores inferiores da área calculada com o fundo polinomial em relação ao fundo reto. Apesar disso, em casos de orientação negativa da concavidade (forma semelhante a $\cap$), torna-se mais difícil inferir se esta diferença ocorre por melhor ajuste ao fundo ou por subestima do pico de TCA. Apesar disso, frequentemente, a área do pico, mesmo com a subestima, continua a ser suficiente para que a rolha seja rejeitada.

Segundo os operadores dos equipamentos, a utilização da reta como função de base obrigava à escolha de limites de integração que não permitiam considerar todo o pico de TCA que identificavam visualmente, a fim de manter a aceitação elevada o suficiente para justificar a produção das rolhas.

Os operadores relatam também a dificuldade de conciliação dos múltiplos fundos cromatográficos existentes em simultâneo no mesmo equipamento, devido a diferenças impercetíveis de montagem, desgaste desigual de peças e diferentes condições das colunas cromatográficas.

A desigualdade de circunstâncias e a fraca representatividade deste tipo de teste, em conjunto com o facto de as rolhas utilizadas neste fazerem parte da produção, tornam impossível cruzar os resultados destes testes com os de uma análise da contaminação que as rolhas analisadas provocariam. Como tal, não é possível aferir a quantidade real de TCA nas rolhas analisadas, ficando clara a necessidade de outros tipos de análise do desempenho deste algoritmo.

3.1.3 Algoritmos na forma automática - métodos e objetivos

Nesta fase de testes fora de linha dos algoritmos na forma automática, pretende-se verificar se existe capacidade da parte do algoritmo de detetar autonomamente os picos de TCA nos cromatogramas. Também se pretende analisar qual o impacto nas medidas nos casos em que a quantidade deste composto é baixa. Serão comparados, para os mesmos dados de outros testes desta secção, a disparidade entre os picos do analito de interesse identificados pelo operador do equipamento e pelo algoritmo.

Não serão consideradas as áreas nesta análise uma vez que devido ao fator de sensibilidade aplicado na medida, dois picos com uma medida semelhante de área podem apresentar assinaturas visuais distintas, cabendo ao algoritmo ser capaz de detetar o pico em ambos os casos.

3.1.4 Algoritmos na forma automática - resultados e discussão

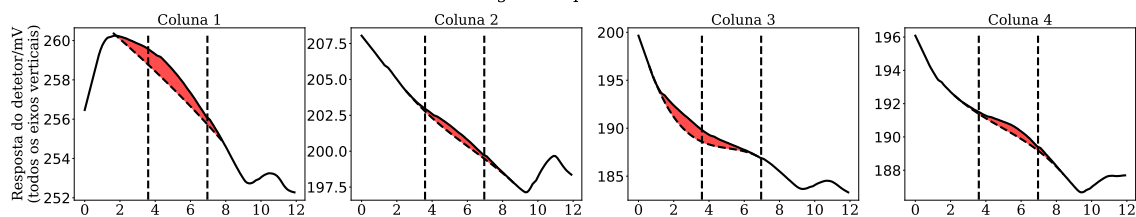
Os resultados dos testes preliminares de desenvolvimento de algoritmos que efetuassem a procura automática, dividem-se, naturalmente, nas duas formas de alcançar este objetivo, as descritas em 2.6.1 e 2.6.2.

Quanto à primeira, cujos cromatogramas podem ser encontrados nas figuras 3.3 e 3.4, onde é possível verificar a falta de convergência entre os resultados da identificação automática e convencional, pela identificação de outras estruturas do sinal (variações de pressão e picos cromatográficos de outros compostos) como sendo o pico de TCA. Tal resulta no cálculo da área destas estruturas pelo algoritmo. Este fenómeno ocorre tão mais frequentemente quanto menor a área do pico. É de salientar também casos de sobrecompensação do fundo, como é o caso da coluna 3 do módulo ¹ 10.2, em que o fundo polinomial aparenta ser desnecessariamente distorcido para que o ajuste do conjunto de funções seja o melhor possível.

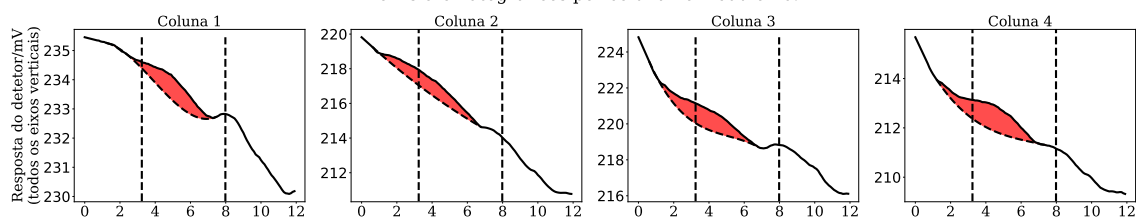
São também identificados casos em que o algoritmo converge num dos mínimos locais da superfície μ , aparentando não resolver condições de tangência - eqs 2.9 - exemplificado (não exclusivamente) pela coluna 4 do equipamento 10.3.

¹ Denominação dos cromatógrafos utilizados em produção

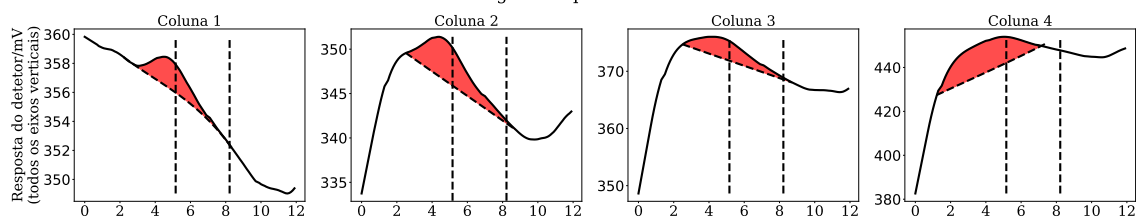
Perfis cromatográficos por coluna no módulo 10.1



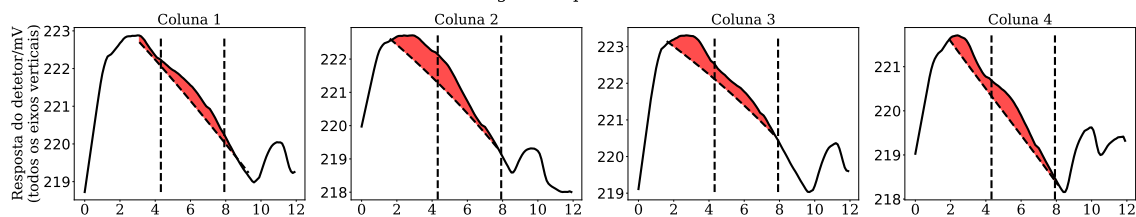
Perfis cromatográficos por coluna no módulo 10.2



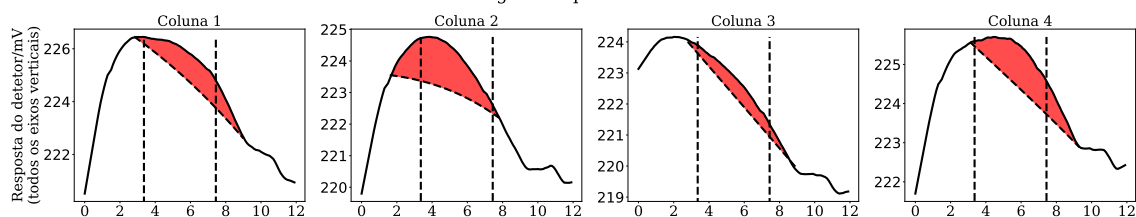
Perfis cromatográficos por coluna no módulo 10.3



Perfis cromatográficos por coluna no módulo 10.4



Perfis cromatográficos por coluna no módulo 10.5



3.1. VALIDAÇÃO FORA DE LINHA DE PRODUÇÃO

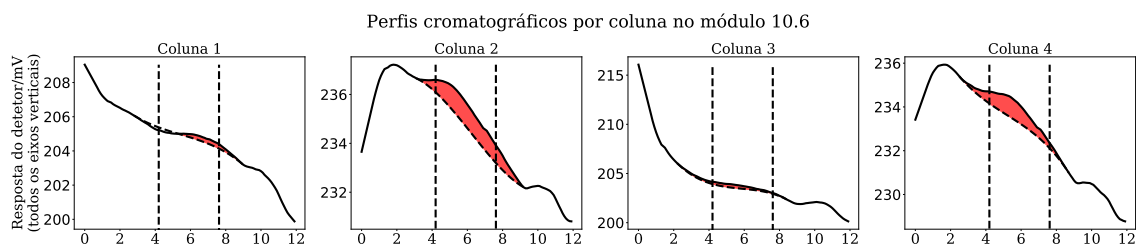
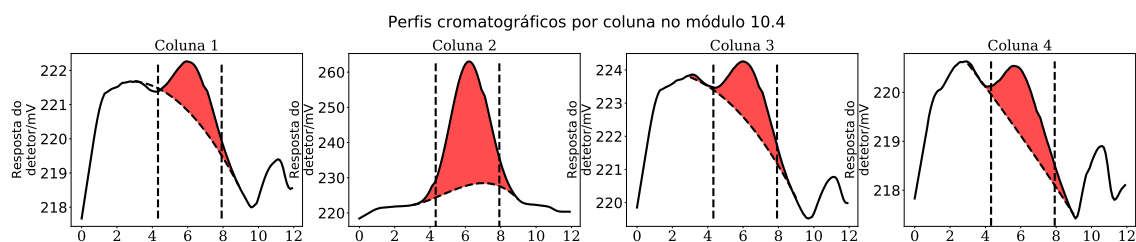
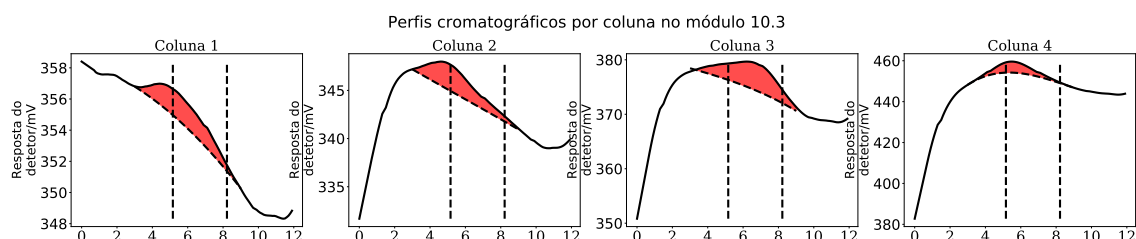
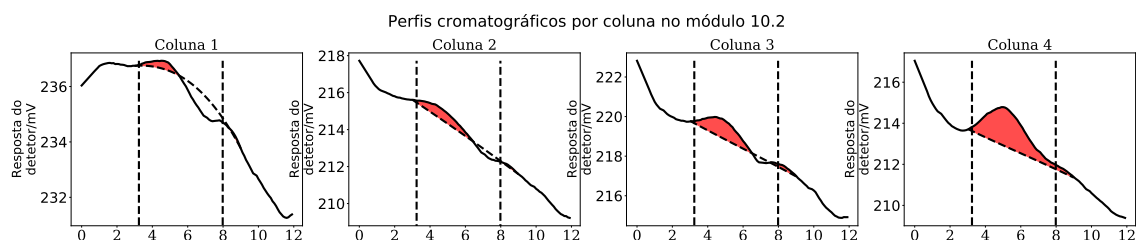
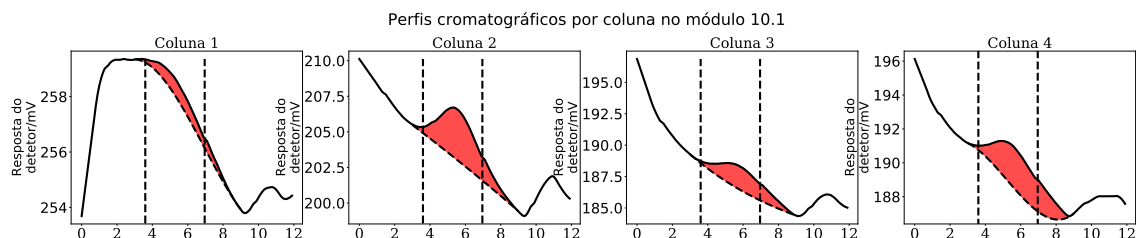


Figura 3.3: Cromatogramas analisados para a área mínima no conjunto de dados (por coluna do equipamento)

A vermelho : área integrada pelo algoritmo ; a preto, a tracejado: limites de integração e fundo cromatográfico

Mesmo com picos cromatográficos de maior dimensão, é possível detetar os mesmos problemas, apesar de aparentar ocorrer uma melhoria da identificação da localização temporal do pico de TCA.



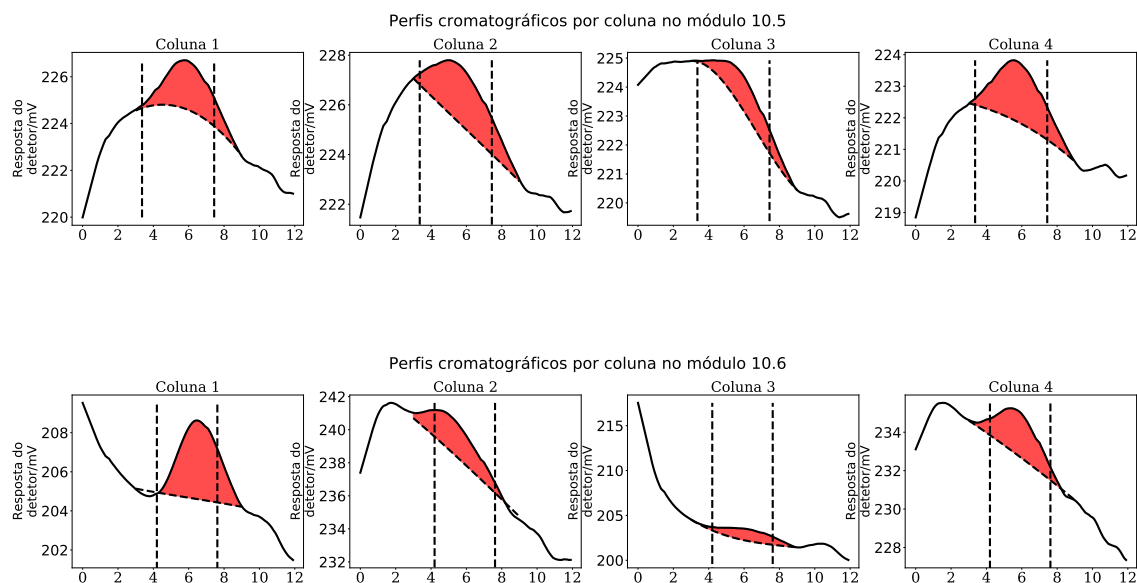


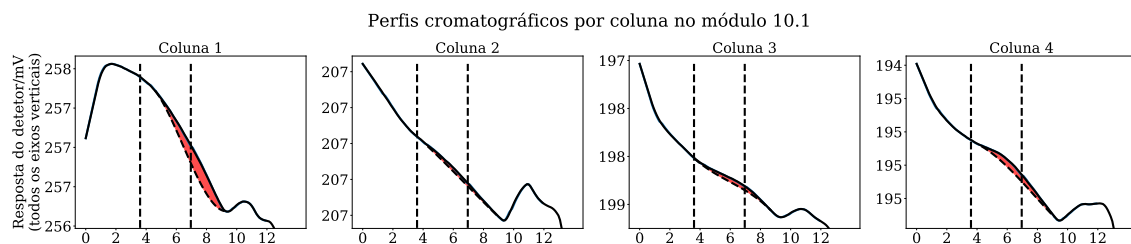
Figura 3.4: Cromatogramas analisados para a área máxima no conjunto de dados (por coluna do equipamento)

A vermelho : área integrada pelo algoritmo ; a preto, a tracejado: limites de integração e fundo cromatográfico

Não obstante de todas as tentativas de melhoria dos parâmetros a utilizar nos problemas de otimização não-linear intrínsecos ao algoritmo, os tempos de execução mantiveram-se elevados, chegando a ultrapassar 1 segundo. Dado estes tempos terem sido obtidos em equipamentos mais recentes do que os do NDtech, percebeu-se que a aplicação deste algoritmo na sua forma atual e sem otimização do ponto de vista computacional seria impraticável.

Esta necessidade motivou a tentativa de aceleração do funcionamento do algoritmo, onde os cálculos envolvessem um menor número de parâmetros a otimizar e um conjunto menor de pontos, resultando no algoritmo descrito em 2.6.2.

Após essa tentativa, verificou-se empiricamente uma diminuição considerável do tempo de execução de cada análise. Apesar disso, esta técnica de identificação continua a produzir frequentemente resultados díspares das decisões do utilizador na decisão dos limites de integração a utilizar e da localização do pico cromatográfico relevante, pelo que se torna impossível colocar um algoritmo automático nesta fase de desenvolvimento em funcionamento com grau aceitável de confiança nos seus resultados.



3.1. VALIDAÇÃO FORA DE LINHA DE PRODUÇÃO

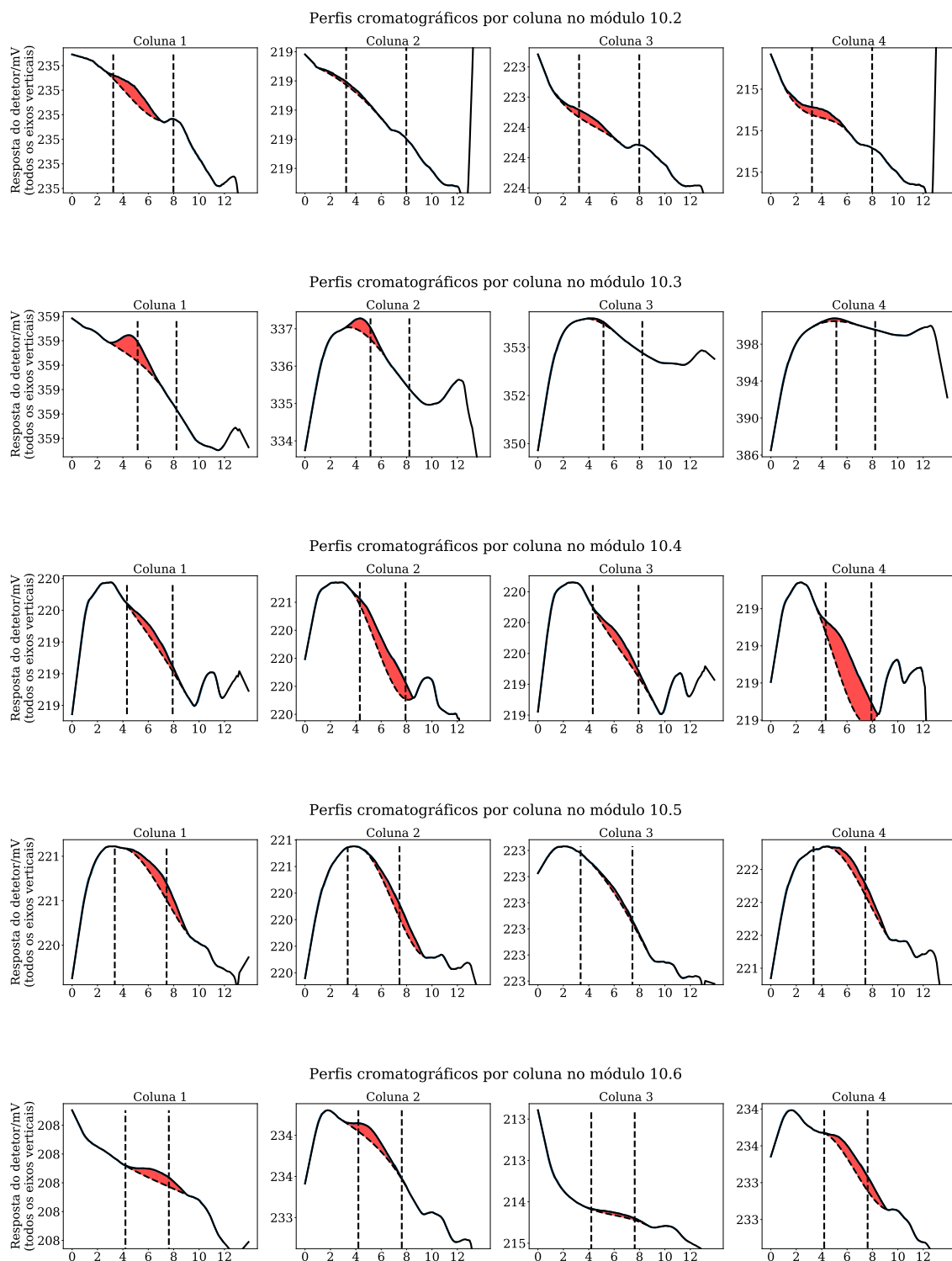


Figura 3.5: Cromatogramas analisados para a área mínima no conjunto de dados (por coluna do equipamento)

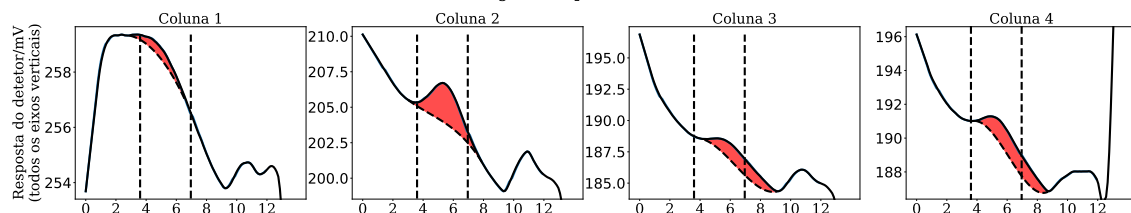
A vermelho : área integrada pelo algoritmo ;

A preto, a tracejado: limites de integração e fundo cromatográfico

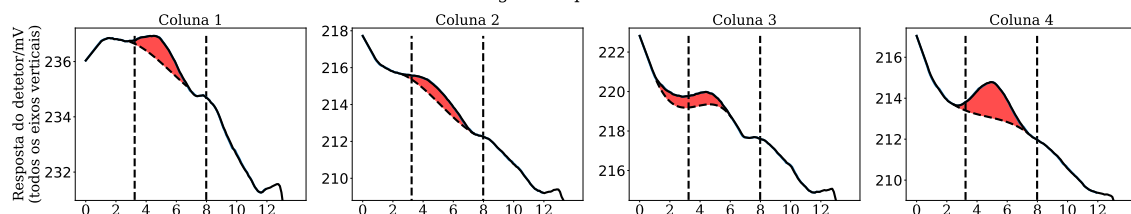
Apesar disso, a utilização de um pico-padrão obtido com a mesma substância e no mesmo sistema permite ter em conta alguma da assimetria do pico, sendo esta mais uma

melhoria em relação à utilização do pico gaussiano, de forma visível nas figuras 3.6 e, em particular, na figura 3.5.

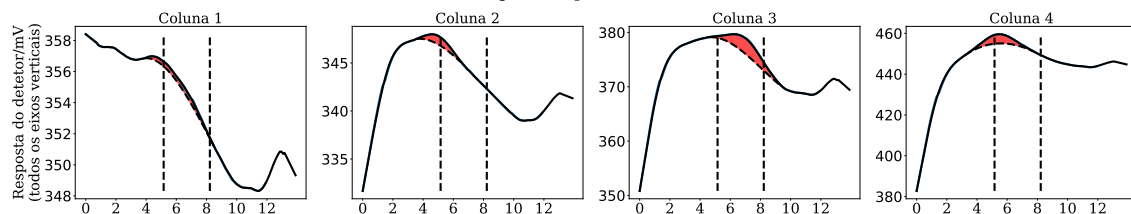
Perfis cromatográficos por coluna no módulo 10.1



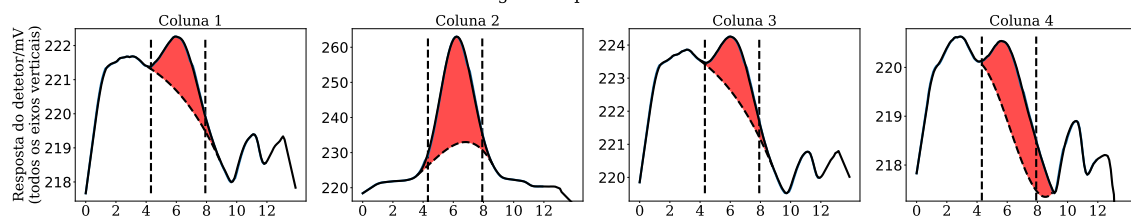
Perfis cromatográficos por coluna no módulo 10.2



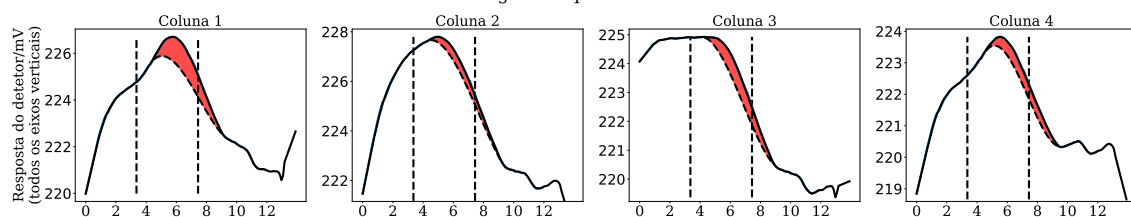
Perfis cromatográficos por coluna no módulo 10.3

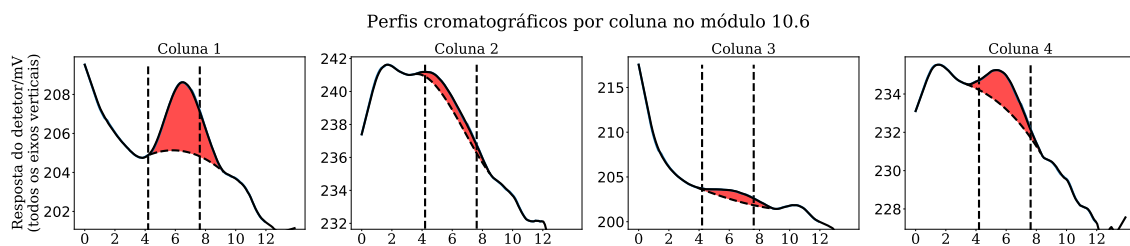


Perfis cromatográficos por coluna no módulo 10.4



Perfis cromatográficos por coluna no módulo 10.5





A vermelho : área integrada pelo algoritmo ; a preto, a tracejado: limites de integração e fundo cromatográfico

Figura 3.6: Cromatogramas analisados para a área máxima no conjunto de dados (por coluna do equipamento)

3.1.5 Conclusão

Apesar de a fase de desenvolvimento descrita nesta secção permitir os passos primordiais do desenvolvimento, o seu âmbito é limitado e os resultados obtidos carecem de qualquer tipo de confirmação no que à quantidade real de TCA diz respeito.

O facto de as decisões não serem tomadas em tempo real criam uma vantagem falaciosa que privilegia o algoritmo deste trabalho, uma vez que as decisões atuam sobre um conjunto de rolhas que já foram produzidas e cujos cromatogramas já foram obtidos na totalidade do seu conjunto.

Apesar disso, a flexibilidade decorrente do uso de um polinómio de grau superior permite, na opinião do autor (com grau de subjetividade não nulo), uma melhor adequação do fundo polinomial escolhido e uma supressão superior das estruturas ocorrentes no cromatograma, tais como variações de pressão e picos cromatográficos de outros compostos

Tais conclusões e a atenuação da subjetividade envolvida até este ponto justificam a persecução de formas mais elucidativas de validação e comparação dos resultados obtidos por este algoritmo com as quantidades reais do contaminante de interesse.

3.2 Validação em linha de produção (sem substituição de resultados) - algoritmo na forma manual

A validação em linha de produção foi efetuada com uma aplicação desenvolvida para os propósitos deste trabalho, denominada *NDfit tester*. Esta foi criada a partir da biblioteca *Tkinter*[32] que efetua a interface entre programas *Python* e *Tk*.

O seu funcionamento detalhado, bem como ilustrações e instruções de utilização encontra-se no apêndice A.

3.2.1 Objetivos da validação em linha de produção

O propósito dos testes do algoritmo na linha de produção sem a substituição de resultados é a possibilidade de replicar em tempo real todas as alterações efetuadas pelos operadores e de fatores como o TR. Assim, garante-se a maior equidade de circunstâncias nas comparações dos métodos de cálculo de ambos os algoritmos. Passa também a ser permitido alterar os fatores de sensibilidade do equipamento de forma síncrona com as alterações nominais destes parâmetros, algo que não era contemplado em 3.1.

Fundamentalmente, torna-se possível comparar as decisões de cada um dos algoritmos para uma dada rolha e analisá-la com recurso a sistemas GC/ECD de laboratório, sendo esta a análise de referência para avaliar a performance de qualquer algoritmo utilizado no NDtech. O fluxo dos ficheiros e da informação neles contida é exemplificado no esquema 3.7, porém uma versão com maior detalhe pode ser consultada na figura A.1 (apêndice A).

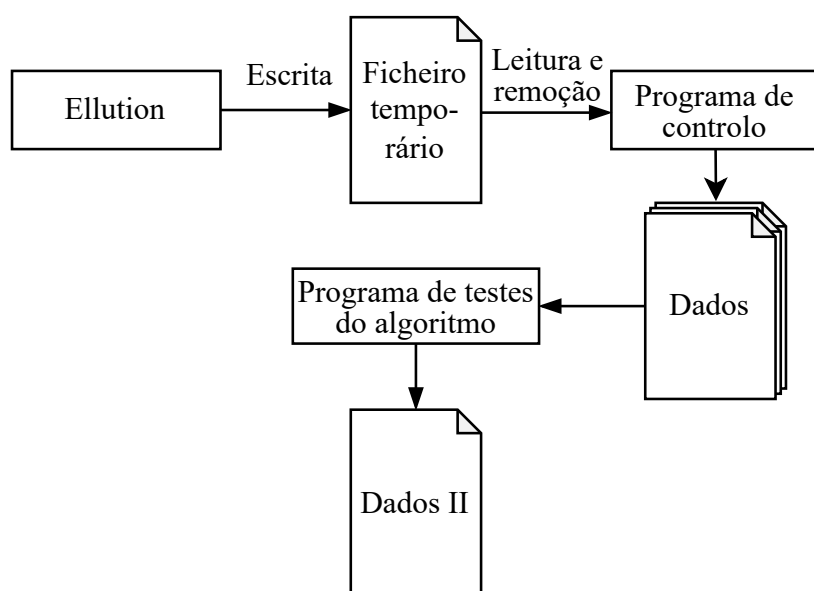


Figura 3.7: Funcionamento da aplicação de teste de algoritmos sem substituição de resultados

A utilização dos ficheiros produzidos aquando da análise das rolhas permite estudar

3.2. VALIDAÇÃO EM LINHA DE PRODUÇÃO (SEM SUBSTITUIÇÃO DE RESULTADOS) - ALGORITMO NA FORMA MANUAL

o funcionamento do algoritmo sobre os cromatogramas em tempo real e realizar análises estatísticas com efeitos mínimos na produção pelos equipamentos. Como consequência, é possível efetuar uma análise paralela à produção - subsecções 3.2.2 a 3.2.4.

Para além desta análise paralela, esta configuração da aplicação foi utilizada na análise de rolhas rastreadas (numeradas)- subsecções 3.2.5 a 3.2.7. Desta forma, para cada rolha analisada, foi possível cruzar a decisão do programa Ellution, com a decisão do algoritmo desenvolvido neste trabalho e com os resultados do laboratório de I&D, tendo sempre presente e recuperável o cromatograma correspondente para análise posterior.

3.2.2 Teste de média escala - Métodos

Para a realização de um teste preliminar, foram analisados os cromatogramas obtidos pelos cromatógrafos do NDtech durante a produção de 9946 rolhas de um lote proposi-tadamente escolhido como não sendo representativo pela elevada quantidade de TCA. A razão dessa escolha prende-se com a maior probabilidade de encontrar rolhas com quan-tidades de tricloroanisol passíveis de gerar desacordos entre os softwares, facilitando a procura de disparidades entre o resultado de ambos.

3.2.3 Teste de média escala - Resultados

Os resultados desta experiência encontram-se expostos nesta secção em grau crescente de granularidade:

O resultado global (tab. 3.1) é o de um ganho significativo da aceitação, na qual se verifica um aumento de $14,6 \pm 0,7$ %.

Tabela 3.1: Comparação entre as classificações dos diferentes algoritmos

	Estimativa da proporção do grupo no total de rolhas (%)	
	Algoritmo atual	Algoritmo deste trabalho - forma manual (NDfitM)
G1	$73,9 \pm 0,5$	$88,5 \pm 0,4$
G2	$24,7 \pm 0,5$	$10,3 \pm 0,4$
G3	$1,3 \pm 0,2$	$1,2 \pm 0,2$

Este resultado aparenta ser sinónimo do cumprimento dos objetivos deste trabalho. Porém, convém lembrar que este aumento de aceitação não pode ocorrer a troco de rolhas contaminadas com TCA no G1. A origem das rolhas que transitam para o G1 é importante no sentido que permite uma auscultação da viabilidade das melhorias, uma vez que transições de rolhas do G3 para o G1 seriam questionáveis pois indicariam uma grande disparidade da medida obtida pelos dois algoritmos. A exposição dos dados numa matriz de confusão (tabela 3.2) permite verificar que a totalidade das rolhas que transita para o G1 na substituição do algoritmo atual pelo deste trabalho fazia parte do grupo G2. Neste último havia sido constatado que uma percentagem significativa das rolhas eram

de facto TCA-negativas[24]. Existem também transições em sentido inverso, que atenuam os ganhos de aceitação na utilização do NDfit, neste caso na sua versão manual.

Tabela 3.2: Classificações das rolhas por ambos os algoritmos, em percentagem do total

		NDfitM		
		G1/%	G2/%	G3/%
Algoritmo atual	G1/%	72,0	1,9	0,0
	G2/%	16,5	8,1	0,1
	G3/%	0,0	0,2	0,1

Uma atenção mais pormenorizada sobre as rejeições subdivididas pelas colunas dos equipamentos (fig. 3.8) indicam, que o algoritmo atual apresenta, nas colunas 2 e 4 do módulo 6.2 e na coluna 4 do módulo 6.3, aceitações várias vezes superiores às dos outros equipamentos.

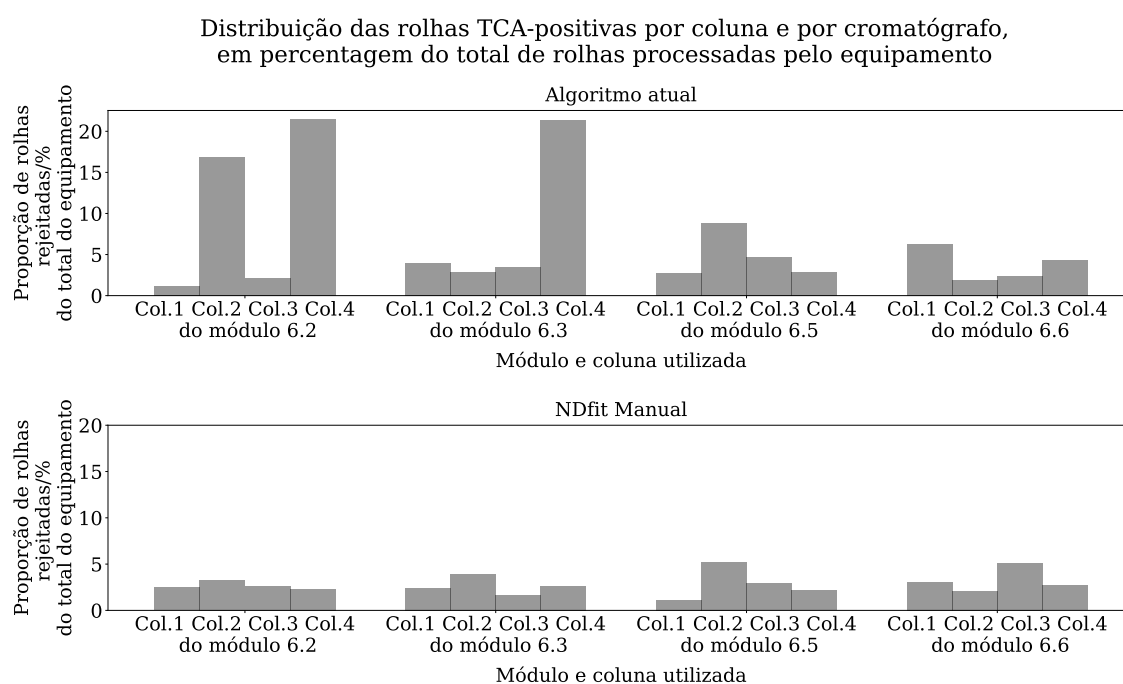


Figura 3.8: Rejeição discriminada pelos equipamentos e pelas suas colunas cromatográficas

A média e desvio padrão da rejeição das colunas segundo ambos os algoritmos tomam valores de $7 \pm 7\%$ para algoritmo atual e de $2,8 \pm 1,1\%$ para algoritmo NDfit Manual.

Já a divisão do lote em grupos pode ser consultada na figura seguinte 3.9):

3.2. VALIDAÇÃO EM LINHA DE PRODUÇÃO (SEM SUBSTITUIÇÃO DE RESULTADOS) - ALGORITMO NA FORMA MANUAL

Proporção de rolhas de cada grupo, estimada por cada equipamento
(com intervalo de confiança de 95%)

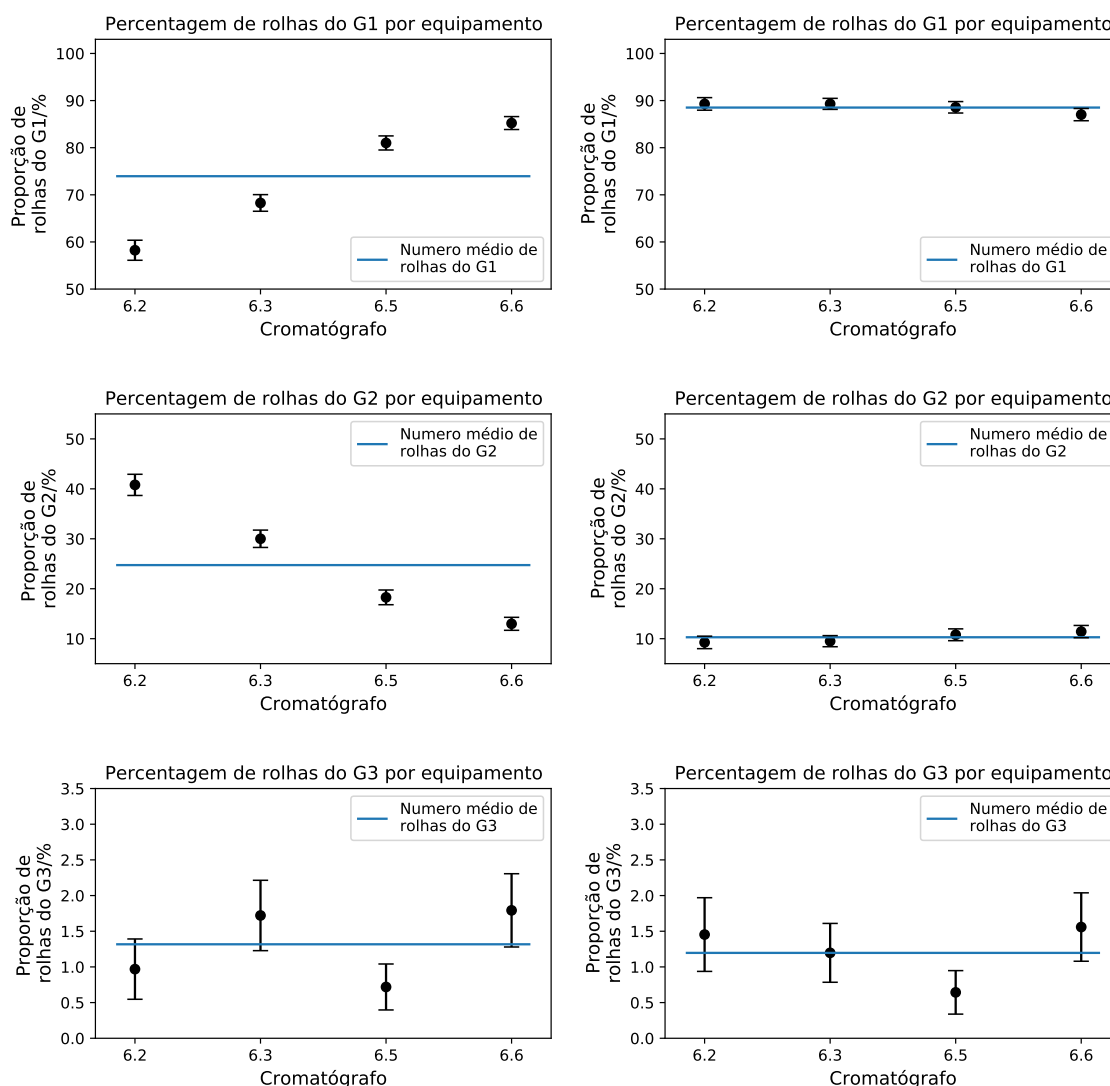


Figura 3.9: Estimativas da proporção global de rolhas para cada um dos grupos de rolhas no NDtech

onde se verifica que um intervalo de confiança de 95 % não é suficiente para fazer concordar a aceitação dos diferentes módulos no valor médio, o que indica variações entre equipamentos superiores às variações estatísticas esperadas de um processo aleatório de seleção de rolhas de um mesmo conjunto.

3.2.4 Teste de média escala - Conclusão

Dado as rolhas que cada um destes módulos processa serem de um mesmo lote, isto é, amostras da mesma população, não seria de esperar tais discrepâncias entre a quantidade

de rolhas rejeitadas em cada uma das colunas. A título de exemplo², a coluna 2 do cromatógrafo 6.2 rejeitou 21,4 % das rolhas que o equipamento analisou. Esta coluna, uma vez que analisou 25 % das rolhas, rejeitou mais de 85 % das rolhas que lhe foram atribuídas, resultando num custo avultado em comparação com a situação de resolução prévia deste transtorno.

Consequentemente, a realização deste teste permitiu identificar este fenómeno de conjugação de perfil cromatográfico adverso com o intervalo de integração escolhido como uma das causas da elevada quantidade de rolhas FP em particular no grupo G2 (referido previamente na subsecção afeta às causas de FP, 2.2.1). A utilização do algoritmo desenvolvido neste trabalho não só reduz a rejeição como também o desvio padrão das medidas, aproximando as aceitações entre os equipamentos e gerando resultados mais próximos do esperado estatisticamente (rejeições similares entre equipamentos após a análise de um elevado número de rolhas).

Este resultado aparenta generalizar-se para todos os grupos de rolhas: a medida da proporção de rolhas dos G1, G2 e G3 atua como estimador da proporção destes grupos na totalidade do lote de produção. Mesmo com um intervalo de confiança de 95 %, os resultados do algoritmo atual apresentam grande disparidade entre si e esta é reduzida com a aplicação do algoritmo a que este trabalho diz respeito e este resultado reflete-se nos dados na figura 3.9.

Na numeração dos equipamentos nos dados expostos nesta secção é notória a numeração sequencial dos cromatógrafos, o que implica a aparente omissão dos equipamentos 6.1 e 6.4. A ausência do primeiro prende-se com motivos de manutenção programada. Já a ausência do equipamento 6.4 merece mais atenção devido a erros na sua configuração inicial que motivaram a rejeição dos dados.

Na ausência de uma escolha adequada de intervalo de integração, o programa permaneceu em funcionamento com os valores selecionados por omissão para os limites deste intervalo. Tal funcionamento resultou na supressão de parte ou da totalidade da área dos picos de TCA pelo algoritmo, devido à conjugação das condições do sistema da eq. 2.9 com os limites de integração inapropriados. A seriedade desta falha é evidente na captura de ecrã (fig. 3.10) que ilustra a omissão de um pico de área considerável pelo algoritmo.

Esta falha por negligência motivou o retorno ao desenvolvimento de uma versão automática do algoritmo e dos seus testes na linha de produção. A pretensão de existência de um algoritmo com características automáticas é poder atuar não só sobre alterações ao TR como também a casos de desatenção, por forma a proteger a marca NDtech® e as suas garantias.

²Exemplo de entre os casos de rejeição largamente superior à mediana

3.2. VALIDAÇÃO EM LINHA DE PRODUÇÃO (SEM SUBSTITUIÇÃO DE RESULTADOS) - ALGORITMO NA FORMA MANUAL

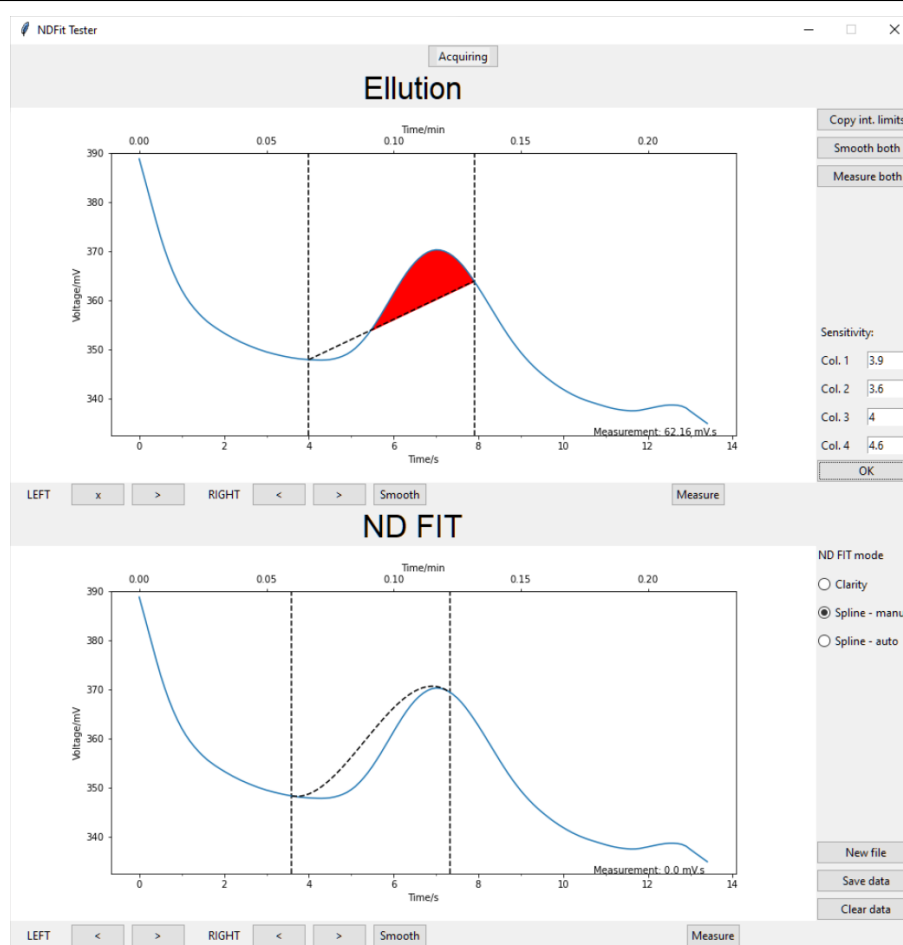


Figura 3.10: Erro de integração do pico de TCA pelo NDFit.

A azul: Traçado do cromatograma.

A vermelho: Área integrada

A preto, a tracejado: Limites de integração e fundo cromatográfico.

3.2.5 Teste de pequena escala - Métodos e objetivos

Em conjunto com o teste mencionado previamente, foi realizada a experiência de analisar a quantidade de TCA das rolhas classificadas em grupos distintos pelos dois algoritmos. Para tal, num equipamento disponível para a realização do teste, foram analisadas 528 rolhas alimentadas manualmente num dos equipamentos utilizados na produção. A combinação dos resultados de ambos os algoritmos e a rastreabilidade entre rolha, cromatograma, posteriormente, análise no laboratório de I&D, permitiu separá-las nas atribuições simultâneas ou não aos grupos 1, 2 e 3 e efetuar uma análise mais cuidada para afirmar erros de um ou de ambos os métodos de cálculo disponíveis.

3.2.6 Teste de pequena escala - Resultados

Após a separação das rolhas por ambos os algoritmos, estas ficaram distribuídas do modo descrito na matriz de confusão abaixo (tabela 3.3) onde se pode verificar a concordância de ambos os algoritmos na grande maioria das rolhas. Às 455 rolhas que o Ellution (algoritmo atual) atribui ao grupo 1, é proposta a adição de 59 rolhas provenientes do G2 e 2 são enviadas em sentido oposto. Sobraram apenas 3 rolhas atribuídas ao G3 de forma unânime, não havendo trocas de rolhas entre os grupos 2 e 3.

Tabela 3.3: Distribuição de classificação das rolhas por grupo (em número de rolhas)

		NDfitM		
		G1	G2	G3
Alg. atual	G1	453	2	0
	G2	59	11	0
	G3	0	0	3

Todas as rolhas que suscitaram desacordos (dos elementos fora da diagonal da tabela), bem como um controlo de 10 rolhas TCA-negativas e 3 rolhas TCA-positivas (duas do G2 e uma do G3) concordantes entre os algoritmos foram enviadas para o laboratório de I&D da AC para uma análise mais cuidada.

O subconjunto dos dados da tab. 3.3 enviados para esse mesmo laboratório pode ser sumariamente representado da mesma forma na tabela 3.4:

Tabela 3.4: Distribuição de classificação das rolhas por grupo (em número de rolhas)

		NDfitM		
		G1	G2	G3
Alg. atual	G1	10	2	0
	G2	59	2	0
	G3	0	0	1

Os resultados encontram-se expostos em detalhe no apêndiceA.2 devido à sua extensão e ao facto de outra representação que não a tabular implicar perda de informação. Poderão, no entanto, ser resumidos nos seguintes pontos:

1. De entre as 10 rolhas em que ambos os algoritmos classificaram como G1 não foi encontrada qualquer rolha contaminada com TCA;
2. De entre as 2 rolhas em que ambos os algoritmos classificaram como G2 apenas uma delas tinha, de facto, uma quantidade de TCA detetável. Já a rolha do G3 apresentava uma quantidade de TCA considerável (de $3,6 \pm 0,5$ ng/L);
3. As 2 rolhas classificadas como G1 pelo algoritmo atual e G2 pelo algoritmo NDfitM encontravam-se TCA-negativas;

4. Das 59 rolhas reaproveitadas do G2, 3 continham TCA em quantidades que geraram contaminações de $0,5 \pm 0,5$ a $0,6 \pm 0,5$ ng/L. Como tal, a utilização do algoritmo NDfitM resultaria na venda de rolhas com contaminação superior à da garantia da marca.

3.2.7 Teste de pequena escala - Conclusão

O resultado deste teste foi o aparente compromisso da garantia da marca NDtech® a troco um ganho aceitação. Dado o posicionamento desta marca no segmento de topo do mercado rolheiro, este compromisso não faz qualquer sentido. Assim sendo, estes resultados põem em causa esta versão do algoritmo. Não obstante, poderá ser utilizado como base para o desenvolvimento de soluções posteriores.

3.3 Alterações ao algoritmo decorrentes dos testes

Para poder cumprir com o conjunto de objetivos anteriormente traçados, foram necessárias alterações ao funcionamento do algoritmo como descrito na secção 2.6.

Em primeiro lugar, poder-se-á indicar a falta de necessidade de alteração do método de integração dos picos de dimensões consideráveis, para os quais os métodos atuais funcionam adequadamente. Foi traçado um limite de 8 mV.s, o qual se considerou que permitiria conciliar o risco de passagem de rolhas contaminadas com o maior aproveitamento das rolhas, por haver uma quantidade relevante de rolhas entre os 6 e os 8 mV.s. Acima deste valor, a decisão que seria tomada antes dos desenvolvimentos deste trabalho foram consideradas inalteráveis devido ao risco de aceitação de rolhas contaminadas, resultando no algoritmo esquematicamente representado na fig. 3.11, abaixo.

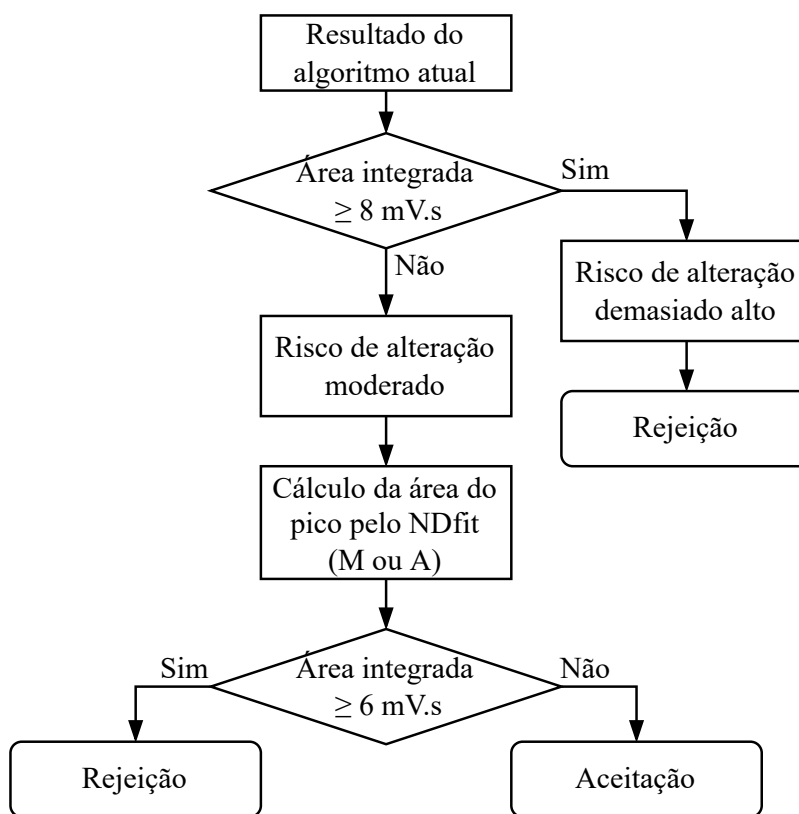


Figura 3.11: Alterações ao funcionamento do algoritmo

Em segundo lugar, de forma a permitir um maior foco do algoritmo na zona de interesse, os limites de integração permaneceram os introduzidos pelo utilizador, mas permitiu-se ao algoritmo procurar um melhor ajuste a um pico cromatográfico ocorrente não só no intervalo selecionado pelo utilizador como para a extensão desse intervalo 1 segundo em cada extremo.

Este modo (automático - NDfitA - e como descrito em 2.6.2), em conjunto com o método já testado que será efetuada a validação na componente seguinte do trabalho.

3.4 Validação em linha de produção (com substituição de resultados)

3.4.1 Objetivos e métodos

Com substituição dos resultados do programa Ellution, atualmente utilizado pela AC, pretende-se aumentar a taxa de aquisição de dados em relação ao método de validação anterior, manipulando a comunicação entre os programas. Esta comunicação ocorre por via de um ficheiro temporário, que é escrito pelo Ellution e, após um curto intervalo de tempo, lido e apagado pelo programa de controlo.

3.4. VALIDAÇÃO EM LINHA DE PRODUÇÃO (COM SUBSTITUIÇÃO DE RESULTADOS)

Durante o curto intervalo de tempo de existência do ficheiro temporário, é possível lê-lo e editá-lo. O programa de testes do algoritmo toma partido deste facto para editar as informações contidas neste ficheiro, de forma a substituir os cálculos efetuados pelo Ellution com o objetivo de manipular a decisão efetuada pelo programa de controlo (de produção) (fig. 3.12).

Para acelerar a forma automática do algoritmo, foi necessária a implementação da reutilização dos parâmetros dos fundos polinomiais anteriores no cálculo dos seguintes, tendo este método reduzido grandemente o número de passos tomados pelo método de cálculo numérico.

Realizando os testes do modo aqui descrito, a taxa de aquisição de resultados é a mesma da taxa de produção de rolas por um dado equipamento, o que permite um acelerar do ritmo de aquisição de resultados.

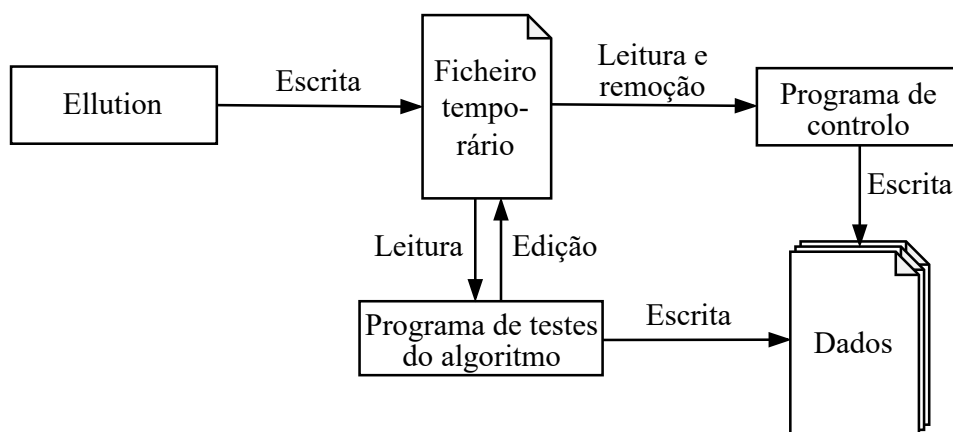


Figura 3.12: Funcionamento da aplicação de teste de algoritmos em modo de substituição de resultados

Como tal, foram analisadas nos equipamentos de produção 5000 rolas com os programas que a AC dispõe atualmente, outras 5000 com o NDfitM e, por constrangimentos de tempo e de disponibilidade de equipamentos, 1884 rolas com o algoritmo NDfitA.

Após a separação, as rolas do grupo 1 foram analisadas por amostragem, devido ao seu número elevado. Já as rolas dos grupos 2 e 3, foram caracterizadas na totalidade, com exceção do grupo 2 selecionado pelo algoritmo atual, que devido à elevada quantidade, foi também analisado por amostragem.

3.4.2 Resultados

Em comparação com o teste prévio, a quantidade de TCA disperso pelas rolas que compõem o lote aparenta ser manifestamente mais baixa, tendo todas as formas de cálculo apresentado aceitação acima dos 90% (tabs. 3.6 a 3.7)

Apesar disso, as duas versões do novo algoritmo oferecem melhorias em termos de aceitação, da ordem dos $6,9 \pm 0,9 \%$ para o NDfitM e de $7,4 \pm 0,9 \%$ para o NDfitA.

Tabela 3.5: Classificação das rolhas em grupos pelo algoritmo atual da AC

Classificação das rolhas pelo algoritmo atual da AC		
Grupo	Quantidade (rolhas)	Estimativa da proporção ³ (%)
G1	4614	92,3 ± 0,8
G2	376	7,5 ± 0,8
G3	10	0,2 ± 0,2

Tabela 3.6: Classificação das rolhas em grupos pelo algoritmo NDfitM

Classificação das rolhas pelo algoritmo NDfitM		
Grupo	Quantidade (rolhas)	Estimativa da proporção ³ (%)
G1	4959	99,2 ± 0,3
G2	35	0,7 ± 0,3
G3	6	0,1 ± 0,1

Tabela 3.7: Classificação das rolhas em grupos pelo algoritmo NDfitA

Classificação das rolhas pelo algoritmo NDfitA		
Grupo	Quantidade (rolhas)	Estimativa da proporção ³ (%)
G1	1878	99,7 ± 0,2
G2	6	0,3 ± 0,2
G3	0	0 ± 0

Apesar deste resultado, cada um dos algoritmos falhou na remoção total de rolhas com tricloroanisol do G1. O resultado da análise em maceração realizada no laboratório de I&D das amostras de 500 rolhas deste grupo para cada um dos algoritmos (tab. 3.8)

Tabela 3.8: Precisão na identificação do G1 pelos algoritmos em análise

Precisão na identificação do G1 pelos algoritmos em análise ³ (%)	
Algoritmo atual	99,8 ± 0,4
NDfitM	99,2 ± 0,8
NDfitA	99,2 ± 0,8

apresenta valores próximos que apesar de penderem no sentido do algoritmo atual, não são disjuntos nos seus intervalos de confiança (de 95%). Estes resultados ocorrem à custa 1 rolha TCA-positiva não detetada pelo método em uso corrente e 4 rolhas TCA-positiva não detetadas por cada um dos algoritmos NDfit (tab. 3.9).

³Intervalo de confiança de 95%

3.4. VALIDAÇÃO EM LINHA DE PRODUÇÃO (COM SUBSTITUIÇÃO DE RESULTADOS)

Tabela 3.9: Rolhas contaminadas no G1 não detetadas pelos algoritmos em análise

Contaminação provocada pelas rolhas TCA-positiva não detetadas pelos algoritmos	
Algoritmo	Contaminação de TCA (ng/L)
Atual	10,6 ± 0,5
NDfitM	6,9 ± 0,5
	2,7 ± 0,5
	2,5 ± 0,5
	0,9 ± 0,5
NDfitA	2,2 ± 0,5
	1,2 ± 0,5
	1,2 ± 0,5
	0,7 ± 0,5

As estatísticas relativas à proporção de rolhas TCA-positiva nos grupos 2 e 3 encontram-se no apêndice B.

3.5 Repetitividade experimental

A fim de avaliar a repetitividade das áreas dos picos de TCA medidos no NDtech, avaliou-se qual a dispersão de valores obtidas na medida de uma dada rolha. Para tal, com o equipamento a funcionar continuamente e com a utilização do programa já em uso pela empresa, foram colocadas rolhas contaminadas em todas as câmaras do equipamento. Os resultados encontram-se expressos nos gráficos das figs. 3.13 e 3.14.

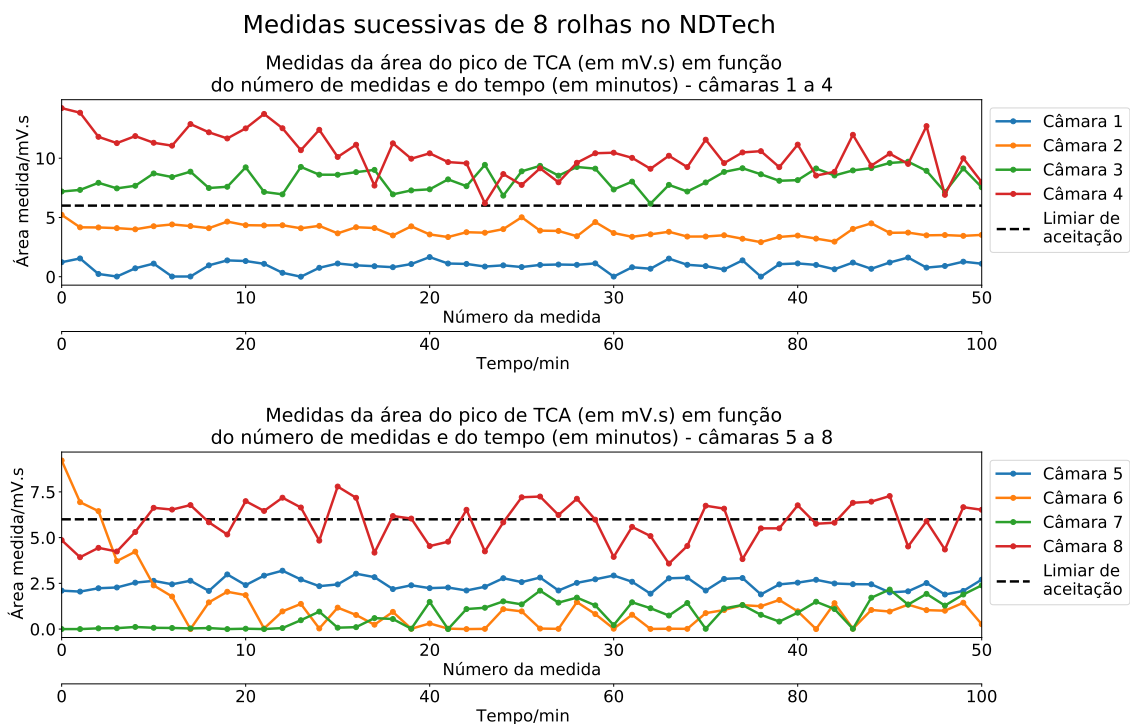


Figura 3.13: Resultados de medidas sucessivas das áreas dos picos de TCA de um conjunto de 8 rolhas no NDtech

Na primeira destas (fig. 3.13), verifica-se a elevada variabilidade das medidas das áreas das rolhas, resultantes de fenómenos aleatórios e ilustrados pelas variações bruscas do valor de área entre medidas. A rolha contida na câmara 8 destaca-se pelo facto de as suas alterações de valor motivarem que, dependendo desta medida, a rolha alterne a sua classificação como apropriada para G1 ou para G2.

A aplicação do filtro de média móvel⁴ sobre os dados permite analisar tendências ao longo de várias medidas. Com exceção da rolha 6 que verifica uma descida abrupta na área do pico de TCA, chegando a ser classificada como G2 nas primeiras medidas⁵ para posteriormente obter classificações exclusivamente no G1, nenhuma das rolhas apresenta grandes variações na quantidade de TCA medida.

⁴Convolução dos dados com um filtro de matriz $1/7 \cdot [1, 1, 1, 1, 1, 1, 1]$

⁵Atenuadas pelo filtro de média mas visíveis na primeira das figuras

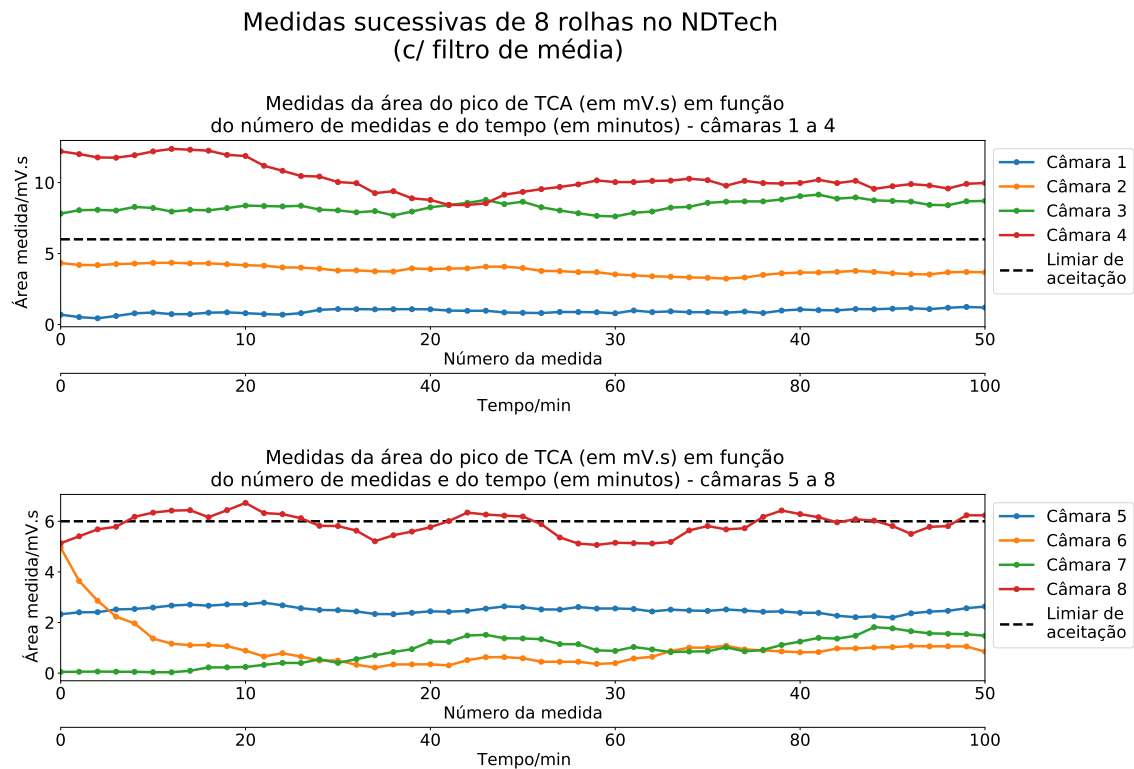


Figura 3.14: Resultados de medidas sucessivas das áreas dos picos de TCA de um conjunto de 8 rolhas no NDtech, com filtro de média aplicado.

A análise da fonte desta variabilidade poderia, na opinião do autor, constituir uma fonte de informação importante de forma a prever falhas no equipamento e atenuar o efeito destas na integridade desta marca.

Trabalho futuro e conclusão

4.1 Trabalho futuro

4.1.1 Produção e manutenção

Apesar de não ser o foco deste trabalho, as análises dos resultados em maior detalhe (para cada coluna cromatográfica) permitiu o diagnóstico de um conjunto de problemas, mesmo que a intenção não fosse a sua colmatação com o algoritmo desenvolvido.

Nesse sentido, a análise das aceitaçãoes relativamente a cada coluna fornece maior informação sobre as causas das baixas de aceitação que contribuem grandemente para o conhecimento das causas de baixas de aceitação na produção de rolhas NDtech®. Por conseguinte, sugere-se a implementação de uma solução que permita discernir as estatísticas das colunas cromatográficas dos diversos equipamentos utilizados na produção.

Sugere-se também a revisão dos fatores de sensibilidade que justificam a substituição das colunas cromatográficas e a redução da margens de aceitação dos processos de controlo referidos em 1.2.3. No entanto, a redução dessas margens, dada a variabilidade do método de introdução da solução de TCA nas câmaras cromatográficas - por pipetagem manual - introduz variabilidade que poderá inviabilizar esta melhoria. De for a mitigar esta variabilidade, propõe-se que seja introduzida uma forma de realizar esta injeção automaticamente em versões futuras dos equipamentos.

A continuação dos esforços de redação destes procedimentos de forma escrita para serem facilmente seguidos pelos operadores e sua formação contínua mostram-se também como esforços que merecem a sua continuidade.

4.1.2 Algoritmo e à aplicação de testes

Não sendo da competência do autor a decisão sobre a implementação no local de qualquer um dos algoritmos desenvolvidos no âmbito deste trabalho (ou outros), sugere-se a utilização da interface de testes (A) para o teste de qualquer opção de algoritmo de solução do problema de integração do pico de TCA no NDtech, uma vez que esta permite a realização de testes a custo quase-nulo, por não exigir necessidade de integração das

funções de cálculo no programa de controlo de produção, poupando custos de I&D.

Sugere-se também a utilização de computação paralela no cálculo do fundo polinomial (vide fig. 4.1). Desta forma, partindo de várias tentativas iniciais, poder-se-ia utilizar a que mais rapidamente convergisse de modo a formar o fundo e permitiria a utilização de uma maior percentagem da capacidade de computação dos equipamentos utilizados.

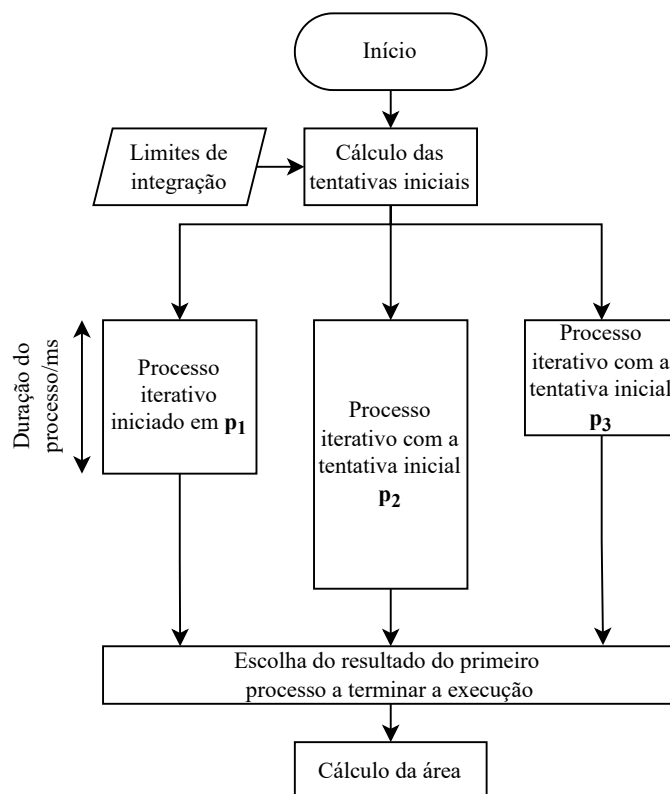


Figura 4.1: Exemplo de paralelização do algoritmo desenvolvido neste trabalho.

Recomenda-se também a o seguimento das normas de engenharia de software mais atuais em qualquer alteração posterior efetuada a este trabalho. A ausência destes, que se não se consideram totalmente integrados no âmbito deste trabalho, devido à sua dimensão, provocou dificuldades que poderiam ter sido resolvidas mais celeremente.

Dentro das recomendações de alternativa ao algoritmo utilizado, serão ainda deixadas duas sugestões para trabalhos posteriores:

Machine Learning (ML) A omissão dos algoritmos de ML até neste trabalho prende-se com a consideração da perspetiva pela parte do autor. O funcionamento não-trivial dos seus algoritmos, díspar da formação científica da maioria dos operadores, poderia resultar numa pioria da relação operador-equipamento com prejuízo da produção e/ou precisão da identificação do G1. Apesar disso, o autor distancia-se de qualquer estigma para com os algoritmos mais convencionais de ML e crê que qualquer tentativa de melhoria dos resultados da identificação do pico de TCA nestes cromatogramas será benéfica.

Alteração dos modelos de sensibilidade Na sua forma mais simples, o resultado do algoritmo de cálculo atual consiste na parte positiva do integral entre o cromatograma e o segmento de reta - $f_1(t)$ - definido pelas interseções do cromatograma com o segmento de reta mencionado. Este integral é presentemente apenas multiplicando por um fator, o fator de sensibilidade, α , obtendo a medida da quantidade de TCA libertável, Q (em mV.s), como descrita na equação 4.1:

$$Q = \alpha \cdot \int_{t_0}^{t_1} (s(t) - f_1(t)) dt \text{ mV.s} \quad (4.1)$$

Devido à similaridade dos cromatogramas de uma dada coluna de um equipamento, a hipótese de adição de um valor de compensação, β poderá ser considerada, da mesma forma da qual o fator de sensibilidade é utilizado presentemente. Tal permitiria a separação da área que de momento apenas contribui para o fundo do cromatograma e poderia ser distinto para os vários fundos cromatográficos de um dado equipamento. Assim, a quantidade de TCA libertável poderia ser avaliada como descreve a equação seguinte (eq. 4.2):

$$Q = \alpha \cdot \left(\int_{t_0}^{t_1} s(t) dt \right) \pm \beta \text{ mV.s} \quad (4.2)$$

O incremento da dificuldade de escolha do parâmetro β de forma a que este parâmetro seja o mais próximo possível de $\alpha \cdot \int_{t_0}^{t_1} s_{\text{sem TCA}}(t) dt$ constitui, na visão do autor, o maior entrave à utilização deste modelo de calibração dos equipamentos. A utilização de uma média móvel deste valor para os últimos cromatogramas poderá auxiliar neste sentido.

4.2 Conclusões

O trabalho descrito neste documento teve como objetivos a atingir a melhoria simultânea da precisão e da aceitação das rolhas naturais utilizadas na produção destas.

A criação de soluções que permitam a validação deste algoritmo em meio de produção (que poderão vir a ser utilizadas noutros trabalhos com o mesmo objetivo) resultou num dispêndio de tempo considerável. Este dispêndio de tempo afetou negativamente os testes do algoritmo nas suas versões mais avançadas uma vez que as próprias ferramentas carecem também das suas fases de verificação e validação. Contemplando esse facto, em conjunto com

1. as baixas probabilidades de ocorrência de FN,
2. o tempo limitado para a realização deste trabalho,
3. a disponibilidade dos equipamentos e o seu estado de funcionamento,
4. a disponibilidade de matéria prima para a realização de testes

e, fundamentalmente, o facto de se estar a examinar um produto de origem natural próximo dos limites da técnica analítica em questão, originaram resultados a partir dos quais não é possível, com grau de confiança satisfatório, afirmar que os ganhos de aceitação não ocorrem à custa de perdas de precisão na identificação das rolhas do G1.

Sendo as rolhas NDtech® o produto topo de gama da AC e um artigo sob o qual assentam a confiança dos consumidores na resolução do problema do TCA pela indústria corticeira, o risco de perda da confiança do consumidor é suficiente para justificar a não inclusão deste algoritmo na produção de imediato.

Apenas a concretização de testes a uma escala suficiente distinguir a quantidade de FN provenientes de ambos os *softwares* ou a introdução faseada do algoritmo (com o parecer dos compradores destas rolhas) poderiam assegurar a manutenção do nível de fiabilidade.

Bibliografia

- [1] J. M. Lourenço. *The NOVAthesis L^AT_EX Template User's Manual*. NOVA University Lisbon. 2021. URL: <https://github.com/joaomlourenco/novathesis/raw/master/template.pdf>.
- [2] APCOR - Associação Portuguesa da Cortiça. *Boletim Estatístico - Mercados da Cortiça*. Abr. de 2022. URL: https://www.apcor.pt/wp-content/uploads/2022/04/apcor_boletim_trimestral_11_2022_web.pdf (acedido em 13/07/2022).
- [3] A. L. de Almeida Pires da Fonseca e O. M. N. D. Teodoro. “Caracterização de microfluxos: medição, concepção e calibração”. Faculdade de Ciências e Tencologia da Universidade Nova de Lisboa, nov. de 2013, pp. 100–101. URL: <http://hdl.handle.net/10362/10930>.
- [4] J. A. Maga e J.-L. Puech. “Cork and Alcoholic Beverages.” Em: *Food Reviews International* 21 (1 2005), pp. 53–53–68. ISSN: 8755-9129. DOI: 10.1081/FRI-200040593. URL: <https://search.ebscohost.com/login.aspx?direct=true&db=a9h\&AN=16133531\&site=eds-live>.
- [5] M. C. Cravero. “Musty and moldy taint in wines: A review”. Em: *Beverages* 6 (2 jun. de 2020), pp. 1–13. ISSN: 23065710. DOI: 10.3390/beverages6020041.
- [6] M. K. Taylor et al. “Supercritical fluid extraction of 2,4,6-trichloroanisole from cork stoppers”. Em: *Journal of Agricultural and Food Chemistry* 48 (6 jun. de 2000), pp. 2208–2211. ISSN: 00218561. DOI: 10.1021/jf991045q.
- [7] M. I. V. Teixeira et al. “2,4,6-Trichloroanisole: A Consumer Panel Evaluation”. Em: *Ciência e Técnica Vitivinícola* 21 (2 2006), pp. 53–65.
- [8] E. Hervé et al. *Chemical Analysis of TCA As a Quality Control Tool For Natural Corks*. URL: <https://docplayer.net/146328650-Chemical-analysis-of-tca-as-a-quality-control-tool-for-natural-corks.html> (acedido em 24/04/2022).
- [9] E. C. Federation. *Código Internacional das Práticas Rolheiras*.
- [10] M. Cabral. “Sistema e dispositivo para extrair compostos arrastados pelo vapor de água em produtos em cortiça”. Pat. 2005.

- [11] M. Cabral, I. Roseira e S. Dahl. “Dispositivo para a extração de compostos voláteis por vaporização de água em rolhas de cortiça e/ou em peças de cortiça utilizadas na produção de rolhas”. Pat. 2009.
- [12] Amorim Cork. *NDtech - Portfólio Produto - Portfólio - Amorim Cork*. URL: <https://www.amorimcork.com/pt/portfolio/d/ndtech/231/> (acedido em 24/06/2022).
- [13] Amorim Cork. *Twin Top Evo - Portfólio Produto - Portfólio - Amorim Cork*. URL: <https://www.amorimcork.com/pt/portfolio/d/twin-top-evo/225/> (acedido em 24/06/2022).
- [14] Amorim Cork. *NDtech Sparkling - Portfólio Produto - Portfólio - Amorim Cork*. URL: <https://www.amorimcork.com/pt/portfolio/d/ndtech-sparkling/232/> (acedido em 24/06/2022).
- [15] Amorim Cork. *Advantec - Portfólio Produto - Portfólio - Amorim Cork*. URL: <https://www.amorimcork.com/pt/portfolio/d/advantec/227/> (acedido em 24/06/2022).
- [16] M. R. dos Santos Serrasqueiro Rossa. “Projeto de equipamento industrial para extração de TCA de rolhas de cortiça natural”. Universidade Nova de Lisboa, out. de 2020. URL: https://run.unl.pt/bitstream/10362/110869/1/Rossa_2020.pdf.
- [17] APCOR - Associação Portuguesa da Cortiça. *Realcork – Combate ao TCA*. URL: <https://www.apcor.pt/combate-ao-tca-tricloroanisole/>.
- [18] International Organization for Standardization. *ISO - ISO 22308-1:2021 - Cork bark selected as bottling product — Part 1: Sensory evaluation — Methodology for sensory evaluation by soaking*. Rel. téc. Jan. de 2021. URL: <https://www.iso.org/standard/73976.html>.
- [19] Wine Spectator e S. Mustaich. *New Wine Technology Spurs Natural Cork Comeback | Wine Spectator*. Dez. de 2016. URL: <https://www.winespectator.com/articles/new-cork-technology-detects-tca-taint>.
- [20] A. C. Portocork Italia. *Una rivoluzione chiamata NDtech*. URL: <https://portocorkitalia.com/una-rivoluzione-chiamata-ndtech/> (acedido em 21/07/2022).
- [21] K. B. Lee. *Laboratoire National Henri Becquerel*. Nov. de 2005. URL: http://www.lnhb.fr/nuclides/Ni-63_tables.pdf (acedido em 08/07/2022).
- [22] F. W. Fifield e D Kealey. *Principles and Practice of Analytical Chemistry*. Ed. por B. S. Ltd. 5ª. Blackwell Science Ltd., 2000, pp. 1–105.
- [23] K MacNamara, R Leardi e A Hoffmann. “Developments in 2-D gas chromatography with heartcutting”. Em: *LC GC NORTH AMERICA* (fev. de 2004), pp. 82–91. ISSN: 1527-5949. URL: <https://search.ebscohost.com/login.aspx?direct=true&db=edswsc&AN=000224024400059&site=eds-live>.
- [24] Amorim Cork. *Indicadores acumulados do NDtech em 2021*. Fev. de 2022. Documento interno não publicado.

-
- [25] V. B. D. Marco e G. G. Bombi. “Mathematical functions for the representation of chromatographic peaks”. Em: *Journal of Chromatography A* 931 (1-2 out. de 2001), pp. 1–30. ISSN: 0021-9673. DOI: 10.1016/S0021-9673(01)01136-0.
- [26] A. Felinger. *5 Peak shape analysis*. Ed. por A. Felinger. 1998. DOI: [https://doi.org/10.1016/S0922-3487\(98\)80024-5](https://doi.org/10.1016/S0922-3487(98)80024-5).
- [27] P. Virtanen et al. “SciPy 1.0—Fundamental Algorithms for Scientific Computing in Python”. Em: *Nature Methods* 17 (261 2010). DOI: 10.1038/s41592-019-0686-2. URL: <http://arxiv.org/abs/1907.10121>.
- [28] M. Newville et al. *LMFIT: Non-Linear Least-Square Minimization and Curve-Fitting for Python*. Versão 0.8.0. Set. de 2014. DOI: 10.5281/zenodo.11813. URL: <https://doi.org/10.5281/zenodo.11813>.
- [29] P. Virtanen et al. *SciPy v1.8.0 Manual*. Fev. de 2021. URL: https://github.com/scipy/scipy/blob/v1.8.0/scipy/optimize/_minpack_py.py#L533-L839.
- [30] K. Levenberg. “A Method for the Solution of Certain Non-linear Problems in Least Squares”. Em: *Source: Quarterly of Applied Mathematics* 2 (2 1944), pp. 164–168.
- [31] D. W. Marquardt. “An Algorithm for Least-Squares Estimation of Nonlinear Parameters”. Em: *Journal of the Society for Industrial and Applied Mathematics* 11 (2 1963), pp. 431–441. DOI: 10.1137/0111030.
- [32] F. Lundh. *An introduction to tkinter*. Out. de 1999. URL: www.pythonware.com/library/tkinter/introduction/index.html (acedido em 13/05/2022).

Interface *NDfit Tester*

A.1 Funcionamento

A interface *NDfittester* funciona com o uso de encadeamentos de execução (*threads*) para realizar duas tarefas em simultâneo.

O primeiro *thread* faz a atualização gráfica da interface e efetua todos os cálculos, ao passo que o segundo monitoriza a criação de ficheiros pelo programa de controlo do cromatógrafo utilizado pela AC, sendo o primeiro o *thread master* e o segundo o *thread slave*. Após a criação de um novo ficheiro, o segundo (ver esquema A.1) inicia o processo de leitura do ficheiro no primeiro, que posteriormente atualiza a interface com o resultado dos cálculos e armazena os dados num ou vários ficheiros.

Alterações aos parâmetros de integração também provocam uma atualização da interface e armazenamento dos dados recém-calculados.

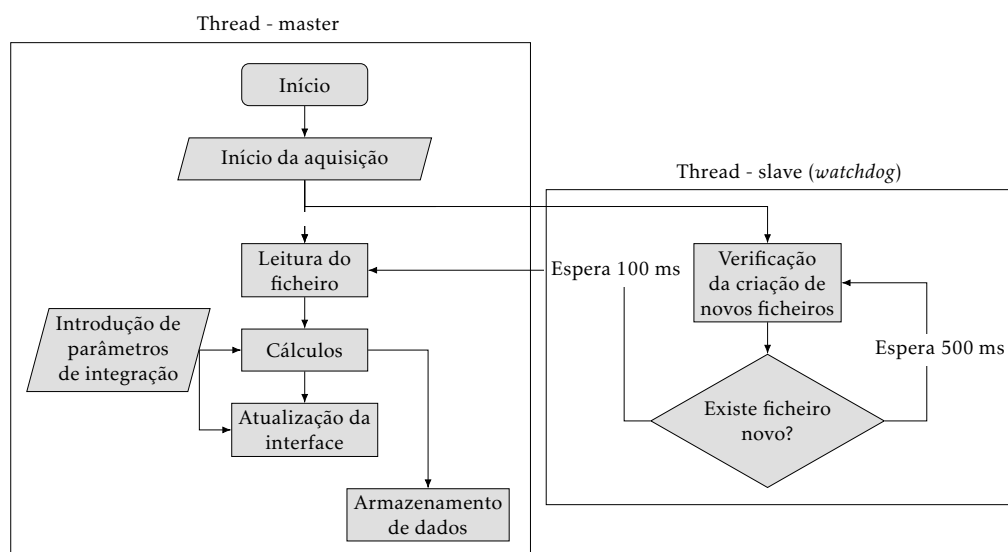


Figura A.1: Esquema de funcionamento da interface *NDfittester*

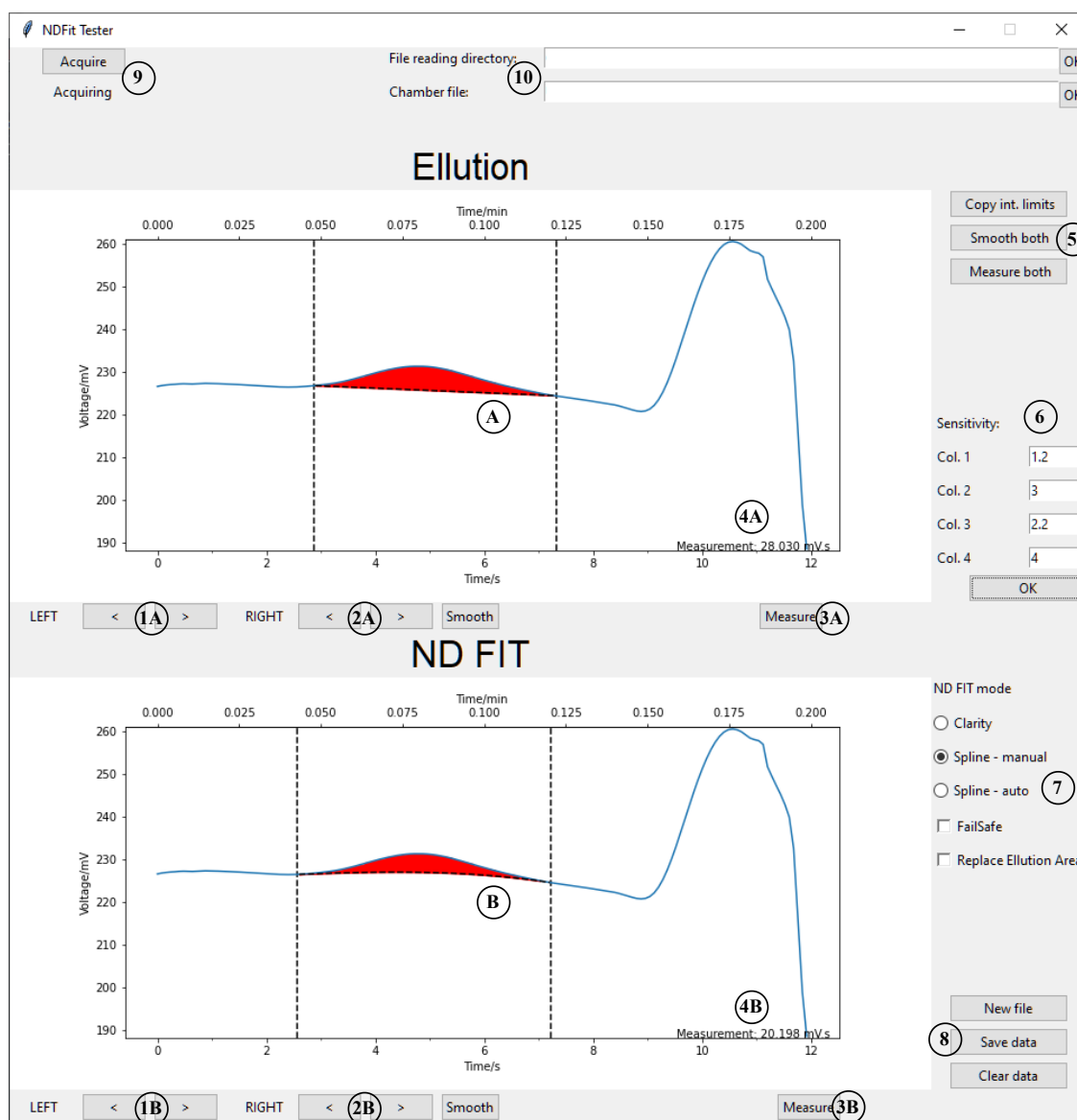


Figura A.2: Interface gráfica para testes dos algoritmos

A.2 Instruções

A interface gráfica *NDFit Tester* dispõe dos seguintes comandos consultáveis na tabela A.1, sendo esta a chave para interpretação da figura A.2.

A utilização inicia-se pela introdução da diretoria onde serão lidos os ficheiros temporários no campo *File reading directory*, efetuando o mesmo para o campo *Chamber file* (no entanto esta segunda diretoria deve incluir o nome do ficheiro) a seleção é confirmada com os botões *OK* e a aquisição poderá então ser iniciada no botão *Acquire*.

Assim que começarem a aparecer cromatogramas, poderá ser utilizado o comando *Measure both* para que se iniciem as medidas, cujo resultado aparece no canto inferior direito dos cromatogramas. Todos os restantes campos encontram-se preenchidos por omissão e, em caso de necessidade, poderão ser alterados pelo utilizador.

Tabela A.1: Instruções da aplicação

Botão	Cromatograma e integração segundo o algoritmo:
A	Área do cromatograma integrado segundo o algoritmo contido no Ellution
B	Área do cromatograma integrado segundo o algoritmo NDfit
1A, 1B	Movimentação do limite inferior de integração em A e B, respetivamente
2A, 2B	Movimentação do limite superior de integração em A e B, respetivamente
3A, 3B	Início da integração em A e B, respetivamente
4A, 4B	Resultado da integração em A e B, respetivamente
	Cópia dos limites de integração de A para B;
5	Aplicação do filtro de Savitzky-Golay em A e B; Início da integração em A e B, simultaneamente.
6	Introdução dos fatores de sensibilidade das colunas;
7	Seleção dos modos de funcionamento do NDfit; Criação de ficheiro de dados;
8	Início da gravação de dados; Limpeza do ficheiro de dados;
9	Iniciar a aquisição de cromatogramas a partir dos ficheiros gerados;
10	Diretoria de leitura dos ficheiros;

(A e B representam, respetivamente, os comandos relativos aos algoritmos atual - contido no programa Ellution - e NDfit)

Resultados experimentais

B.1 Resultados do teste de pequena escala

Tabela B.1: Comparação entre as áreas medidas pelo algoritmo atual e pelo e a quantidade de TCA medida no laboratório para as rolhas TCA-negativas

Área de integração segundo o Ellution (mV.s)	Área de integração segundo o NDfitM (mV.s)	Quantidade de TCA segundo o laboratório de I&D (ng/L)
$4,1 \pm 0,9$	$4,0 \pm 0,9$	$0,0 \pm 0,5$
$4,3 \pm 0,9$	$1,4 \pm 0,7$	$0,0 \pm 0,5$
$4,4 \pm 0,9$	$1,2 \pm 0,6$	$0,0 \pm 0,5$
$4,5 \pm 0,9$	$1,5 \pm 0,7$	$0,0 \pm 0,5$
$4,6 \pm 0,9$	$2,7 \pm 0,8$	$0,0 \pm 0,5$
$4,9 \pm 0,9$	$0,9 \pm 0,6$	$0,0 \pm 0,5$
5 ± 1	$0,2 \pm 0,6$	$0,0 \pm 0,5$
5 ± 1	$1,9 \pm 0,7$	$0,0 \pm 0,5$
6 ± 1	$1,0 \pm 0,6$	$0,0 \pm 0,5$
6 ± 1	$2,4 \pm 0,7$	$0,0 \pm 0,5$

Tabela B.2: Comparação entre as áreas medidas pelo algoritmo atual e pelo NDfitM e a quantidade de TCA medida no laboratório para as rolhas do G1

Área de integração segundo o Ellution (mV.s)	Área de integração segundo o NDfitM (mV.s)	Quantidade de TCA segundo o laboratório de I&D (ng/L)
7 ± 2	7 ± 2	$1,6 \pm 0,5$
8 ± 2	8 ± 2	$0,0 \pm 0,5$
41 ± 4	59 ± 6	$3,6 \pm 0,5$

Tabela B.3: Comparação entre as áreas medidas pelo algoritmo atual e pelo NDfitM e a quantidade de TCA medida no laboratório para as rolhas G2 em ambos os algoritmos

Área de integração segundo o Ellution (mV.s)	Área de integração segundo o NDfitM (mV.s)	Quantidade de TCA segundo o laboratório de I&D (ng/L)
7 ± 1	7 ± 1	$1,6 \pm 0,5$
8 ± 2	8 ± 2	$0,0 \pm 0,5$

Tabela B.4: Comparação entre as áreas medidas pelo algoritmo atual e pelo NDfitM e a quantidade de TCA medida no laboratório para as rolhas do G1

Área de integração segundo o Ellution (mV.s)	Área de integração segundo o NDfitM (mV.s)	Quantidade de TCA segundo o laboratório de I&D (ng/L)
6 ± 1	$0,1 \pm 0,6$	$0,0 \pm 0,5$
6 ± 1	$1,8 \pm 0,7$	$0,0 \pm 0,5$
6 ± 1	$0,0 \pm 0,5$	$0,0 \pm 0,5$
6 ± 1	$2,0 \pm 0,7$	$0,0 \pm 0,5$
6 ± 1	$0,0 \pm 0,6$	$0,0 \pm 0,5$
6 ± 1	$2,1 \pm 0,7$	$0,0 \pm 0,5$
6 ± 1	$2,4 \pm 0,7$	$0,0 \pm 0,5$
6 ± 1	$1,2 \pm 0,6$	$0,0 \pm 0,5$
6 ± 1	$0,1 \pm 0,6$	$0,0 \pm 0,5$
6 ± 1	$4,1 \pm 0,9$	$0,0 \pm 0,5$
6 ± 1	$0,8 \pm 0,6$	$0,0 \pm 0,5$
6 ± 1	$2,2 \pm 0,7$	$0,0 \pm 0,5$
6 ± 1	$2,4 \pm 0,7$	$0,0 \pm 0,5$
6 ± 2	$0,0 \pm 0,6$	$0,0 \pm 0,5$
6 ± 2	$0,2 \pm 0,6$	$0,0 \pm 0,5$
6 ± 2	$2,3 \pm 0,7$	$0,0 \pm 0,5$
6 ± 2	$1,9 \pm 0,7$	$0,0 \pm 0,5$
6 ± 2	$1,2 \pm 0,6$	$0,0 \pm 0,5$
6 ± 2	$0,3 \pm 0,6$	$0,0 \pm 0,5$
6 ± 2	$1,7 \pm 0,7$	$0,0 \pm 0,5$
6 ± 2	$1,8 \pm 0,7$	$0,0 \pm 0,5$
7 ± 2	$2,5 \pm 0,8$	$0,0 \pm 0,5$
7 ± 2	$2,5 \pm 0,7$	$0,0 \pm 0,5$
7 ± 2	$2,6 \pm 0,8$	$0,0 \pm 0,5$
7 ± 2	$2,3 \pm 0,7$	$0,4 \pm 0,5$

B.1. RESULTADOS DO TESTE DE PEQUENA ESCALA

7 ± 2	$2,7 \pm 0.8$	$0,0 \pm 0,5$
7 ± 2	$0,0 \pm 0.6$	$0,0 \pm 0,5$
7 ± 2	$0,9 \pm 0.6$	$0,0 \pm 0,5$
7 ± 2	$2,2 \pm 0.7$	$0,0 \pm 0,5$
7 ± 2	$2,7 \pm 0.8$	$0,0 \pm 0,5$
7 ± 2	$0,0 \pm 0.6$	$0,0 \pm 0,5$
7 ± 2	$1,1 \pm 0.6$	$0,0 \pm 0,5$
7 ± 2	$0,6 \pm 0.6$	$0,0 \pm 0,5$
7 ± 2	$1,6 \pm 0.7$	$0,0 \pm 0,5$
7 ± 2	$2,1 \pm 0.7$	$0,0 \pm 0,5$
7 ± 2	$0,0 \pm 0.6$	$0,0 \pm 0,5$
7 ± 2	$0,1 \pm 0.6$	$0,0 \pm 0,5$
7 ± 2	$3,3 \pm 0.8$	$0,0 \pm 0,5$
7 ± 2	$0,1 \pm 0.6$	$0,0 \pm 0,5$
7 ± 2	$3,1 \pm 0.8$	$0,0 \pm 0,5$
7 ± 2	$0,3 \pm 0.6$	$0,0 \pm 0,5$
8 ± 2	$0,2 \pm 0.6$	$0,0 \pm 0,5$
8 ± 2	$2,8 \pm 0.8$	$0,0 \pm 0,5$
8 ± 2	$0,4 \pm 0.6$	$0,0 \pm 0,5$
8 ± 2	$0,8 \pm 0.6$	$0,0 \pm 0,5$
8 ± 2	$4,0 \pm 0.9$	$0,0 \pm 0,5$
8 ± 2	$4,5 \pm 0.9$	$0,0 \pm 0,5$
8 ± 2	$2,9 \pm 0.8$	$0,0 \pm 0,5$
8 ± 2	$0,7 \pm 0.6$	$0,0 \pm 0,5$
8 ± 2	$3,9 \pm 0.9$	$0,1 \pm 0,5$
8 ± 2	$0,0 \pm 0.6$	$0,0 \pm 0,5$
8 ± 2	$0,8 \pm 0.6$	$0,3 \pm 0,5$
9 ± 2	$4,4 \pm 0.9$	$0,0 \pm 0,5$
9 ± 2	$1,0 \pm 0.6$	$0,5 \pm 0,5$
9 ± 2	$1,9 \pm 0.7$	$0,2 \pm 0,5$
9 ± 2	$2,3 \pm 0.7$	$0,6 \pm 0,5$
10 ± 2	$5,0 \pm 0.9$	$0,6 \pm 0,5$
10 ± 2	$3,8 \pm 0.9$	$0,0 \pm 0,5$
10 ± 2	$4,4 \pm 0.9$	$0,0 \pm 0,5$

B.2 Resultados da validação em linha de produção (Grupos 2 e 3)

Tabela B.5: Percentagem de rolhas TCA-positivas nos G2 dos vários algoritmos

	Número de rolhas TCA-positivas	Percentagem de rolhas TCA-positivas/%	Total de rolhas no grupo
Ellution	7	14	50
NDfitM	19	54	35
NDfitA	9	64	14

Tabela B.6: Percentagem de rolhas TCA-positiva e TCA-negativa nos G3 dos vários algoritmos

	Número de rolhas TCA-positiva	Percentagem de rolhas TCA-positiva/%	Total de rolhas no grupo
Ellution	5	50	10
NDfitM	2	33	6
NDfitA	0	N.A.	0





2022 GO RIGHT INTO GRADUATION

2022 GO RIGHT INTO GRADUATION