

Universidade de São Paulo
Instituto de Matemática e Estatística
Bacharelado em Ciência da Computação

Isabela Blücher

**Filtragem de ruído telúrico
em sinais astronômicos**

São Paulo
2019

Filtragem de ruído telúrico em sinais astronômicos

Monografia final da disciplina
MAC0499 – Trabalho de Formatura Supervisionado.

Supervisora: Profa. Dra. Paula Rodrigues Teixeira Coelho (IAG-USP)

Co-supervisor: Prof. Dr. Marcelo Gomes de Queiroz (IME-USP)

São Paulo
2019

Agradecimentos

À Universidade de São Paulo e ao Instituto de Matemática e Estatística pela oportunidade de cursar o Bacharelado em Ciência da Computação e por proporcionar um ambiente questionador e criativo para o desenvolvimento de pesquisas acadêmicas.

Aos meus orientadores Profa. Dra. Paula Rodrigues Teixeira Coelho e Prof. Dr. Marcelo Gomes de Queiroz pela oportunidade de elaborar um trabalho que mescla minhas áreas preferidas do conhecimento e pelo paciente trabalho de ensino e revisão que fizeram comigo ao longo do ano.

À Profa. Dra. Nina Sumiko Tomita Hirata, por coordenar a disciplina MAC0499 e pelo apoio no meu envolvimento com a Astronomia e o IAG-USP durante a graduação.

À minha família pelo apoio incondicional durante todas as fases da minha vida e por investirem e acreditarem tanto na minha educação.

Aos amigos que me acompanharam de perto nesta jornada, por sempre estarem presentes nos momentos difíceis.

A todos que direta ou indiretamente fizeram parte da minha formação, o meu muito obrigada.

Resumo

Na área da astronomia observacional, um dos principais dados usados em pesquisa são os espectros estelares, devido à grande quantidade de informação contida neles. A maioria das observações estelares capturadas por espectrógrafos são feitas a partir do solo e estão sujeitas à interferência da atmosfera, um fenômeno denominado contaminação telúrica. Este trabalho usa dados estelares reais e dados telúricos sintéticos para testar o método de correção telúrica atual e usa o algoritmo *Dynamic Time Warping* como uma abordagem para propor uma melhoria na correção dos espectros. Os resultados mostram que o algoritmo não funciona idealmente quando aplicados nos espectros estelar e telúrico inteiros, mas que poderia ser aplicado em intervalos menores contendo linhas de absorção bem separadas. Além disso, o texto aponta para várias questões interessantes de pesquisa futura usando métodos computacionais em astronomia observacional.

Palavras-chave: espectroscopia astronômica, contaminação telúrica, processamento de sinalis digitais, realinhamento de sinalis.

Abstract

In the context of observational astronomy, one of the main data used in research are stellar spectra, due to the amount of information they contain about observed stars. Most stellar observations are captured by ground-based spectrographs and therefore are subject to atmospheric interference, which is commonly called telluric contamination. This work uses real stellar data and synthetic telluric data to test the currently used method for telluric correction, and uses the Dynamic Time Warping algorithm as an approach to potentially improve the current spectral correction. The results show that the algorithm does not perform optimally when applied to the full stellar and telluric spectra, but could potentially be used on smaller regions containing well separated absorption lines. Besides, the text points to several interesting questions for future research using computational methods in observational astronomy.

Keywords: astronomical spectroscopy, telluric contamination, digital signal processing, signal realignment.

Sumário

1	Introdução	1
1.1	Contextualização	1
1.2	Estrutura do Trabalho	2
2	Fundamentação Teórica	3
2.1	Espectroscopia Astronômica	3
2.2	Redução de Dados	6
2.3	Contaminação Telúrica	7
2.4	Similaridade e alinhamento de sinais	10
3	Desenvolvimento Experimental	19
3.1	Formato de dados	19
3.2	Ambiente e ferramentas	20
3.3	Conjuntos de dados	21
3.4	Divisão do sinal estelar pelo telúrico	25
3.5	DTW no sinal estelar	27
3.6	DTW no sinal atmosférico	31
4	Conclusão	35
4.1	Fechamento	35
4.2	Trabalhos futuros	36
	Referências Bibliográficas	37

Capítulo 1

Introdução

A Computação é a área do conhecimento que estuda técnicas, metodologias e instrumentos computacionais com o objetivo de automatizar processos e desenvolver soluções para uma variedade de problemas. Um dos grandes usos da Computação atualmente é como ferramenta no avanço de pesquisas científicas em diversas áreas, por exemplo em ciências naturais, nas áreas de física, biologia, geografia e etc.

A ênfase deste trabalho é na aplicação de métodos computacionais em Astronomia. Com o passar do tempo e com uma constante evolução da tecnologia, as ferramentas astronômicas mudaram, influenciando desde como o universo e os objetos celestes são observados até como estes dados observacionais são processados e armazenados. Atualmente existem diversos desafios em astronomia que envolvem áreas da computação como aprendizado de máquina, ciência de dados, processamento de sinais digitais e *Big Data*.

O recorte específico da computação na astronomia tratado nesta monografia é referente à espectroscopia estelar e à remoção de contaminação telúrica nestes dados, ou seja, a limpar o sinal estelar de contaminações provenientes da atmosfera terrestre. Apresentamos os métodos usados atualmente para fazer a correção telúrica dos sinais e experimentos que testam métodos de processamento de sinais digitais de forma a melhorar a qualidade desta correção.

1.1 Contextualização

A astronomia pode ser dividida em duas grandes áreas de estudo: a astronomia teórica e a astronomia observacional. Em contraste com a astronomia teórica, que tem como foco a modelagem de fenômenos astronômicos e o cálculo de suas implicações no universo, a astronomia observacional tem como objetivo a observação e o registro de dados sobre o universo observável e o estudo e interpretação destas observações.

Dentro do cenário da astronomia observacional, é estudada uma grande variedade de objetos celestes através de suas fontes de informação: a emissão de luz ou radiação eletrromagnética. Para observar e capturar esta radiação, astrônomos usam uma variedade de instrumentos que possuem a capacidade de transformar um sinal luminoso em informação

radiativa cientificamente interpretável.

Um exemplo de instrumento observacional muito usado em pesquisas astronômicas é o espectrógrafo, que transforma uma observação celeste em um espectro, através da separação da radiação eletromagnética emitida em seus diversos comprimentos de onda. Em termos computacionais, um espectro é um sinal unidimensional que representa o fluxo de energia emitido em função do comprimento de onda. O espectro de um objeto celeste, e no caso deste trabalho, das estrelas, é uma informação muito relevante para a astronomia observational, pois está relacionado diretamente a diversas características sobre a estrela, como sua composição química, temperatura e luminosidade.

No entanto, a observação de espectros estelares não é um processo perfeito e o sinal capturado pode sofrer distorções tanto instrumentais quanto físicas. A radiação eletromagnética emitida por uma estrela, quando entra em contato com a atmosfera terrestre interage com as moléculas nela presentes, o que resulta em um espectro contaminado pela atmosfera terrestre. Esta interação da atmosfera com a emissão estelar é chamada de contaminação telúrica e é um fenômeno que interfere com todas as observações estelares feitas a partir do solo terrestre.

O objetivo geral deste trabalho consiste em estudar métodos existentes usados por astrônomos para corrigir espectros estelares de contaminação telúrica e investigar métodos computacionais com o potencial de auxiliar este processo de correção.

1.2 Estrutura do Trabalho

O capítulo 2 apresenta a base teórica astronômica por trás da espectroscopia estelar e seu sistema de classificação, o processo de captura e redução de dados e a ciência da contaminação telúrica de espectros estelares. Além disso, são apresentados conceitos e algoritmos associados a similaridade e alinhamento de sinais digitais usados neste campo.

O capítulo 3 apresenta especificações sobre o formato de dados utilizado para representar os espectros estelares e as ferramentas computacionais utilizadas para a leitura e manipulação destes dados, bem como os conjuntos de dados públicos ou fornecidos por comunicação particular nos quais se basearam os experimentos. Ainda nesse capítulo são descritos os experimentos realizados durante o desenvolvimento deste trabalho e seus resultados.

O capítulo 4 traz algumas conclusões sobre a pesquisa realizada e as dificuldades encontradas, confronta os resultados com os objetivos iniciais e expõe potenciais desenvolvimentos futuros para este trabalho.

Capítulo 2

Fundamentação Teórica

O problema de remoção de ruído telúrico em sinais astronômicos é de natureza interdisciplinar, o que implica na necessidade de compreender tanto aspectos da teoria astronômica de espectros estelares, quanto aspectos relativos às representações e manipulações computacionais relacionadas ao problema. Neste capítulo são descritos conceitos básicos em espectroscopia astronômica, principalmente estelar, e algoritmos de processamento de sinais digitais utilizados na pesquisa, que em conjunto caracterizam o problema da contaminação telúrica de espectros estelares e sua possível solução através de técnicas de filtragem.

2.1 Espectroscopia Astronômica

Estrelas são corpos celestes que consistem de esferas de gás ionizado que são mantidas íntegras pela sua gravidade. Uma fonte fundamental de energia das estrelas são as reações nucleares, principalmente a fusão de hidrogênio em hélio. A energia resultante desta fusão nuclear é emitida para o espaço em grande parte como radiação eletromagnética (Kepler e Saraiva, 2014).

A radiação da estrela se propaga pelo meio interestelar como um fluxo de partículas denominadas fôtons. Quando no espaço, estes fôtons atingem o limite de velocidade universal, a velocidade da luz $c \approx 3 \times 10^8$ m/s. Os fôtons se aproximam da Terra nesta velocidade, mas assim que entram em contato com a atmosfera terrestre começam a interagir com moléculas como o oxigênio e o vapor de água (Wikipedia contributors, 2019a). Estas moléculas de gás podem absorver os fôtons do sinal estelar, e consequentemente, alterar a quantidade de informação emitida por uma estrela que alcança a superfície terrestre (Wikipedia contributors, 2019b).

Os fôtons que alcançam a superfície terrestre são medidos instrumentalmente, e essa captura de sinal é fundamental para estudos em astronomia observacional. Um dos instrumentos utilizados para fazer observações a partir do solo são os espectrógrafos, presentes em muitos telescópios.

Um espectrógrafo tem a capacidade de dividir a radiação eletromagnética de um objeto

celeste em seus respectivos comprimentos de onda, resultando em um espectro: um mapa de radiação em função do comprimento de onda. Na astronomia, diversos corpos celestes podem ser o objeto de estudo de observações espetrais, como estrelas, planetas, nebulosas, galáxias e núcleos galácticos ativos.

Essa variedade de objetos celestes produz diferentes espetros, que podem ser divididos em três tipos principais: contínuo, de absorção e de emissão. Um espetro contínuo é uma função que tem como domínio todos os comprimentos de ondas do espetro eletromagnético. Ele é gerado pela observação de um corpo opaco quente, que seria o equivalente a observar diretamente o núcleo de uma estrela sem intervenção de matéria entre estrela e observador. Um espetro de absorção, ou espetro de linhas escuras, é gerado por um gás transparente e comparativamente mais frio em frente ao corpo opaco quente, e está associado à absorção dos fótons da radiação estelar em determinados comprimentos de onda. E por último, o espetro de emissão, ou espetro de linhas brilhantes, é gerado por um gás transparente que foi excitado por uma fonte de energia próxima, o que resulta na emissão de fótons de comprimentos de onda específicos. A figura 2.1 ilustra os três tipos de espetros.

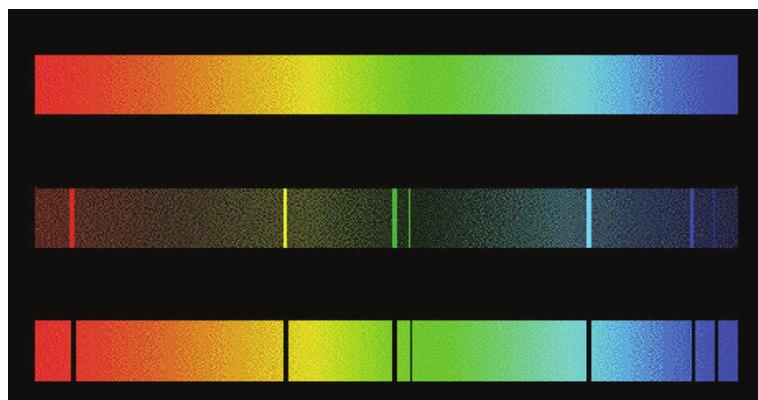


Figura 2.1: Os três tipos de espetro: contínuo, de emissão e de absorção (McLaughlin, 2013)

Dentre esta grande variedade de objetos celestes, as estrelas são especialmente estudadas pelos seus espetros. Por serem objetos quentes, rodeados de gases mais frios (as chamadas atmosferas estelares), estrelas emitem um espetro de linhas de absorção sobreposto a um espetro contínuo. As linhas de absorção têm comprimentos de onda característicos dos elementos químicos que geraram o espetro. A figura 2.2 ilustra o espetro da estrela Vega, cujo formato segue o de um espetro contínuo com a sobreposição de linhas de absorção de diversos elementos químicos.

No final do século XIX, astrônomos perceberam que a análise cuidadosa do espetro de uma estrela fornece uma riqueza de detalhes sobre ela, incluindo sua temperatura efetiva, velocidade de rotação, velocidade de translação, densidade, composição química e metalicidade. Com a observação rotineira de espetros estelares em grandes números, estes pesquisadores notaram que era possível agrupar as estrelas com base em suas características espetrais, e assim surgiram diversos sistemas de classificação estelar.

O sistema moderno de classificação de estrelas foi adotado em 1910 e foi criado por um

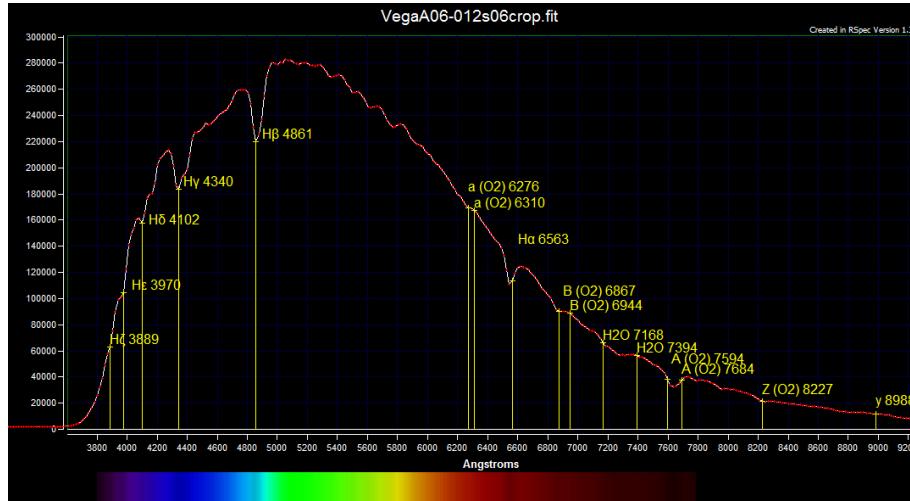


Figura 2.2: O espectro da estrela Vega(Haworth, 2011)

time do observatório da Universidade de Harvard. Este sistema baseia-se nas intensidades relativas das linhas de absorção presentes no espectro. As variações nas linhas espectrais para diferentes estrelas são devidas principalmente à diferença de temperatura das camadas externas de gás na estrela, logo o sistema de classificação espectral é baseado na temperatura efetiva da estrela.

As classes espectrais do sistema, em ordem decrescente de temperatura efetiva da estrela são O, B, A, F, G, K e M, como pode ser visto na tabela 2.1. Cada uma dessas classes se divide em 10 subclasses (indicadas pelos números de 0 a 9), sendo 0 a mais quente dentro da classe e 9 a mais fria. Na figura 2.3 é possível observar como diferentes linhas de absorção e bandas moleculares variam em intensidade conforme muda a temperatura efetiva, ou tipo espectral, da estrela.

Tipo Espectral	Temperatura efetiva (K)
O	30.000
B	20.000
A	10.000
F	7000
G	6000
K	4000
M	3000

Tabela 2.1: A temperatura superficial estelar para cada tipo espectral (Hetem e Pereira, 2010)

Os tipos espectrais das estrelas continuam sendo uma parte essencial para a pesquisa em astronomia, seja para auxiliar na descoberta de exoplanetas ou na interpretação da história da evolução das galáxias (Charbonneau *et al.*, 2002).

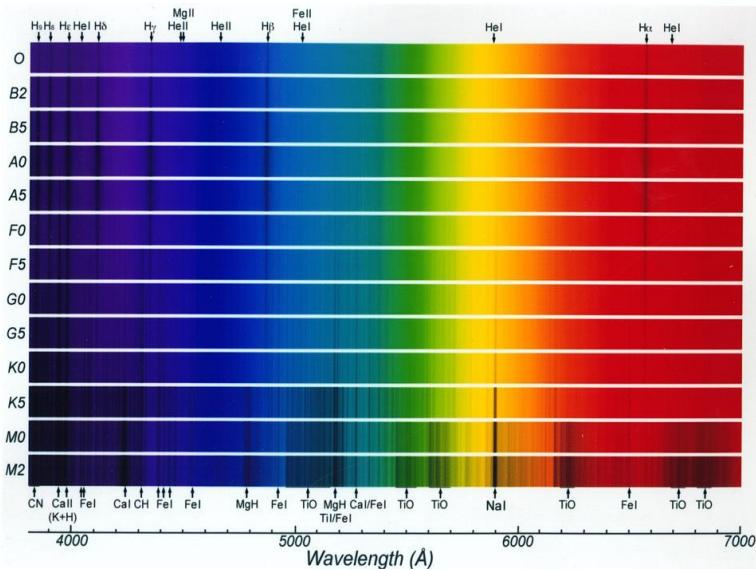


Figura 2.3: Tipos de espectro estelar (Burrows, 2012)

2.2 Redução de Dados

Pesquisas na área de espectroscopia astronômica são dependentes de uma sólida estrutura instrumental. Isto significa que para coletar dados de objetos celestes é necessário ter acesso a instrumentos de observação, equipamentos de *hardware* sofisticado que são capazes de transformar a radiação eletromagnética que alcança a superfície terrestre em dados cientificamente interpretáveis.

Estes instrumentos de observação muitas vezes estão munidos de um CCD ou *charge-coupled device*, um sensor eletrônico constituído por vários quadrados fotossensíveis, que registram o sinal em pixels de uma imagem. CCDs respondem linearmente ao sinal que os atinge e registram uma imagem da região do céu que está sendo observada. A chegada de cada fóton em um de seus pixels gera uma pequena carga elétrica armazenada para leitura posterior, e esta aumenta de forma proporcional à quantidade de fótons que atingem o aparelho (Davenhall *et al.*, 2001).

Os benefícios de se usar um CCD para capturar imagens celestes é que eles automaticamente fornecem uma imagem digitalizada capaz de ser exibida e processada por um computador. Porém, estes dados brutos contêm uma série de assinaturas instrumentais que devem ser eliminadas antes que o dado possa ser usado para fins científicos (Davenhall *et al.*, 2001).

O processo de corrigir as imagens capturadas pelo CCD de suas assinaturas instrumentais é chamado de **redução de dados** pela comunidade astronômica, e consiste de uma série de transformações aplicadas à imagem bruta. No final deste processo é obtida uma imagem científica utilizável e pronta para ser analisada.

No caso dos espectros estelares, a redução de dados implica na extração de um espectro unidimensional a partir de uma imagem de CCD da estrela observada, também chamada de estrela de ciência. A figura 2.4, ilustra as imagens de CCD antes e depois do processo de

redução (painéis à esquerda e ao meio, respectivamente), e as representações dos espectros extraídos (painéis à direita).

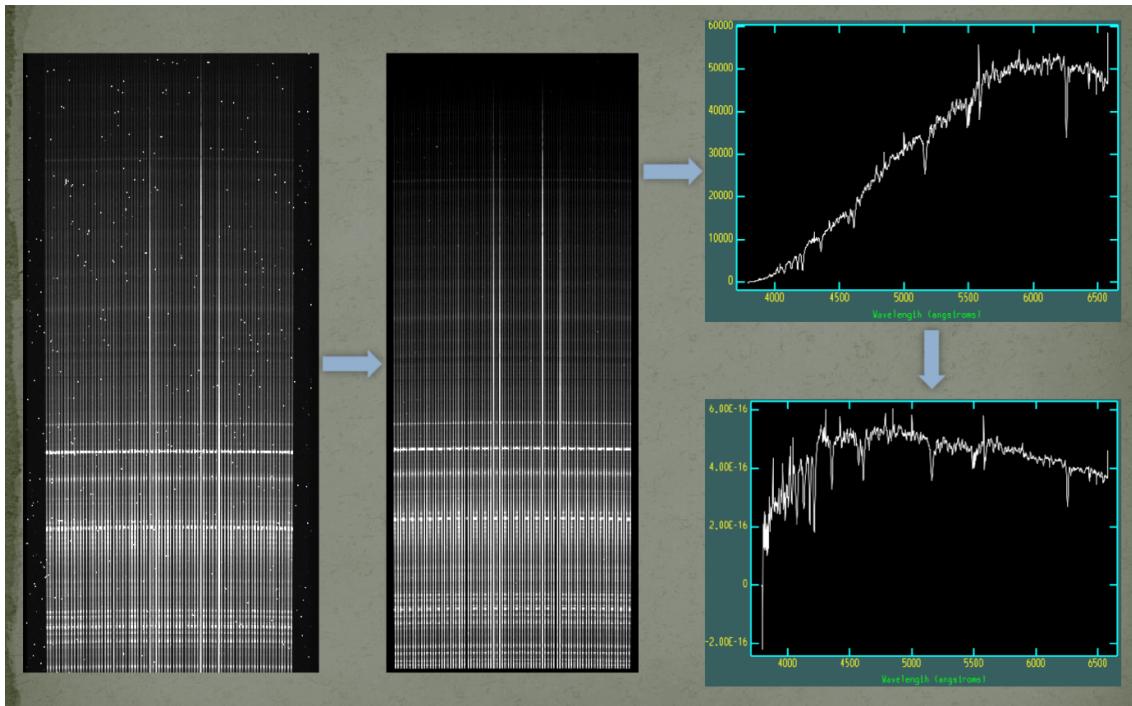


Figura 2.4: Redução de dados de uma imagem de CCD a um espectro unidimensional (Moravec, 2004)

O espectro resultante deste processo é unidimensional no domínio do comprimento de onda da estrela e tem sua intensidade representada por um fluxo, cuja unidade é $\text{erg}/\text{cm}^2/\text{s}/\text{\AA}$, que representa a intensidade de energia por unidade de área, tempo e comprimento de onda (Djorgovski, 2004).

2.3 Contaminação Telúrica

Como mencionado na seção 2.1, a maioria das observações astronômicas é feita a partir do solo e, neste caso, nem toda a luz irradiada por uma estrela consegue ser capturada por instrumentos de observação.

A principal consequência da interação da radiação eletromagnética estelar com a atmosfera terrestre é que novas linhas espectrais são formadas no espectro capturado instrumentalmente. As moléculas presentes na atmosfera terrestre formam linhas de absorção no espectro observado, denominadas linhas telúricas.

No espectro estelar, regiões de diferentes comprimentos de onda possuem diferentes sensibilidades à atmosfera. Isso implica que em certas regiões do espectro de ciência existe uma maior presença das linhas telúricas do que em outras. Na figura 2.5, é possível ver estas linhas telúricas e suas respectivas moléculas responsáveis em uma grande extensão de comprimentos de onda.

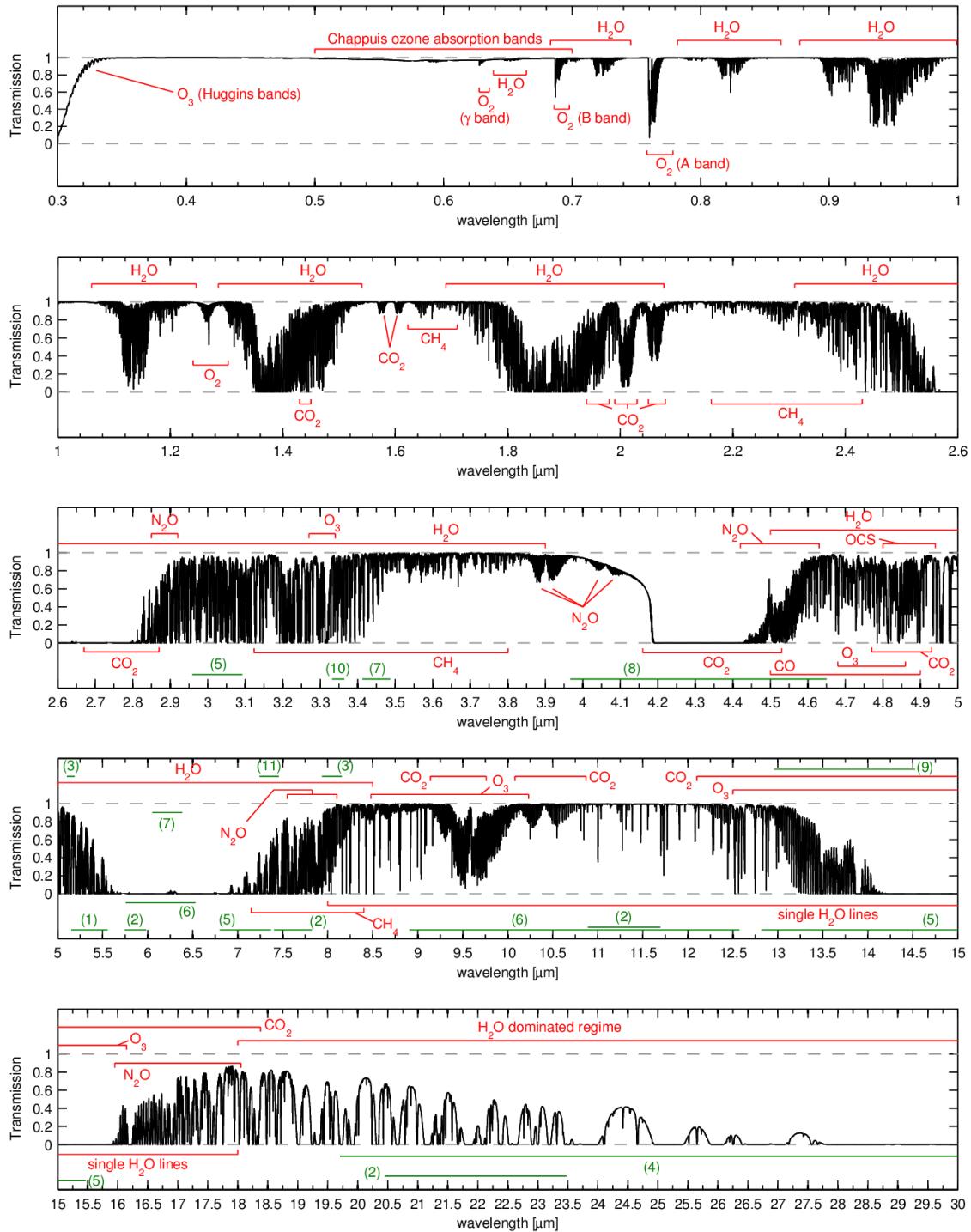


Figura 2.5: Espectro de absorção sintético criado pelo Molecfit que ilustra como diferentes moléculas interagem com diferentes regiões de comprimento de onda (Smette et al., 2015)

As linhas telúricas contribuem para a criação de uma observação distorcida ou contaminada. A menos que seja corrigida, esta contaminação pode produzir erros de interpretação e ruídos que reduzem a precisão dos dados observados, e consequentemente, dificultam o avanço de diversas pesquisas astronômicas.

Existem alguns procedimentos típicos usados hoje em dia para remover as linhas telúricas de um espectro de ciência. Um método popular consiste em aplicar uma divisão simples

entre o espectro da observação e uma referência telúrica, que consiste em uma função representando a transmissão radiativa da atmosfera terrestre para cada comprimento de onda, desde que satisfeita a condição de que o referencial telúrico tenha as mesmas assinaturas instrumentais do espectro de ciência. Esta referência telúrica pode ser obtida de duas maneiras: através da observação de uma estrela padrão, ou pela simulação de um espectro atmosférico teórico, conforme detalhado a seguir.

As estrelas padrão são estrelas quentes em rotação rápida de tipo espectral B ou A. Elas são escolhidas pois seus espectros não possuem características marcantes além de fortes linhas de hidrogênio. Para que o espectro de uma estrela padrão seja usado como uma referência telúrica, é necessário observá-la em condições muito próximas (tempo, posição no céu e condições atmosféricas) daquelas referentes ao espectro de ciência. Contudo, existem várias limitações fundamentais no nível de correção que pode ser obtido com o método da estrela padrão. Primeiramente, outras características estelares além das linhas de hidrogênio ainda podem estar presentes no espectro da estrela padrão, evidenciando que ela não é um modelo preciso da transmissão radiativa da atmosfera. Além disso, o tempo e a direção de observação das duas estrelas nunca será exatamente o mesmo, o que significa que a assinatura atmosférica de ambos os espectros também será necessariamente diferente.

Devido às limitações no uso das estrelas padrão, foi criado um método mais sofisticado e automatizado para produzir uma referência telúrica: softwares de simulação do espectro atmosférico. Estes softwares utilizam modelos de transmissão radiativa da atmosfera para criar um espectro telúrico sintético, cujas condições de simulação são muito próximas às condições de observação da estrela de ciência. O modelo de transmissão atmosférica mais utilizado hoje em dia é o *Line-By-Line Radiative Transfer Model (LBLRTM)* (Clough *et al.*, 2005), que fornece cálculos de radiância espectral com precisão e eficiência. Este modelo utiliza o HITRAN (Rothman *et al.*, 2009), um acrônimo para *High-resolution Transmission molecular absorption database*, uma base de dados de linhas e parâmetros espectroscópicos.

Um exemplo de software de simulação do espectro atmosférico é o **Molecfit** (Smette *et al.*, 2015), uma ferramenta para a modelagem de linhas telúricas desenvolvida por astrônomos do *European Southern Observatory*¹. De acordo com seus criadores, o **Molecfit** recupera o perfil atmosférico mais compatível com o instante da observação da estrela de ciência. Isto envolve a utilização de um modelo de transmissão radiativa da atmosfera com uma base de dados de condições atmosféricas para recuperar atributos como a variação na temperatura, pressão e umidade em função da altitude da observação.

Independentemente de como é gerado o referencial telúrico, considera-se que ele representa coeficientes de transmissão da radiação através da atmosfera, no sentido que, para cada comprimento de onda, a quantidade de radiação propagada desde a estrela que consegue atravessar a atmosfera e chegar ao instrumento é dada em percentual, representado pelo coeficiente de transmissão².

¹<https://www.eso.org/public/>

²Isso está de acordo com o princípio da superposição e equivale à hipótese de linearidade do processo de

Em termos matemáticos, consideramos os espectros de ciência, telúrico e estelar sem contaminação como vetores o (observado), a (atmosfera) e s (estrela),

$$o = (o_1, o_2, \dots, o_n) \quad a = (a_1, a_2, \dots, a_n) \quad s = (s_1, s_2, \dots, s_n),$$

representando valores discretos do fluxo estelar ou atmosférico em n comprimentos de onda específicos (geralmente equi-espaçados). A contaminação telúrica do espectro de ciência se refere então à equação

$$o = a \circ s \quad (\text{ou equivalentemente: } o_i = a_i \cdot s_i, \forall i),$$

onde o símbolo \circ representa o produto de Hadamard entre o espectro estelar original e o referencial telúrico, e a equação pressupõe que as três sequências estejam perfeitamente alinhadas, ou seja, que cada índice i represente o mesmo comprimento de onda nos três espectros. Esse é um possível ponto de partida para o tratamento dos espectros estelares com técnicas computacionais associadas ao processamento de sinais digitais, abordadas a seguir.

2.4 Similaridade e alinhamento de sinais

O processamento de sinais digitais tem como foco extrair informações significativas dos sinais, bem como modificá-los a fim de atingir determinados objetivos. Uma das operações importantes na análise de sinais digitais é a mensuração da similaridade entre dois sinais. Medidas de similaridade possuem diversas aplicações em áreas como reconhecimento de fala, para saber se um sinal de voz corresponde a um determinado texto, ou em economia, para comparar investimentos em períodos diferentes, entre outras aplicações.

Como foi mencionado na seção 2.3, é possível assumir que o problema da contaminação telúrica pode ser operacionalizado como uma multiplicação simples partindo do pressuposto de que os sinais se encontram em uma situação ideal. Estabelece-se como uma situação ideal os sinais que: (i) possuem mesma intensidade, ou seja, que foram observados com as mesmas condições atmosféricas (pressão, temperatura, umidade, turbulência etc); (ii) possuem mesma assinatura instrumental; (iii) são amostrados nos mesmos comprimentos de onda, ou seja, estão perfeitamente alinhados pixel a pixel.

Nesse trabalho, optamos por estudar a questão do alinhamento do sinal, motivados por uma dificuldade real que foi encontrada por um grupo de pesquisadores envolvidos na *X-Shooter Spectral Library* (Chen *et al.*, 2014, XSL). A XSL é uma coleção de estrelas observada pelo espetrógrafo de resolução média *X-Shooter*³, do ESO. Essa biblioteca de estrelas tem diversas aplicações em Astronomia, desde o estudo de abundâncias químicas de estrelas

transmissão radiativa no vácuo, sob a qual a transmissão pode ser modelada independentemente para cada frequência ou comprimento de onda.

³<https://www.eso.org/sci/facilities/paranal/instruments/xshooter.html>

da nossa Galáxia até a modelagem evolutiva de outras galáxias. Pesquisadores que usaram estes dados para suas pesquisas relataram desalinhamentos nos comprimentos de onda dos espectros (Gonneau *et al.*; Moehler, 2015), que podem ser observados na figura 2.6.

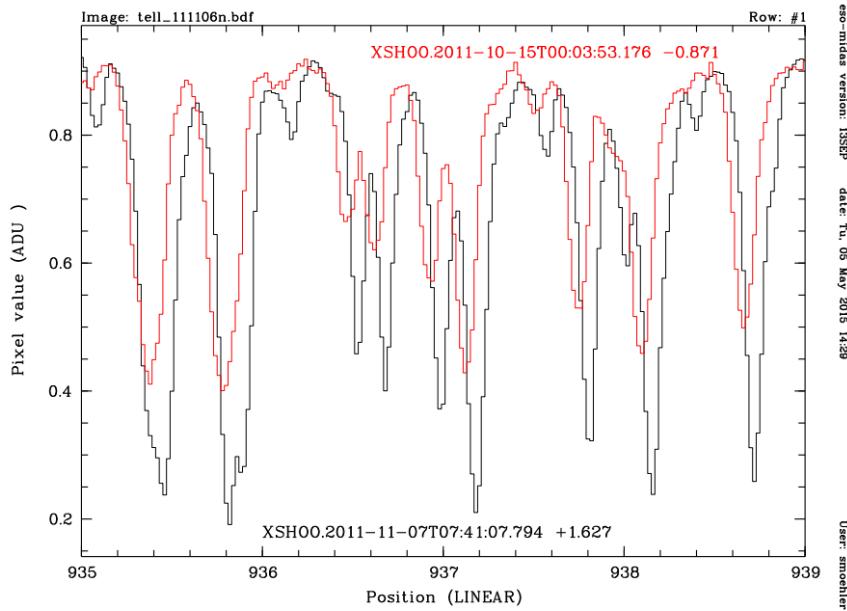


Figura 2.6: Desalinhamentos observados nos espectros telúricos do XSL (Moehler, 2015)

Diante disso, é possível entender porque o estudo da similaridade entre sinais pode ser útil em espectros estelares: para alinhar regiões de contaminação atmosférica de moléculas específicas em um espectro de ciência com as regiões correspondentes no espectro telúrico. Em teoria, isto melhoraria o resultado da divisão entre o espectro de ciência observado e o telúrico, e consequentemente, a qualidade do espectro estelar sem contaminação.

Uma medida tradicional de diferença entre informações numéricas que poderia ser usada como medida de similaridade entre dois sinais digitais é a distância euclidiana entre os vetores correspondentes. Porém em sinais com flutuações rápidas é fácil perceber que mesmo um pequeno desalinhamento poderia provocar um grande aumento na distância entre os sinais.

Um algoritmo robusto utilizado para determinar a similaridade entre dois sinais digitais desalinhados é o *Dynamic Time Warping* (DTW). Este algoritmo acha uma correspondência ótima entre os índices de dois vetores, de tal maneira a minimizar as distâncias entre os índices correspondentes. Essa correspondência equivale a um alinhamento entre as sequências que torna os sinais realinhados mais similares (Shou *et al.*, 2005).

A figura 2.7 ilustra o alinhamento ótimo encontrado pelo DTW entre um sinal cuja frequência aumenta com o tempo e uma sinusoide. O alinhamento encontrado corresponde à distância euclidiana mínima entre as duas sequências, de forma que regiões com formatos similares estejam o mais próximas o possível no alinhamento.

A formulação do problema (Salvador e Chan, 2007) do DTW é a seguinte: dadas duas

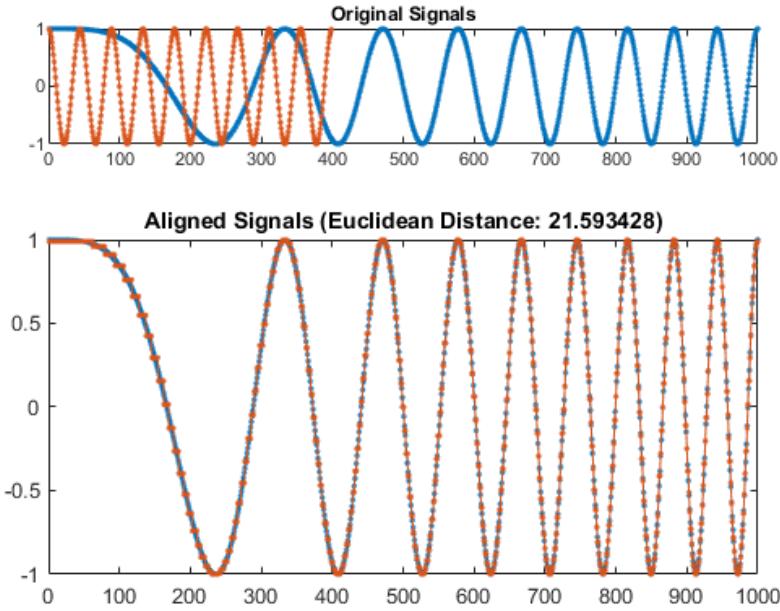


Figura 2.7: Sinais originais e alinhados com a DTW com a distância euclideana final (MATLAB)

sequências X e Y , de comprimentos respectivos n e m ,

$$X = (x_1, x_2, \dots, x_n)$$

$$Y = (y_1, y_2, \dots, y_m),$$

construa um caminho W de comprimento k

$$W = (w_1, w_2, \dots, w_k), \quad \max(n, m) \leq k < n + m,$$

que satisfaça as seguintes condições:

- o k -ésimo elemento do caminho W é $w_k = (i_k, j_k)$, onde i_k, j_k são índices das sequências X e Y , respectivamente;
- o caminho W deve começar em $w_1 = (1, 1)$ e deve terminar em $w_k = (n, m)$, de forma a assegurar que o alinhamento considera as duas sequências completas;
- o caminho W força i_k e j_k a serem monótonos não-decrescentes como funções de k ; em termos matemáticos isso equivale a

$$i_k \leq i_{k+1} \leq i_k + 1 \quad \text{e} \quad j_k \leq j_{k+1} \leq j_k + 1,$$

garantindo que todos os índices das duas sequências serão usados;

- o caminho W minimiza a função

$$\text{dist}(W) = \sum_{n=1}^k \text{dist}(x_{i_n}, y_{j_n})$$

onde $\text{dist}(x_{i_n}, y_{j_n})$ é uma medida de distância, tipicamente a euclídea, entre um par de elementos das sequências X e Y .

As três primeiras condições definem uma enorme quantidade de caminhos/alinhamentos possíveis, ao passo que a quarta condição estabelece um critério de comparação entre dois alinhamentos distintos.

A implementação do algoritmo utiliza o paradigma da programação dinâmica, onde cada instância do problema é resolvida combinando soluções de subproblemas da instância original. O diferencial desta abordagem é que os resultados dos subproblemas são armazenados em uma tabela ou matriz. No caso da DTW é criada uma matriz de distâncias acumuladas D , de dimensão $n \times m$, onde $D(i, j)$ armazena o valor da distância total do caminho mínimo que pode ser construído entre as subsequências $X' = \{x_1, \dots, x_i\}$ e $Y' = \{y_1, \dots, y_j\}$. Logo, o valor armazenado em $D(n, m)$ representa a distância do caminho mínimo que comprehende as duas sequências inteiras. O preenchimento da matriz D é feito da seguinte forma:

$$D(i, j) = \begin{cases} \sum_{k=1}^j \text{dist}(x_1, y_k) & \text{se } i = 1 \\ \sum_{k=1}^i \text{dist}(x_k, y_1) & \text{se } j = 1 \\ \text{dist}(x_i, y_j) + \min\{D(i-1, j), D(i, j-1), D(i-1, j-1)\} & \text{caso contrário} \end{cases}$$

Esta abordagem funciona pois, ao se calcular $D(i, j)$, todos os caminhos ótimos de subsequências menores do que as sequências de tamanhos i e j já foram considerados, logo, a distância armazenada em $D(i, j)$ corresponde à distância mínima entre todos os caminhos possíveis para subsequências com apenas um índice de diferença em relação a i ou j (ou ambos), mais a distância entre o par x_i e y_j . A inicialização da linha $i = 1$ e da coluna $j = 1$ explora o fato de que para alinhar uma subsequência qualquer de X (ou Y) com apenas um elemento do outro vetor, o único caminho possível é aquele que alinha todos os elementos da subsequência ao único elemento possível.

Depois que a matriz D está completa, é necessário reconstruir um caminho ótimo entre $D(1, 1)$ e $D(n, m)$, pois a matriz só traz esse caminho de forma implícita, através dos mínimos que foram considerados no cálculo de $D(i, j)$. Essa reconstrução é feita na ordem inversa: a partir do final do caminho $D(n, m)$, consideram-se os elementos à esquerda $D(n, m - 1)$, abaixo $D(n - 1, m)$ e na diagonal inferior esquerda $D(n - 1, m - 1)$; dentre estes, o elemento com o menor valor é adicionado ao caminho ótimo como antecessor de $D(n, m)$. Este procedimento é repetido até que o caminho chegue na posição $D(1, 1)$.

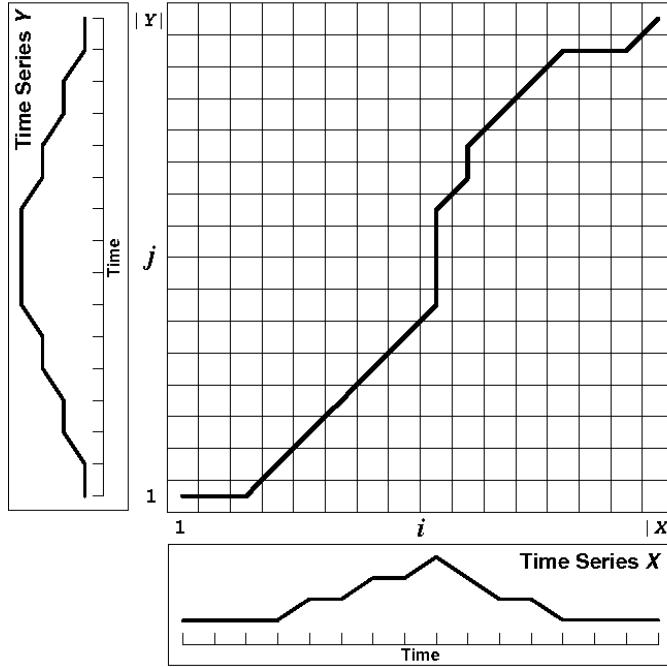


Figura 2.8: Uma matriz de similaridade com o caminho ótimo da DTW entre as séries X e Y (Salvador e Chan, 2007)

O pseudo-código dos procedimentos de preenchimento da matriz (Fong, 2012) e de *backtracking* que compõem a DTW são dados pelos algoritmos 1 e 2.

É possível analisar a complexidade de tempo e espaço da DTW com o algoritmo de preenchimento da matriz de similaridade. Cada elemento da matriz $n \times m$ é escrito exatamente uma vez em tempo constante, e a matriz inteira é armazenada na memória. Logo, ambas as complexidades são $\mathcal{O}(nm)$, que no caso de $n = m$, equivale a uma complexidade assintótica $\mathcal{O}(n^2)$. A complexidade de espaço é especialmente proibitiva, visto que sequências obtidas de espectros estelares com apenas 177 mil elementos geram uma matriz de similaridades que exige *terabytes* de memória para ser armazenada.

A complexidade quadrática tanto em tempo quanto em espaço da DTW cria a demanda por métodos que consigam acelerar este procedimento. No artigo de Salvador e Chan (2007), é proposta uma abordagem multinível inspirada no problema da partição de grafos. Essa técnica tende a funcionar bem para problemas muito grandes e de difícil solução exata. Neste caso, soluções de problemas menores ou parciais podem ser refinadas e combinadas para que se encontre uma boa aproximação da solução do problema original.

O algoritmo proposto, chamado de *FastDTW*, depende de algumas operações-chave: a redução das séries temporais, a projeção do caminho da DTW para maiores resoluções e o refinamento do caminho encontrado. A redução das séries temporais é feita tirando a média de pontos adjacentes e o resultado são vetores com a metade do seu comprimento original. Repete-se esta operação várias vezes para se criar todas as resoluções das séries temporais que serão usadas, de maneira similar à transformada Wavelet. O algoritmo quadrático padrão da DTW é utilizado para as séries temporais com menor resolução e o caminho ótimo encontrado

Algorithm 1 DTW matrix filling

```

1: procedure DTW( $x, y$ )                                ▷ Seja  $M$  uma matriz bidimensional  $n \times m$ 
2:    $M[0, 0] \leftarrow 0$ 
3:   for  $i \leftarrow 1$  to  $m$  do
4:      $M[0, i] \leftarrow \infty$ 
5:   end for
6:   for  $i \leftarrow 1$  to  $n$  do
7:      $M[i, 0] \leftarrow \infty$ 
8:   end for
9:   for  $i \leftarrow 1$  to  $n$  do
10:    for  $j \leftarrow 1$  to  $m$  do
11:       $d = \text{dist}(x[i], y[j])$ 
12:       $M[i, j] = d + \min(M[i - 1, j], M[i, j - 1], M[i - 1, j - 1])$ 
13:    end for
14:  end for
15:  return  $M[n, m]$                                 ▷ Distância do caminho ótimo
16: end procedure

```

Algorithm 2 DTW path backtracking

```

1: procedure DTW_PATH( $M$ )                                ▷ Seja  $P$  uma lista de tuplas
2:    $i \leftarrow n$ 
3:    $j \leftarrow m$ 
4:   while  $i \neq 0$  and  $j \neq 0$  do
5:     P.add( $i, j$ )
6:     if  $M[i - 1, j - 1] \leq M[i - 1, j]$  and  $M[i - 1, j - 1] \leq M[i, j - 1]$  then
7:        $i \leftarrow i - 1$ 
8:        $j \leftarrow j - 1$ 
9:     else if  $M[i - 1, j] \leq M[i - 1, j - 1]$  and  $M[i - 1, j] \leq M[i, j - 1]$  then
10:       $i \leftarrow i - 1$ 
11:    else
12:       $j \leftarrow j - 1$ 
13:    end if
14:   end while
15:   return  $P$                                          ▷ Índices do caminho ótimo
16: end procedure

```

é então projetado para a próxima maior resolução que foi computada. Na projeção aplica-se uma versão restrita da DTW apenas na vizinhança próxima do caminho resultante da projeção, o que não garante que o caminho ótimo será encontrado. Para aumentar as chances de encontrar a solução ótima existe um parâmetro de raio r , que representa o número de células adjacentes ao caminho projetado que farão parte da DTW restrita. Este procedimento é repetido para todas as resoluções até se obter um caminho na resolução original.

O algoritmo usa uma estratégia recursiva, na qual o caso base é o uso da DTW quadrática padrão em uma resolução bastante baixa dos sinais. Uma análise da complexidade de tempo e espaço do pior caso em que ambas as séries temporais possuem n elementos resulta em $n(8r+14)$ para tempo e $n(4r+7)$ para espaço, ou seja, se o parâmetro r é um valor constante pequeno ($\ll n$), podemos afirmar que as complexidades assintóticas são $\mathcal{O}(n)$.

Como pode ser visto nos algoritmos 1 e 2, as saídas do DTW são a distância e o caminho ótimo entre duas sequências. Neste trabalho foi produzido um algoritmo que tem como objetivo produzir sequências alinhadas. Para isso é utilizado o caminho resultante da DTW e é feita uma aproximação do realinhamento de uma das sequências em relação à outra. O algoritmo implementado usa os índices repetidos no caminho da DTW e tira a média entre os elementos correspondentes para criar um valor representativo da correspondência de uma região maior de uma sequência com uma região menor de outra.

Em termos matemáticos temos as sequências a e o , definidas na seção 2.3, e queremos construir um \bar{a} que representa a sequência a realinhada em relação à sequência o , a partir de um caminho P obtido pelo DTW. Os caminhos construídos pelo algoritmo DTW podem associar cada índice de uma sequência a vários índices adjacentes da outra sequência. Para contornar essa questão na construção de \bar{a} , foi definido o seguinte procedimento: para cada índice $l = 0, \dots, n$ da sequência o , define-se

$$\bar{a}_l = \text{média}\{a_k\}_{k \in L}$$

onde

$$L = \{m | (m, l) \in P\}$$

e P representa o caminho resultante da DTW entre a e o . O pseudo-código correspondente a essa definição pode ser visto no algoritmo 3 abaixo.

Algorithm 3 Sequence alignment

```

1: procedure ALIGN_SEQUENCE_DTW_PATH( $P, s$ )  $\triangleright$  Seja  $P$  uma lista de tuplas e  $s$  o
   vetor a ser realinhado
2:    $d \leftarrow \emptyset$   $\triangleright$  Seja  $d$  uma associação chave-valor vazia
3:    $aligned \leftarrow \emptyset$ 
4:   for  $(i, j)$  in  $P$  do
5:     insere  $i$  na lista  $d[j]$ 
6:   end for
7:   for  $(j, lista)$  do
8:      $m \leftarrow$  média dos valores de  $s$  associados aos índices na lista  $d[j]$ 
9:      $aligned[j] \leftarrow m$ 
10:    end for
11:   return  $aligned$   $\triangleright$  Sequência realinhada
12: end procedure

```

Os procedimentos descritos acima serão explorados nos experimentos apresentados no próximo capítulo.

Capítulo 3

Desenvolvimento Experimental

Este trabalho tem o objetivo de explorar possíveis soluções computacionais que sejam capazes de remover ou atenuar a contaminação telúrica de espectros astronômicos capturados a partir do solo. Para fazer isso é necessário um conhecimento aprofundado dos dados coletados, desde os procedimentos adotados na aquisição e redução de dados até interpretações astronômicas de características morfológicas desses sinais.

Neste capítulo são descritos os dados utilizados e os experimentos realizados durante o desenvolvimento do trabalho. Os experimentos foram úteis para permitir a visualização e compreensão dos espectros de ciência e telúrico, o teste do realinhamento de sinais como possível coadjuvante na remoção da contaminação telúrica, e o aprendizado de características fundamentais dos espectros na solução do problema da contaminação telúrica.

3.1 Formato de dados

O formato de dados utilizado nos experimentos é o FITS ou *Flexible Image Transfer System* (Pence *et al.*, 2010). Este formato de arquivo digital facilita o armazenamento, processamento e transmissão de dados científicos e foi projetado para armazenar conjuntos de dados n -dimensionais que consistem em matrizes e tabelas. Este é o formato de dados mais utilizado na astronomia, possuindo recursos específicos para incluir informações de calibração fotométrica e espacial e outros metadados astronômicos relativos aos sinais armazenados.

Um arquivo FITS é composto por segmentos chamados de HDU (*Header-Data Units*) e ele deve necessariamente ter um *Primary HDU*, que armazena os dados científicos principais. O *PrimaryHDU* pode ser acompanhado de HDUs adicionais, categorizados entre extensões de imagens, extensões de tabelas ASCII e extensões de tabelas binárias (McGlynn, 2014).

O *header* de um arquivo FITS contém uma lista de palavras-chave em letra maiúscula, associadas a um valor e possivelmente seguidas de um comentário. Estas palavras-chave representam elementos dos dados que são importantes, como a data de observação do arquivo, seu autor, seu histórico de processamento, seu telescópio de origem e também podem representar elementos relacionados à física da observação, como a massa de ar no momento e

local de observação, temperatura ambiente, umidade e etc. A Fig. 3.1 ilustra o *header* típico de estrelas de ciência.

```

SIMPLE = T / Fits standard
BITPIX = -64 / Bits per pixel
NAXIS = 1 / Number of axes
NAXIS1 = 595493 / Axis length
EXTEND = T / File may contain extensions
ORIGIN = 'NOAO-IRAF FITS Image Kernel July 2003' / FITS file originator
IRAF-TLM= '2013-10-10T19:32:18' / Time of last modification
OBJECT = 'RED NONMERGED SCI POINT REDU' / Name of the object observed
COMMENT FITS®(Flexible Image Transport System) format is defined in 'Astronomy
COMMENT and Astrophysics', volume 376, page 359; bibcode: 2001A&A...376..359H
DATE = '2013-10-08T17:59:44' / file creation date (YYYY-MM-DDThh:mm:ss UT)
CTYPE1 = 'LINEAR '
BUNIT = 'ADU '
CRVAL1 = 3052.0000
CRPIX1 = 1.
CDELT1 = 0.012379642365
CUNIT1 = 'Angstrom' / Wavelength units
LAMRMS = 0.002193 / RMS of wavelength solution [CUNIT1]
LAMNLLIN = 262 / No. of lines used in wavelength solution
CRDER1 = 0.000135 / Wavelength uncertainty [CUNIT1]
CSYER1 = 0.000311 / Typical systematic wavelength error [CUNIT1]
TELESCOP= 'ESO-VLT-U2' / ESO Telescope Name
INSTRUME= 'UVES ' / Instrument used.
RA = 190.413633 / 12:41:39.2 RA (J2000) pointing (deg)
DEC = -1.44935 / -01:26:57.6 DEC (J2000) pointing (deg)
EQUINOX = 2000. / Standard FK5 (years)
RADECSYS= 'FK5 ' / Coordinate reference frame
MJD-OBS = 55677.07266342 / MJD start (2011-04-26T01:44:38.119)
DATE-OBS= '2011-04-26T01:44:38.119' / Date of observation
UTC = 6273. / 01:44:33.000 UTC at start (sec)
LST = 40669.384 / 11:17:49.384 LST at start (sec)
PI-COI = 'UNKNOWN ' / PI-COI name.
OBSERVER= 'UNKNOWN ' / Name of observer.
DATAMD5 = '7f14c9c9f9127e3ff7303f7254d41775' / MD5 checksum
PIPEFILE= 'red_nonmerged_science_redu.fits' / Filename of data product
AIRMASS = 1.164 / Airmass at start
IMAGETYP= 'OBJECT,POINT' / Observation type
UT = '01:44:33.000' / UT at start
ST = '11:17:49.383' / ST at start

```

Figura 3.1: Trecho do cabeçalho do arquivo FITS da estrela HD110379

Dentro de uma HDU em um arquivo FITS, o componente em sequência ao *header*, quando presente, possui um vetor que pode ter desde 1 até 999 dimensões. Esta seção representa o dado astronômico capturado, como os valores do fluxo de um espectro estelar ou uma imagem de um objeto celeste. Os dados podem ter um de 5 possíveis tipos de dados (Pence *et al.*, 2010):

1. Inteiros de 8 bits
2. Inteiros de 16 bits
3. Inteiros de 32 bits
4. Números reais de ponto flutuante de 32 bits
5. Números reais de ponto flutuante de 64 bits

3.2 Ambiente e ferramentas

A realização dos experimentos deste trabalho foi dependente de um conjunto de bibliotecas especializadas. Esta seção descreve as ferramentas utilizadas e detalhes da implementação dos experimentos.

A linguagem de programação de escolha para a implementação dos experimentos foi o *Python*, por esta permitir o uso de bibliotecas especializadas para o processamento de sinais e manipulação de dados astronômicos.

Os experimentos foram implementados em *notebooks* do *Jupyter*¹ (Kluyver *et al.*, 2016), um projeto de código-aberto com o objetivo de oferecer suporte à ciência de dados e computação científica interativa. Neste ambiente computacional é possível escrever código em células que podem ser executadas de forma independente. Esta interatividade faz com que os *notebooks* sejam ambientes perfeitos para uma programação de natureza exploratória, já que é possível executar experimentos e apresentar resultados e gráficos entre várias células de código.

Estabelecidos a linguagem e o ambiente dos experimentos, foram usadas diversas bibliotecas de *Python* para ler, manipular e visualizar os dados estelares. Os pacotes utilizados foram: o *NumPy*² (Oliphant, 2006), uma biblioteca eficiente para computação científica com suporte para vetores e matrizes grandes e multidimensionais, além de diversas funções matemáticas de alto nível; o *Matplotlib*³ (Hunter, 2007), uma biblioteca gráfica que permite a visualização de dados em diversos tipos de gráficos; e finalmente o *Astropy*⁴ (Price-Whelan *et al.*, 2018), uma biblioteca que é amplamente usada para processar computacionalmente diversos dados astronômicos, possuindo funcionalidades como leitura de arquivos FITS, conversões de unidades entre quantidades físicas e computações cosmológicas.

3.3 Conjuntos de dados

Para o desenvolvimento dos experimentos deste trabalho, foram obtidos alguns conjuntos de dados estelares. Todos os arquivos destes conjuntos estão no formato FITS. Estes conjuntos de dados estelares englobam tanto observações reais de estrelas quanto espectros telúricos sintéticos. A lista a seguir descreve estes conjuntos.

- UVES: dois pares de arquivos FITS contendo observações das estrelas HD110379 e HD186791 e seus respectivos referenciais telúricos. Estes dados foram capturados pelo espectrógrafo UVES⁵ (Dekker *et al.*, 2000), instalado no *Very Large Telescope Array*⁶, no Cerro Paranal, Chile. Os dados foram obtidos como parte do *proposal* de observação 087.B-0308(B), PI P. Coelho;
- FTS: uma tabela FITS contendo observações reais do Sol, da estrela Arcturus e um referencial telúrico para ambas (Hinkle *et al.*, 2000), observados pelo *Fourier Transform Spectrometer* instalado no telescópio KPNO no National Optical Astronomy Observatory⁷;

¹<https://jupyter.org/>

²<https://numpy.org/>

³<https://matplotlib.org/>

⁴<http://www.astropy.org>

⁵<https://www.eso.org/sci/facilities/paranal/instruments/uves.html>

⁶<https://www.eso.org/public/brazil/teles-instr/paranal-observatory/vlt/?lang>

⁷https://www.noao.edu/image_gallery/html/im0449.html

- XSL: dados de observações do *X-Shooter* (Chen *et al.*, 2014; Gonneau *et al.*), outro espectrógrafo que compõe o *Very Large Telescope Array*. Mais detalhes sobre este conjunto de dados são explicados nesta seção (Anaïs Gonneau, comunicação privada);
- **Molecfit**: dados telúricos sintéticos gerados com a ferramenta **Molecfit** (Smette *et al.*, 2015).

Antes de qualquer teste foi necessário um processo de ambientação e familiarização com os dados astronômicos. Para a leitura e manipulação de dados no formato FITS foi usada a biblioteca *Astropy*.

A combinação dos dados com o ferramental adequado, possibilitou analisar e visualizar o espectro de diversas estrelas. Dois exemplos de estrelas de ciência e seus referenciais telúricos estão nas figuras 3.2 e 3.4, e representam, respectivamente, a estrela HD110379 e o Sol. Além dos espectros completos, a figura 3.3 ilustra dois painéis aumentados dos espectros observado e telúrico de HD110379, assim é possível ver as linhas espectrais com mais detalhes e compreender qual informação é perdida para a contaminação telúrica.

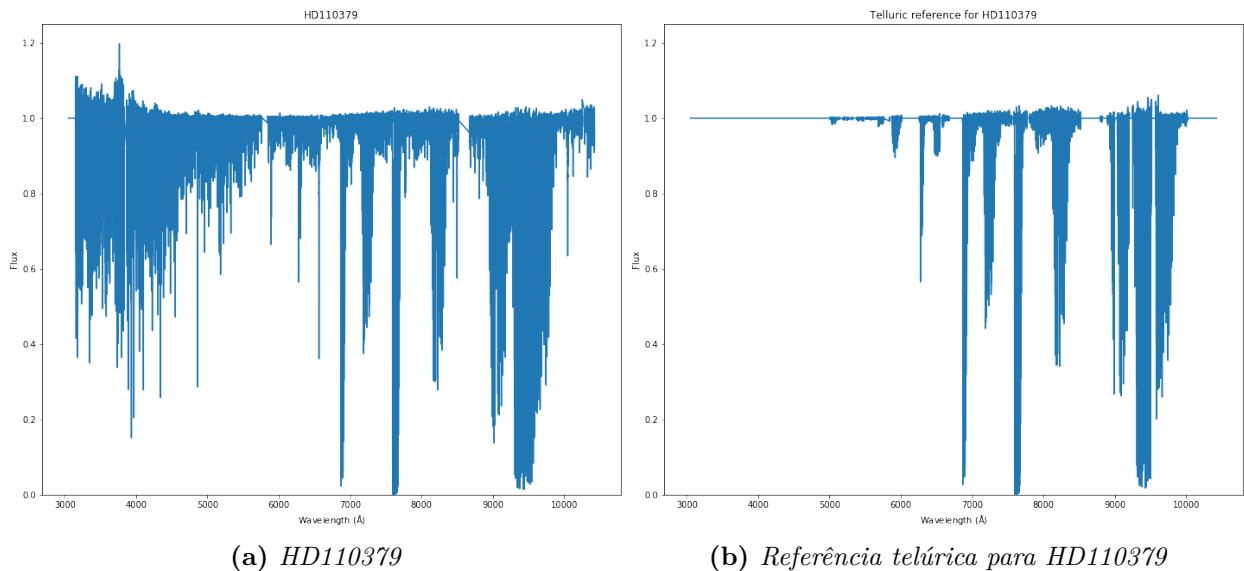


Figura 3.2: Espectro observado com contaminação e espectro telúrico de HD110379

Além das estrelas acima, como foi mencionado na seção 2.4, o principal conjunto de dados utilizado nos experimentos deste trabalho são estrelas do *X-Shooter Spectral Library*. O diferencial deste conjunto de dados é que além das observações estelares e referenciais telúricos, foram fornecidos os espectros resultantes da correção telúrica realizada por especialistas. Desta maneira, foi possível estabelecer um *ground truth* como base de comparação para os resultados dos experimentos. Neste conjunto de dados foram fornecidas quatro estrelas (X0319, X0386, X0538, X0771), com nove arquivos cada, distribuídos da seguinte maneira:

- *X_N_E.fits ou *X_O_E.fits: o espectro da estrela com contaminação telúrica, separado em 3 arquivos que cobrem intervalos de comprimento de ondas diferentes (X no nome do arquivo será U = ultravioleta, V = visível ou N = infravermelho);

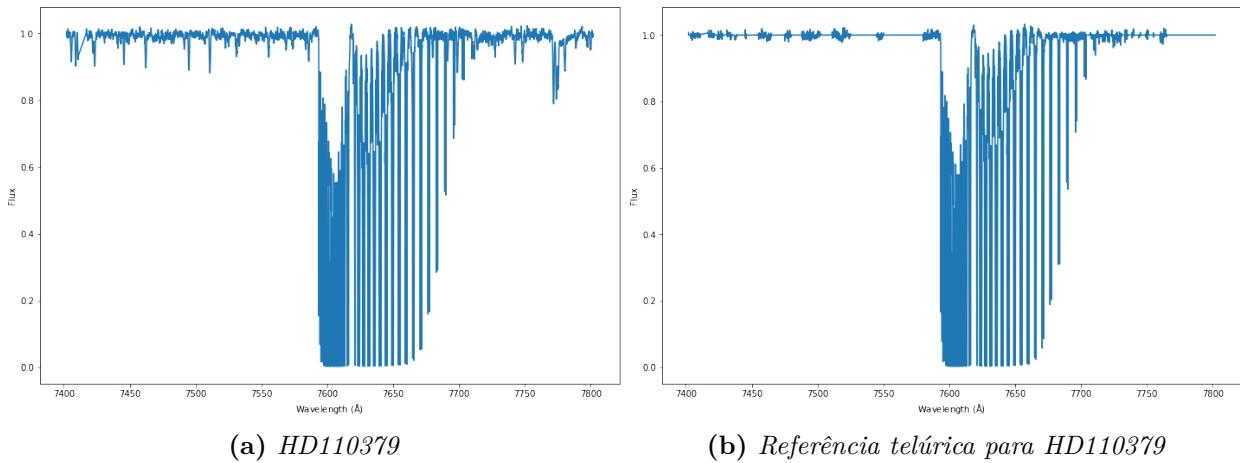


Figura 3.3: Parte do espectro observado com contaminação e espectro telúrico de HD110379 no intervalo de comprimento de onda de 7400 \AA a 7800 \AA

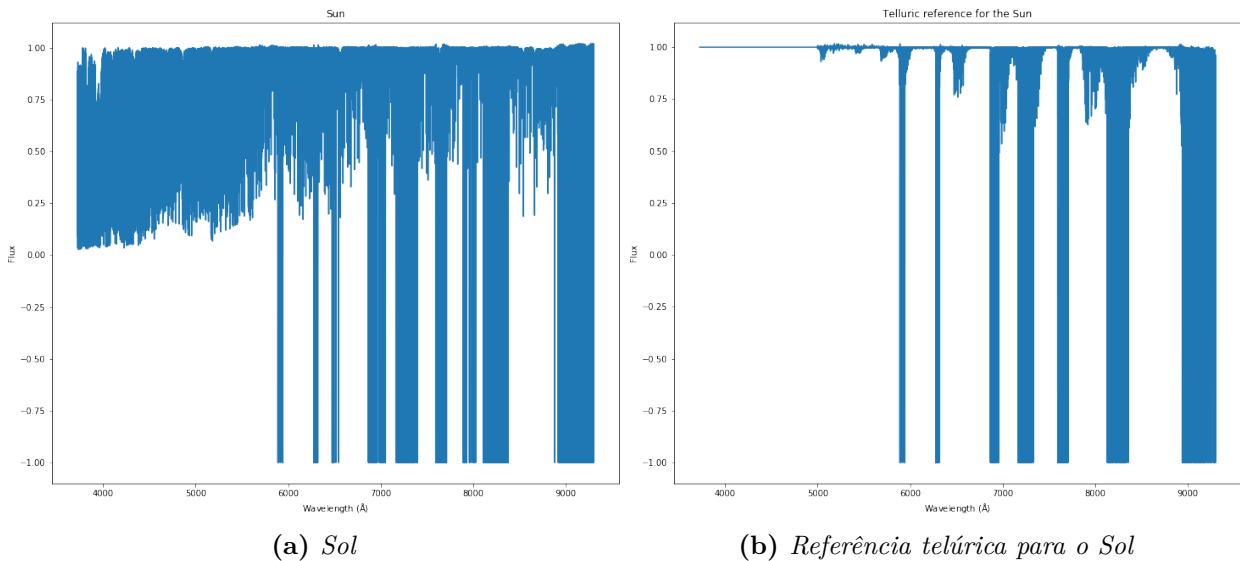


Figura 3.4: Espectro observado com contaminação e espectro telúrico do Sol

- *TRA.fits: modelo **Molecfit** que representa o espectro telúrico correspondente ao dia da observação, também separado em 3 arquivos que cobrem comprimentos de ondas diferentes;
- *TAC.fits ou *TAC_final.fits: espectros das estrelas corrigidos pelo modelo **Molecfit** e já considerando que as distorções observadas em comprimento de onda foram corrigidas, também separados em três arquivos.

As figuras 3.5, 3.6 e 3.7 ilustram exemplos de alguns espectros do conjunto de dados do *X-Shooter*, tanto de diferentes estrelas quanto de diferentes regiões de comprimentos de onda.

Uma das diferenças entre o espectro observado e o telúrico de uma mesma estrela é a diferença de ordem de magnitude da escala do eixo y dos gráficos. Como foi mencionado

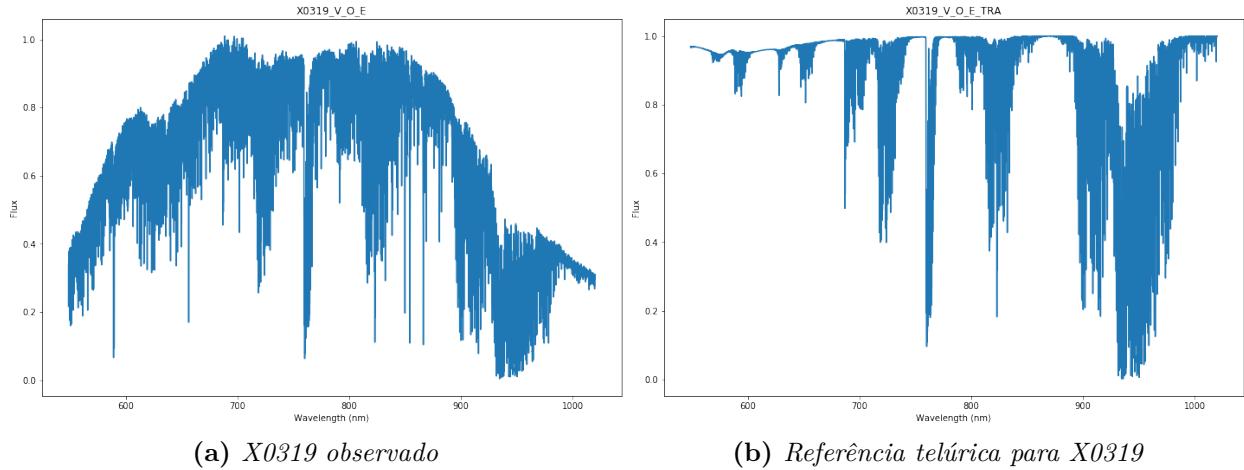


Figura 3.5: Espectro observado com contaminação e espectro telúrico de X0319 na região visível de comprimento de onda

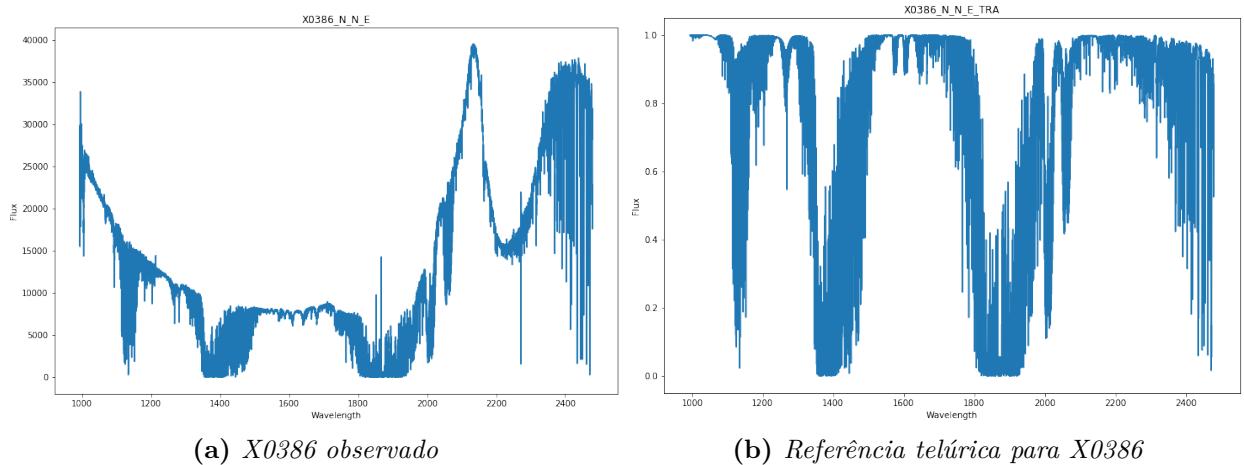


Figura 3.6: Espectro observado com contaminação e espectro telúrico de X0386 na região infravermelho de comprimento de onda

na seção 2.2, os espectros observados possuem uma unidade física ao final do processo de redução de dados. No caso das observações do conjunto de dados do *X-Shooter* utilizadas nesse trabalho, os espectros não passaram pelo processo de redução de dados por completo. Nesse caso o eixo y não possui o mesmo significado físico de intensidade por unidade de área, tempo e comprimento de onda, mas é uma quantidade adimensional relacionada ao número de fótons detectado pelo CCD.

Como o fluxo é capturado pelo CCD através da corrente elétrica produzida por uma certa contagem de fótons, o processo de conversão entre a tensão produzida no instrumento e finalmente uma unidade fisicamente interpretável é extremamente sofisticado em termos de *hardware* e não será abordado neste trabalho. Para os experimentos realizados neste capítulo é importante notar que o eixo y é relacionado tanto ao fluxo da estrela quanto à assinatura instrumental do espetrógrafo.

Já nos espectros telúricos sintéticos do *Molecfit*, o eixo y do gráfico é normalizado para

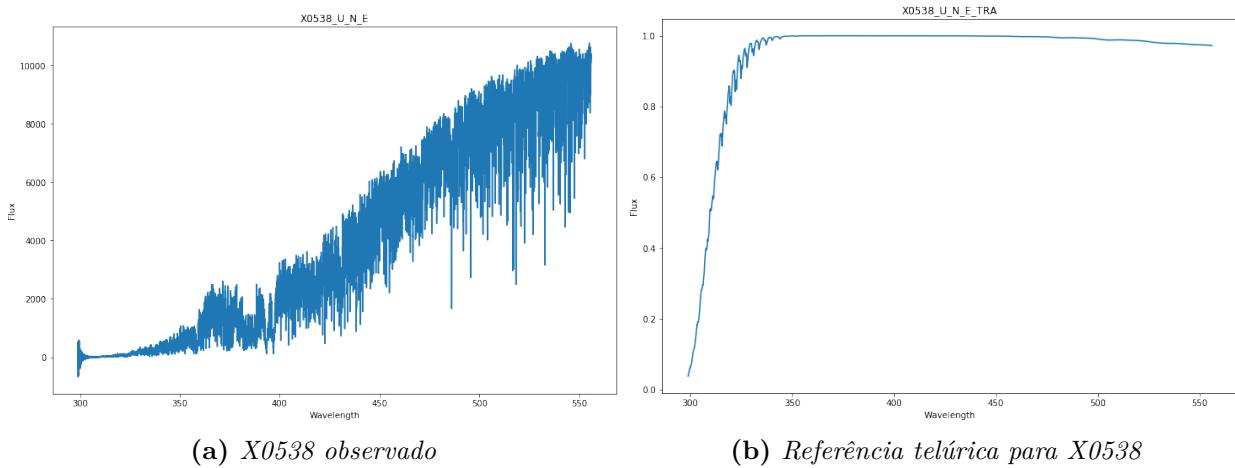


Figura 3.7: Espectro observado com contaminação e espectro telúrico de X0538 na região ultravioleta de comprimento de onda

o intervalo $[0, 1]$ e representa um mapa de transmitância da atmosfera, ou seja, uma série de fatores de absorção que quando interagem com o sinal estelar capturam parte da sua emissão. Para facilitar o restante dos experimentos, as observações do *X-Shooter* também tiveram seus intervalos normalizados de forma a simplificar a manipulação dos dados.

3.4 Divisão do sinal estelar pelo telúrico

Após o processo de exploração e familiarização com os dados astronômicos fornecidos, o primeiro experimento idealizado neste trabalho foi a divisão dos espectros estelares observados pelas suas respectivas referências telúricas, a fim de evidenciar as limitações desse processo. Como mencionado na seção 2.3, por causa da natureza multiplicativa do modelo físico da transmissão da radiação estelar através da atmosfera, a correção telúrica corresponderia de fato à divisão simples entre os espectros de ciência e telúrico. A motivação por trás deste experimento é observar que na prática esta solução não é ideal e cria artefatos no espectro estelar.

Algumas explicações do porquê o método da correção telúrica pela divisão simples de espectros não funciona como proposto no modelo teórico provêm das distorções do sinal que possuem origem instrumental. Alguns exemplos de distorções esperadas pelos astrônomos em suas observações são o alargamento das linhas espectrais, desconsiderando o alargamento Doppler (Levenhagen e Kunzel, 2008), e a criação de deslocamentos no domínio de comprimento de onda provenientes de artefatos de *hardware* (e.g. Moehler, 2015). A primeira divisão realizada foi do conjunto de dados do espectrógrafo UVES, na estrela HD110379.

Na figura 3.8, temos o espectro de ciência original da estrela HD110379 e o resultado da correção telúrica que é a divisão simples entre os espectros. Ambos tiveram seus eixos y redimensionados de modo a facilitar a inspeção visual dos resultados. A correção telúrica neste caso de fato diminui algumas das linhas de absorção do espectro, principalmente na

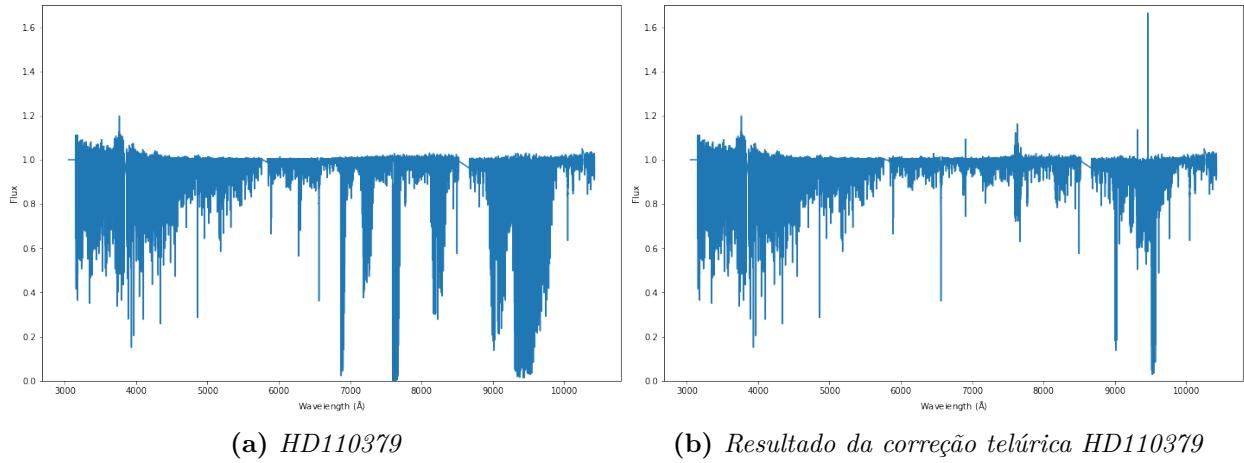


Figura 3.8: Espectros de ciência e dividido de HD110379 pelo seu referencial telúrico

região dos 6500Å aos 8500Å de comprimento de onda. Porém, é visivelmente perceptível que ela também cria artefatos como picos no espectro resultante, sendo alguns menores e outros maiores quando comparados ao intervalo [0, 1] do fluxo estelar.

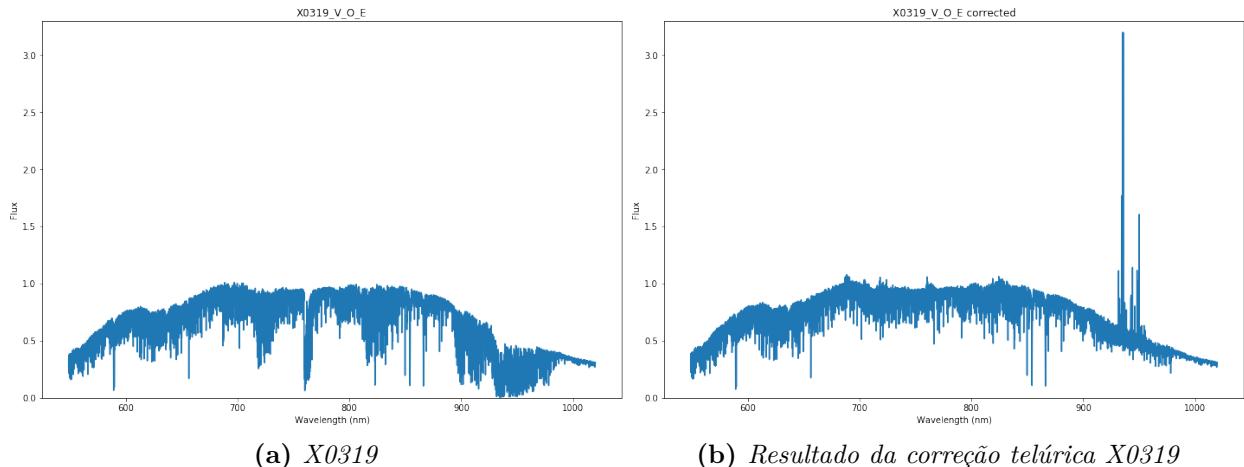


Figura 3.9: Espectros de ciência e dividido de X0319 pelo seu referencial telúrico

A figura 3.9 mostra mais um exemplo de correção telúrica feita nos dados fornecidos, desta vez na estrela X0319 do espectrógrafo *X-Shooter*. Novamente, os eixos dos gráficos dos espectros foram redimensionados, de modo a facilitar a visualização dos efeitos da divisão do observado pelo telúrico. Neste caso, é possível observar tanto a diminuição das linhas de absorção ao longo do espectro, quanto a criação de picos no mesmo, principalmente na região entre 900nm e 1000nm, na qual o pico resultante tem uma escala aproximadamente três vezes maior que a do espectro original.

Os artefatos criados pelas correções telúricas das figuras acima não deveriam existir na situação ideal do problema, onde, elemento a elemento, o fluxo do espectro de ciência deveria ser maior ou igual ao fluxo após a absorção atmosférica representada pelo espectro telúrico. Isto indica a presença de um dos efeitos esperados deste tipo de operação: o desalinhamento

entre o espectro estelar e o telúrico. Os resultados das correções telúricas e os seus efeitos resultantes fazem sentido, principalmente no caso da estrela do *X-Shooter*, onde pesquisadores observaram estes mesmos problemas (Gonneau *et al.*; Moehler, 2015).

Os resultados das divisões nesta seção ilustram as dificuldades conhecidas no processo atual de correção telúrica e deixam explícita a necessidade de propor soluções alternativas que melhorariam a qualidade dos dados observados.

3.5 DTW no sinal estelar

Os resultados do experimento de correção telúrica através da divisão de um espectro estelar pelo seu referencial telúrico esclarecem a necessidade de considerar melhorias ao processo atual. Como foi visto, um dos problemas nos dados é a falta de alinhamento entre os sinais, e a consequente criação de artefatos e distorções nos espectros corrigidos. O experimento desta seção propõe uma solução alternativa a uma das dificuldades conhecidas das observações: o desalinhamento dos espectros.

Como foi explicado na seção 2.4, podemos usar o algoritmo *Dynamic Time Warping*, representado pelos algoritmos 1 e 2, para encontrar o alinhamento ótimo entre dois sinais, e o algoritmo 3 para realinear os sinais em relação ao resultado do DTW. Assim, seria possível realinear o espectro da observação estelar com o espectro telúrico para resultar em uma divisão de maior qualidade e menor perda de informação e criação de artefatos.

Para este experimento o conjunto de dados utilizado foi o XSL e a implementação do DTW foi a *FastDTW*, de modo a permitir a aplicação do algoritmo nos espectros completos (ou seja, em todas as faixas de comprimentos de onda).

O primeiro exemplo da aplicação do DTW nos dados da XSL considera todo o intervalo visível de comprimento de onda da estrela X0319, cujos espectros são ilustrados na figura 3.5. O algoritmo foi aplicado nos espectros estelar e telúrico, e o caminho gerado pelo DTW está na figura 3.10.

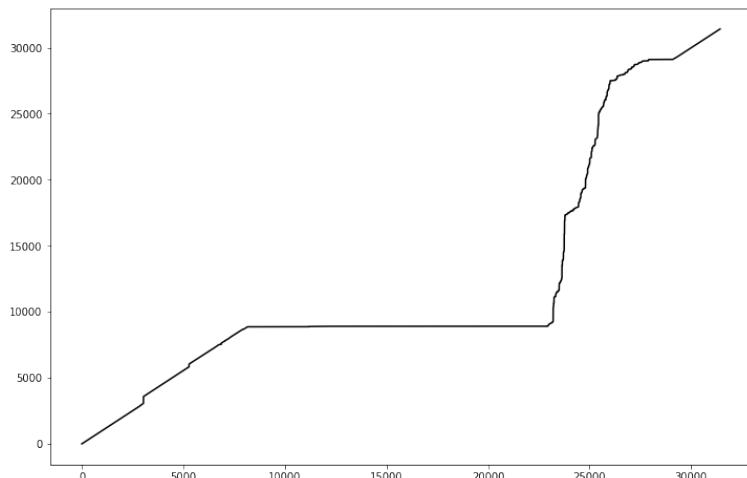


Figura 3.10: Caminho do DTW para o espectro estelar e telúrico de X0319

O caminho resultante do algoritmo já indica uma série de regiões problemáticas dos espectros que não puderam ser alinhadas de maneira razoável. Idealmente o caminho deveria se distanciar pouco da diagonal da matriz de similaridade, mas neste caso são claras as regiões onde as sequências simplesmente não são similares, como a linha horizontal que ocupa a maioria do caminho, seguida de um trecho quase vertical.

O espectro telúrico original do **Molecfit** e o realinhado em relação à observação estelar que foi reconstruído a partir do caminho do DTW estão apresentados na figura 3.11.

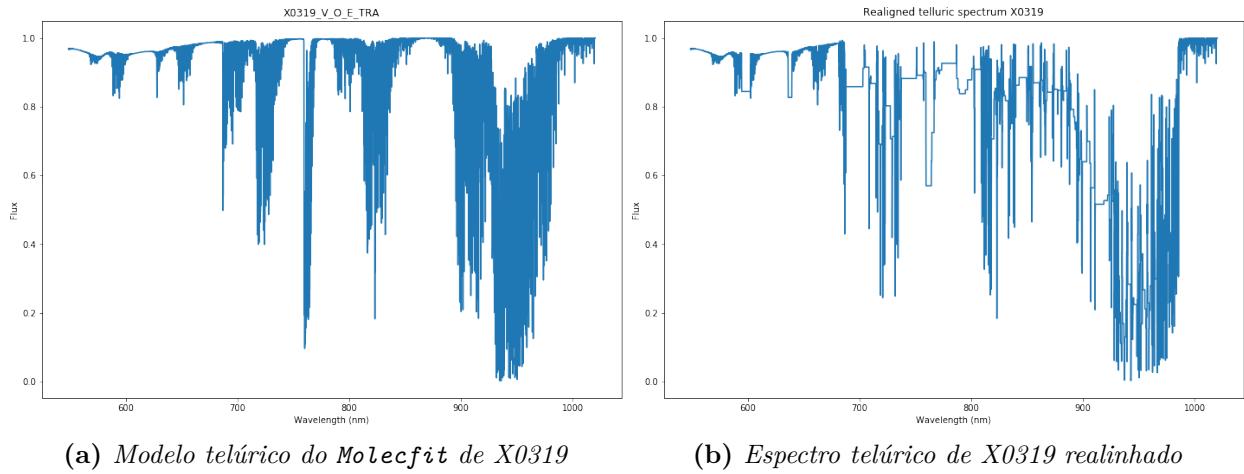


Figura 3.11: Espectro telúrico de X0319 e espectro telúrico realinhado a partir do caminho do DTW

É possível ver que, por mais que o espectro telúrico realinhado da estrela X0319 tenha o formato geral do modelo do **Molecfit**, ele possui uma forte distorção em relação às linhas espectrais originais e exibe a formação de ângulos retos que não são característicos deste tipo de dado.

O próximo teste também foi feito em uma estrela do XSL, a estrela X0386, no intervalo visível de comprimento de onda do espectro. A escolha desta região espectral se deve ao fato de que a contaminação telúrica está presente, porém é menos intensa e de mais difícil correção do que na região de comprimento de onda infravermelho. Quanto à região ultravioleta, esta parece ser menos sensível aos efeitos da atmosfera terrestre, como pode ser visualizado na figura 2.5.

A figura 3.12 a seguir ilustra a observação estelar de X0386, para que se possa perceber o formato do espectro contaminado. Nesta observação e no seu referencial telúrico foi aplicado o DTW, e a figura 3.13 contém o caminho resultante do algoritmo. Neste caso também é possível observar que o caminho apresenta uma grande distorção em relação à diagonal ideal, o que significa que o sinal estelar e telúrico não possuem um alinhamento razoável que permitiria a divisão simples dos espectros.

Os efeitos da distorção causada pelo realinhamento de espectro telúrico usando o caminho resultante do DTW podem ser vistos na figura 3.14. O espectro telúrico realinhado sofre uma grande perda de informação nas suas linhas espectrais e é possível ver que ele se assemelha

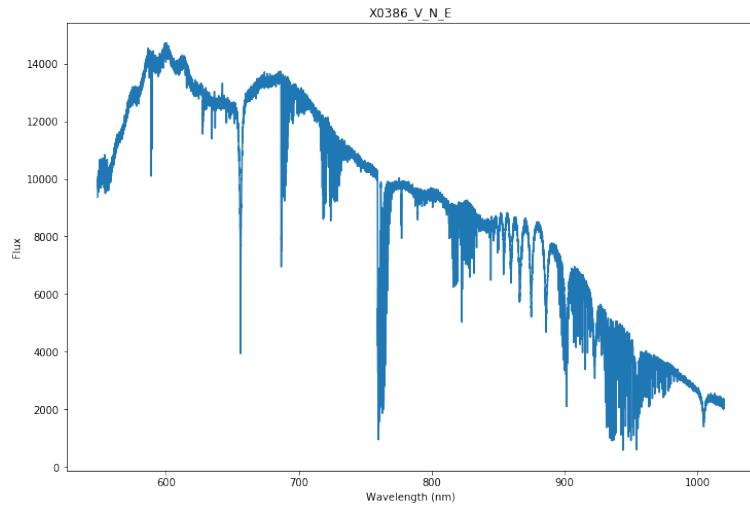


Figura 3.12: Espectro observado com contaminação telúrica de X0386

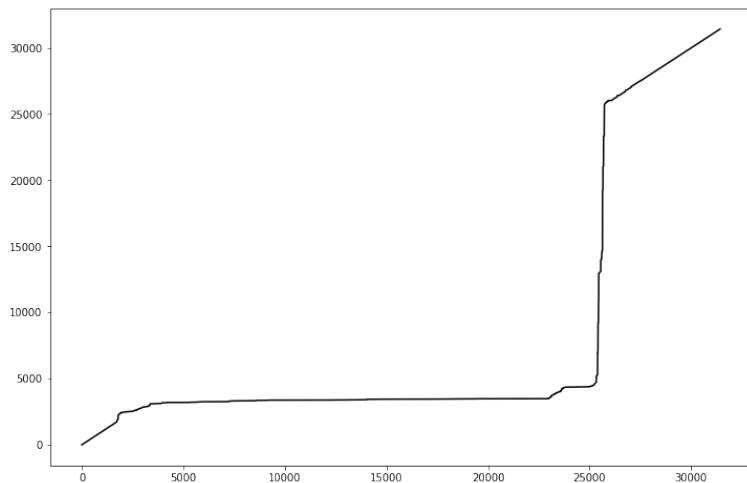


Figura 3.13: Caminho do DTW para o espectro estelar e telúrico de X0386

ao modelo original apenas no intervalo entre 900nm e 1000nm, a mesma região do caminho do DTW que se aproxima da diagonal secundária da matriz de similaridade.

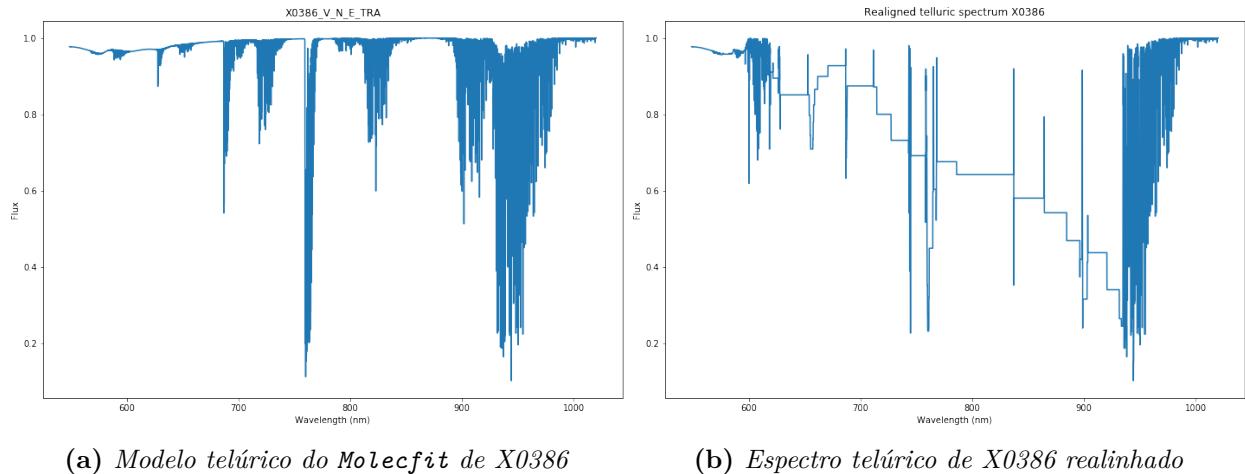


Figura 3.14: Espectro telúrico de X0386 e espectro telúrico realinhado a partir do caminho do DTW

Uma hipótese para explicar o baixo desempenho da aplicação do DTW nos espectros completos seriam as grandes diferenças de perfil global que se observam entre os espectros de ciência e os espectros telúricos: a intuição que levou à seleção do DTW como ferramenta de alinhamento entre esses espectros diferentes correspondia à similaridade observável nesses espectros entre as linhas de absorção correspondentes a elementos químicos específicos. Assim, um próximo passo natural correspondia a testar o algoritmo de realinhamento em regiões menores do espectro.

Para verificar essa nova hipótese, foi aplicado o DTW no espectro da estrela X0319, no intervalo de comprimento de onda entre 834nm e 841nm. As figuras 3.15 e 3.16 mostram o caminho resultante na matriz de similaridade e o espectro telúrico original e realinhado.

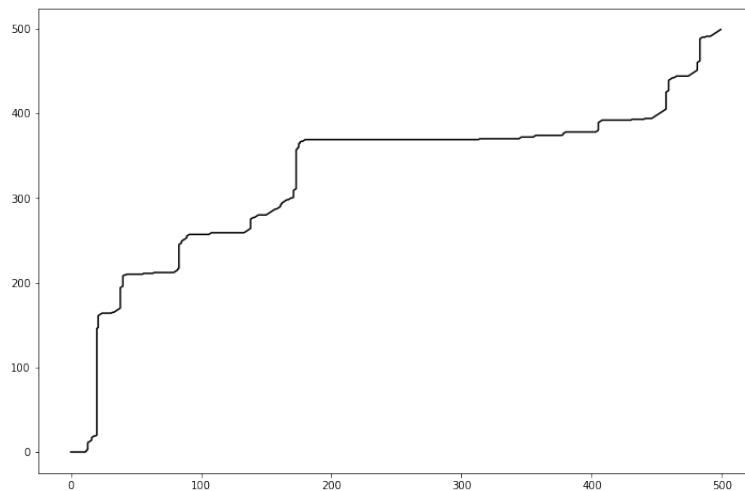


Figura 3.15: Caminho do DTW para o espectro estelar e telúrico de X0319 no intervalo de 834nm a 841nm

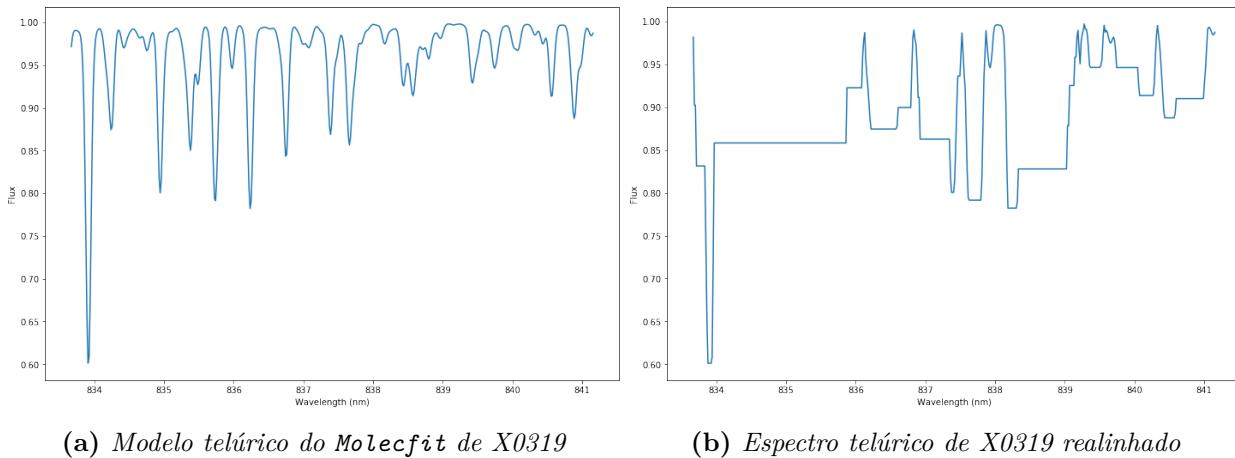


Figura 3.16: Espectro telúrico de X0319 e espectro telúrico realinhado a partir do caminho do DTW no intervalo de 834nm a 841nm

Neste experimento são observados os mesmos problemas dos experimentos anteriores em menor escala. Em um intervalo de comprimento de onda que abrange 9nm de espectro estelar ocorrem os mesmos deslocamentos horizontais e verticais no caminho da matriz de similaridades, e no espectro telúrico realinhado também ocorre a formação de ângulos retos nas linhas de absorção.

A aplicação do *Dynamic Time Warping* entre o sinal estelar e o modelo telúrico não se mostrou adequado para alinhar as linhas de absorção correspondentes nesses espectros permitindo assim a correção da distorção telúrica. O algoritmo, que se baseia na premissa de forte similaridade entre os conteúdos a serem alinhados, não produz alinhamentos razoáveis quando aplicado em sinais que não representam a mesma informação, embora compartilhem regiões com linhas de absorção de origens equivalentes. O DTW ainda poderia funcionar, hipoteticamente, em um caso onde fosse possível isolar setores muito pequenos, com linhas de absorção similares e claramente separadas.

3.6 DTW no sinal atmosférico

Os experimentos de uso do *Dynamic Time Warping* como ferramenta de realinhamento entre o sinal estelar e o referencial telúrico contribuíram para uma maior familiarização em relação aos dados usados neste trabalho e também introduziram questionamentos sobre o sinal atmosférico obtido do *Molecfit*.

Os pesquisadores do XSL, ao identificarem os desalinhamentos nos espectros resolveram o problema dividindo o sinal em sub-espectros, e para cada um destes utilizaram o *Molecfit* para recomputar o modelo telúrico neste intervalo de comprimento de onda. Esta solução tem o objetivo de empiricamente obter o *shift* único de cada sub-espectro. A partir deste procedimento foi realizada a correção telúrica nos espectros do XSL, como é descrito no artigo de [Gonneau et al.](#).

Com os dados fornecidos dos espectros estelares corrigidos que são considerados como o *ground-truth* neste trabalho, é possível resgatar o espectro telúrico realinhado a partir da divisão do espectro estelar observado pelo espectro estelar corrigido. Desta forma, se temos os vetores o e c que representam a observação contaminada e o espectro final corrigido, teríamos o espectro real de transmitância da atmosfera dado por

$$a' = o \oslash c \quad (\text{ou equivalentemente:} \quad a'_i = o_i/c_i, \forall i),$$

onde a' representa o espectro telúrico com correção dos desalinhamentos observados (\oslash denota a divisão de Hadamard).

O estudo do sinal atmosférico corrigido em relação ao modelo telúrico do **Molecfit** permite obter mais informações sobre os dados e entender o quão grande é o desalinhamento entre estes espectros. Os experimentos desta seção se propõem a resgatar o sinal atmosférico correspondente ao *ground-truth* e compará-lo com o sinal atmosférico obtido com o conjunto de dados do XSL.

As estrelas utilizadas para este experimento são a X0319 e a X0386 do *X-Shooter Spectral Library*, na região visível de comprimento de onda. As figuras 3.17 e 3.18 mostram os modelos telúricos dados e os seus respectivos espectros telúricos realinhados a partir da divisão entre a observação estelar com contaminação telúrica e o espectro corrigido ([Gonneau et al.](#)).

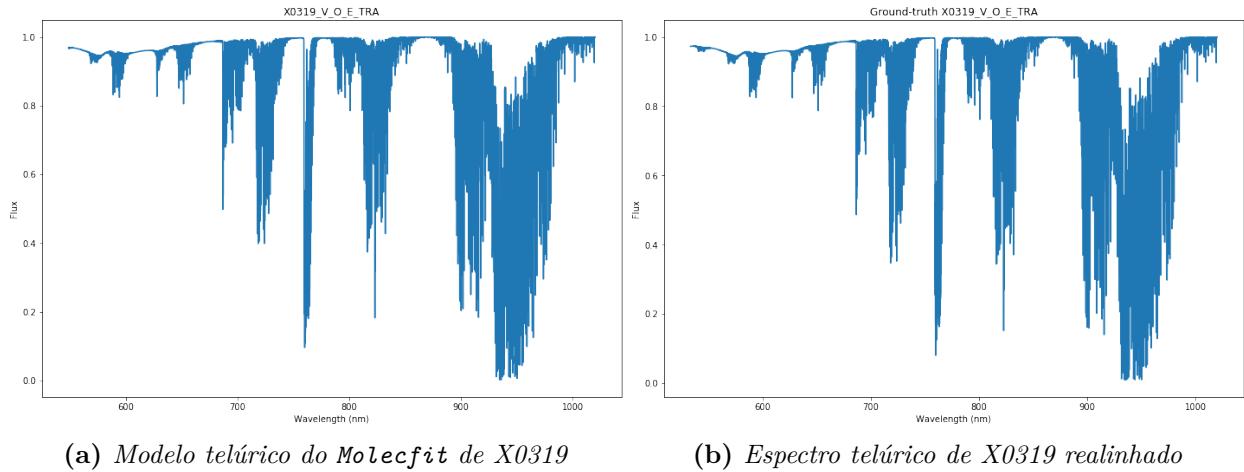


Figura 3.17: Espectro telúrico dado e espectro telúrico corrigido de shifts de X0319

Por inspeção visual, é possível observar que os espectros corrigidos de desalinhamentos não possuem distorções perceptíveis em relação aos seus correspondentes modelos do **Molecfit**. Isto indica que se existem desalinhamentos entre estes sinais, estes devem ser pequenos.

A presença de pequenos desalinhamentos entre os sinais sugere que estes dados seriam adequados à aplicação do algoritmo do *Dynamic Time Warping*. Desta forma seria possível observar regiões com maior desalinhamento, e portanto, quais regiões do espectro tiveram correções de *shifts* feitas pelo grupo de pesquisadores que lidaram com estes espectros.

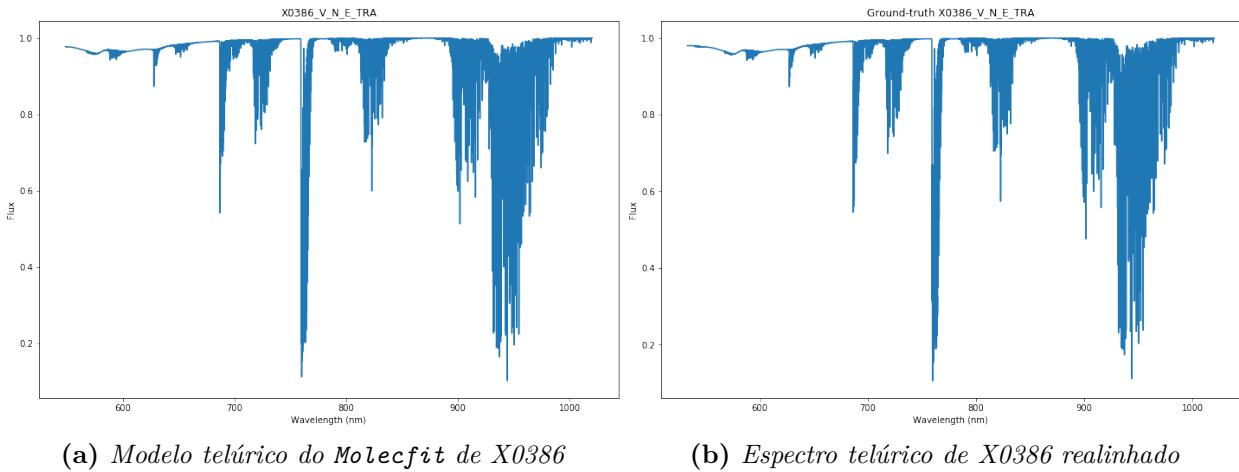


Figura 3.18: Espectro telúrico dado e espectro telúrico corrigido de shifts de X0386

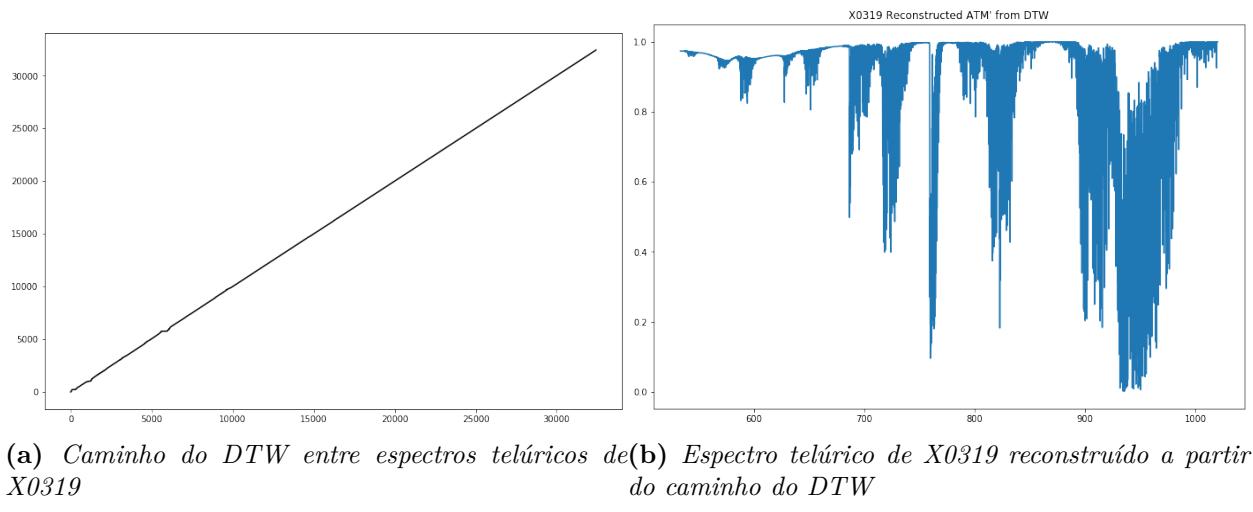
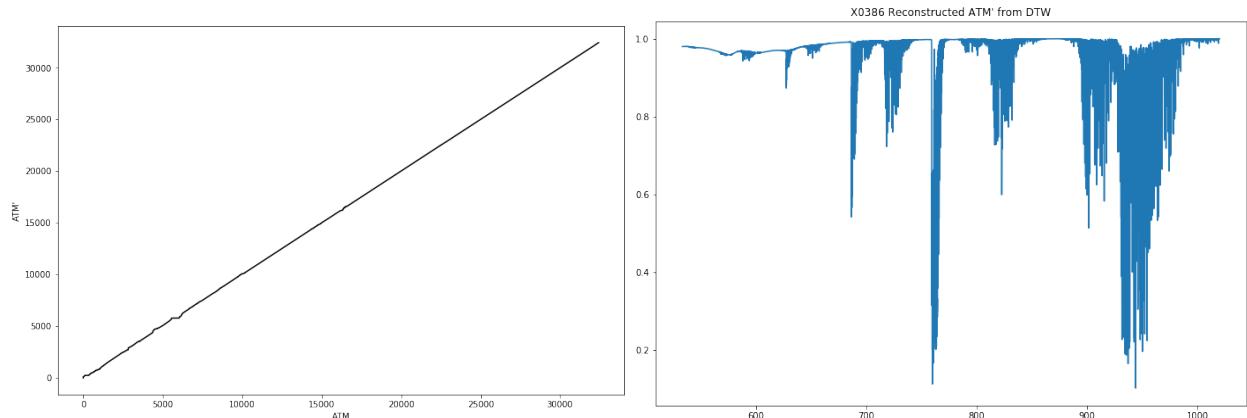


Figura 3.19: Caminho do DTW e espectro telúrico reconstruído de X0319

As figuras 3.19 e 3.20 mostram os caminhos resultantes da aplicação do DTW nos espectros telúricos do *Molecfit* e nos espectros telúricos *ground-truth* de X0319 e X0386. Diferentemente dos caminhos resultantes do DTW quando aplicados no sinal estelar e telúrico, neste caso quase não existem distorções e o caminho se aproxima do caso ideal. Os pequenos desvios em relação à diagonal no caminho resultante indicam desalinhamentos no sinal estelar que foram manualmente corrigidos pelo grupo de pesquisadores responsável pelo espectro final. Para melhor compreender as regiões do espectro telúrico mais afetadas por desalinhamentos, foi computada uma função *shift* ao longo do caminho do DTW, tal que, para cada par de índices (i, j) desse caminho, o valor do *shift* é representado por $j - i$. Desta forma é possível ver os desalinhamentos em termos de quantidade de índices de desvio entre os respectivos sinais telúricos.

A figura 3.21 ilustra a magnitude dos desalinhamentos entre o espectro dado do *Molecfit* e o espectro telúrico realinhado da estrela X0386. Os maiores desalinhamentos são adiantamentos ou atrasos em 200 elementos dos vetores que representam os espectros, o que



(a) Caminho do DTW entre espectros telúricos de X0386
 (b) Espectro telúrico de X0386 reconstruído a partir do caminho do DTW

Figura 3.20: Caminho do DTW e espectro telúrico reconstruído de X0386

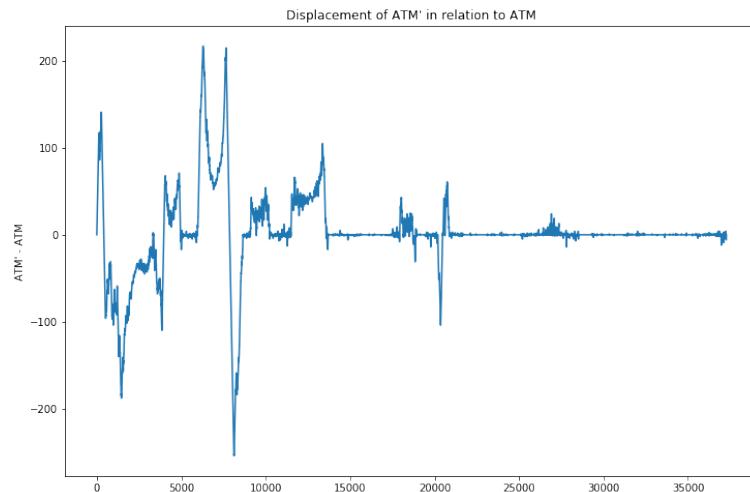


Figura 3.21: Desalinhamentos em número de elementos entre o modelo do Molecfit e o espectro telúrico realinhado de X0386

representa desalinhamentos em comprimento de onda de aproximadamente 3 nm. Isto acontece principalmente entre o elemento 5000 e 10000 dos vetores, que é a região de 608 nm a 683 nm.

Este resultado ilustra uma aplicação possível do algoritmo DTW em vetores que representam a mesma informação, nesse caso entre dois espectros telúricos antes e depois de um realinhamento realizado por especialistas. Tal aplicação poderia ser de interesse em tarefas de aprendizado automático relacionadas com correção de desalinhamentos instrumentais, por exemplo.

Capítulo 4

Conclusão

4.1 Fechamento

Este trabalho possibilitou explorar um problema que une as áreas da Astronomia e da Computação, ao estudar observações de espectros estelares desde uma perspectiva computacional. O problema da contaminação telúrica em espectros estelares é uma fonte de distorções em observações astronômicas feitas do solo, e continua sendo um impedimento para o avanço da pesquisa com estes dados. Linhas de absorção de diferentes perfis e origens moleculares podem ser difíceis de remover com precisão, porém, essa remoção é fundamental para interpretar os espectros capturados.

A investigação exploratória dos dados aqui apresentada é preliminar e objetivou identificar problemas e possíveis soluções com o intuito de minimizar a interferência atmosférica nestas observações. O desenvolvimento desta monografia tornou possível a exploração de dados reais de estrelas e a tentativa de propor uma solução para um dos problemas conhecidos destas observações, os desalinhamentos encontrados entre os espectros estelares e telúricos.

A abordagem escolhida com o algoritmo do *Dynamic Time Warping* tinha como objetivo explorar as similaridades entre os sinais estelar e telúrico, especificamente a presença de linhas de absorção em comprimentos de onda específicos, para reconstruir um espectro telúrico realinhado, de modo a minimizar a perda de informação e os artefatos criados na divisão simples dos espectros. Embora as linhas de absorção possam fornecer pistas visuais importantes para os olhos de um pesquisador treinado, as diferenças importantes entre os espectros estelar e telúrico revelou as limitações daquele algoritmo para o alinhamento deste tipo de sequências, visto que o DTW distorce fortemente o espectro telúrico para minimizar a distância à referência estelar observada, o que não corresponde a uma solução ideal e generalizável para as observações espetrais e seus respectivos referenciais telúricos.

Apesar das dificuldades encontradas, o processo pretendido de familiarização com os dados dos espectros estelares, com algumas das interpretações associadas às configurações neles encontradas, bem como com as técnicas de manipulação desse tipo de informação foi muito proveitoso, permitindo a identificação de uma série de questões para pesquisa que

serão exploradas em trabalhos futuros.

Esta pesquisa tem interesse, entre outras razões, pela grande quantidade de problemas observacionais e instrumentais conhecidos – além dos desalinhamentos nos sinais – que impactam a qualidade dos dados disponíveis para pesquisas astronômicas. Além disso, esta pesquisa tem importância pela sua natureza interdisciplinar, pelo potencial explorável na aplicação de teorias e métodos computacionais na solução de problemas de interesse da comunidade astronômica, e por ajudar a criar pontes de comunicação através de um vocabulário compartilhado entre essas duas importantes áreas do conhecimento.

4.2 Trabalhos futuros

Durante o desenvolvimento deste trabalho interdisciplinar, uma série de ideias de pesquisa surgiram, passíveis de serem exploradas no futuro. A lista abaixo pretende servir como memória para possíveis extensões da pesquisa aqui iniciada.

- O DTW, pela sua característica de maximizar a similaridade entre dois sinais através de realinhamentos, poderia ter utilidade em uma etapa da redução de dados chamada de calibração em comprimento de onda. Nessa etapa procura-se encontrar uma função – possivelmente não-linear – que converte os índices do espectro na escala física dos comprimentos de onda.
- Uma das dificuldades maiores no alinhamento dos espectros observado e telúrico diz respeito às diferenças de perfil global dos espectros, ou seja, dos níveis de onde saem as linhas de absorção, bem como às diferenças de profundidade e largura das linhas de absorção. Seria possível pensar em diminuir essas diferenças através de técnicas de *pre-whitening*, de filtros do tipo passa-altas ou de filtros de remoção da mediana, antes de tentar alinhar partes dos espectros considerados.
- Na modelagem do processo de aquisição através de espectrografia astronômica, seria plausível tentar modelar a assinatura instrumental como um filtro, e buscar estimar parâmetros desse filtro através de pares de espectros observados e corrigidos, ou a partir de observações obtidas por instrumentos diferentes em condições similares (posição, direção, condições climáticas).
- Pode-se buscar uma descrição *data-driven* do espectro telúrico, em contrapartida à modelagem física de transferência radiativa. Nessa abordagem, pode-se partir de uma grande quantidade de espectros telúricos disponíveis em *archives* de astronomia, para generalizar um *template* telúrico em termos de filtros gaussianos. As variações de intensidade do sinal telúrico (ligadas às condições atmosféricas) e de resolução espectral (ligadas ao instrumentos) poderiam ser parametrizadas em termos de funções agnósticas às particularidades físicas do problema.

Referências Bibliográficas

- Burrows(2012)** Adam Burrows. The classification of stellar spectra, 2012. URL <https://www.astro.princeton.edu/~burrows/classes/204/stellar.atmospheres.HR.pdf>. [Online; accessed September 14, 2019]. Citado na pág. 6
- Charbonneau et al.(2002)** David Charbonneau, Timothy M. Brown, Robert W. Noyes e Ronald L. Gilliland. Detection of an extrasolar planet atmosphere. *The Astrophysical Journal*, 568(1):377–384. doi: 10.1086/338770. URL <https://doi.org/10.1086/pág. 5>
- Chen et al.(2014)** Yan-Ping Chen, Scott C. Trager, Reynier F. Peletier, Ariane Lançon, Alexandre Vazdekis Vazdekis, Ph. Prugniel, D. Rodrigues Silva e Anais Gonneau. The x-shooter spectral library (XSL) - i. dr1: Near-ultraviolet through optical spectra from the first year of the survey. Citado na pág. 10, 22
- Clough et al.(2005)** S. A. Clough, M. W. Shephard, E. J. Mlawer, J. S. Delamere, M. J. Iacono, K. Cady-Pereira, S. Boukabara e P. D. Brown. Atmospheric radiative transfer modeling: a summary of the AER codes. *JQSRT*, 91:233–244. doi: 10.1016/j.jqsrt.2004.05.058. Citado na pág. 9
- Davenhall et al.(2001)** AC Davenhall, GJ Privett e MB Taylor. The 2-d ccd data reduction cookbook. Citado na pág. 6
- Dekker et al.(2000)** H. Dekker, S. D'Odorico, A. Kaufer, B. Delabre e H. Kotzlowski. Design, construction, and performance of UVES, the echelle spectrograph for the UT2 Kueyen Telescope at the ESO Paranal Observatory. Em M. Iye e A. F. Moorwood, editors, *Optical and IR Telescope Instrumentation and Detectors*, volume 4008 of , páginas 534–545. doi: 10.1117/12.395512. Citado na pág. 21
- Djorgovski(2004)** George Djorgovski. Astronomical instruments and measurements, 2004. URL <http://www.astro.caltech.edu/~george/ay20/Ay20-Lec3x.pdf>. [Online; accessed October 30, 2019]. Citado na pág. 7
- Fong(2012)** Simon Fong. Using hierarchical time series clustering algorithm and wavelet classifier for biometric voice classification. *Journal of biomedicine & biotechnology*, 2012: 215019. doi: 10.1155/2012/215019. Citado na pág. 14
- Gonneau et al.()** Anaïs Gonneau, M. Lyubenova, A. Lançon, S. C. Trager, R. F. Peletier, A. Arentsen, Y.-P. Chen, P. R. T. Coelho, M. Dries, J. Falcón-Barroso, P. Prugniel, P. Sánchez-Blázquez, A. Vazdekis e K. Verro. The x-shooter spectral library (XSL): Data release 2. submitted to *Astronomy & Astrophysics*. Citado na pág. 11, 22, 27, 31, 32
- Haworth(2011)** David Haworth. Vega low resolution spectrum, 2011. URL http://www.stargazing.net/david/NSI2011/VegaA06-012s06_20110802_1104.html. [Online; accessed October 19, 2019]. Citado na pág. 5

- Hetem e Pereira(2010)** Jane Cristina Gregorio Hetem e Vera Jatenco Silva Pereira. Estrelas: Classificação espectral, 2010. URL <http://www.astro.iag.usp.br/~jane/aga215/apostila/cap10.pdf>. [Online; accessed October 19, 2019]. Citado na pág. 5
- Hinkle et al.(2000)** K. Hinkle, L. Wallace, J. Valenti e D. Harmer. *Visible and Near Infrared Atlas of the Arcturus Spectrum 3727-9300 A*. Visible and Near Infrared Atlas of the Arcturus Spectrum 3727-9300 A ed. Kenneth Hinkle, Lloyd Wallace, Jeff Valenti, and Dianne Harmer. (San Francisco: ASP) ISBN: 1-58381-037-4, 2000. Citado na pág. 21
- Hunter(2007)** J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95. doi: 10.1109/MCSE.2007.55. Citado na pág. 21
- Kepler e Saraiva(2014)** S. O. Kepler e M.F.O Saraiva. Estrelas, 2014. URL <http://astro.if.ufrgs.br/estrelas/estrelas.htm>. [Online; accessed October 27, 2019]. Citado na pág. 3
- Kluyver et al.(2016)** Thomas Kluyver, Benjamin Ragan-Kelley, Fernando Pérez, Brian Granger, Matthias Bussonnier, Jonathan Frederic, Kyle Kelley, Jessica Hamrick, Jason Grout, Sylvain Corlay, Paul Ivanov, Damián Avila, Safia Abdalla e Carol Willing. Jupyter notebooks – a publishing format for reproducible computational workflows. Em F. Loizides e B. Schmidt, editors, *Positioning and Power in Academic Publishing: Players, Agents and Agendas*, páginas 87 – 90. IOS Press. Citado na pág. 21
- Levenhagen e Kunzel(2008)** R.S. Levenhagen e R. Kunzel. A rotação estelar e seu efeito sobre os espectros. *Revista Brasileira de Ensino de Física*, 30:4701.1 – 4701.4. ISSN 1806-1117. URL http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1806-11172008000400016&nrm=iso. Citado na pág. 25
- MATLAB()** MATLAB. dtw - distance between signals using dynamic time warping. URL <https://www.mathworks.com/help/signal/ref/dtw.html>. [Online; accessed November 2, 2019]. Citado na pág. 12
- McGlynn(2014)** Thomas A. McGlynn. A primer on the FITS data format, 2014. URL https://fits.gsfc.nasa.gov/fits_primer.html. [Online; accessed October 14, 2019]. Citado na pág. 19
- McLaughlin(2013)** Brendan McLaughlin. *McGraw Hill Science Yearbook 2013*. Citado na pág. 4
- Moehler(2015)** S. Moehler. *Analysis of wavelength shifts reported for X-shooter spectra*. European Southern Observatory, 2015. Citado na pág. 11, 25, 27
- Moravec(2004)** Emily Moravec. The study of merger induced galaxy evolution, 2004. URL http://www.astro.wisc.edu/~moravec/data_reduction.html. [Online; accessed October 30, 2019]. Citado na pág. 7
- Oliphant(2006)** Travis E Oliphant. *A guide to NumPy*, volume 1. Trelgol Publishing USA. Citado na pág. 21
- Pence et al.(2010)** William D Pence, Lucio Chiappetti, Clive G Page, Richard A Shaw e Elizabeth Stobie. Definition of the flexible image transport system (FITS), version 3.0. *Astronomy & Astrophysics*, 524:A42. Citado na pág. 19, 20

Price-Whelan *et al.*(2018) A. M. Price-Whelan, B. M. Sipőcz, H. M. Günther, P. L. Lim, S. M. Crawford, S. Conseil, D. L. Shupe, M. W. Craig, N. Dencheva, A. Ginsburg, J. T. VanderPlas, L. D. Bradley, D. Pérez-Suárez, M. de Val-Borro, (Primary Paper Contributors, T. L. Aldcroft, K. L. Cruz, T. P. Robitaille, E. J. Tollerud, (Astropy Coordination Committee, C. Ardelean, T. Babej, Y. P. Bach, M. Bachetti, A. V. Bakanov, S. P. Bamford, G. Barentsen, P. Barmby, A. Baumbach, K. L. Berry, F. Biscani, M. Boquien, K. A. Bostroem, L. G. Bouma, G. B. Brammer, E. M. Bray, H. Breytenbach, H. Buddelmeijer, D. J. Burke, G. Calderone, J. L. Cano Rodríguez, M. Cara, J. V. M. Cardoso, S. Cheedella, Y. Copin, L. Corrales, D. Crichton, D. D'Avella, C. Deil, É. Depagne, J. P. Dietrich, A. Donath, M. Droettboom, N. Earl, T. Erben, S. Fabbro, L. A. Ferreira, T. Finethy, R. T. Fox, L. H. Garrison, S. L. J. Gibbons, D. A. Goldstein, R. Gommers, J. P. Greco, P. Greenfield, A. M. Groener, F. Grollier, A. Hagen, P. Hirst, D. Homeier, A. J. Horton, G. Hosseinzadeh, L. Hu, J. S. Hunkeler, Ž. Ivezić, A. Jain, T. Jenness, G. Kanarek, S. Kendrew, N. S. Kern, W. E. Kerzendorf, A. Khvalko, J. King, D. Kirkby, A. M. Kulkarni, A. Kumar, A. Lee, D. Lenz, S. P. Littlefair, Z. Ma, D. M. Macleod, M. Mastropietro, C. McCully, S. Montagnac, B. M. Morris, M. Mueller, S. J. Mumford, D. Muna, N. A. Murphy, S. Nelson, G. H. Nguyen, J. P. Ninan, M. Nöthe, S. Ogaz, S. Oh, J. K. Parejko, N. Parley, S. Pascual, R. Patil, A. A. Patil, A. L. Plunkett, J. X. Prochaska, T. Rastogi, V. Reddy Janga, J. Sabater, P. Sakurikar, M. Seifert, L. E. Sherbert, H. Sherwood-Taylor, A. Y. Shih, J. Sick, M. T. Silbiger, S. Singanamalla, L. P. Singer, P. H. Sladen, K. A. Sooley, S. Sornarajah, O. Streicher, P. Teuben, S. W. Thomas, G. R. Tremblay, J. E. H. Turner, V. Terrón, M. H. van Kerkwijk, A. de la Vega, L. L. Watkins, B. A. Weaver, J. B. Whitmore, J. Woillez, V. Zabalza e (Astropy Contributors. The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package. *AJ*, 156:123. doi: 10.3847/1538-3881/aabc4f. Citado na pág. 21

Rothman *et al.*(2009) Laurence S Rothman, Iouli E Gordon, Alain Barbe, D Chris Benner, Peter F Bernath, Manfred Birk, Vincent Boudon, Linda R Brown, Alain Campargue, J-P Champion *et al.* The hitran 2008 molecular spectroscopic database. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 110(9-10):533–572. Citado na pág. 9

Salvador e Chan(2007) Stan Salvador e Philip Chan. Toward accurate dynamic time warping in linear time and space. *Intelligent Data Analysis*, 11(5):561–580. Citado na pág. 11, 14

Shou *et al.*(2005) Yutao Shou, Nikos Mamoulis e David W Cheung. Fast and exact warping of time series using adaptive segmental approximations. *Machine Learning*, 58(2-3):231–267. Citado na pág. 11

Smette *et al.*(2015) A Smette, Hugues Sana, S Noll, H Horst, W Kausch, S Kimeswenger, M Barden, C Szyszka, AM Jones, A Gallenne *et al.* Molecfit: A general tool for telluric absorption correction-i. method and application to ESO instruments. *Astronomy & Astrophysics*, 576:A77. Citado na pág. 8, 9, 22

Wikipedia contributors(2019a) Wikipedia contributors. Photon — Wikipedia, the free encyclopedia, 2019a. URL <https://en.wikipedia.org/w/index.php?title=Photon&oldid=922319569>. [Online; accessed 29-October-2019]. Citado na pág. 3

Wikipedia contributors(2019b) Wikipedia contributors. Telluric contamination — Wikipedia, the free encyclopedia, 2019b. URL https://en.wikipedia.org/w/index.php?title=Telluric_contamination&oldid=912083597. [Online; accessed 29-October-2019]. Citado na pág. 3