



DOI: 10.5335/rbca.v11i2.9179 Vol. 11, Nº 2, pp. 86-99

Homepage: seer.upf.br/index.php/rbca/index

ARTIGO ORIGINAL

Estratégia de dupla filtragem utilizando o filtro de Savitzky-Golay em dados espectrais de supernovas

Luis Ricardo Arantes Filho¹, Lamartine Nogueira Frutuoso Guimarães^{1,2}, Francisca Joamila Brito do Nascimento¹ and Reinaldo Roberto Rosa¹

¹Instituto Nacional de Pesquisas Espaciais – INPE and ²Instituto de Estudos Avançados – IEAv *luisricardoengcomp@gmail.com; lamar.guima@gmail.com; francisca.nascimento@inpe.br; rrrosa.inpe@gmail.com····

Recebido: 08/03/2019. Revisado: 26/06/2019. Aceito: 05/07/2019.

Resumo

O desenvolvimento de aplicações que lidam com processamento de sinais deve considerar a qualidade dos dados. Técnicas de aprendizado de máquina e técnicas estatísticas requerem ajustes e normalizações no conjunto de dados antes da análise de um dado fenômeno. Quando um conjunto de dados não é tratado para reduzir inconsistências e ruídos fornecidos por instrumentos ou por condições naturais, a análise acrescenta uma tendência, ou seja, os resultados não podem ser reproduzidos porque o conjunto de dados recebe inconsistências condicionadas pelo ruído. Neste sentido, o trabalho a seguir apresenta um sistema para processamento de sinais e ajuste de dados, utilizando como estudo de caso a aplicação em dados espectrais de supernovas, para configurar uma normalização automática e uniforme em grandes conjuntos de dados. Este trabalho propõe uma estratégia de dupla filtragem utilizando o filtro Savitzky–Golay para otimização da redução de ruído. Este sistema produz um sinal filtrado capaz de garantir a otimização em alguns parâmetros, como o deslocamento mínimo do comprimento de onda comparado ao sinal original, a manutenção da forma do sinal original, a atenuação do ruído no sinal e a qualidade na busca por mínimos e máximos locais. A comparação com outras estratégias da literatura salienta a eficiência do sistema e sua aplicabilidade para diversos tipos de sinais e espectros

Palavras-Chave: dados espectrais; dupla filtragem; filtro de Savitzky-Golay; redução de ruído;

Abstract

The development of applications that deal with signal processing should consider data quality. Machine Learning and Statistical techniques require adjustments and normalizations on the dataset before the analysis of a given phenomenon. When a dataset is not managed to reduce inconsistencies and noises provided by instruments or natural conditions the analysis carries a tendency, i. e, the results cannot be reproduced because the dataset receives several inconsistencies. In this sense, the following work presents a system for signal processing and dataset's adjustment, using as a case study an application on supernovae spectral data, to set up an automatic and uniform normalization in a large and diverse dataset. This work proposes a double filtering strategy using the Savitzky–Golay filter to optimize the noise reduction process. This system produces a filtered signal and ensures the optimization in a few parameters, such as minimum wavelength shifts, the shape maintenance of the original signal, the noise reduction and the quality in the search for local minima and maxima points. The comparison with other strategies from literature highlights the system efficiency and the applicability for several types of signals and spectra.

Key words: double filtering; noise reduction; Savitzky-Golay filter; spectral data;

1 Introdução

Em muitos casos relacionados à aplicação intensiva de técnicas de aprendizado de máquina e inteligência artificial, a qualidade dos dados submetidos a uma análise pode interferir fatalmente na capacidade do modelo em predizer, classificar e identificar. Em um dataset inconsistente com a falta de informações, ruídos e até mesmo informações erradas sobre determinado fenômeno, é necessária uma etapa de pré-processamento (processamento e tratamento de dados para normalizações e remoções de inconsistências antes da submissão a um modelo).

O trabalho de mineração de dados permite facilmente identificar, visualizar e manipular estas informações antes que algum modelo seja utilizado para abstrair conhecimento de determinado fenômeno. Entretanto, quando as informações ruidosas são intrínsecas ao dataset é difícil a manipulação e extração de conhecimento. Exemplos de dados que possuem esta característica estão relacionados a sinais digitais, provindos de instrumentos que extraem informações e dados de diversos tipos de fenômenos da natureza. Alguns fenômenos podem ser relacionados, por exemplo, a séries temporais de ocorrência de precipitação, séries temporais relacionadas a manchas solares (Hathaway et al.; 1999; Chattopadhyay et al.; 2011), espectros de radiação solar, séries de conflitos armados (Pettersson and Wallensteen; 2015), espectros de estrelas (Von Hippel et al.; 1994; Singh et al.; 1998), entre outros.

A discretização destes fenômenos não é simples e é feita por diferentes instrumentos que estão sujeitos a variados tipos de interferências. Determinados instrumentos são mais sensíveis do que outros e desta forma um mesmo tipo de informação, como por exemplo, um mesmo sinal coletado por instrumentos diferentes, pode carregar ruídos e interferências de maneiras distintas. Esta diferença pode ser responsável por uma ambiguidade em que se identificam erradamente o mesmo fenômeno de duas maneiras diferentes. Estas imperfeições inerentes aos dados brutos e aos instrumentos, quando não tratadas, podem se propagar no decorrer da análise ocasionando em uma falsa interpretação do fenômeno. Batista et al. (2003) atentam para a importância do tratamento do dataset no aprendizado de máquina supervisionado e não-supervisionado de forma a não dimensionar a análise de um fenômeno onde nos dados existam inconsistências, erros, e ruídos.

Ainda sob o enfoque das técnicas em aprendizado de máquina e estatística é necessário um processamento sobre um grande volume de informações (Big Data) para que um fenômeno seja avaliado corretamente. Essas técnicas necessitam de muitos exemplos e padrões de fenômenos para inferir uma determinada classificação ou reconhecimento. Em aprendizado de máquina é apropriado citar as Redes Neurais Artificiais (RNA), essas redes necessitam de uma grande quantidade de exemplos (sinais, características, e outras abstrações feitas por um especialista) para identificar e classificar corretamente os dados. Existem na literatura diversos tipos de aplicações das RNA's, no tratamento e classificação de sinais espectrais são citados (Von Hippel et al.; 1994; Singh et al.; 1998; Módolo; 2016). Em técnicas estatísticas, o modelo ARIMA (Autoregressive Integrated Moving Average)

também é sensível ao nível de ruído presente em um sinal ou série temporal, sendo incapaz de predizer ou simular um fenômeno (como uma série de manchas solares) se nas amostras do sinal no tempo t-1 houver a falta de informação ou ruídos (Hathaway et al.; 1999; Chattopadhyay et al.; 2011). Essas abordagens são sensíveis a qualidade dos dados e, por isso, um ruído aleatório pode condicionar a uma classificação ou predição errada.

A criação e desenvolvimento de estratégias que permitam o ajuste de dados, especificamente, em dados em que o ruído é intrínseco e provindo de interferências causadas nos instrumentos é um dos focos deste trabalho. Estratégias desenvolvidas sobre este enfoque tratam do ajuste de dados em diversos tipos de fenômenos, para suavização de séries ruidosas, normalização de séries temporais, ajuste de curvas (curve fitting), etc. Em aplicações voltadas ao setor aeroespacial, alguns trabalhos podem ser citados, como: o ajuste de curvas e suavização de dados de radar para indicar a trajetória de aeronaves e identificar dados ruidosos nas amostras (Júnior et al.; 2014); Kennedy (2016) propõe o desenvolvimento de um sistema para a atenuação de ruído em sistemas de controle; em (Chen et al.; 2004; Bradley et al.; 2007; Hird and Mc-Dermid; 2009) os autores descrevem a aplicação de filtros para o ajuste de dados e reconstrução de séries temporais NDVI (Normalized Difference Vegetation *Index*) para identificação de níveis de desmatamento em dados de satélites.

Esse tipo de análise e normalização de dados, também está presente em aplicações de diversas áreas, como por exemplo: A aplicação de filtros digitais para redução de ruído aleatório e otimização da análise em dados de fenômenos geofísicos (Liu et al.; 2016). A otimização de dados pelo filtro de Savitzky-Golay também é aplicada em Staggs (2005) para a manipulação de dados provindos de calorímetro de cone, para determinar com maior exatidão as taxas de liberação de calor e calor efetivo de combustão de materiais. Linares-Pérez et al. (2014) propõe a otimização e filtragem para otimizar a preparação de uma amostra de dados de espectroscopia a laser.

Em aplicações relacionadas à análise espectral no campo da cosmologia e astronomia, tanto a suavização de dados como a normalização são necessárias para ampliar a acurácia na detecção de um objeto ou para melhorar a análise do espectro óptico em estrelas e explosões de supernovas (Rudnick; 2002; Madgwick et al.; 2003). Tal precisão é requerida para a avaliação correta de fenômenos que podem ser de difícil detecção como a ocorrência de Ondas Gravitacionais, ou serem de ocorrência rara como a explosão de supernovas. Tanto a análise espectral de supernovas para a medição de distâncias astronômicas (Riess et al.; 1998; Perlmutter et al.; 1999) como a detecção de ondas gravitacionais (Abbott et al.; 2016) corresponderam a importantes saltos científicos sendo reconhecidos com o prêmio Nobel de Física em 2011 e 2017 respectivamente. Desta forma, qualquer inconsistência nos dados, mesmo que mínima, acarreta em prejuízo analítico.

Em relação aos modelos para análise de dados astronômicos gerados por aprendizado de máquina e técnicas estatísticas, citamos alguns trabalhos. Módolo (2016) apresenta a classificação espectral de supernovas utilizando RNA's no sistema denominado

CIntIa (Classificador Inteligente de supernovas do tipo Ia), os resultados são expressivos para a classificação de supernovas do tipo Ia (SNIa), entretanto a robustez do modelo não é atingida devido a uma quantidade expressiva de dados que foram eliminados da análise por possuírem altos níveis de ruído e inconsistências, a base de dados original correspondia a 3218 espectros, entretanto foram utilizados apenas 649 espectros (20% de toda a amostra). No âmbito estatístico Harutyunyan et al. (2008) desenvolveram um sistema para classificação de supernovas de forma a combinar modelos de espectros para classificação de tipos, neste trabalho o problema causado pela falta de normalização dos dados é agravado, pois a técnica de correlação estatística para combinar os espectros necessita de dados com mínimos índices de ruído, e, ainda mais, que espectros normalizados (ou seja, espectros suavizados para eliminação de ruído) mantenham a sua característica original de formato, ou seja, não sejam deformados pela filtragem. Tal tarefa mostra-se complexa para unir essas duas características e é neste cenário que se desenvolve o presente trabalho.

Este trabalho apresenta uma estratégia de suavização e remoção de ruídos para o tratamento de dados espectrais garantindo um processamento de dados automático e de maneira uniforme. O sistema desenvolvido foi denominado Sistema de Dupla Filtragem pelo filtro de Savitzky-Golay (SDF-SG). Como estudo de caso aplicamos este esquema em uma base de dados de espectros de supernovas, para propiciar dados mais seguros para uma análise e resolver problemas de inconsistências, como visto nos trabalhos de (Harutyunyan et al.; 2008; Módolo; 2016).

O SDF-SG tem como objetivo gerar um espectro filtrado que garanta a combinação de algumas características, como por exemplo:

- deslocamentos mínimos nos comprimentos de onda do espectro filtrado em relação aos comprimentos de onda do espectro original;
- a manutenção das características de formato do sinal original.
- redução do ruído;
- a qualidade de uma busca gradiente para encontrar apenas os reais mínimos e máximos locais do espectro.

O filtro de Savitzky-Golay foi escolhido por suas características relacionadas à sua simples implementação e custo computacional. Conforme aponta Fernandes (2017), este filtro apresenta uma aritmética simples devido aos poucos coeficientes inteiros gerados no decorrer da implementação. Este baixo custo computacional contribui para a aplicação deste filtro em sistemas embarcados, que, por sua vez, podem ser acoplados diretamente a outros dispositivos de coleta de dados, como espectrômetros e telescópios. Esta característica é importante quando o sistema coleta e analisa dados de maneira autônoma. Outra questão importante reside na fácil padronização do sistema de filtragem, que pode ser feita baseada em uma resposta de frequência específica. A escolha deste tipo de filtragem em detrimento de outras estratégias foi feita mediante a testes para verificar se outras estratégias de filtragem permitiam a combinação da manutenção de características originais dos sinais

aliada a redução do ruido com a correta identificação de picos e vales. Conforme observado nos testes o filtro de Savitzky-Golay foi capaz de combinar todas as características citadas.

A escolha de parâmetros e a construção das etapas de filtragem foram baseadas em algumas estratégias da literatura (Howell et al.; 2001; Harutyunyan et al.; 2008; Shappee et al.; 2012; Módolo; 2016) buscando encontrar os parâmetros que melhor se ajustassem aos dados espectrais. Os resultados mostram comparações com outras estratégias de filtragem e suavização de dados espectrais. Aplicamos esta abordagem para providenciar um sistema genérico e adaptável a diversos tipos de problemas que envolvam a otimização e o tratamento de sinais espectrais.

Este trabalho é organizado da seguinte maneira: A seção 2 descreve brevemente os conceitos matemáticos relacionados ao filtro de Savitzky-Golay e ao processamento de sinais. A seção 3 discute os conceitos de supernovas, indicando a importância do fenômeno, descrevendo o tipo de dados utilizados neste trabalho. A seção 4 refere-se ao desenvolvimento do sistema de dupla filtragem SDF-SG indicando os principais métodos e materiais utilizados. A seção 5 refere-se aos experimentos e resultados da aplicação do SDF-SG e da sua comparação com outras estratégias de filtragem. Finalmente, na seção 6 encontram-se as discussões e conclusões sobre a abordagem.

2 Conceitos sobre o Filtro de Savtizky-

O uso de sistemas de filtragem relaciona-se a um dos problemas clássicos no processamento de sinais. Este problema reside na compensação da distorção e modificação de forma em sinais surgida quando um sinal é capturado e transmitido através de um sistema físico. O objetivo do tratamento de sinais pela filtragem é a questão da transmissão e manipulação de dados sem distorção. Este tratamento propicia uma forma mais segura de avaliação de sinais para a representação de fenômenos físicos.

O filtro de Savitzky-Golay (Savitzky and Golay; 1964; Luo et al.; 2005; Schafer et al.; 2011) é caracterizado por ser um filtro digital de resposta ao impulso finita (FIR, Finite Impulse Response) onde sua resposta ao impulso unitário tem um número finito de amostras não nulas. Este filtro foi proposto em 1964 por Savitzky and Golay (1964) para a construção de um sistema de tratamento de ruídos e análise de componentes químicos em espectros. O filtro Savitzky Golay é desenvolvido segundo a combinação de duas características nos filtros FIR: a equivalência entre a filtragem digital passa-baixa e a suavização de ruído por polinômios. Esta abordagem consiste da interpolação de um conjunto de dados pelo método dos mínimos quadrados.

A execução de um ajuste polinomial por mínimos quadrados é feita pela convolução dos dados de entrada (sinal/espectros) em uma janela de tamanho 2m+1, determinando o valor suavizado do ponto central do conjunto de dados através de uma regressão polinomial de grau k deslocando a janela ponto a ponto até que todo o conjunto de dados seja suavizado. O processo de filtragem é constituído pelo desenvolvimento do método de média ponderada, conforme

visto em (1).

$$s_{j}^{*} = \frac{\sum_{j=-m}^{j=m} c_{i} s_{j+i}}{N}$$
 (1)

Onde:

- s: é o sinal original;
- s_i*: é o resultado da suavização, ou seja, é o valor do sinal suavizado;
- c_i: é o coeficiente para a i-ésima suavização;
- N: é o número de pontos da janela de suavização, este número é impar dado por 2m+1;
- m: é a metade do valor da janela de pontos, ou seja,
- $s_i + i$: é o valor dos coeficientes do filtro de Savitzky-Golay;
- · j: O índice j representa o índice de execução dos dados ordenados no conjunto de dados original.

O valor dos coeficientes de suavização são obtidos pela aplicação do método de mínimos quadrados e pela adoção de um polinômio em uma janela deslizante para se ajustar ao sinal original, ponto a ponto. A modelagem do polinômio pelo metódo dos mínimos quadrados é descrito em (2).

$$f_k(i) = b_0 + b_1 i + b_2 i^2 + \dots + b_k i^k = \sum_{n=0}^k b_n i^n, i \le 2m$$
 (2)

A efetiva suavização é construída extraindo o sinal original do ambiente de ruído enquanto mantém a forma do sinal o mais próximo possível do original. O processo de filtragem obtido pela adoção dos parâmetros de tamanho de janela e de grau de polinômio obedece a seguinte relação: Quanto maior o parâmetro da janela e menor o grau de polinômio o sinal de saída resultante é apresentado como uma suavização rígida, ou seja, o sinal de saída apresenta distorção; Quanto menor o tamanho da janela de pontos e maior o grau do polinômio o sinal resultante de saída é apresentado como um sinal ruidoso.

As escolhas dos parâmetros de janela e de grau de polinômio influenciam no desempenho do filtro, a escolha destes parâmetros para o projetista do filtro deve ser baseada pela resposta em frequência previamente determinada. Em relação ao parâmetro de grau de polinômio, esta escolha deve ser baseada na banda de passagem do filtro, isto é, quanto maior o grau do polinômio maior a largura da banda passante.

Neste trabalho a escolha dos parâmetros de janela e grau de polinômio se deve a análise das estratégias de filtragem pesquisadas (Howell et al.; 2001; Harutyunyan et al.; 2008; Shappee et al.; 2012; Módolo; 2016). As estratégias compreendem diferentes parâmetros de filtragem e resultados diferentes para os mesmos tipos de dados, ou seja, espectros de supernovas. Desta forma, os parâmetros do sistema SDF-SG foram escolhidos no sentido de gerar uma otimização. As estratégias e parâmetros pesquisados são descritos na seção 4.

O aumento da banda passante propicia a ineficiência do filtro para o tratamento de sinais de distribuição gaussiana. Deve-se adotar um equilíbrio

que permita uma suavização coerente garantindo a permanência de informações importantes no sinal. Esta escolha pode ser relacionada: ao tipo de sinal tratado, a resposta do filtro a uma dada frequência preestabelecida, ao tamanho da banda passante, e ao custo computacional Fernandes (2017).

Este trabalho apresenta a construção do sistema **SDF-SG** sob um enfoque computacional, e por isso a conceituação do filtro de Savitzky-Golay com maior profundidade não é tratada. Entretanto providenciamos referências que se aprofundam suficientemente, (Savitzky and Golay; 1964; Staggs; 2005; Schafer et al.; 2011; Liu et al.; 2016; Fernandes; 2017).

Supernovas e Características Espec-3

A classificação espectral de supernovas é um dos desafios encontrados ao longo dos anos na busca pelo entendimento do universo. Supernovas são explosões catastróficas que ocorrem no ciclo final de evolução de uma estrela, esta explosão pode destruir completamente uma estrela dispersando material pelo espaço.

As estrelas possuem um mecanismo que permite o equilíbrio entre os processos de fusão do Hidrogênio e a força gravitacional. Quando uma estrela consome boa parte do Hidrogênio (combustível para os processos de fusão internos), cria-se uma instabilidade entre a pressão (causada pela fusão nuclear) e a força gravitacional contribuindo para expulsão de matéria no espaço e pela fusão de outros elementos como o Hélio, Carbono, Oxigênio até atingir o núcleo de Ferro. Quando se atinge o processo de fusão do núcleo de Ferro a estrela entra em colapso, tal processo é irreversível e todo o envoltório da estrela camadas exteriores compostas pelo Hidrogênio, Hélio, Carbono, Oxigênio, etc.) colide contra o núcleo, que por sua vez ricocheteia toda a matéria para o espaço. Este tipo explosão é denominado supernova de colapso de núcleo e ocorre em estrelas massivas \approx 10 $M\odot$ ($M\odot$, Massas Solares), esta explosão pode ter maior luminosidade do que uma galáxia inteira, supernovas formadas desta maneira são classificadas em três tipos principais: SNIb, SNIc (espectro escasso em Hidrogênio) e SNII (abundante em Hidrogênio) (Filippenko; 1997; Blondin et al.; 2011; Modjaz et al.; 2014). A formação de elementos nas camadas de uma estrela massiva são vistas na Fig. 1.

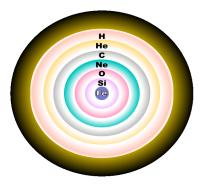
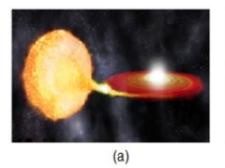
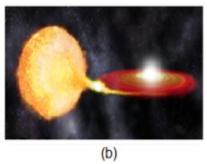


Figura 1: Camadas de Elementos em Estrelas Massivas





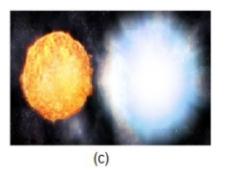


Figura 2: Formação de SNIa. Onde (a) é a interação entre a estrela gigante vermelha e a anã branca, (b) é o acréscimo de massa provindo da absorção do Hidrogênio da gigante vermelha pela anã branca depositando-se no disco de acresção, (c) representa a anã branca atingindo o limite de Chandrasekhar e entrando e colapso para a detonação da SNIa.

Este processo de morte de uma estrela é responsável pela formação de elementos mais complexos, como o Ferro e o Cálcio, essenciais na composição química de organismos complexos no planeta Terra.

Existem formas pela qual uma estrela pode entrar em colapso e explodir em supernova, uma destas formas compreende as reações de acréscimo de massa em estrelas anas brancas (compostas essencialmente de Carbono e Oxigênio em condições degeneradas), este tipo de reação ocorre em sistemas múltiplos onde uma ana branca absorve massa de uma companheira (que pode ser uma estrela na sequência principal, gigantes vermelhas ou anãs brancas). Em determinado ponto do processo, quando a anã branca atinge a massa de \approx 1.4 M \odot , a estrela entra em colapso desencadeando reações termonucleares que destroem completamente a estrela, esta explosão é denominada supernova termonuclear, classificada como SNIa (Filippenko; 1997). Os processos conhecidos de formação de SNIa podem ser vistos na Fig. 2.

A formação das SNIa é bem definida, pois considera o cenário do colapso gerado pelo acréscimo de massa em anãs brancas. Desta forma, é evidenciado o colapso de estrelas de composição e massa bem similares, esta similaridade é vista na análise da curva de luz (luminosidade em função do tempo), e no espectro (fluxo de radiação em função do comprimento de onda). Na Fig. 3 são mostradas as curvas de luz para diferentes SNIa indicando o quanto a luminosidade é parecida nestes objetos (Das; 2008).

A análise de curvas de luz em SNIa demanda um tempo significativo, esta análise pode durar aproximadamente de 40 a 60 dias. A curva de luz revela o início da explosão seguida de um pico de luminosidade máxima e após isso um decaimento gradativo na luminosidade. Em contrapartida, é possível avaliar a composição química da supernova pelo espectro óptico assim que a explosão é detectada, em um período curto de dias. O espectro das SNIa é rico em Silício (Si) e Enxofre (S) nas regiões de comprimento de onda de $\approx \lambda$ 5400 Å a λ 6500 Å. Desta forma, é possível classificar esse objeto pelo espectro identificando as linhas de emissão e absorção (picos e vales) nos comprimentos de onda destes elementos, conforme aponta Filippenko (1997).

Conforme Blondin et al. (2011, 2012), as características espectrais ficam mais evidentes em um período próximo a luminosidade máxima da explosão,

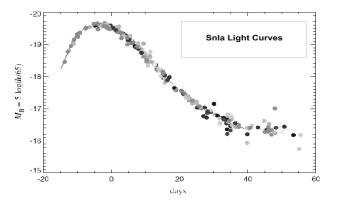


Figura 3: Curvas de Luz de Supernovas SNIa Das (2008)

ou seja, quando a curva de luz atinge seu pico. Após este período as características dos espectros mudam drasticamente, tornando a avaliação fora do período de luminosidade máxima imprecisa. A Fig. 4 ilustra espectros originais (não tratados) de diferentes SNIa da base de dados de Blondin et al. (2011), estes espectros estão em um periodo de +2.5 a -2.5 dias em relação ao pico de luminosidade máxima da curva de luz de SNIa.

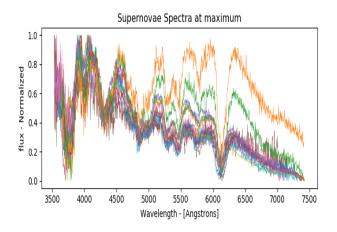


Figura 4: Espectros de Supernovas SNIa Blondin et al. (2011)

A importância da análise das supernovas SNIa está na constatação deste fenômeno como uma fonte de luminosidade bem definida o que caracteriza este objeto como uma vela padrão. Este tipo de constatação é evidenciado na análise da curva de luz, que se apresenta de forma homogênea na maioria dos casos conhecidos. As implicações cosmológicas a respeito deste fenômeno permitiram a constatação da expansão cósmica acelerada do universo. Esta contribuição resultou no prêmio Nobel de Física de 2011 para os astrofísicos Saul Perlmutter, Brian P. Schmidt e Adam G. Riess (Perlmutter et al.; 1999; Riess et al.; 1998).

A complexa análise deste fenômeno no contexto de classificação e avaliação espectral de um grande conjunto de dados necessita de métodos e ferramentas computacionais de forma a executar as operações necessárias em um curto período de tempo sem a necessidade de um especialista para avaliar o modelo. Esta abordagem permite a aplicação dos modelos computacionais em sistemas autônomos operados remotamente, como telescópios localizados em regiões isoladas, garantindo que mesmo na ausência do especialista um grande volume de dados possa ser avaliado da melhor maneira possível. Para que tal abordagem seja bem sucedida é necessário que os dados coletados sejam tratados para garantir a eficiência da avaliação espectral.

Materiais e Métodos

Aplicamos a estratégia SDF-SG em dados espectrais de supernovas por que estes dados estão sujeitos a diversos tipos de inconsistências relacionadas aos instrumentos de coleta (telescópios, espectrógrafos) e interferências naturais como condições atmosféricas, raios cósmicos, x-rays, etc. Essas interferências podem gerar classificações ou identificações condicionadas a um ruído, que muitas vezes pode ser aleatório, esse tipo de característica leva um modelo computacional a uma falha sistemática em que novos dados podem ser descartados por não apresentarem a característica específica do ruído que está presente no processamento do modelo.

Um dos tipos de inconsistências em dados espectrais é a ocorrência de picos e vales em regiões com muitas distorções. Para ilustrar este problema, foi desenvolvido um algoritmo simples em linguagem Python utilizando a biblioteca argrelextrema (Bressert; 2012) para a detecção de picos e vales no sinal. Este algoritmo integra o sistema SDF-SG, como uma busca dos pontos extremos dos espectros para verificar picos e vales sem qualquer restrição de busca, ou seja, são encontrados todos os pontos caracterizados como picos e vales. O código na Fig. 5 indica os passos do algoritmo.

Na Fig. 6 é ilustrada uma busca feita pelos picos e vales na supernova SNIa 1999cc indicando a quantidade de pontos encontrados que prejudicam a análise de elementos do espectro. Foi escolhida uma região no espectro da SNIa em que o Si está presente em abundância, representado por um vale largo próximo ao comprimento de onda de λ 6150 Å.

O SDF-SG busca reduzir a quantidade de pontos de forma a encontrar apenas os picos e vales mais evidentes dos espectros. No espectro de supernova na Fig. 6, a linha espectral mais importante é o

ponto em λ 6150 Å, entretanto como é mostrado na figura, 146 picos e vales foram encontrados. Tais picos e vales são correspondentes ao ruído do espectro. É evidente que outras técnicas de busca podem ser aplicadas, entretanto neste trabalho o objetivo é obter bom resultado na redução do ruído dos espectros e por isso a busca de gradientes não deve ter nenhuma restrição.

Na Tab. 1 é descrita a base de dados utilizada para experimentos e testes do sistema SDF-SG. Esta base de espectros foi coletada do acervo The CfA Supernova Archive (CfA; 2018). Os espectros são dispostos em arquivos no formato ASCII com as colunas para o comprimento de onda em [Å, angstroms] e o fluxo de radiação em [1e-15 erg/s/cm2/A].

Tabela 1: Base de dados de Espectros de Supernovas utilizada nos experimentos do SDF-SG

Principal referência	Tipo de SN	Qtde. Espectros
Blondin et al. (2011)	SNIa	47
Blondin et al. (2012)	SNIa	2603
Matheson et al. (2008)	SNIa	432
Modjaz et al. (2014)	SNIb	217
Modjaz et al. (2014)	SNIc	282
Modjaz et al. (2014)	SNII	116
Total	4 Tipos	3697

Tipo de SN: Tipo de Supernova (SN) existente no banco de dados contendo supernovas de colapso de núcleo e supernovas termonucleares.

Otde. Espectros: Quantidade de espectros de cada super-nova existente no banco de dados.

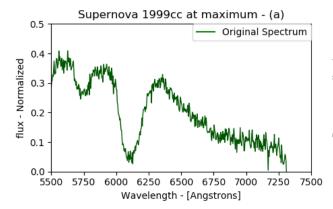
Os dados da Tab. 1 foram utilizados em (Harutyunyan et al.; 2008; Módolo; 2016) e foram introduzidos em diferentes modelos de classificação espectral por técnicas estatísticas e RNA's.

Os espectros contidos nas bases de dados possuem diversidade em relação a:

- quantidades de instrumentos de coleta de dados;
- variação nos comprimentos de onda dos espectros;
- variação em relação as fases das supernovas, compreendendo longos períodos de observação.

```
from scipy.signal import argrelextrema
##Busca de pontos extremos para determinar picos e vales def gradient_Search(x,y):
                                        em x para reorganizar os dados em y
      worstald = np.argsort(x)
x = x[sortId]
y = y[sortId]
# Desta forma, o eixo x corresponde ao indice de x
peak = 0
valley = 0
#Encontre Local Max
       #Encontre Local Max
peak = np.asarray(argrelextrema(y, np.greater))
       #Encontre Local min
valley = np.asarray(argrelextrema(y, np.less))
      #Transpor os vetore
peak = peak.T
valley = valley.T
return peak, valley
```

Figura 5: Código fonte em Python para a pesquisa de pontos extremos do sistema SDF-SG



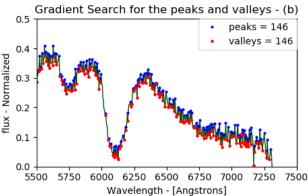


Figura 6: Busca de linhas espectrais. Em (a) é mostrado o espectro original da supernova 1999cc sem tratamento pelo SDF-SG. Em (b) uma busca feita pelo algoritmo descrito na Fig. 5, indicando a quantidade de pontos que representam as linhas espectrais (picos e vales) encontradas.

Esquema de Dupla Filtragem de dados SDF-SG

O sistema SDF-SG foi desenvolvido em linguagem de programação Python e possui três etapas fundamentais: Normalização, Filtragem Simples e Filtragem Dupla.

• Etapa de Normalização: consiste em ajustar os dados brutos de espectros de supernovas para as etapas de filtragem. Inicialmente o espectro é amostrado em 1000 pontos e submetido a uma interpolação linear. Após a interpolação o espectro têm seus valores de fluxo de radiação normalizados na faixa de valores de o a 1 conforme é descrito em (3).

Normalization(y) =
$$\frac{y - min(y)}{max(y) - min(y)}$$
 (3)

Onde y é referente ao valor do fluxo de radiação do espectro. As funções max() e min() correspondem ao maior e menor valor do vetor de y.

- **Etapa de Filtragem Simples**: consiste na primeira filtragem do espectro normalizado na etapa anterior. O espectro é submetido ao filtro de Savitzky-Golay (Savitzky and Golay; 1964) com o valor de janela de pontos igual a 71 (janela ímpar) e com o grau de polinômio igual a 9.
- **Etapa de Filtragem Dupla**: consiste na filtragem do espectro filtrado na etapa anterior. O espectro filtrado é submetido ao filtro de Savitzky-Golay (Savitzky and Golay; 1964) com os mesmos parâmetros da filtragem simples (janela = 71; grau de polinômio = 9).

Na Fig. 7 é ilustrado como funciona o fluxo de informações quando um espectro de supernova é submetido ao sistema SDF-SG.

Os parâmetros escolhidos na modelagem do esquema de filtragem foram inspirados no que foi utilizado nos trabalhos de Howell et al. (2001); Harutyunyan et al. (2008); Shappee et al. (2012); Módolo (2016), nestes trabalhos foram construídas estratégias que permitiram suavizar os espectros a fim de se manter as principais características do sinal para

gerar comparações e indicar correlações com espectros sem que existissem erros nas amostras de dados. A estratégia de filtragem desenvolvida neste trabalho busca otimizar o que foi feito nestes trabalhos. Na Tab. 2 são descritos os filtros e parâmetros utilizados em cada estratégia citada.

Em Howell et al. (2001) (estratégia Howell) foi utilizado um esquema de filtragem para suavizar os parâmetros dos espectros principalmente relacionados ao índice SNR do sinal original, neste trabalho o autor evidencia a dificuldade em fazer a avaliação dos espectros com os altos níveis de ruído e por isso, executa uma estratégia de suavização.

Em Harutyunyan et al. (2008), definidas aqui como estratégias Haruty_A e Haruty_B, os autores utilizaram a suavização do tipo boxcar que usa o filtro moving average. Esta mesma estratégia foi utilizada em Módolo (2016), em que o sistema de filtragem necessitou de ajustes pontuais para cada espectro. O que difere nas estratégias Haruty_A e Haruty_B é o tamanho da janela de pontos, pois cada estratégia foi usada em espectros diferentes nas abordagens propostas. Nesses trabalhos além do processamento de dados indesejáveis (ruído e mudanças morfológicas no sinal), a suavização foi feita de forma diferente para regiões específicas e alguns espectros precisaram de parâmetros diferentes na filtragem, isto é, esse processamento não foi uniforme. Para o desenvolvimento de um sistema automático, seja ele baseado tanto em técnicas de aprendizado de máquina, como em técnicas estatísticas um processamento homogêneo de dados é necessário, principalmente quando o especialista não está disponível para análises e ajustes "manuais".

Em Shappee et al. (2012) (estratégia Shappee), foi utilizada uma estratégia para o cálculo do espectro contínuo de forma a identificar corretamente as características de emissão e absorção do hidrogênio em espectros de supernovas. Para a suavização dos dados foi utilizado o filtro de Savtizky-Golay com uma janela de 61 pontos e polinômio de 2ª ordem, esta estratégia evidencia uma comparação direta ao SDF-SG, pois é baseado no mesmo filtro, entretanto é executada apenas uma filtragem simples.

Desta forma, buscamos solucionar o problema referente a falta de uniformidade na suavização de dados

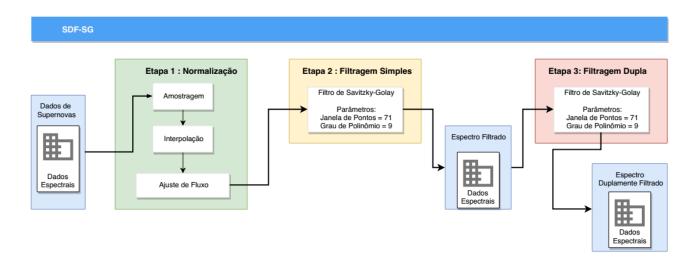


Figura 7: Esquema de Dupla Filtragem SDF-SG

Tabela 2: Estratégias de Filtragem utilizadas na comparação do sistema SDF-SG

Estratégia	Principal Referência	Tipo de Filtro	Janela	Polinômio
Howell	Howell et al. (2001)	Box-Car	65	_
Haruty_A	Harutyunyan et al. (2008); Módolo (2016)	Box-Car	70	-
Haruty_B	Harutyunyan et al. (2008); Módolo (2016)	Box-Car	10	-
Shappee	Shappee et al. (2012)	Savitzky-Golay	61	2

Estratégia: nome dado a estratégia para fins de simplificação.

Janela: valor da janela de pontos utilizada no filtro.

Polinômio: valor do grau de polinômio aplicado ao filtro. neste caso, apenas o filtro de Savitzky-Golay utiliza este atributo.

e remoção do ruído nos espectros de supernovas encontrados nos trabalhos de Harutyunyan et al. (2008); Módolo (2016).

4.2 Métricas Utilizadas

Esta seção descreve as principais métricas para avaliar o desempenho das estratégias de filtragem pesquisadas em comparação com a estratégia desenvolvida neste trabalho. As métricas são descritas nos seguintes itens:

i Correlação cruzada: correlação entre os sinais filtrados e o sinal original que mede a similaridade dos sinais como uma função do deslocamento de um em relação a outro. A correlação cruzada entre os sinais é dada pela convolução de duas funções (sinais). Em (4) e (5) descrevemos a função e sua respectiva normalização para valores de o a 1.

$$(f*g)[n] = \sum_{m=-\infty}^{\infty} f*[m]g[n+m]$$
 (4)

$$\rho_{XY} = \frac{(f * g)[n]}{\sqrt{(f * f)[0].(g * g)[0]}}$$
(5)

Onde (f * f)[0] é o primeiro valor da autocorrelação cruzada do sinal original, e (g * g)[0] é o primeiro valor da autocorrelação cruzada do sinal modificado (filtrado).

ii Correlação de Spearman e de Pearson: estes ín-

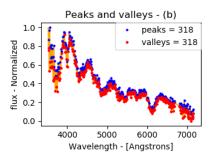
dices de correlação foram aplicados para identificar a manutenção das características originais dos espectros, avaliando se no decorrer das filtragens houveram a deformações do sinal original. A correlação de Spearman é uma medida não paramétrica da relação de monotonicidade entre dois conjuntos de dados. A correlação de Pearson mede as relações lineares entre dois conjuntos de dados (Hauke and Kossowski; 2011).

- iii Distância entre os sinais: este índice é medido de acordo com a correlação cruzada entre os sinais ((4) e (5)). Este índice revela o quanto o sinal filtrado se deslocou em relação ao sinal original.
- iv Root Mean Square Error (RMSE): este indice revela a comparação entre os sinais filtrados (simples e duplamente filtrados) e o sinal original. Essa métrica é utilizada para verificar as diferenças entre os valores preditos por um modelo (sinal filtrado) e os valores realmente observados do ambiente que está sendo modelado (Willmott; 1982).

Experimentos e Resultados

Esta seção aborda um experimento para o espectro de supernova SN 1998dx, este espectro teve seu índice SNR calculado como 5.099[dB] contendo 318 picos e 318 vales no sinal original. O sinal original e a busca de picos e vales são ilustrados na Fig.8. Na Fig. 8 também consta a região de comprimento de onda de λ 5500 Å λ 7500 Å do espectro para ilustrar como estão localizados picos e vales em regiões ruidosas.

Na Tab. 3 constam as respostas do SDF-SG em relação às métricas citadas na seção 4.2. Essas métri-



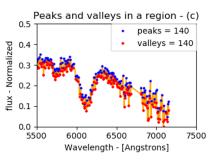


Figura 8: Supernova utilizada no experimento com o sistema SDF-SG

cas visam relacionar o desempenho das estratégias de filtragem em relação aos parâmetros de qualidade.

O desempenho medido em relação aos índices de correlação revelam a similaridade do espectro filtrado em relação ao espectro original, ou seja, quanto mais próximo de 100% maior é a similaridade entre os sinais. É possível inferir ainda, que, estes índices revelam o quanto o sinal sofreu deformações após a filtragem. O sistema SDF-SG com a dupla filtragem obteve os índices de 99,6%, para a correlação cruzada entre o sinal original e o sinal duplamente filtrado. Os índices de Pearson e de Sperman que medem a relação de similaridade entre conjuntos de dados também obtiveram os valores de 98,47% e de 98,81%, respectivamente. Neste ponto verifica-se que o sinal filtrado obteve modificações mínimas em relação ao espectro original.

A comparação com as outras estratégias revelam que os sinais sofreram modificações importantes em seu formato original. A estratégia de Shappee que se refere a uma filtragem simples com o filtro de Savitzky-Golay apresenta índices de correlação na faixa de 70% a 80% de modificação em relação ao sinal original. Esta consideração revela que a filtragem dupla permite contornar modificações importantes do sinal removendo apenas o ruído. As estratégias relacionadas aos filtros boxcar (estratégia Howell, estratégia Haruty_A e estratégia Haruty_B) demonstram a existência de modificações no sinal.

Em relação as estratégias propostas por Harutyunyan et al. (2008) e utilizadas por Módolo (2016)(Haruty_A e Haruty_B), a modificação do sinal original é uma característica que prejudica o desempenho das abordagens de classificação de supernovas. No sistema gerado por Harutyunyan et al. (2008), o sinal modificado deveria ser correlacionado e combinado com sinais "modelo"em uma base de dados para gerar classificações de tipos de espectros. Desta forma, quando são correlacionados sinais com as características de espectros de supernovas deformados pela suavização existe um erro sistemático. Na Fig. 9 é ilustrado o resultante das suavizações feitas pelas

estratégias de filtragem no experimento.

As deformações geradas pelas estratégias de $Haruty_A$ e de $Haruty_B$ indicam que o tamanho da janela de pontos influenciou a transformação do sinal original. Na estratégia de Howell onde a janela de pontos é igual a 65, o sinal filtrado possui melhor correlação com o sinal original, entretanto não é capaz de atenuar o ruído existente. Na Fig. 10 é ilustrado o resultante das estratégias de filtragem na região de λ 5500 Å a λ 7500 Å, esta região foi escolhida por apresentar forte absorção do Silício em SNIa.

Conforme observado na Tab. 3 a filtragem feita pelo **SDF-SG** é a que apresenta a melhor performance em atenuar o ruído e disponibilizar um sinal com as características principais dos espectros de supernova em relação ao formato das linhas espectrais de emissão e absorção. A filtragem dupla também proporciona o menor deslocamento em relação ao sinal original. A estratégia de Howell, além de indicar uma deformação do sinal ainda é capaz de gerar um deslocamento. Em relação ao tratamento do ruído no espectro filtrado é verificado o índice RMSE. A estratégia de Shappee e a estratégia de Haruty_B por possuírem tamanho de janela menor, indicam a capacidade de deformar o espectro. A estratégia SDF-SG possui o melhor desempenho, pois o sinal filtrado obtém o menor erro em relação ao sinal original sendo a única estratégia capaz de combinar e otimizar os parâmetros de correlação, distância e de erro.

Para verificar a qualidade na atenuação do ruído, foi feita uma busca pelas linhas espectrais (picos e vales) na mesma região ilustrada na Fig. 10. A Fig. 11 revela a capacidade de se escolher apenas os mínimos e máximos locais do sinal, não escolhendo pontos caracterizados como ruído como ocorre na Fig. 8 e na Fig. 6.

Os resultados descritos pelas métricas revelam a importância da estratégia de filtragem construída no sistema SDF-SG em localizar apenas os picos e vales mais importantes no sinal, ou seja, picos e vales importantes para o especialista no momento da análise de características das linhas espectrais como

Tabela 3: Desempenho das Estratégias de Filtragem para o Espectro de supernova SN 1998dx

Métrica	SDF-SG	Howell	$Haruty_A$	$Haruty_B$	Shappee
Correlação Cruzada	99,63%	98,60%	94,85%	93,69%	93,99%
Pearson	98,47%	94,15%	78,68%	75,01%	76,06%
Spearman	98,81%	96,81%	81,20%	78,47%	79,35%
Distância Entre os Sinais	1,53e-02	5,85e-02	2,13e-01	2,50e-01	2,39e-01
RMSE	3,69e_02	7,20e_02	1,40 <i>e</i> _01	1,57e_01	1,54e_01

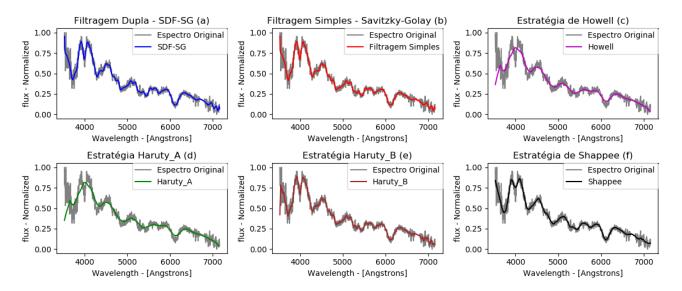


Figura 9: Resultante da aplicação das estratégias de filtragem. Onde (a) é referente ao sistema SDF-SG, (b) é referente a uma filtragem simples pelo filtro de Savitzky-Golay com os mesmos parâmetros definidos na Fig. 7. Os quadros (c,d,e,f) correspondem ao resultante das estratégias da Tab. 2.

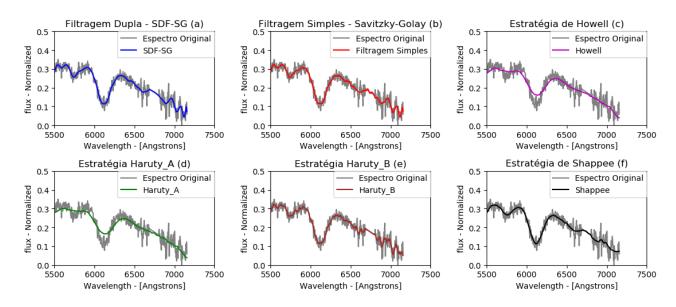


Figura 10: Resultante da aplicação das estratégias de filtragem na região de λ 5500 Å a λ 7500 Å. Onde (a) é referente ao sistema SDF-SG, (b) é referente a uma filtragem simples pelo filtro de Savitzky-Golay com os mesmos parâmetros definidos na Fig. 7. Os quadros (c,d,e,f) correspondem ao resultante das estratégias da Tab. 2.

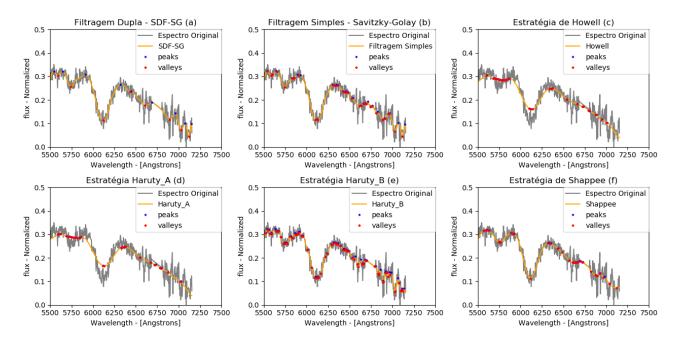


Figura 11: Resultante da busca de picos e vales na região do espectro de λ 5500 Å a λ 7500 Å. Onde (a) é referente ao sistema SDF-SG, (b) é referente a uma filtragem simples pelo filtro de Savitzky-Golay com os mesmos parâmetros definidos na Fig. 7. Os quadros (c,d,e,f) correspondem a busca de picos e vales referentes às estratégias da Tab. 2.

suas larguras e intensidades. Como é observado na Fig. 11, as outras estratégias falham consideravelmente em relacionar corretamente os picos e vales do espectro, estas estratégias de filtragem em sinais espectrais de supernova não permitem que exista a relação pico-vale-pico, isto é, não permitem identificar computacionalmente a localização das linhas de emissão e absorção.

A estratégia de busca usada neste trabalho não elimina nenhum pico ou vale existente no sinal, indicando todos os mínimos e máximos locais. Outras técnicas são utilizadas para este fim, entretanto a escolha de parâmetros para buscar os mínimos e máximos globais acaba eliminado informações que estão fora dos intervalos de busca. Por isso, pontos importantes podem ser removidos da análise. A estratégia SDF-SG é a única estratégia que permite observar apenas os reais picos e vales em um sinal, desprezando pontos em regiões ruidosas, todas as outras estratégias pesquisadas encontram pontos ruidosos.

5.1 Experimento para Todo Conjunto de dados

Em relação ao conjunto completo de dados (3697 espectros) 1, também foram computados a média, mediana, o desvio padrão e a variância do desempenho das estratégias de filtragem em relação aos índices de correlação cruzada. Foi verificado o deslocamento médio dos sinais filtrados em relação aos sinais originais além da capacidade de remoção de inconsistências medida, por sua vez, pelo índice RMSE. A Tab. 4, indica o desempenho total referente às métricas obtidos pelas estratégias para a correlação de similaridades.

Os valores resultantes para os 3697 espectros demonstram que o sistema SDF-SG não deforma os sinais, mantendo a similaridade média em 99,63% em relação aos sinais originais. As outras estratégias não possuem esta característica, logo o SDF-SG possui um resultado superior. Destaca-se que a amostra de dados compreende espectros de diversos tipos de supernovas e ainda assim o SDF-SG permite a correta suavização sem a deformação do sinal.

Este experimento indica o bom desempenho em relação às métricas descritas do SDF-SG para uma grande variedade de tipos de sinais tornando a aplicabilidade deste sistema adequada quando o foco da abordagem é remover o ruído e escolher corretamente os picos e vales do sinal. Na Tab 5 são descritos os resultados das estratégias em relação ao deslocamento dos sinais após a filtragem.

Os resultados referentes ao deslocamento provocado pela filtragem indica que o SDF-SG se destaca em relação às outras estratégias. Isto significa que o processo de filtragem mantém o sinal com deslocamentos mínimos reduzindo o erro na filtragem. As outras estratégias também revelam deslocamentos mínimos no sinal, entretanto não existe a combinação entre a manutenção das características originais do espectro. Na espectroscopia este deslocamento é um fator crucial em determinar a existência de elementos químicos em comprimentos de onda específicos, logo, grandes deslocamentos tornam esta análise imprecisa. Em relação a medir o quanto as estratégias de filtragem removem ruído da amostra foi utilizado o índice RMSE. Na Tab. 6 são descritos os resultados das estratégias em relação a este índice.

Em relação a Tab. 6, algumas estratégias como a de $Haruty_A$, Howell e Shappee demonstram erros muito baixos (RMSE), entretanto a estratégia de dupla filtragem é evidenciada como possuindo o menor índice médio. Isto significa que os sinais produzidos pelo SDF-SG possuem menos inconsistências que os

Tabela 4: Resultado Médio da Correlação Cruzada nas Estratégias de Filtragem para Todo Conjunto de Dados

Correlação Cruzada	SDF-SG	Howell	$Haruty_A$	$Haruty_B$	Shappee
Média	99,63%	98,45%	86,75%	85,71%	85,79%
Mediana	99,82%	98,68%	87,95%	86,63%	86,70%
Desvio Padrão	1,52e-02	1,97e-02	8,31e-02	8,54e-02	8,59e-02
Variância	2,30e_04	3,90e_04	6,91e_03	7,30e_03	2,30e_04

Tabela 5: Resultado Médio da Distância Entre os Sinais nas Estratégias de Filtragem para Todo Conjunto de Dados

Distância Entre Sinais	SDF-SG	Howell	$Haruty_A$	Haruty _B	Shappee
Média	4,81e-02	1,22e-01	6,65e-01	6,79e-01	6,81e-01
Mediana	9,69e_03	7,57e_02	5,63e_01	5,92e_01	5,95e_01
Desvio Padrão	9,07e_02	1,32e_01	4,60e_01	4,50e_01	4,55e_01
Variância	8,22e_03	1,75e_02	$2,12e_{-}01$	2,03e_01	$2,07e_{-}01$

Tabela 6: Resultado Médio do RMSE nas Estratégias de Filtragem para Todo Conjunto de Dados

RMSE	SDF-SG	Howell	$Haruty_A$	Haruty _B	Shappee
Média	3,21e_02	7,36e_02	2,39e_01	2,50e_01	2,49e_01
Mediana	2,81e_02	7,50e_02	2,32e_01	2,45e_01	2,45e_01
Desvio Padrão	1,89e_02	1,49e_02	9,24e-02	9,35e_02	9,42e_02
Variância	3,56e-04	2,21e-04	8,54e-03	8,74e-03	8,86e-03

sinais produzidos pelas outras estratégias.

Conclusões

O que se propôs neste trabalho foi o desenvolvimento de uma estratégia de filtragem capaz de se adequar aos seguintes itens:

- a manutenção da forma original do sinal;
- · deslocamento mínimo dos comprimentos de onda do sinal original;
- · a atenuação do ruído no sinal;
- · e a qualidade da busca de gradientes relacionada aos mínimos e máximos locais, ou seja, para a análise correta de cada pico e vale do sinal.

As estratégias utilizadas nos trabalhos (Howell et al.; 2001; Harutyunyan et al.; 2008; Shappee et al.; 2012; Módolo; 2016), não produziram bons resultados em relação aos itens citados. Nos trabalhos de Harutyunyan et al. (2008) e Módolo (2016) foi revelado que a estratégia de suavização boxcar gerou deformações importantes no sinal, nestes trabalhos existe o aprendizado e combinações de padrões para classificações de tipos de espectros. Deformações desta natureza comprometem seriamente este tipo de análise, pois os sinais aprendidos não possuem as suas características naturais e sim as artificias geradas pela filtragem. A dupla filtragem gerada neste trabalho é uma alternativa para a suavização nestes trabalhos pois produz sinais filtrados com características próximas dos sinais originais removendo as inconsistências e ruídos.

Um dos pontos mais importantes da estratégia de dupla filtragem também é encontrar corretamente os mínimos e máximos locais em um sinal. Este tipo de abordagem quando aplicada na área de espectroscopia permite que se avalie correta e unicamente as linhas espectrais importantes do sinal que carregam informações sobre a composição química de um espectro.

Em relação a correta manipulação de dados espectrais de supernovas a estratégia de dupla filtragem SDF-SG contribuiu para o desenvolvimento e aprimoramento do sistema CIntIa (Módolo; 2016) que avaliou apenas 649 espectros devido a inconsistências na base de dados. O sistema CINTIA 2 (do Nascimento et al.; 2019) analisa e classifica uma base de dados muito ampla e compreende um universo de discurso mais robusto do que o utilizado na versão inicial. CINTIA 2 avalia cerca de 9156 espectros dos quatro tipos principais de supernovas (SNIa, SNIb, SNIc e SNII). A eliminação de ruído e a otimização dos espectros pelo SDF-SG possibilitou que espectros não fossem descartados e que o treinamento por RNA's fosse feito de maneira mais segura, pois os dados filtrados possuíam índices mínimos de ruído e ainda sim mantinham as características principais dos espectros originais. Desta forma, os resultados obtidos com a classificação de supernovas pela CINTIA 2 possuem uma qualidade estatística superior, atingindo uma acurácia média de 95% (do Nascimento et al.; 2019).

Pelos índices médios e pela variação do dataset utilizado, é possível evidenciar o desempenho do SDF-SG para uma grande variedade de sinais. Este trabalho encoraja testes para sinais de outras naturezas, como referentes às áreas de clima espacial, lasers, telecomunicações, etc.

Devido a abordagem proporcionada e aos resultados obtidos, este trabalho cumpre seus objetivos. Haja vista que propicia contribuição referente à estratégias de filtragem e suavização em sinais e espectros.

Agradecimentos

Agradecemos ao CNPq e a CAPES pelo apoio financeiro. Agradecemos ao Instituto Nacional de Pesquisas Espaciais (INPE) e ao Instituto de Estudos Avançados (IEAv) pela estrutura e suporte computacional.

Referências

- Abbott, B. P., Abbott, R., Abbott, T., Abernathy, M., Acernese, F., Ackley, K., Adams, C., Adams, T., Addesso, P., Adhikari, R. et al. (2016). Observation of gravitational waves from a binary black hole merger, Physical review letters 116(6): 061102.
- Batista, G. E. d. A. P. et al. (2003). Pré-processamento de dados em aprendizado de máquina supervisionado, PhD thesis, Universidade de São Paulo.
- Blondin, S., Mandel, K. S. and Kirshner, R. P. (2011). Do spectra improve distance measurements of type ia supernovae?, Astronomy & Astrophysics **526**: A81.
- Blondin, S., Matheson, T., Kirshner, R. P., Mandel, K., Berlind, P., Calkins, M., Challis, P., Garnavich, P. M., Jha, S., Modjaz, M. et al. (2012). The spectroscopic diversity of type ia supernovae, The Astronomical Journal **143**(5): 126.
- Bradley, B. A., Jacob, R. W., Hermance, J. F. and Mustard, J. F. (2007). A curve fitting procedure to derive inter-annual phenologies from time series of noisy satellite ndvi data, Remote sensing of environment **106**(2): 137-145.
- Bressert, E. (2012). SciPy and NumPy: an overview for developers, "O'Reilly Media, Inc."
- CfA (2018). The CfA Supernova Archive, Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge. Available at https://www. cfa.harvard.edu/supernova/SNarchive.html (Accessed 19 June 2018).
- Chattopadhyay, S., Jhajharia, D. and Chattopadhyay, G. (2011). Trend estimation and univariate forecast of the sunspot numbers: development and comparison of arma, arima and autoregressive neural network models, Comptes Rendus Geoscience **343**(7): 433-442.
- Chen, J., Jönsson, P., Tamura, M., Gu, Z., Matsushita, B. and Eklundh, L. (2004). A simple method for reconstructing a high-quality ndvi time-series data set based on the savitzky-golay filter, Remote sensing of Environment **91**(3-4): 332-344.
- Das, S. (2008). Aspects of quintessence matter the driver of the late time acceleration of the universe.
- do Nascimento, F. J. B., Arantes Filho, L. R. and Guimarães, L. N. F. (2019). Intelligent classification of supernovae using artificial neural networks, Inteligencia Artificial **22**(63): 39–60. DOI: https: //doi.org/10.4114/intartif.vol22iss63pp39-60.
- Fernandes, B. d. A. (2017). Utilização do filtro de savitzky-golay no cancelamento de interferência de 60-hz em sinais de eletrocardiograma.
- Filippenko, A. V. (1997). Optical spectra of supernovae, Annual Review of Astronomy and Astrophysics **35**(1): 309-355.
- Harutyunyan, A., Pfahler, P., Pastorello, A., Taubenberger, S., Turatto, M., Cappellaro, E., Benetti, S., Elias-Rosa, N., Navasardyan, H., Valenti, S. et al. (2008). Esc supernova spectroscopy of non-esc targets, Astronomy & Astrophysics 488(1): 383–399.

- Hathaway, D. H., Wilson, R. M. and Reichmann, E. J. (1999). A synthesis of solar cycle prediction techniques, Journal of Geophysical Research: Space Physics 104(A10): 22375-22388.
- Hauke, J. and Kossowski, T. (2011). Comparison of values of pearson's and spearman's correlation coefficients on the same sets of data, Quaestiones geographicae **30**(2): 87-93.
- Hird, J. N. and McDermid, G. J. (2009). Noise reduction of ndvi time series: An empirical comparison of selected techniques, Remote Sensing of Environment **113**(1): 248-258.
- Howell, D. A., Höflich, P., Wang, L. and Wheeler, J. C. (2001). Evidence for asphericity in a subluminous type ia supernova: spectropolarimetry of sn 1999by, The Astrophysical Journal 556(1): 302.
- Júnior, B., Bezerra, A. and Pires, P. S. d. M. (2014). An approach to outlier detection and smoothing applied to a trajectography radar data, Journal of Aerospace Technology and Management 6(3): 237–248.
- Kennedy, H. L. (2016). An adaptive digital filter for noise attenuation in sampled control systems, *In*ternational Journal of Adaptive Control and Signal Processing 30(6): 928-938.
- Linares-Pérez, J., Caballero-Águila, R. and García-Garrido, I. (2014). Optimal linear filter design for systems with correlation in the measurement matrices and noises: Recursive algorithm and applications, International Journal of Systems Science **45**(7): 1548-1562.
- Liu, Y., Dang, B., Li, Y., Lin, H. and Ma, H. (2016). Applications of savitzky-golay filter for seismic random noise reduction, Acta Geophysica **64**(1): 101–
- Luo, J., Ying, K., He, P. and Bai, J. (2005). Properties of savitzky-golay digital differentiators, Digital Signal Processing **15**(2): 122–136.
- Madgwick, D. S., Hewett, P. C., Mortlock, D. J. and Wang, L. (2003). Spectroscopic detection of type ia supernovae in the sloan digital sky survey, The Astrophysical Journal Letters **599**(1): L33.
- Matheson, T., Kirshner, R. P., Challis, P., Jha, S., Garnavich, P., Berlind, P., Calkins, M. L., Blondin, S., Balog, Z., Bragg, A. et al. (2008). Optical spectroscopy of type ia supernovae, The Astronomical Journal 135(4): 1598.
- Módolo, M. (2016). Classificação Automática de Supernovas Usando Redes Neurais Artificiais, PhD thesis, Instituto Nacional de Pesquisas Espaciais. **URL:** http://mtc-m21b.sid.inpe.br/col/sid.inpe.br/mtcm21b/2016/04.20.20.44/doc/publicacao.pdf
- Modjaz, M., Blondin, S., Kirshner, R. P., Matheson, T., Berlind, P., Bianco, F. B., Calkins, M. L., Challis, P., Garnavich, P., Hicken, M. et al. (2014). Optical spectra of 73 stripped-envelope core-collapse supernovae, The Astronomical Journal **147**(5): 99.
- Perlmutter, S., Aldering, G., Goldhaber, G., Knop, R., Nugent, P., Castro, P., Deustua, S., Fabbro, S.,

- Goobar, A., Groom, D. et al. (1999). Measurements of ω and λ from 42 high-redshift supernovae, The Astrophysical Journal **517**(2): 565.
- Pettersson, T. and Wallensteen, P. (2015). med conflicts, 1946–2014, Journal of peace research **52**(4): 536-550.
- Riess, A. G., Filippenko, A. V., Challis, P., Clocchiatti, A., Diercks, A., Garnavich, P. M., Gilliland, R. L., Hogan, C. J., Jha, S., Kirshner, R. P. et al. (1998). Observational evidence from supernovae for an accelerating universe and a cosmological constant, The Astronomical Journal 116(3): 1009.
- Rudnick, L. (2002). Simple multiresolution filtering and the spectra of radio galaxies and supernova remnants, Publications of the Astronomical Society of the Pacific **114**(794): 427.
- Savitzky, A. and Golay, M. J. (1964). Smoothing and differentiation of data by simplified least squares procedures., Analytical chemistry **36**(8): 1627–1639.
- Schafer, R. W. et al. (2011). What is a savitzky-golay filter, IEEE Signal processing magazine **28**(4): 111–117.
- Shappee, B. J., Stanek, K., Pogge, R. and Garnavich, P. (2012). No stripped hydrogen in the nebular spectra of nearby type ia supernova 2011fe, The Astrophysical Journal Letters 762(1): L5.
- Singh, H. P., Gulati, R. K. and Gupta, R. (1998). Stellar spectral classification using principal component analysis and artificial neural networks, Monthly Notices of the Royal Astronomical Society 295(2): 312-
- Staggs, J. (2005). Savitzky-golay smoothing and numerical differentiation of cone calorimeter mass data, Fire Safety Journal **40**(6): 493-505.
- Von Hippel, T., Storrie-Lombardi, L., Storrie-Lombardi, M. and Irwin, M. (1994). Automated classification of stellar spectra-i. initial results with artificial neural networks, Monthly Notices of the Royal Astronomical Society **269**(1): 97–104.
- Willmott, C. J. (1982). Some comments on the evaluation of model performance, Bulletin of the American *Meteorological Society* **63**(11): 1309–1313.