

Relatório MC040 - Estudo do Espaço de Busca

José Ribeiro Neto¹, Edson Borin¹,
Tiago A. Coimbra¹, Jorge H Faccipieri¹

¹Centro de Estudos de Petróleo (CEPETRO)
Universidade Estadual de Campinas (UNICAMP)
Caixa Postal 6052 – 13.083-970 – Campinas – SP – Brasil

j176665@g.unicamp.br, edson@ic.unicamp.br,
tiago@ggaunicamp.com, jorge@ggaunicamp.com

Resumo

A meta-heurística *Differential Evolution* é usada na análise de dados sísmicos para buscar parâmetros de equações de tempos de trânsito de forma rápida e eficiente. Entretanto, a confiabilidade dos parâmetros obtidos por essa heurística pode estar comprometida devido aos problemas que surgem em situações de baixa razão sinal/ruído. Por outro lado, uma confiança elevada nessas parâmetros torna mais assertivas as decisões de um geofísico. Neste trabalho, o estudo dos motivos que geram essa baixa confiança é feita através da construção de uma nova medida. O estudo é realizado com o auxílio de ferramentas desenvolvidas em *Python*, tendo como principal funcionalidade a geração dos espaços de busca desses parâmetros e a correlação dos mesmos com os seus vizinhos. Por consequência, tal correlação resulta em um novo espaço de busca, sendo este mais estável em relação aos originais, em que tanto a forma como os valores que o compõe são avaliados. Portanto, existe uma semelhança direta entre seu formato e alguns de seus vizinhos, tornando o formato do espaço de busca diretamente relacionado ao número de eventos que intersectam o mesmo ponto sobre o dado. Possibilitando-nos, desse modo, o aumento da confiança do valor obtido pela heurística. Por fim, resultados numéricos confirmam que a abordagem proposta tem potencial para elevar a qualidade dos resultados nas tomadas de decisões.

1 Introdução

No processamento sísmico diversas técnicas são desenvolvidas com o intuito de melhorar a razão sinal / ruído (S / N) do dado sísmico, tendo-se particular atenção para duas dessas: *CRS* [1] e *CRP* [2]. Tais técnicas fazem uso de meta-heurísticas com o intuito de encontrar rapidamente os parâmetros que melhor se aproximam aos eventos presentes no dado, de forma a maximizar a razão S / N, sem que para tanto sejam introduzidos muitos artefatos no resultado final. E dentre as possíveis heurísticas, é comum o uso de uma já amplamente conhecida na literatura, chamada *Differential Evolution (DE)* [3]. O DE apresenta uma alta taxa de convergência, na busca dos parâmetros associados as técnicas CRS e CRP, contudo uma baixa certeza de obtenção de parâmetros associados a eventos do dado - condição esta justificada pelo seu caráter estocástico e pelo desconhecimento à priori do espaço no qual a mesma atua.

Portanto, com base em tal problemática, neste trabalho foram explorados os principais motivos que acarretam na possível baixa confiabilidade dos dados gerados pelo DE, focando principalmente no estudo do formato do espaço de busca e em uma primeira tentativa da geração de uma medida de qualidade para tais dados obtidos.

2 Objetivo

Neste trabalho, o espaço de busca sobre o qual o Differential Evolution atua será estudado, de forma a entender o impacto da presença de máximos locais e de ruídos sobre o mesmo. Além disso, o grau de confiabilidade dos valores encontrados pela meta-heurística será quantificado.

3 Métodos

Para alcançar os objetivos descritos acima, foi desenvolvida uma ferramenta em *Python 3.6.4*, utilizando como biblioteca central o *Matplotlib* [4], com o objetivo de analisar pontos¹ de um dado sísmico. Para estes estudos, foi considerado o dado sintético *Sol Nascente* no qual foram selecionados oito pontos de estudo, indicados na figura 1 abaixo.

¹Definidos aqui por um índice de *CDP* e um valor de *amostras de tempo*

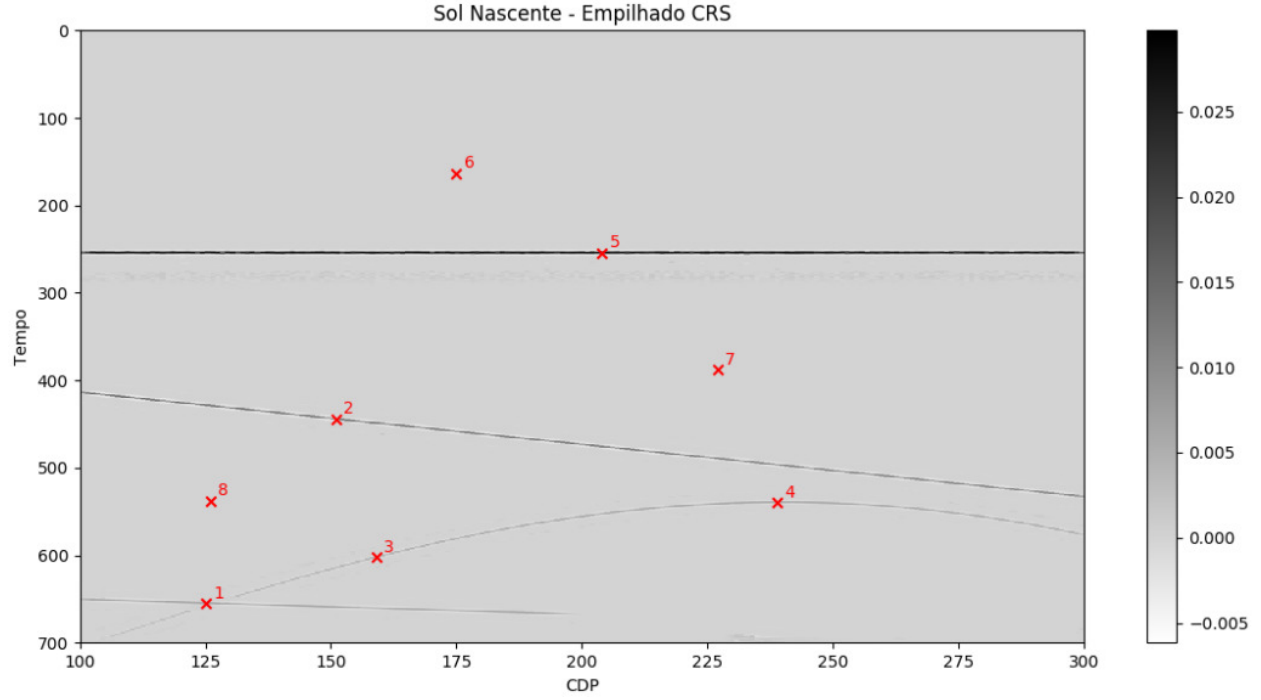


Figura 1: Empilhado CRS do dado "Sol Nascente". Os pontos considerados neste estudo encontram-se numerados e destacados pelos "x"vermelhos.

A partir da ferramenta desenvolvida, foi criado um evento de clique que permite mostrar em uma nova janela o espaço de busca do ponto selecionado. Espaço de busca este definido por uma matriz retangular E com linhas (de índice k), colunas (de índice l) e células ($E[k][l]$) representando as discretizações dos parâmetros do método CRS (a saber: velocidade, angulação e os valores de semblance² [5]. Em seguida, ainda para o mesmo ponto clicado, foi executado o DE com uma população de trinta indivíduos evoluída durante trinta gerações, de forma a obter o indivíduo P_S (sendo este o ponto do espaço com o maior valor de semblance). E para exemplificar, a figura 2 mostra o espaço de busca após o clique sobre o ponto 1 indicado na figura 1, onde o indivíduo P_S é indicado por um X em vermelho (sendo possível também definir a posição de tal indivíduo de forma manual, ao clicar sobre o plot).

² Valores estes obtidos pelo uso da velocidade e ângulação da respectiva célula sobre as equações de tempo de trânsito, sendo que para qualquer outro parâmetro necessário foi adotado o valor zero.

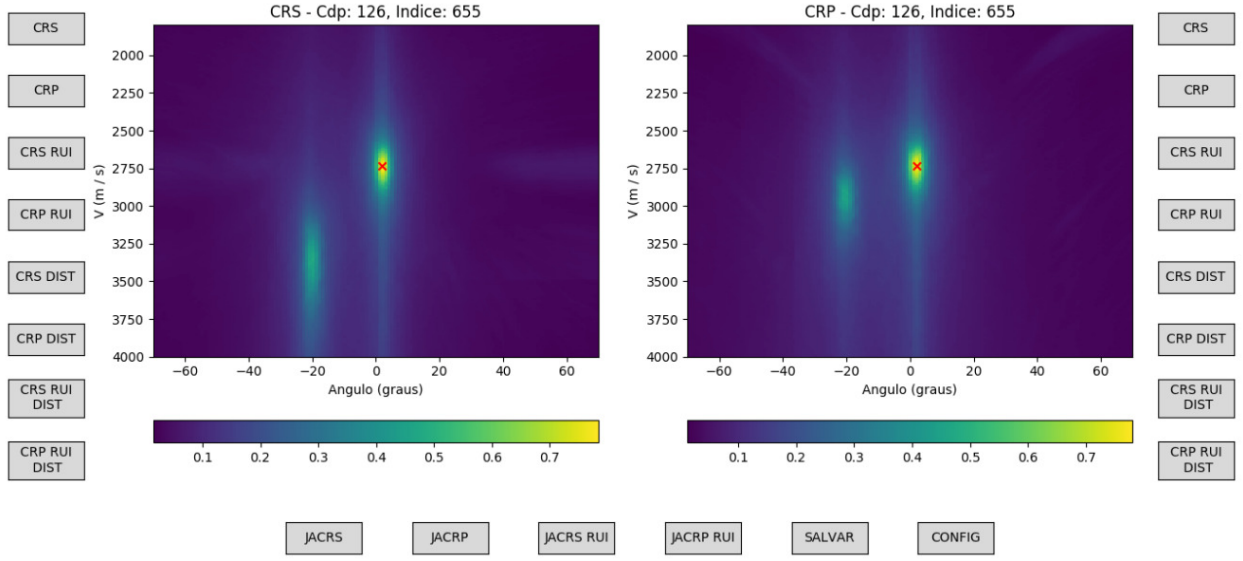


Figura 2: Espaço de busca relacionado ao ponto 1 da figura 1, onde os botões da direita e esquerda atuam respectivamente sobre os plots da direita e esquerda.

Com o espaço de busca em mãos, aos botões *CRS DIST* e *CRP DIST* foram associados a aplicação da seguinte métrica sobre o dado: construir um novo espaço de busca com centro em P_S e dimensões estipuladas por uma abertura de $(P_S.Angulacao \pm 5)^\circ$ e $(P_S.Velocidade \pm 5) m/s$ para cada ponto vizinho distante do ponto clicado em no máximo 10 *cdps* e 10 *amostras de tempo* (denotando por E_P o espaço com tais configurações relacionado ao ponto clicado), e aplicar uma transformação sobre os mesmos, dada pela seguinte expressão:

$$E_{ij}[k][l] = \begin{cases} \frac{E_{ij}[k][l]-a}{\mu-a} - 1 & \text{caso } E_{ij}[k][l] < \mu \\ \frac{E_{ij}[k][l]-\mu}{b-\mu} & \text{do contrário} \end{cases} \quad (1)$$

onde E_{ij} representa o espaço de busca, com as mesmas configurações citadas acima, para um ponto vizinho ao clicado, i e j o índice em tempo e a CDP do ponto ao qual o espaço de busca pertence, a , b e μ o menor, o maior e a média, dos valores de semblance dentre todas as células de todos os espaços de busca vizinhos em relação ao clicado. Por fim, para finalizar o processo de geração da métrica, é criada uma matriz retangular C com linhas (de índice i), colunas (de índice j) e células ($C[i][j]$) representando respectivamente, abertura em índice de tempo a partir do ponto clicado, abertura em cdp a partir do ponto clicado, e a distância entre E_P e E_{ij} . Distância esta estipulada por uma das seguintes fórmulas:

$$C[i][j] = \sum_{k,l} \frac{|E_{ij}[k][l] - E_P[k][l]|}{|E_{ij}[k][l]| + |E_P[k][l]|} \quad (1. \text{ Canberra})$$

$$C[i][j] = \max_{k,l} |E_{ij}[k][l] - E_P[k][l]| \quad (2. \text{ Chebyshev})$$

$$C[i][j] = \sum_{k,l} |E_{ij}[k][l] - E_P[k][l]| \quad (3. \text{ Manhattan})$$

$$C[i][j] = \frac{\sum_{k,l} |E_{ij}[k][l] - E_P[k][l]|}{\sum_{k,l} |E_{ij}[k][l] + E_P[k][l]|} \quad (4. \text{ Bray-Curtis})$$

$$C[i][j] = \sqrt{\sum_{k,l} |E_{ij}[k][l] - E_P[k][l]|^2} \quad (5. \text{ Euclidiana})$$

$$C[i][j] = 1 - \frac{\sum_{k,l} (E_{ij}[k][l] \cdot E_P[k][l])}{\sqrt{\sum_{k,l} |E_{ij}[k][l]|^2} \cdot \sqrt{\sum_{k,l} |E_P[k][l]|^2}} \quad (6. \text{ Coseno})$$

$$C[i][j] = \sum_{k,l} |(E_{ij}[k][l] - E_P[k][l])|^2 \quad (7. \text{ Quadrado Euclidiano})$$

$$C[i][j] = \begin{cases} \overline{E_{ij}} = \text{m\'edia dos elementos de } E_{ij} \\ \overline{E_P} = \text{m\'edia dos elementos de } E_P \\ 1 - \frac{\sum_{k,l} (E_{ij}[k][l] - \overline{E_{ij}}) \cdot (E_P[k][l] - \overline{E_P})}{\sqrt{\sum_{k,l} |E_{ij}[k][l] - \overline{E_{ij}}|^2} \cdot \sqrt{\sum_{k,l} |E_P[k][l] - \overline{E_P}|^2}} \end{cases} \quad (8. \text{ Correla\c{c}o})$$

escolhidas pelo usu\'ario antes de clicar em um dos bot\~oes *CRS DIST* e *CRP DIST*, como indicado na figura 3 abaixo.

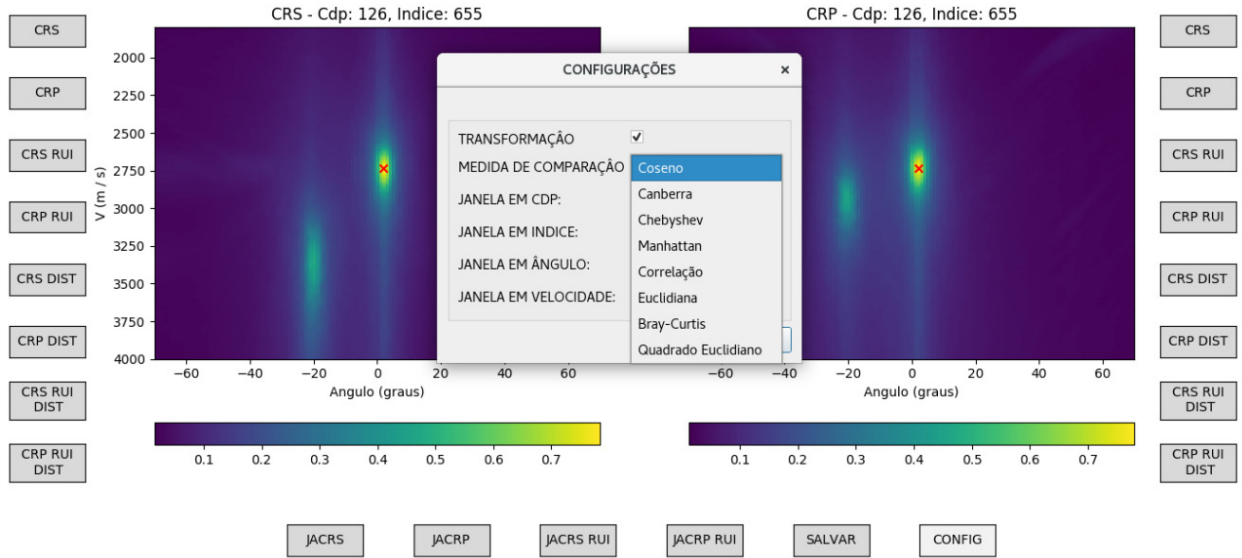


Figura 3: Evento de clique sobre o bot\~ao CONFIG na ferramenta com a op\c{c}o de escolha da medida de dist\~ancia a ser utilizada

Por fim, apesar das fórmulas de distância acima, recorremos a implementação das mesmas via biblioteca *SciPy* [6]. Além disso, o computador onde foram realizados os testes possui as seguintes configurações: *Intel(R) Core(TM) i7 – 6500U CPU @ 2.50GHz, x86_64, 4 cores com S.O. Arch Linux*.

4 Resultados e Discussão

4.1 Análise do espaço de busca

Com base no exposto na metodologia, seguem os resultados dos espaços para cada um dos pontos destacados na figura 1:

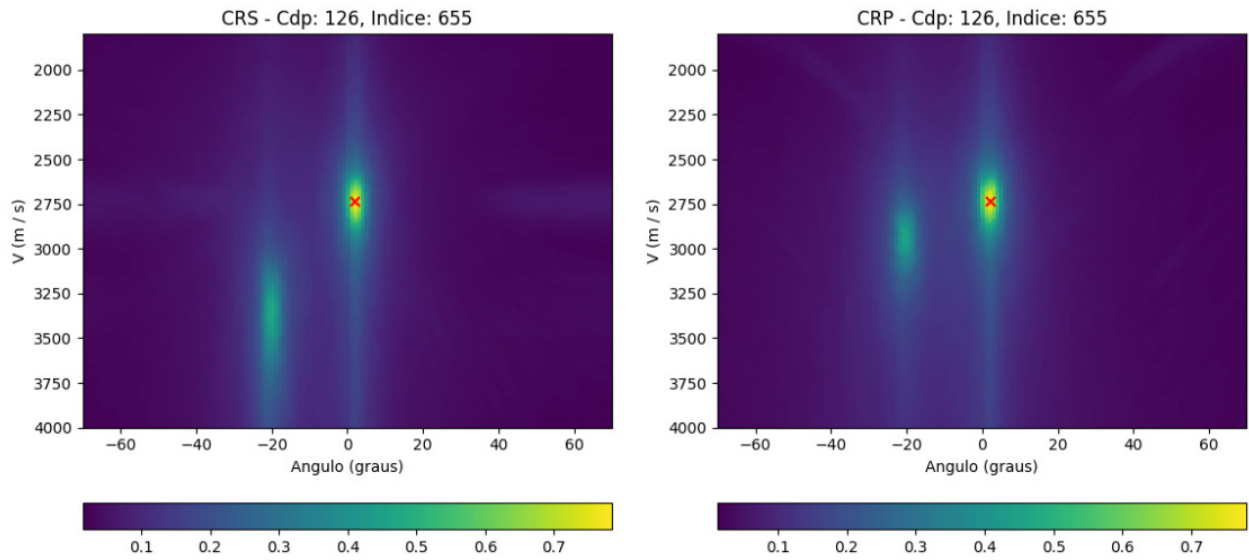


Figura 4: Espaço de busca do ponto 1 da figura 1

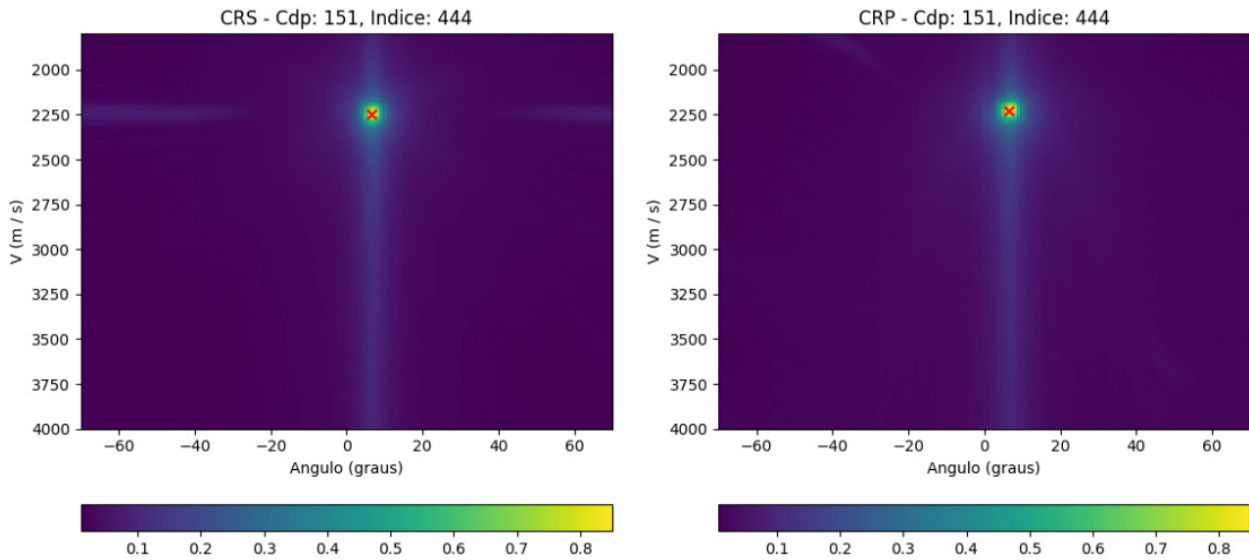


Figura 5: Espaço de busca do ponto 2 da figura 1

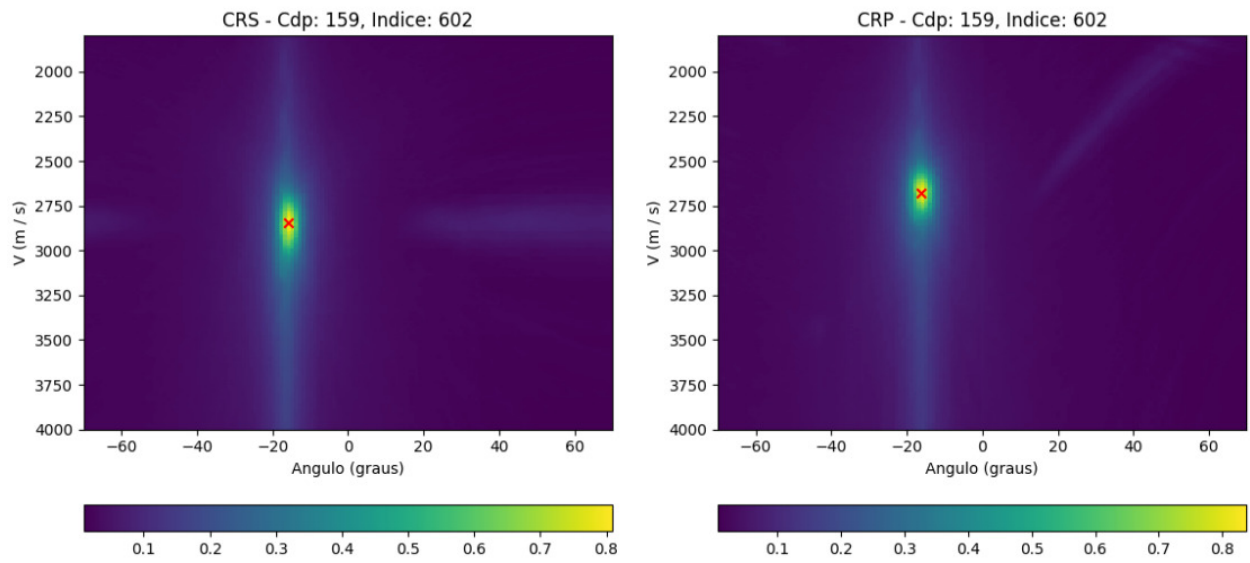


Figura 6: Espaço de busca do ponto 3 da figura 1

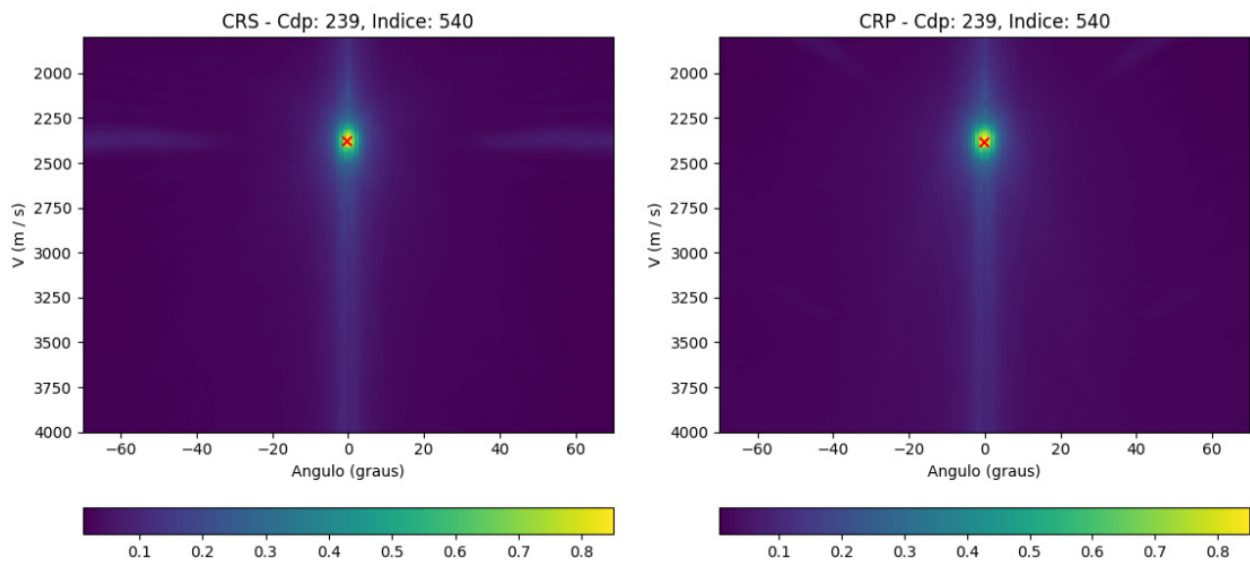


Figura 7: Espaço de busca do ponto 4 da figura 1

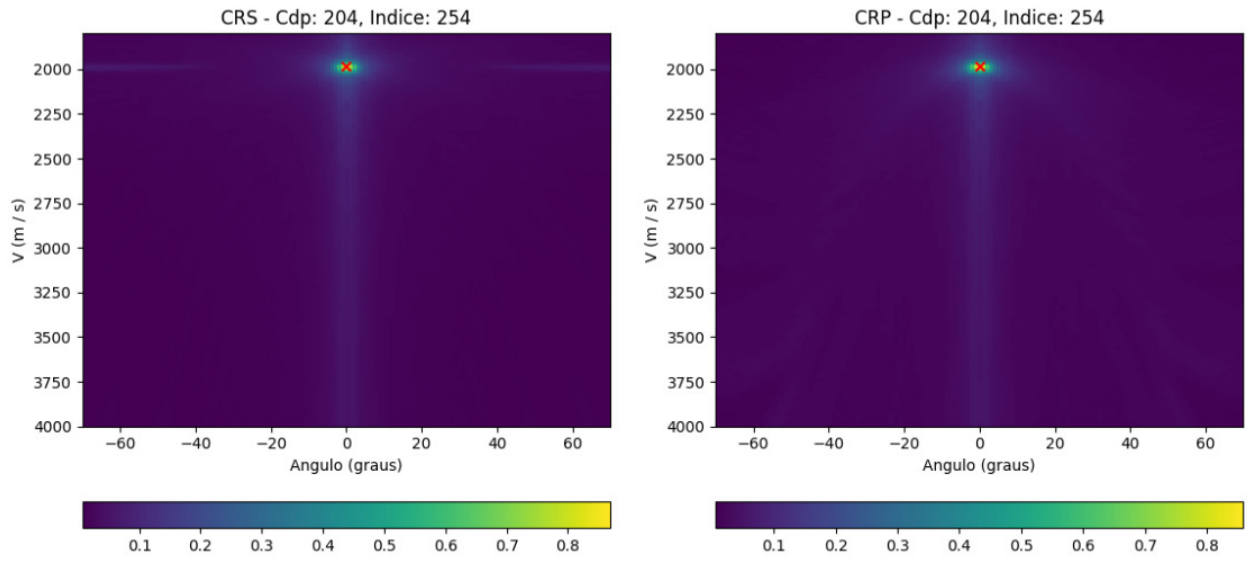


Figura 8: Espaço de busca do ponto 5 da figura 1

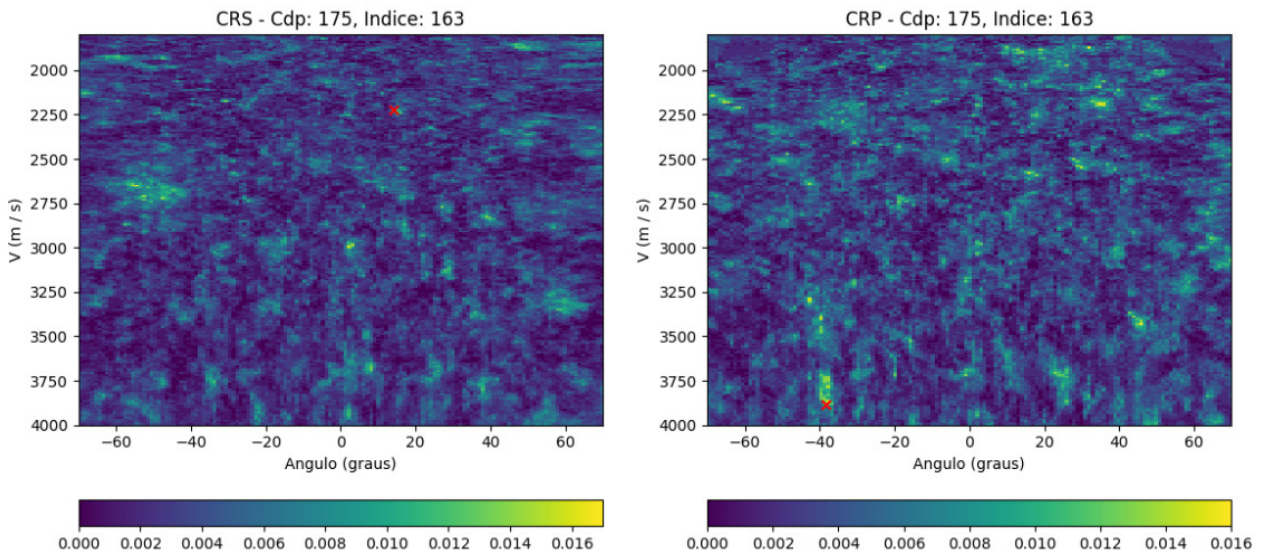


Figura 9: Espaço de busca do ponto 6 da figura 1

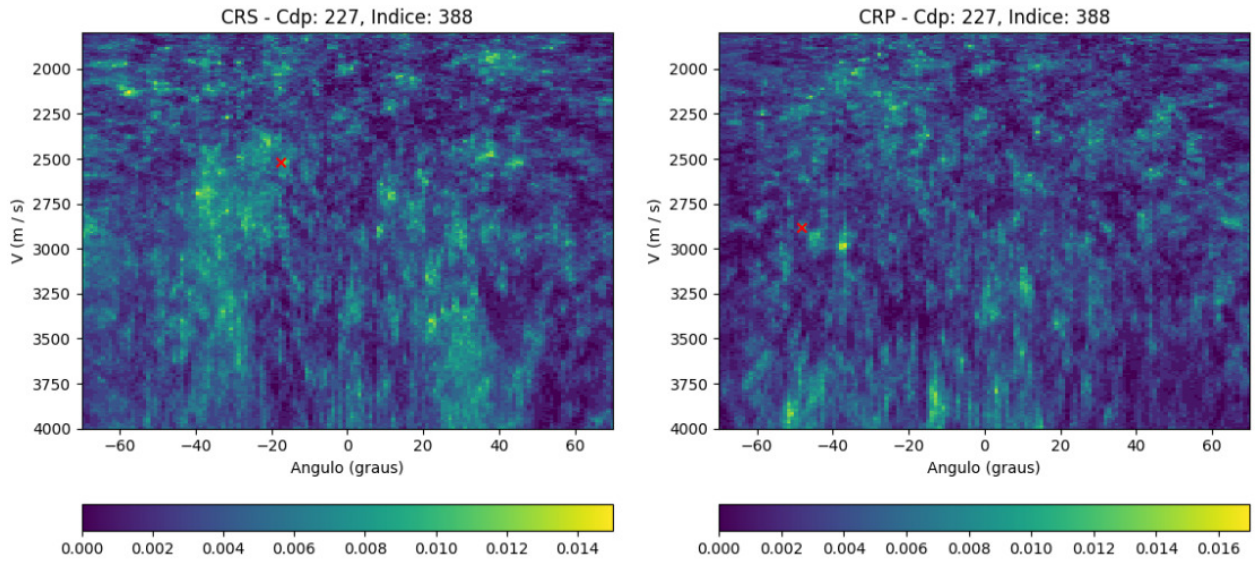


Figura 10: Espaço de busca do ponto 7 da figura 1

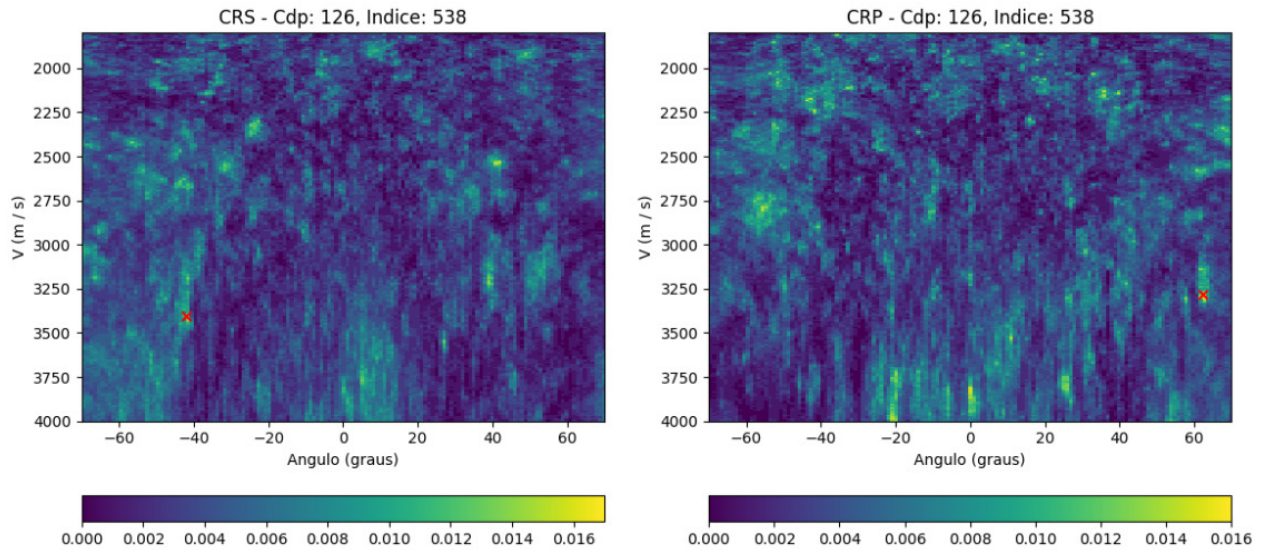


Figura 11: Espaço de busca do ponto 8 da figura 1

A partir de tais resultados, é possível relacionar os espaços com três categorias distintas:

- Sem evento
- Com evento
- Com cruzamento de eventos

No caso de um espaço relacionado a um ponto por onde passa um único evento, há apenas um ponto de máximo relevante sobre o mesmo, como é verificado ao comparar as figuras 5-8 com a 1. Fato que possibilita inferir uma boa confiança do valor obtido pelo DE.

Em contrapartida, para espaços relacionados com pontos por onde passa evento algum, há uma quantidade imensa de pontos de máximo distribuídos de forma aparentemente aleatória sobre o

espaço, sendo igualmente demonstrável ao comparar as figuras 9-11 com a 1. Fato que possibilita inferir uma baixa confiança do valor obtido pelo DE.

Por fim, para espaços relacionados com pontos por onde passa múltiplos eventos, há mais do que um ponto de máximo relevante, como é verificado ao comparar a figura 4 com a 1. Neste caso, há uma certa problemática, devido ao fato do DE escolher o indivíduo com maior valor de semblance, sem que para tanto haja uma correlação dos pontos vizinhos do dado com o central sendo analisado. Desta forma, segue a abertura para a discussão da métrica descrita na metodologia.

4.2 Correlação com pontos vizinhos

Como descrito na subseção anterior, para que se possa ter uma maior confiança do valor obtido pelo DE, em alguns casos não é satisfatório analisar apenas o formato do espaço de busca; é necessário também fazer uma correlação direta de parte do espaço relacionado ao ponto sendo analisado com o de seus vizinhos, de forma a obter informações mais precisas sobre o(s) evento(s) passando por tal ponto.

Neste caso, para ilustrar tal correlação, tomamos o espaço de busca do ponto 1 da figura 1, além de P_{S_1} e P_{S_2} , indicados respectivamente, na esquerda e direita da figura 12 abaixo.

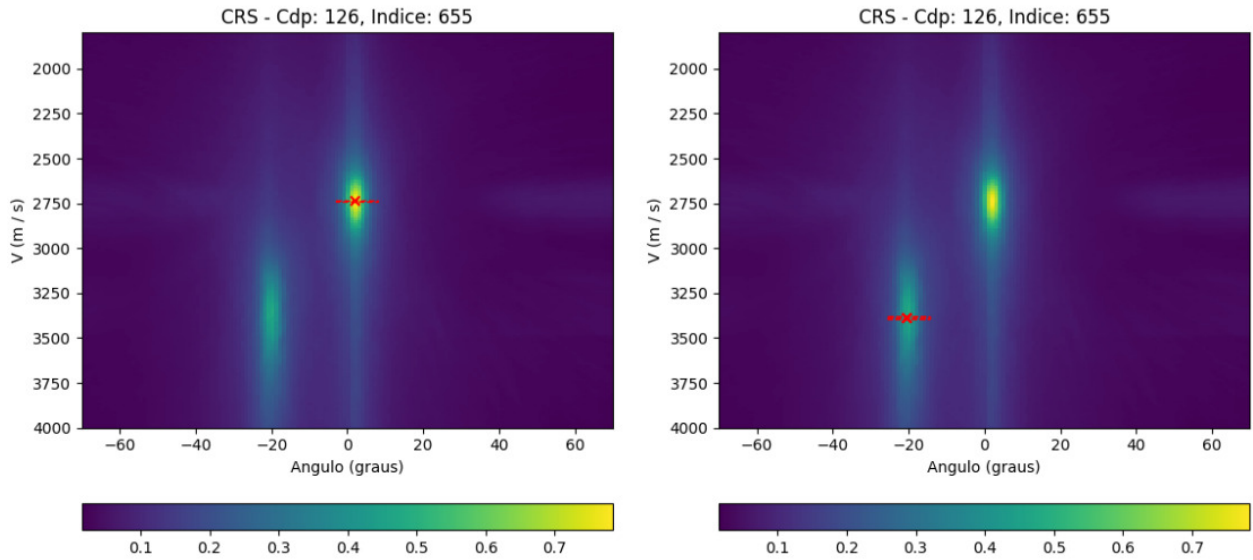


Figura 12: Espaços de busca CRS do ponto 1 da figura 1 e seus respectivos indivíduos P_{S_1} e P_{S_2} (selecionados manualmente)

Fazendo uso de P_{S_1} e P_{S_2} , aplicando a métrica descrita na metodologia, utilizando o método CRS e fazendo uso da fórmula de distância 6 (Coseno), obtenho as seguintes matrizes C_1 (esquerda na figura 13) e C_2 (direita na figura 13), relacionadas diretamente com P_{S_1} e P_{S_2} :

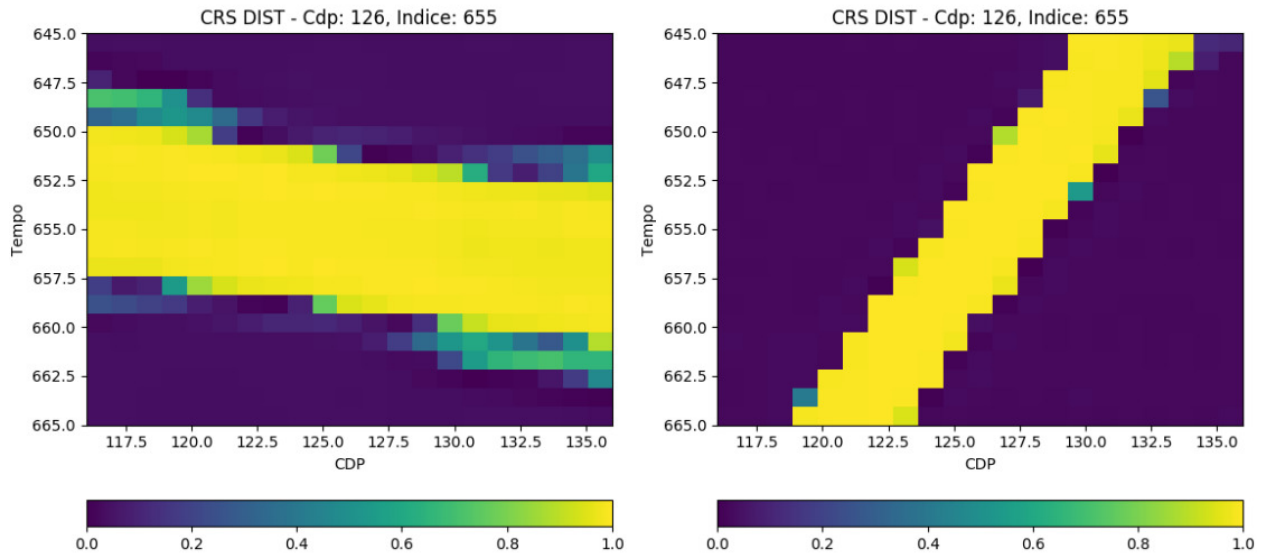


Figura 13: Matriz baseada no coseno dos espaos CRS de pontos vizinhos ao ponto 1 da figura 1

E repetindo o mesmo processo para o mtodo CRP, obtenho:

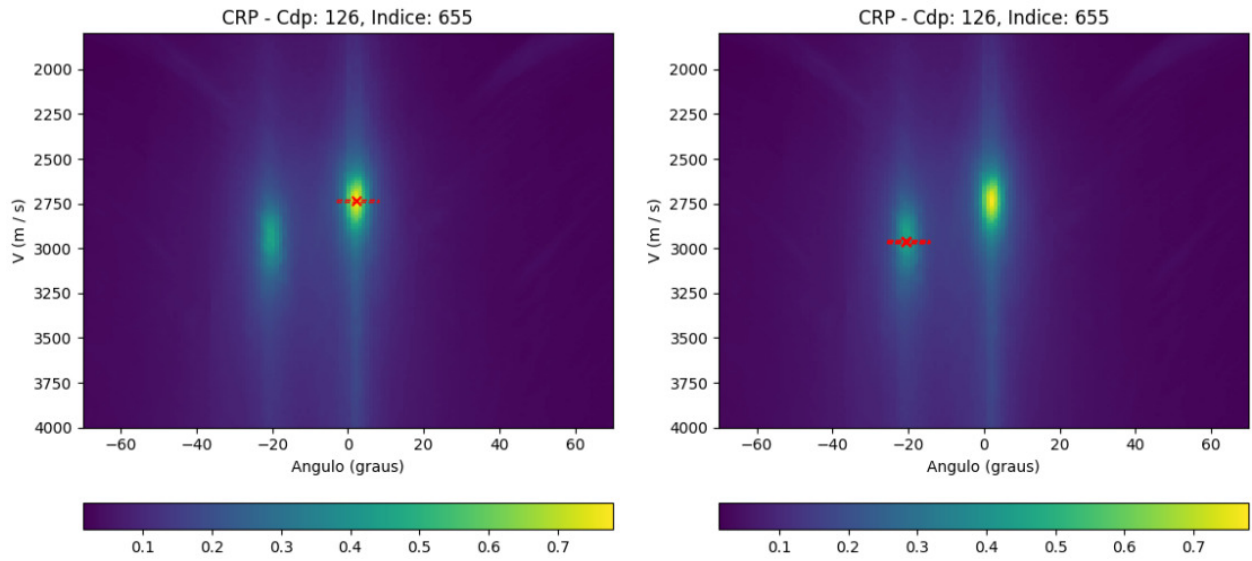


Figura 14: Espaos de busca CRP do ponto 1 da figura 1 e seus respectivos indivduos P_S (selecionados manualmente)

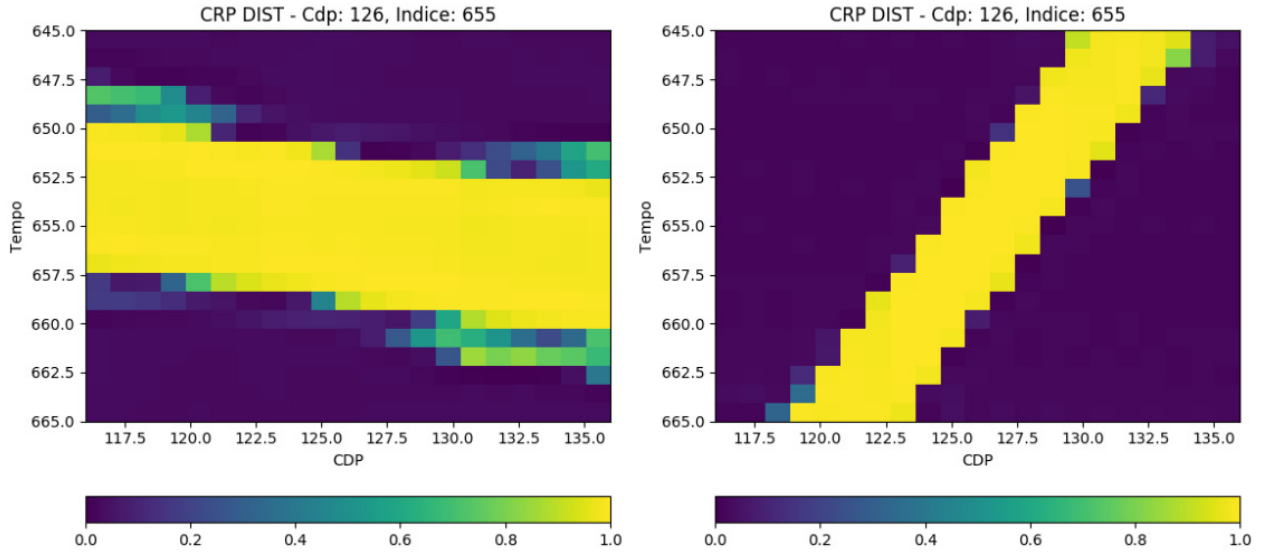


Figura 15: Matriz baseada no coseno dos espaços CRP de pontos vizinhos ao ponto 1 da figura 1

Em ambos os métodos (CRS e CRP), é possível notar uma certa sofisticação da métrica, que permite que um evento seja completamente atenuado dependendo da região do espaço de busca sendo analisada; e mais, há também uma indicação da direção do evento não atenuado, correlacionado diretamente com o dado empilhado, como pode ser notado ao comparar as figuras 13 e 15 com a 1. Com isto, é possível inferir se de fato o ponto P_S (selecionado pelo DE) corresponde ao evento de interesse.

4.3 Considerações sobre os parâmetros da métrica

Apesar da métrica apresentada na subseção anterior apresentar bons resultados para as configurações estabelecidas, é importante ressaltar a dependência da mesma com relação aos seguintes parâmetros: abertura em CDP, em tempo, em ângulo e em velocidade, transformação utilizada e medida de distância aplicada.

Quanto as aberturas em CDP e em tempo, foi observado que existe uma certa dependência de seus valores e a qualidade da matriz C obtida, visto que, ao adicionar novos pontos vizinhos, há uma alteração direta no valor de a , b e μ e por consequência nos valores obtidos pela distância dos dois espaços citados antes. Contudo, em testes realizados, tal dependência foi baixa para o caso da distância 6, não havendo uma grande degradação de C . Para o caso de outras distâncias, cabe uma averiguação maior.

No que diz respeito as aberturas em ângulo e velocidade, nota-se uma sensibilidade maior da métrica, onde para valores acima de 20 para ambos os casos, temos uma diminuição dos valores antes altos e aumento dos valores antes baixos de C . Contudo, para tal dado sintético, aumentar a abertura implica no aumento ou diminuição de valores, porém as configurações do mesmo se mantêm, permitindo inferir ainda a direção do evento. De qualquer forma, vale o adendo que tal conclusão se manteve para as outras medidas de distância. E para ilustrar tal fato, segue na figura 16 abaixo um resultado parecido ao da figura 13, onde foi modificado a abertura de 5 para 30.

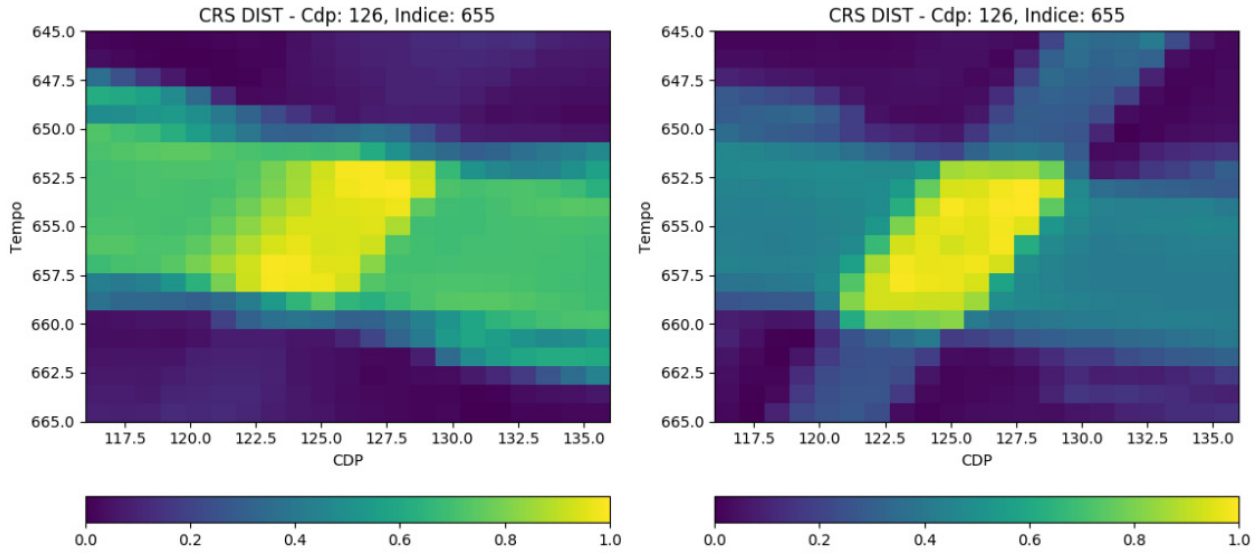


Figura 16: Matriz baseada no cosseno dos espaços (com abertura de 30 para velocidade e angulação) CRS de pontos vizinhos ao ponto 1 da figura 1

Para a transformação, há uma sensibilidade maior ainda, visto que a mesma é aplicada visando o bom funcionamento do último item, ou seja, visando a utilização da medida de distância 6, sendo que ao remover tal transformação, sem que para tanto a medida de distância seja alterada, nada de muito conclusivo pode ser inferido, pois boa parte da matriz C estará próxima de 1, sem permitir indicar a direção de um evento com um alto grau de confiabilidade. Porém, apesar de necessária, a interpolação por duas retas (descrita na equação 1) não é a única ³, podendo ser utilizado uma interpolação por uma parábola, por uma sigmoid, etc., e ainda obter resultados parecidos ao das figuras 13 e 15. A justificativa do uso de retas é devido ao fato de ter sido a que produziu melhores resultados para C dentre as funções testadas, contudo fica em aberto a averiguação de outras transformações que possibilitem uma melhor correlação com espaços vizinhos (principalmente ao expandir tal métrica para o regime de um dado real, onde a presença e formato de ruídos é bem menos previsível). De qualquer forma, segue na figura 17 abaixo o resultado CRS da matriz C para as mesmas configurações da figura 13, tendo-se como exceção a remoção da transformação citada.

³Vale de ressalva que apesar desta não ser única, não é qualquer transformação que produz valores satisfatórios, visto que em testes, ao alterar o valor de μ como sendo a média somente do espaço de busca E do ponto central analisado, tive uma queda grande na qualidade de C , com uma intensidade alta na região central e uma degradação rápida para pontos que se afastam do central

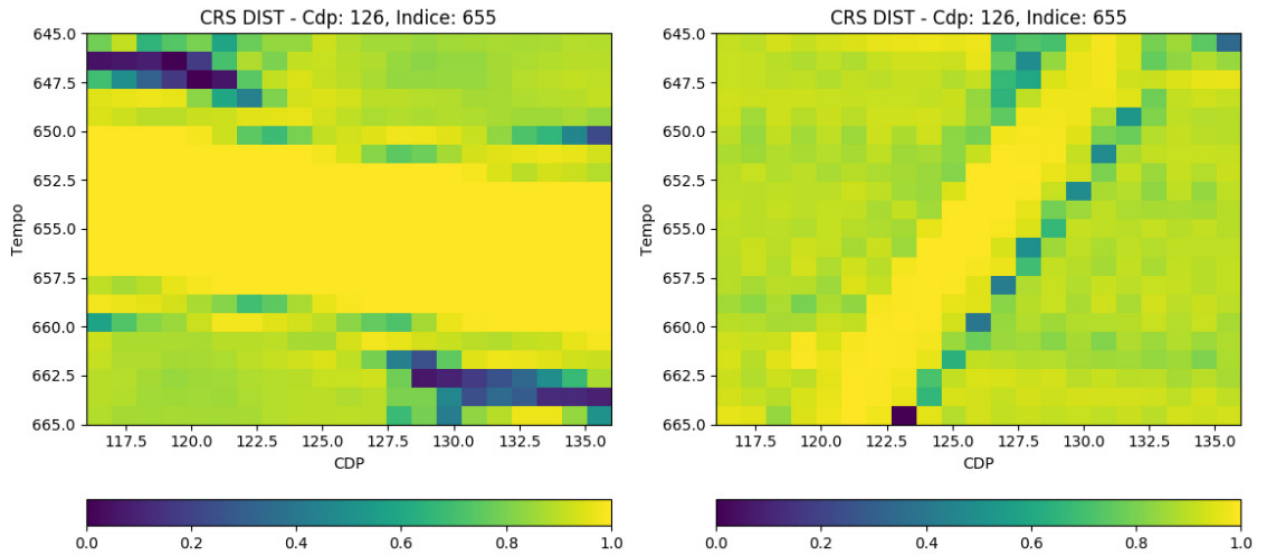


Figura 17: Matriz baseada no coseno dos espaços (sem transformação) CRS de pontos vizinhos ao ponto 1 da figura 1

Por fim, pode-se notar também que a medida de distância 6 utilizada (coseno) não é a única que produziu bons resultados; e mais, como a transformação é utilizada justamente para o coseno, ao mudar de medida de distância, não somente obtive valores parecidos com os das figuras 13 e 15, como também pude remover o uso da transformação e ainda sim continuar obtendo valores razoáveis em comparação ao obtido na figura 17 acima. Fato este verificado nas figuras que se seguem abaixo, onde cada uma das medidas de distância é posta à prova. No mais, assim como no caso da transformação, o tópico da medida de distância fica em aberto também, visto que o coseno com a transformação foi sim um dos que gerou a melhor matriz C para o ponto dado pela figura 13, porém nada se sabe do comportamento das outras medidas para outros pontos de dados sintéticos e reais.

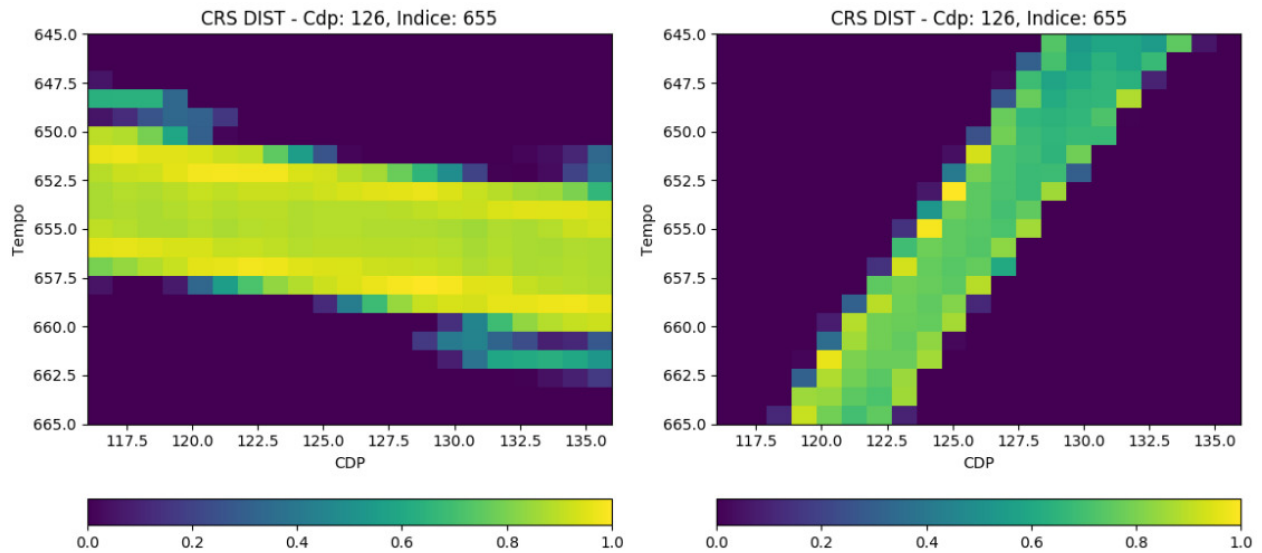


Figura 18: Matriz baseada na medida de distância 1 (Canberra) dos espaços, com o uso da transformação 1, via método CRS de pontos vizinhos ao ponto 1 da figura 1

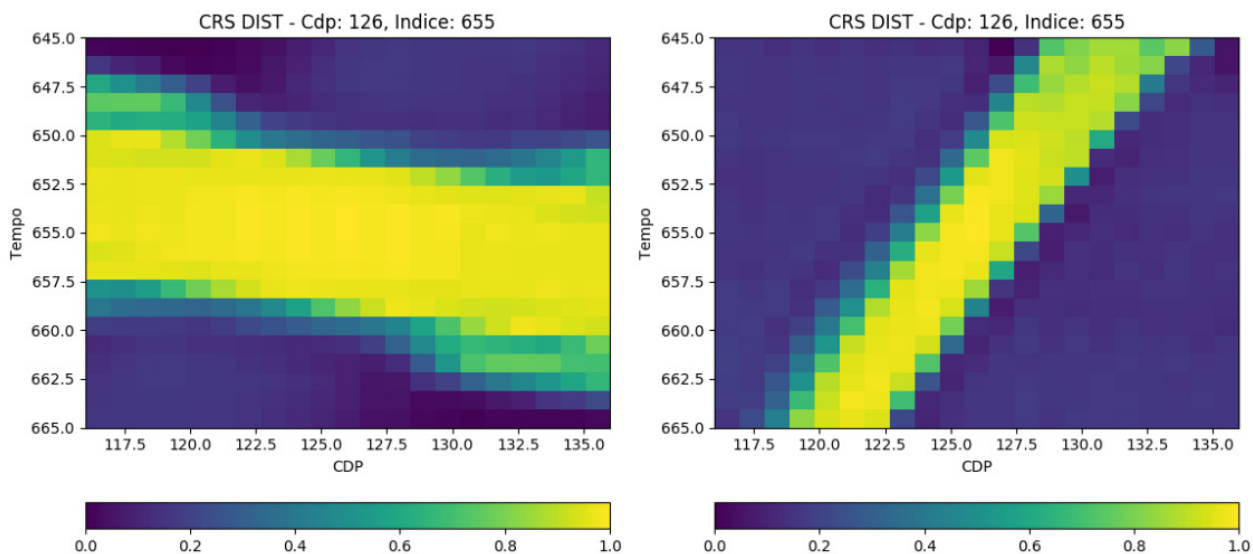


Figura 19: Matriz baseada na medida de distância 1 (Canberra) dos espaços, sem o uso da transformação 1, via método CRS de pontos vizinhos ao ponto 1 da figura 1

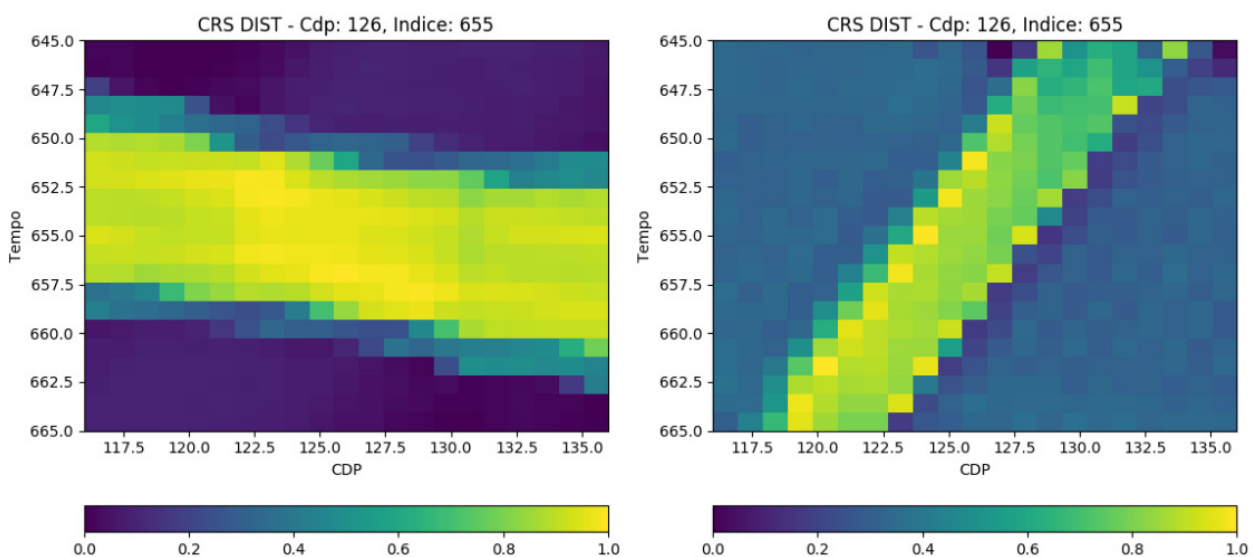


Figura 20: Matriz baseada na medida de distância 2 (Chebyshev) dos espaços, com o uso da transformação 1, via método CRS de pontos vizinhos ao ponto 1 da figura 1

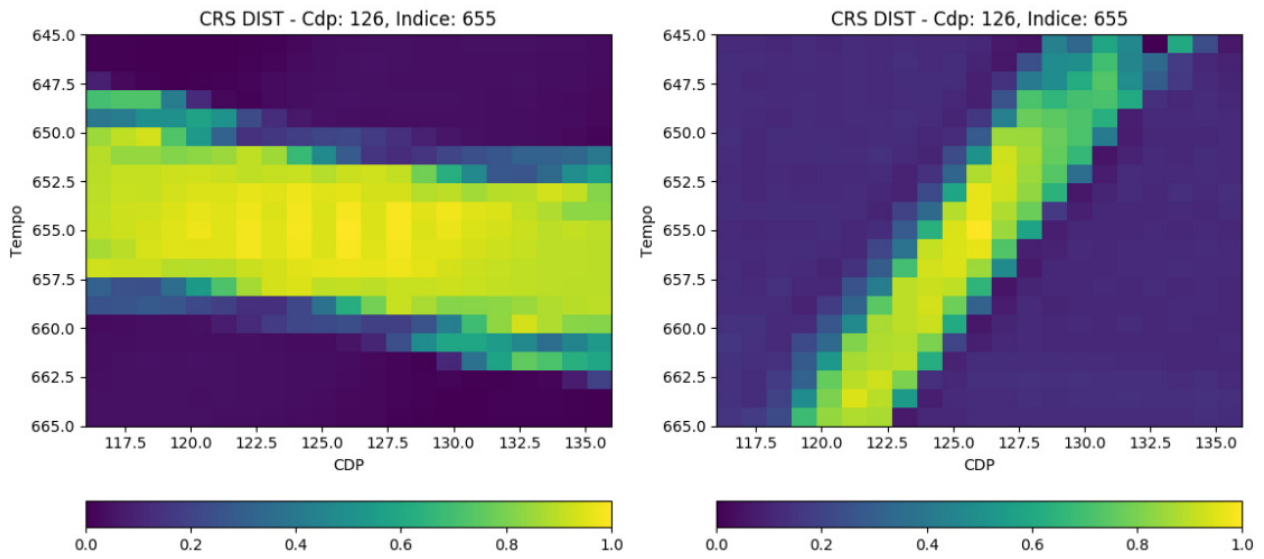


Figura 21: Matriz baseada na medida de distância 2 (Chebyshev) dos espaços, sem o uso da transformação 1, via método CRS de pontos vizinhos ao ponto 1 da figura 1

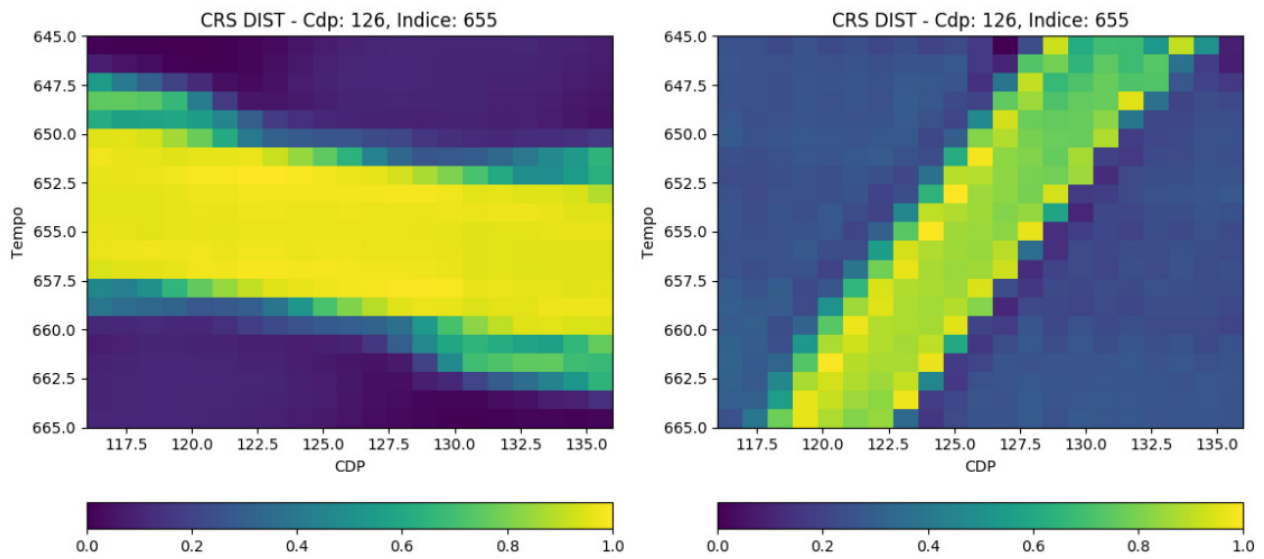


Figura 22: Matriz baseada na medida de distância 3 (Manhattan) dos espaços, com o uso da transformação 1, via método CRS de pontos vizinhos ao ponto 1 da figura 1

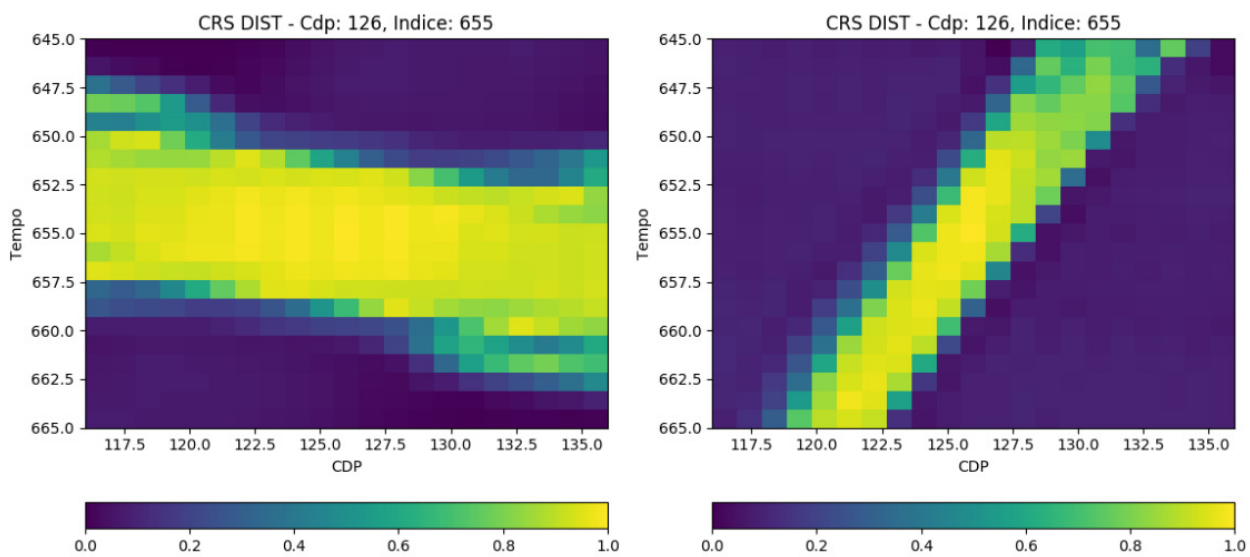


Figura 23: Matriz baseada na medida de distância 3 (Manhattan) dos espaços, sem o uso da transformação 1, via método CRS de pontos vizinhos ao ponto 1 da figura 1

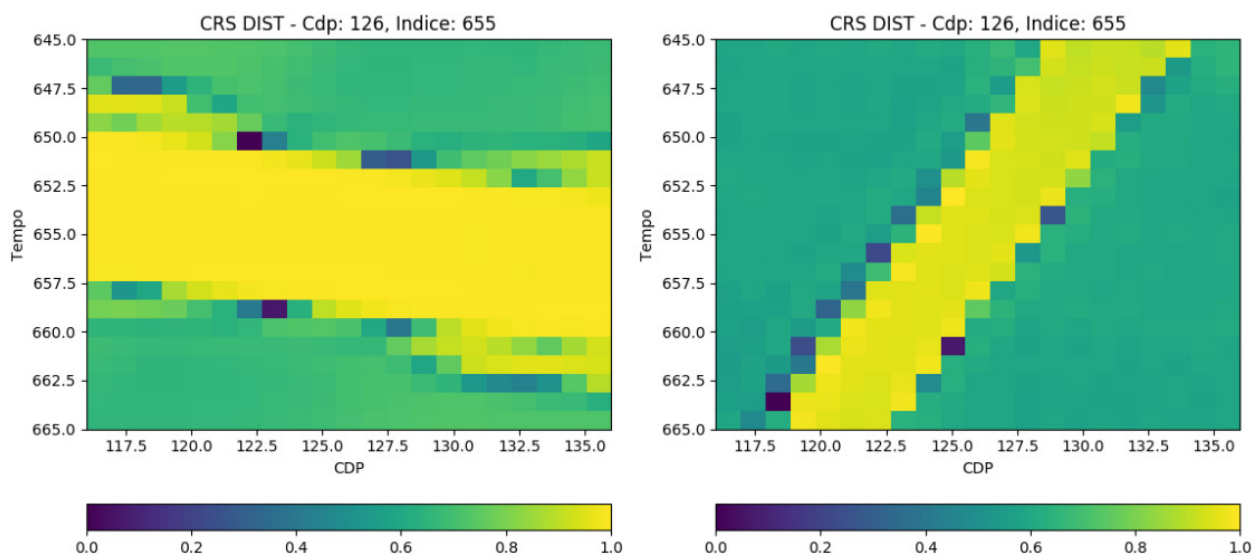


Figura 24: Matriz baseada na medida de distância 4 (Bray-Curtis) dos espaços, com o uso da transformação 1, via método CRS de pontos vizinhos ao ponto 1 da figura 1

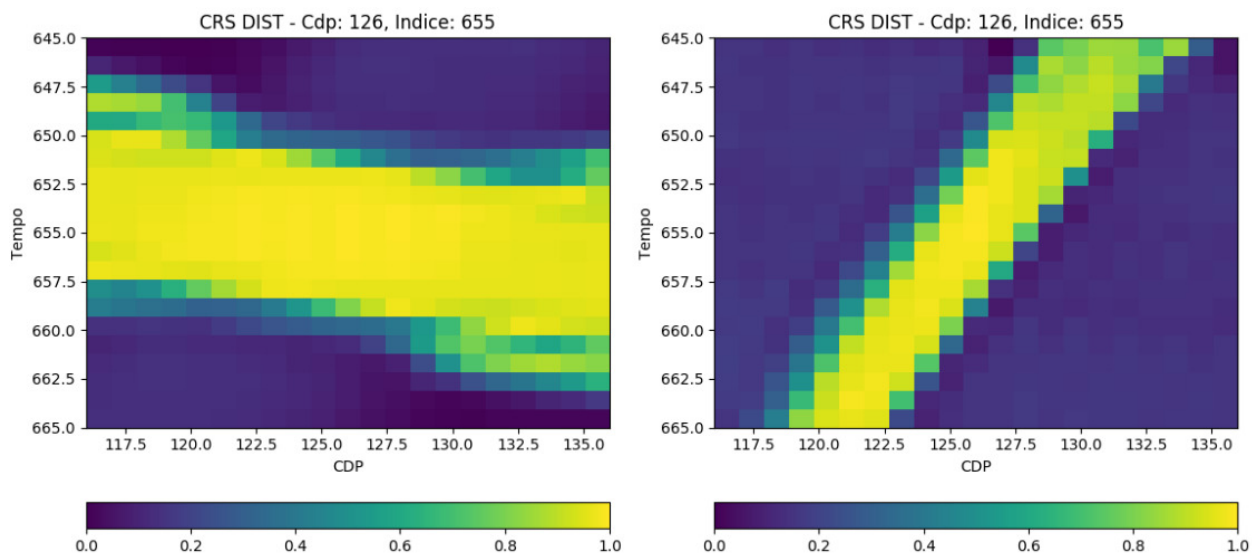


Figura 25: Matriz baseada na medida de distância 4 (Bray-Curtis) dos espaços, sem o uso da transformação 1, via método CRS de pontos vizinhos ao ponto 1 da figura 1

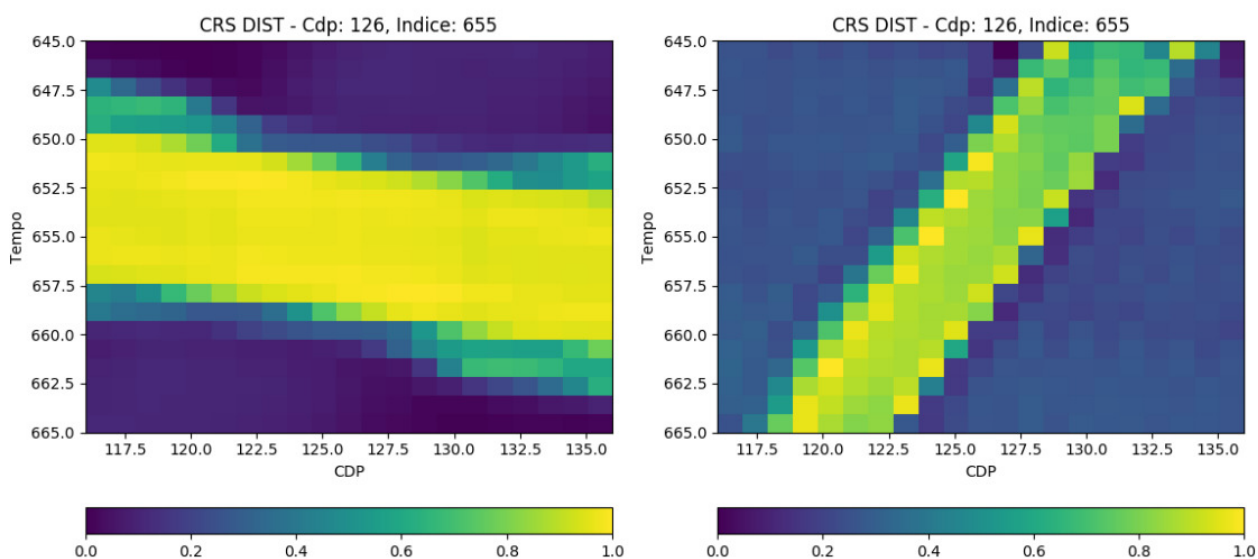


Figura 26: Matriz baseada na medida de distância 5 (Euclidiana) dos espaços, com o uso da transformação 1, via método CRS de pontos vizinhos ao ponto 1 da figura 1

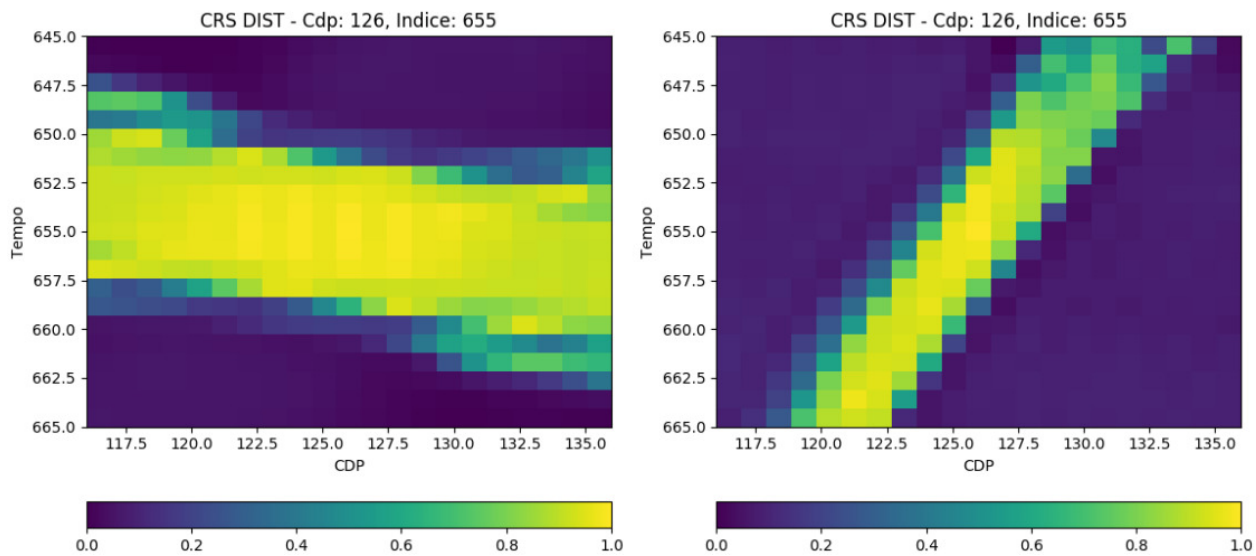


Figura 27: Matriz baseada na medida de distância 5 (Euclidiana) dos espaços, sem o uso da transformação 1, via método CRS de pontos vizinhos ao ponto 1 da figura 1

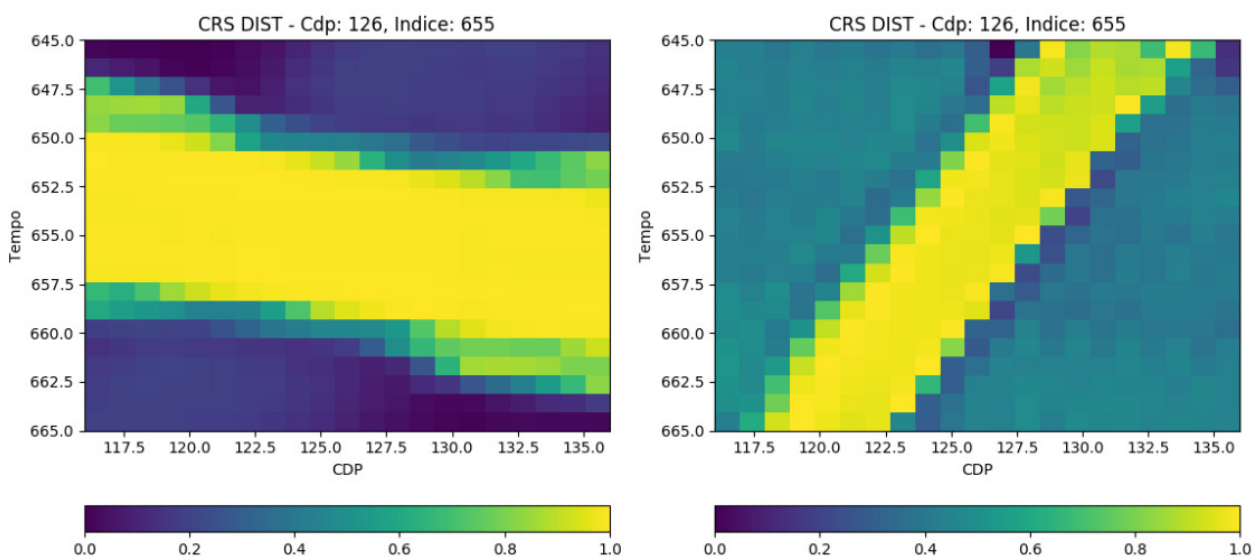


Figura 28: Matriz baseada na medida de distância 7 (Quadrado Euclidiano) dos espaços, com o uso da transformação 1, via método CRS de pontos vizinhos ao ponto 1 da figura 1

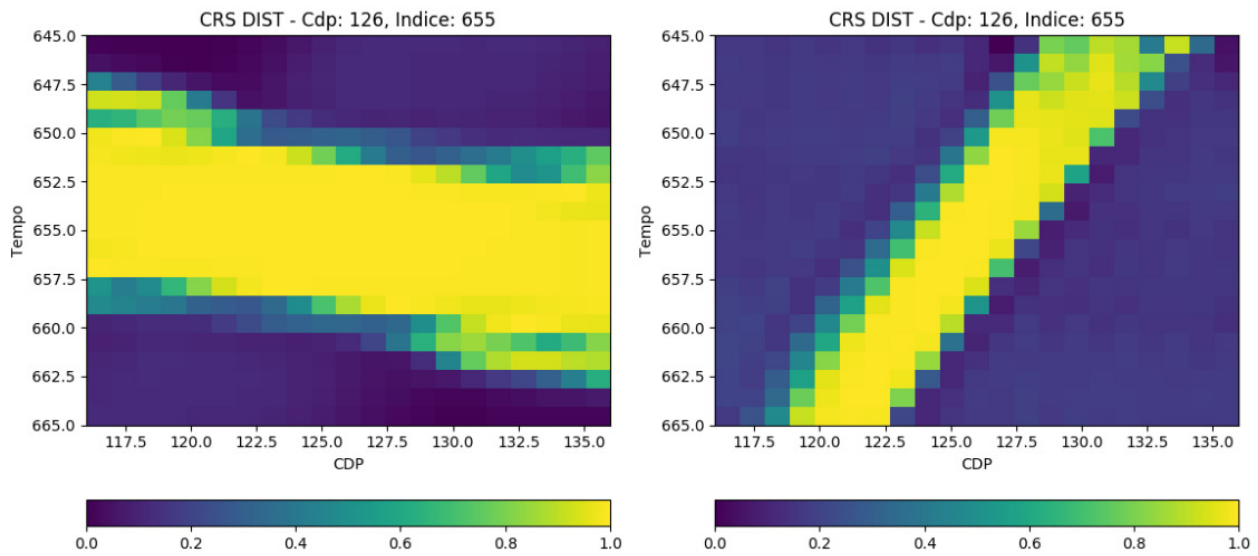


Figura 29: Matriz baseada na medida de distância 7 (Quadrado Euclidiano) dos espaços, sem o uso da transformação 1, via método CRS de pontos vizinhos ao ponto 1 da figura 1

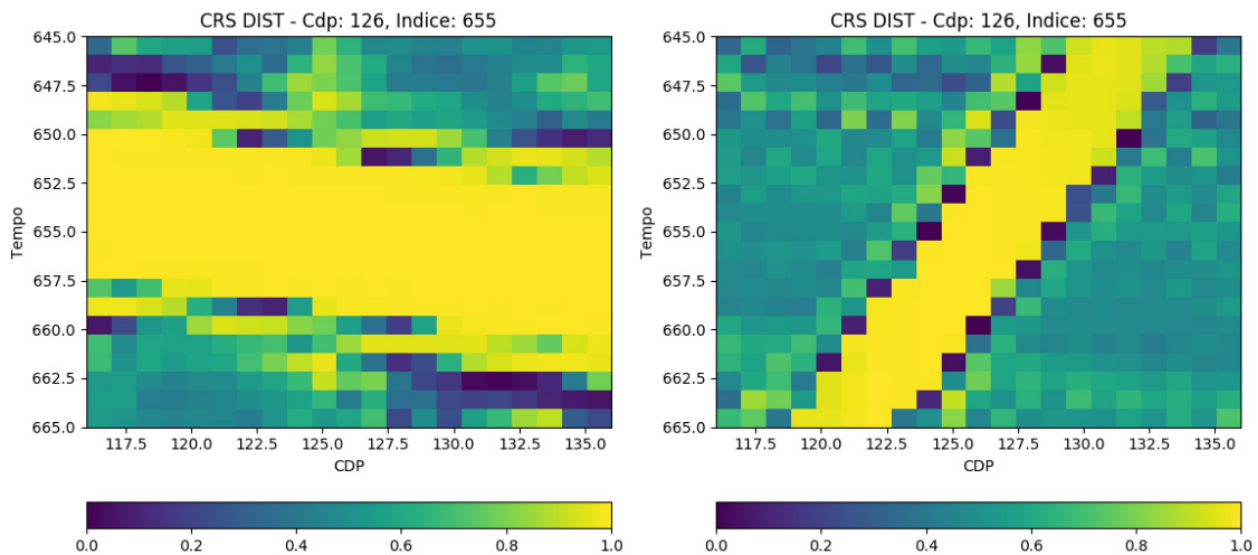


Figura 30: Matriz baseada na medida de distância 8 (Correlação) dos espaços, com o uso da transformação 1, via método CRS de pontos vizinhos ao ponto 1 da figura 1

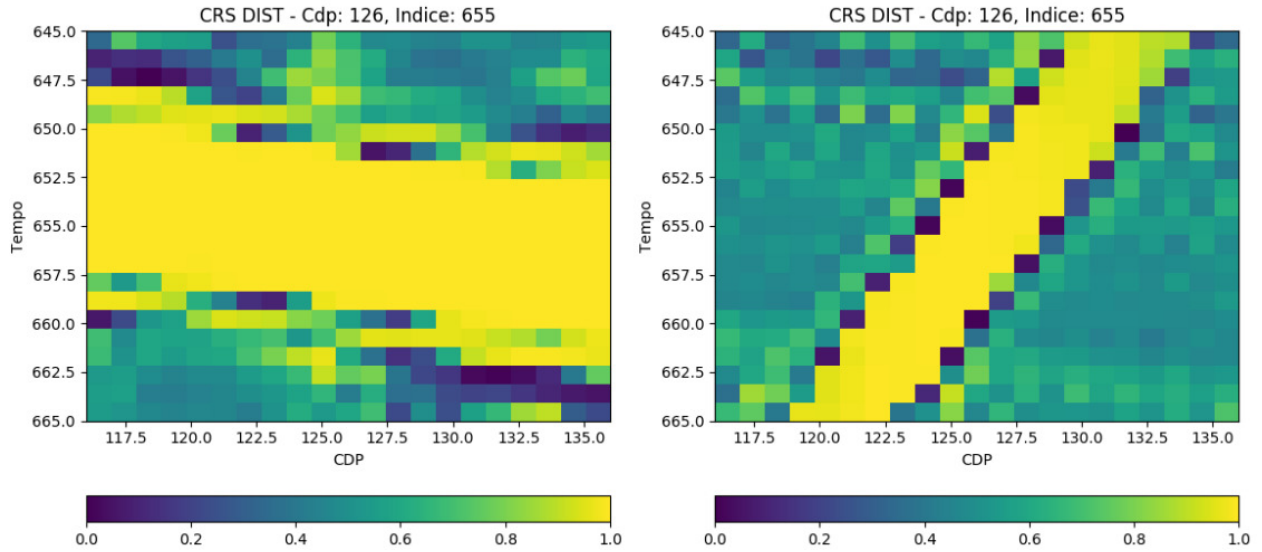


Figura 31: Matriz baseada na medida de distância 8 (Correlação) dos espaços, sem o uso da transformação 1, via método CRS de pontos vizinhos ao ponto 1 da figura 1

5 Conclusão

Diante dos resultados expostos na sessão de resultados e discussão, pode-se concluir que duas informações relevantes podem ser inferidas a partir da análise do espaço de busca: (1) seu formato está diretamente relacionado com o número de eventos passando sobre o ponto do dado ao qual o mesmo pertence, e (2) há uma semelhança direta entre seu formato e o de alguns de seus espaços vizinhos.

Com estes dois itens, foi possível obter uma forma direta de qualificar melhores indivíduos obtidos pelo DE, de forma a aumentar a confiabilidade dos resultados. Ou seja, analisando o espaço de busca no qual o indivíduo pertence, foi possível inferir se o mesmo foi obtido por um fator aleatório devido a presença de ruído, ou se de fato, foi obtido devido a presença de um evento (isto é, seus parâmetros de velocidade e angulação de fato se relacionam com a existência do evento). E mesmo após analisar o respectivo espaço de busca, em casos onde houve dúvidas da autenticidade do indivíduo, foi possível utilizar a métrica gerada de forma a verificar a direção do evento de interesse (utilizando de uma medida de distância entre espaços amostrais vizinhos).

Contudo, apesar de um primeiro entendimento do espaço amostral, novas investigações devem ser realizadas, principalmente no quesito de melhoras da métrica, sendo um possível primeiro passo a expansão para dados reais, onde a mesma realmente terá seu valor.

Bibliografia

- [1] FACCIPIERI, J. H., Separação e processamento de difrações em dados geofísicos de reflexão: MSc thesis, Unicamp, 2012.
- [2] COIMBRA, T. A., OPERAÇÃO PARA CONTINUAÇÃO DO AFASTAMENTO: OPERADOR DIFERENCIAL, COMPORTAMENTO DINÂMICO E EMPILHAMENTO MULTI-PARAMÉTRICO: PhD thesis, Unicamp, 2014.
- [3] Storn, R. and Price, K. (1995). Differential evolution—a simple and efficient adaptive scheme for global optimization over continuous spaces. international computer science institute, berkeley. Technical report, CA, 1995, Tech. Rep. TR-95-012.
- [4] <<https://matplotlib.org/>>. Acesso em: 28 jun. 2018.
- [5] NEIDELL, N. S.; TANER, M. T. Semblance and other coherency measures for multichannel data. Geophysics, v. 36, n. 3, p.482–497, 1971.
- [6] <<https://docs.scipy.org/doc/scipy/reference/spatial.distance.html>>. Acesso em: 28 jun. 2018.