

Combinando análise fractal e mineração de séries temporais para identificação de extremos climáticos regionais

Santiago Augusto Nunes¹, Luciana A. S. Romani², Ana M. H. Avila³,
Elaine P. M. de Sousa¹, Agma J. M. Traina¹

¹ICMC - Universidade de São Paulo - São Carlos - Brasil

²Embrapa Informática Agropecuária - Campinas - Brasil

³Cepagri - Universidade Estadual de Campinas - Brasil

bynael@grad.icmc.usp.br, luciana@cnptia.embrapa.br, avila@cpa.unicamp.br,
{parros, agma}@icmc.usp.br

Abstract. *In the last few decades, huge amounts of climate data from ground-based stations and other different sensors have been gathered and stored by several private and public institutions. The analysis of these data has become an important task due to worldwide climate changes and the consequent social and economic effects. In this work, we propose an approach to analyze multiple climate time series in order to identify intrinsic temporal patterns. By dealing with multiple time series as multidimensional data streams, we can integrate different climate variables and discover general behavior changes over time. Experimental studies on real climate time series collected from different regions of Brazil show the applicability of our approach.*

Resumo. *Nas últimas décadas, grandes quantidades de dados climáticos provenientes de estações meteorológicas e de outros tipos de sensores têm sido coletadas e armazenadas por diversas instituições. A análise desses dados tornou-se uma tarefa importante devido às mudanças climáticas e seus efeitos sociais e econômicos. Este trabalho propõe um processo de análise de múltiplas séries temporais climáticas para identificar padrões temporais intrínsecos aos dados. Considerando múltiplas séries como uma data stream, é possível integrar diferentes variáveis climáticas e detectar mudanças de comportamento ao longo do tempo. Estudos em séries climáticas reais coletadas em diferentes regiões do Brasil mostram o potencial de aplicação dessa abordagem.*

1. Introdução

Os avanços tecnológicos relacionados à coleta de dados meteorológicos fizeram com que conjuntos de dados climáticos crescessem rapidamente, trazendo novos desafios para a pesquisa em climatologia e sua aplicação em agricultura. Dada a importância da análise de dados históricos e recentes para compreender e caracterizar fenômenos climáticos, esses dados têm sido armazenados, gerando assim longas séries temporais históricas.

Os meteorologistas têm estudado esses conjuntos de dados com o objetivo de entender e caracterizar o clima global e prever possíveis cenários futuros em diferentes regiões do planeta. Dentre as técnicas estatísticas utilizadas, destacam-se análise de correlação, regressão, análise de componentes principais, distribuição de frequência, geoestatística e estatística não-paramétrica.

Atualmente, os pesquisadores não têm dúvida de que o clima mundial está mudando, com o aumento das temperaturas e a mudança na distribuição de chuvas. Estudos recentes preparados para o relatório do Painel Intergovernamental de Mudanças Climáticas sugerem um aumento na média global de temperatura de 1,4 a 8,5 graus até o fim deste século. Isso pode levar a uma aceleração do ciclo hidrológico e conseqüentemente intensificar eventos extremos [Alexander et al. 2006, Ganguly and Steinhäuser 2008]. Logo, eventos como deslizamentos de terra, inundações e tempestades podem ocorrer com maior frequência e intensidade [IPCC 2007], requerendo assim mais estudos para compreender melhor o clima tanto regional quanto mundial.

Nesse cenário, o estudo de dados climáticos é importante para entender as tendências de fenômenos extremos de modo a preparar especialistas e populações para situações adversas, criando condições para atenuar as conseqüências e auxiliar nas tomadas de decisões. De acordo com Ganguly e Steinhäuser [Ganguly and Steinhäuser 2008], há ainda inúmeros desafios na área de meteorologia e climatologia que podem ser apoiados pela aplicação de técnicas conhecidas de mineração de dados, bem como pelo desenvolvimento de novas técnicas. Eles afirmam ainda que mesmo técnicas simples podem resultar em descobertas científicas com importante impacto social. À medida que o volume de dados cresce, passando da ordem de gigabytes para petabytes, torna-se fundamental o uso de técnicas computacionais para tornar possível análises mais abrangentes e eficientes, como por exemplo:

1. Análise eficiente de múltiplas séries temporais climáticas visando encontrar padrões e mudanças de tendência.
2. Identificação dos eventos climáticos extremos que indiquem alterações do clima regional e global.

Com o objetivo de auxiliar os especialistas do domínio na execução dessas tarefas, neste trabalho de Iniciação Científica foi proposto um processo para a análise de séries temporais climáticas, seguindo uma abordagem que avalia múltiplas séries temporais como uma *data stream* multidimensional, de maneira que cada série temporal define um atributo da *data stream*. Assim, é possível integrar múltiplas variáveis climáticas em um processo unificado que combina:

1. monitoramento da *data stream* para descoberta de padrões e detecção de mudanças de comportamento, utilizando conceitos da Teoria dos Fractais;
2. agrupamento visando encontrar similaridade e desvio de padrões revelados quando os dados são analisados com diferentes granularidades temporais;
3. e uma análise estatística para identificar mudanças em variáveis individuais associadas a uma variação significativa de comportamento.

Estudos experimentais realizados em séries temporais climáticas reais de diferentes regiões do Brasil indicam que essa abordagem pode ser uma ferramenta útil para auxiliar os especialistas na análise de grandes conjuntos de dados climáticos. O trabalho apresentado neste artigo foi desenvolvido no escopo do projeto “Agrodatamine: Desenvolvimento de métodos e técnicas de mineração de dados para apoiar pesquisas em mudanças climáticas, com ênfase em agrometeorologia”¹. O processo de análise proposto

¹Projeto apoiado pelo instituto Microsoft Research - FAPESP

foi desenvolvido como trabalho de iniciação científica, com o apoio de especialistas de domínio da Embrapa Informática Agropecuária - Campinas e do Cepagri-Unicamp, parceiros do projeto, para análise e validação dos resultados.

Este artigo está organizado da seguinte maneira: Seção 2 apresenta conceitos da Teoria de Fractais e sua aplicação na análise de *data streams*. Seção 3 descreve a abordagem proposta para análise de séries climáticas. Resultados experimentais em dados reais são discutidos na Seção 4. A Seção 5 apresenta as considerações finais.

2. Conceitos Relacionados

Um fractal pode ser definido pela propriedade de auto-similaridade, isto é, um objeto que apresenta as mesmas características em diferentes escalas [Schroeder 1991]. Assim, um conjunto de dados reais exibe um comportamento fractal se é exatamente ou estatisticamente auto-similar, de modo que as partes de qualquer tamanho dos dados apresentam as mesmas características do conjunto de dados inteiro.

Na Teoria de Fractais, a dimensão fractal é particularmente útil para a análise dos dados, pois fornece uma estimativa da dimensão intrínseca D de conjuntos de dados reais. A dimensão intrínseca fornece a dimensionalidade do objeto representado pelos dados, independentemente da dimensão E do espaço em que está inserido. Em outras palavras, D mede o comportamento intrínseco de dados reais [Faloutsos and Kamel 1994, Traina et al. 2005]. Por exemplo, um conjunto de pontos que definem um plano imerso em um espaço tridimensional ($E = 3$) tem dois atributos independentes e um terceiro correlacionado com os outros, resultando em $D = 2$.

A dimensão fractal de conjuntos de dados reais pode ser determinada pela Dimensão Fractal de Correlação D_2 . Uma abordagem eficiente para medir a dimensão fractal de conjuntos de dados imersos em espaços E -dimensional é o método *BoxCounting* [Schroeder 1991], que define D_2 como apresentado na Equação 1, onde r é o lado das células de um hiper-reticulado que divide o espaço de endereçamento do conjunto de dados e $C_{r,i}$ é a contagem de pontos na célula i .

$$D_2 \equiv \frac{\partial \log(\sum_i C_{r,i}^2)}{\partial \log(r)} \quad r \in [r_1, r_2] \quad (1)$$

Um algoritmo eficiente (com custo linear sobre o número de elementos no conjunto de dados) para calcular D_2 foi proposto em [Traina et al. 2000]. Assim, D_2 pode ser uma ferramenta útil para estimar a dimensão intrínseca D de conjuntos de dados reais, com custo computacional viável. Conceitos da Teoria dos Fractais têm sido aplicados a diversas tarefas em mineração e análise de dados, tais como estimativa de seletividade [Baioco et al. 2007], detecção de agrupamentos [Barbará and Chen 2000], previsão em séries temporais [Chakrabarti and Faloutsos 2002], detecção de correlação [Sousa et al. 2007b] e análise da distribuição de dados [Traina et al. 2005].

As informações de comportamento intrínseco fornecidas pela dimensão fractal D_2 também podem ser aplicadas para detectar mudanças de comportamento em *data streams* evolutivas. Basicamente, a ideia é medir continuamente a dimensão fractal de uma *data stream* ao longo do tempo visando monitorar seu comportamento. Assim, variações significativas nas medidas sucessivas de D_2 podem indicar as mudanças nas características intrínsecas dos dados.

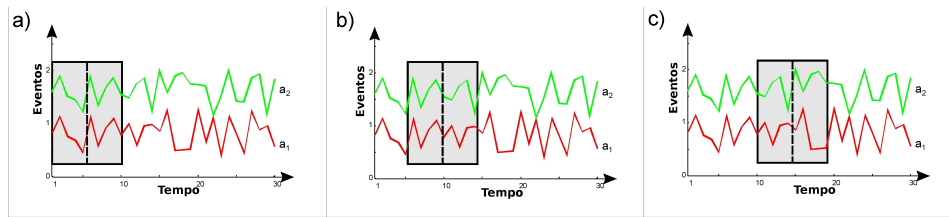


Figura 1. Janela deslizante sobre uma *data stream* bi-dimensional.

Em [Sousa et al. 2007a] é proposta uma técnica para acompanhar as mudanças de comportamento de *data streams* evolutivas, com o algoritmo *SID-meter* para medir continuamente D_2 ao longo do tempo. O *SID-meter* considera uma *data stream* como uma sequência de eventos $\langle e_1, e_2, \dots \rangle$ ordenada e potencialmente infinita onde cada evento é representado por um conjunto de E medidas, isto é, $e_i = (a_1, a_2, \dots, a_E)$.

Para medir a dimensão fractal de uma *data stream* ao longo do tempo, o *SID-meter* define uma janela deslizante para limitar eventos sucessivos que serão considerados no cálculo de D_2 . A janela é dividida em n_c períodos de tempo (chamados de períodos de contagem), cada qual contendo um número pré-determinado de eventos (n_i). Portanto, $n_i * n_c$ determina o tamanho da janela deslizante e n_i representa o tamanho do deslocamento. O valor de D_2 é, então, continuamente calculado considerando os eventos contidos em cada janela e atualizado quando n_i novos eventos são lidos, descartando os n_i eventos mais antigos. A Figura 1 ilustra uma *data stream* bi-dimensional (atributos a_1, a_2) processados por meio de uma janela deslizante dividida em 2 períodos de contagem ($n_c = 2$). Na Figura 1a), pode-se notar que no intervalo temporal de 1 a 10 (a primeira janela), os atributos a_1 e a_2 apresentam comportamento semelhante, pois são linearmente correlacionados. Nas Figuras 1b) e 1c), os eventos antigos são descartados enquanto novos dados são processados. A partir do tempo 10, o comportamento dos atributos a_1 e a_2 torna-se diferente, ou seja, eles não estão mais correlacionados. Quando ocorre alteração na correlação entre os atributos, o valor de D_2 reflete essas mudanças. Portanto, com o monitoramento da dimensão fractal é possível acompanhar a evolução do comportamento dos dados.

3. Processo de Análise

O processo de análise proposto para detectar padrões de interesse em múltiplas séries temporais combina diferentes técnicas: monitoramento da dimensão fractal de *data streams*, agrupamento de dados e análise estatística.

A Figura 2 ilustra o processo completo. O primeiro passo é associar múltiplas séries temporais em uma *data stream* multidimensional, ou seja, cada série temporal determina um atributo da *data stream*. Por exemplo, séries temporais de temperatura e de precipitação podem ser agregadas em uma única *data stream* bi-dimensional. Essa abordagem permite uma análise integrada das diferentes variáveis climáticas, a fim de descobrir as alterações de comportamento global ao longo do tempo. Em outras palavras, é possível avaliar como as variáveis estão correlacionadas e como essas correlações variam, especialmente quando mudanças de comportamento significativas são identificadas.

O segundo passo é a análise fractal de uma *data stream* utilizando a abordagem proposta no *SID-meter*. O algoritmo *SID-meter* foi inicialmente projetado para trabalhar

com uma única janela deslizante com tamanho pré-definido. Uma extensão do *SID-meter* [Nunes et al. 2010] foi projetada para manipular janelas com diferentes configurações de períodos de contagens (n_c) e número de eventos (n_i) simultaneamente em uma única leitura da *data stream*. Como resultado, pode-se encontrar padrões temporais semelhantes ou distintos que ocorrem em diferentes granularidades de tempo, por exemplo, mensalmente ou anualmente.

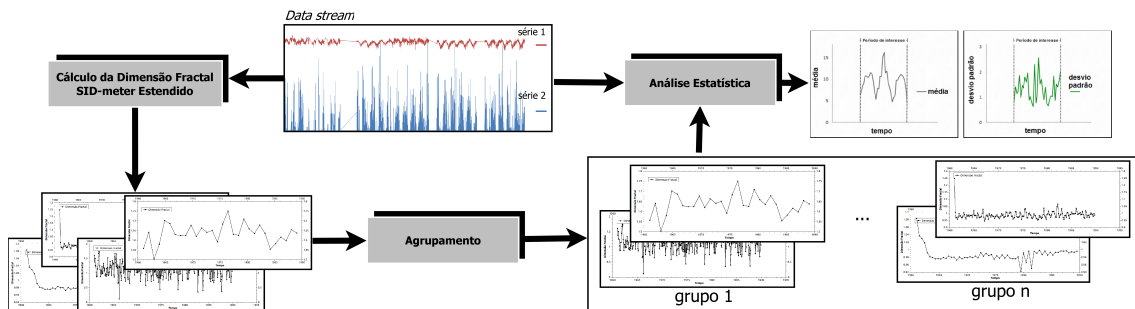


Figura 2. Processo de análise de múltiplas séries temporais.

O *SID-meter* estendido gera janelas deslizantes de tamanhos diferentes a partir de parâmetros iniciais. Basicamente, o menor e o maior tamanho de janela devem ser determinados por meio das seguintes configurações:

- o menor e o maior valor de n_c (períodos de contagem) e de n_i (número de eventos);
- o valor de incremento de n_c e de n_i .

Por exemplo, considere uma *data stream* composta de medidas diárias: configurando n_c variando de 3 a 6 com incremento de 1, e n_i variando de 30 (um mês para cada período de contagem, aproximadamente) a 365 (1 ano para cada período de contagem) com incremento de 30, define-se janelas variando de $3 * 30 = 3$ meses a $6 * 365 = 6$ anos. Esses parâmetros devem ser determinados considerando a granularidade temporal dos dados e os objetivos da análise a ser realizada. Portanto, os especialistas do domínio são fundamentais nessa tarefa.

A saída do *SID-meter* estendido é um conjunto de gráficos de D_2 gerados para as janelas de diferentes tamanhos. Assim, na terceira etapa do processo de análise, mapeia-se cada gráfico em uma série temporal de medidas de D_2 . As séries temporais de D_2 são, então, agrupadas com o objetivo de encontrar padrões semelhantes em granularidades de tempo diferentes e padrões distintos que são detectados em apenas algumas janelas temporais. Além disso, os elementos centrais dos agrupamentos fornecem algumas informações adicionais para os especialistas do domínio sobre como escolher janelas de interesse para uma análise mais aprofundada, ou seja, um elemento central representa o padrão geral do grupo (*cluster*) e, portanto, a janela temporal correspondente (com uma granularidade de tempo específica) pode ser analisada em mais detalhes.

A tarefa de agrupamento combina o *K-Medoids* [Kaufman and Rousseeuw 1990], que é um método de agrupamento baseado em partições, com a função de distância *Dynamic Time Warping* (DTW) [Berndt and Clifford 1994] para medir a similaridade entre as séries temporais de D_2 . Essas técnicas são bastante utilizadas em trabalhos de mineração de séries temporais e apresentaram resultados satisfatórios nos primeiros estudos empíricos.

O último passo do processo é a análise estatística, que pode ser aplicada considerando-se apenas as janelas temporais representadas pelos elementos centrais dos agrupamentos. Medidas de estatística básica, como média e desvio padrão, podem indicar a variação individual das variáveis climáticas em períodos com alterações significativas na dimensão fractal. O objetivo principal é descobrir como cada variável climática influencia as mudanças de correlação, em especial aquelas relacionadas a eventos extremos.

Este processo foi aplicado em séries temporais georreferenciadas provenientes de diferentes regiões do Brasil, com o objetivo de identificar mudanças de comportamento significativas ao longo do tempo e relacioná-las aos eventos climáticos extremos e relevantes. Os resultados são apresentados na Seção 4.

4. Experimentos

Foram realizados experimentos com séries temporais reais para validar a abordagem proposta. Dois conjuntos de dados de diferentes regiões do Brasil foram utilizados:

1. Sul: séries climáticas reais, fornecidas pelo Agritempo², constituídas de medidas diárias de precipitação e temperatura média mensuradas por 19 estações meteorológicas da região Sul no período 1994-2008.
2. Sudeste: séries climáticas reais, também fornecidas pelo Agritempo, constituídas de medidas diárias de precipitação e temperatura média mensuradas por 10 estações meteorológicas do estado de São Paulo, no período de 1995 a 2009.

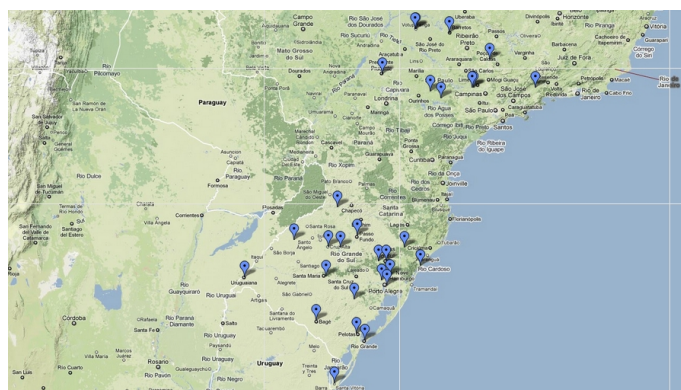


Figura 3. Localização das estações meteorológicas - SP e região Sul.

As séries temporais analisadas nos experimentos são georreferenciadas, uma vez que a posição geográfica das estações meteorológicas afeta significativamente a análise dos resultados. Em geral, o clima apresenta características diferentes em regiões geograficamente distintas. A Figura 3 mostra a localização das estações meteorológicas nas regiões Sul e Sudeste. Vale ressaltar que o estado de São Paulo tem um número significativo de estações, quando comparado aos outros estados da região Sudeste e, portanto, é analisado separadamente.

Para cada conjunto de dados gerou-se uma *data stream* bi-dimensional composta pelos atributos de precipitação e temperatura média. Como parâmetros iniciais do *SID-meter* estendido definiu-se o número de períodos de contagem (n_c) de 2 a 5, com

²www.agritempo.gov.br/

deslocamentos (n_i) de 1 mês a 1 ano. Em outras palavras, foram definidas janelas com tamanho variando de dois meses a cinco anos.

Os gráficos que descrevem a dimensão fractal D_2 medida ao longo do tempo pelo *SID-meter* estendido para as duas *data streams* apresentam tendências significativamente diferentes, embora representem as mesmas variáveis climáticas (temperatura e precipitação) e o mesmo intervalo de tempo. As diferenças nos padrões observados nos gráficos mostram as características climáticas distintas das duas regiões. Também observa-se que, para ambos os conjuntos de dados, o valor de D_2 permaneceu em torno de 1, o que indica que temperatura média e precipitação são variáveis correlacionadas. Esse comportamento é consistente com as expectativas dos especialistas, uma vez que a correlação entre essas variáveis alterna de uma correlação mais forte em um determinado período a uma menor correlação em outros.

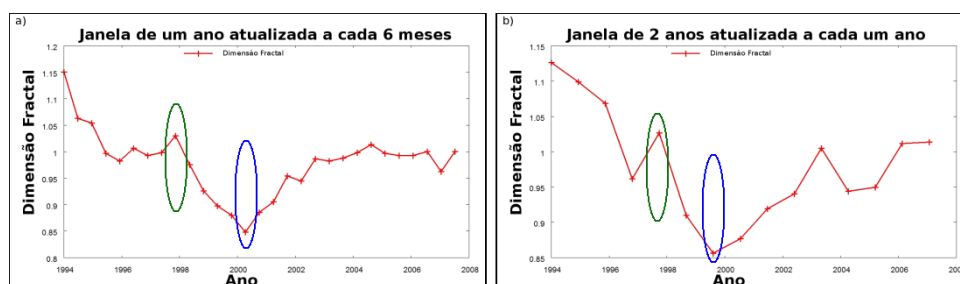


Figura 4. Variação de D_2 para a região Sul: (a) janela de um ano com deslocamento de 6 meses; (b) janela de 2 anos com deslocamento de um ano.

A Figura 4 mostra dois gráficos de D_2 para a região Sul, considerando janelas temporais de diferentes tamanhos. É possível observar que, embora os gráficos estejam relacionados a diferentes granularidades temporais, há um padrão similar de variação de D_2 . Ambos os gráficos mostram um pico em 1998 e um declínio no valor de D_2 em torno do ano 2000. Ao avaliar a flutuação da média e do desvio padrão para as variáveis de precipitação e temperatura durante 1997/1998 e 2000/2001, pode-se observar maiores valores de desvio padrão da precipitação (variação 14,7 a 19,8) no primeiro período e maiores valores de desvio padrão da temperatura (variação 6,2 a 8,9) no segundo período. De acordo com especialistas e registros de dados de temperatura, os invernos de 2000 e 2001 foram marcados por baixas temperaturas, clima muito frio e neve no Rio Grande do Sul. Além disso, o período de 1999/2001 coincide com a ocorrência de La Niña, um fenômeno climático caracterizado pelo resfriamento das águas do Pacífico.

Durante o período 1997-1998 ocorreu um El Niño muito forte. El Niño é um fenômeno de grande escala que ocorre no Oceano Pacífico, ao longo da costa da América do Sul, e é caracterizado por um aumento na temperatura da água do mar nessa região. Além disso, os Complexos Convectivos de Meso-escala (CCMs) influenciam fortemente a precipitação ao longo do noroeste do Rio Grande do Sul durante a primavera [Velasco and Fritsch 1987]. Durante o fenômeno El Niño, eventos extremos podem mover-se para o sul ou para o norte, mas ambos os casos indicam a importância da parte ocidental do Sul do Brasil no aumento de eventos extremos durante anos de El Niño. CCMs são frequentes nessa área e aproveitam as anomalias de circulação durante os episódios de El Niño [Grimm and Tedeschi 2008].

A Figura 5 mostra dois gráficos de D_2 para São Paulo, considerando duas janelas temporais de diferentes tamanhos com um padrão semelhante de comportamento. A Figura 5a) corresponde a uma janela de dois anos atualizada a cada ano. É possível observar que o valor de D_2 diminui no período de 2000 a 2002, que coincide com a ocorrência de La Niña. Nesse período (2000/2001), houve um alerta em São Paulo sobre a possibilidade de racionamento de eletricidade no centro Sul, provavelmente devido a uma diminuição de precipitação na região central do Brasil. No Sudeste, os anos de 2000 e 2001 foram caracterizados por uma estação seca pronunciada. Em 2000, os meses de abril, maio e junho foram muito secos com atraso na chegada da estação chuvosa.

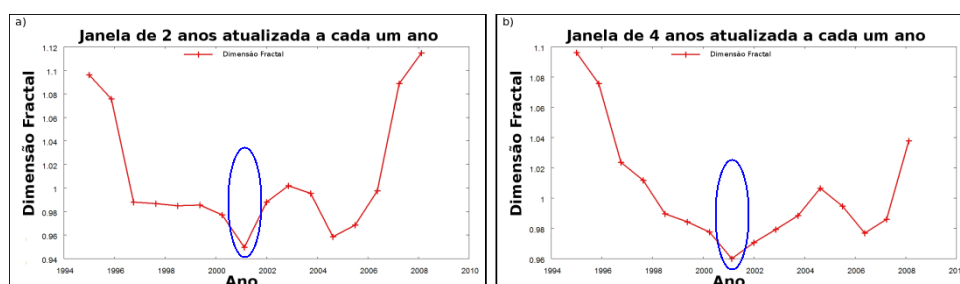


Figura 5. Variação de D_2 para o estado de São Paulo: (a) janela de 2 anos com deslocamento de um ano; (b) janela de 4 anos com deslocamento de um ano.

Os gráficos das Figuras 5a) e 5b) mostram uma mudança de tendência entre 2004 e 2009. Uma análise estatística desse período indica que o ano de 2007 apresentou uma grande variação nos valores da média e desvio padrão para as variáveis temperatura e precipitação. Em 2007, a região teve um verão quente e chuvoso, e um outono mais úmido e um pouco mais quente. Segundo os especialistas, em anos de El Niño há um aumento nas ocorrências de chuvas entre outubro e fevereiro, com uma pausa em janeiro. No entanto, no caso de precipitação mensal total, essa variação é menos pronunciada, ocorrendo em algumas regiões e em outras não.

Em geral, observa-se que, para ambas as regiões analisadas, os anos de El Niño apresentaram uma diminuição na correlação entre a temperatura e a precipitação (maior D_2). Por outro lado, as ocorrências de La Niñas estão relacionadas com um aumento desta correlação (menor D_2). Segundo meteorologistas, em anos de La Niña, as chuvas são causadas pelos CCMs, que são chuvas intensas, tempestades severas, o que pode causar uma descontinuidade na relação entre as variáveis.

O processo de análise inclui também uma etapa de agrupamento sobre os gráficos gerados pelo *SID-meter* estendido. O objetivo principal nesta etapa é encontrar padrões similares que são revelados mesmo que os dados analisados tenham granularidades de tempo diferentes, tais como as variações relevantes de D_2 identificadas nas Figuras 4a) e 4b). Além disso, o especialista também recebe como saída os gráficos selecionados como elementos centrais dos grupos. Assim, pode-se determinar o tamanho da janela apropriada a ser usada para destacar os fenômenos climáticos que estão sendo avaliados. Para ilustrar, a Figura 6 mostra os elementos centrais de cinco agrupamentos gerados para os gráficos de São Paulo. Com os elementos centrais, o especialista pode ter uma visão geral do conjunto de gráficos dentro de cada agrupamento e definir quais grupos poderiam ser utilizados para análises posteriores. Essa abordagem facilita a tarefa de análise, já que o número de possíveis combinações de configurações de janela pode ser elevado.

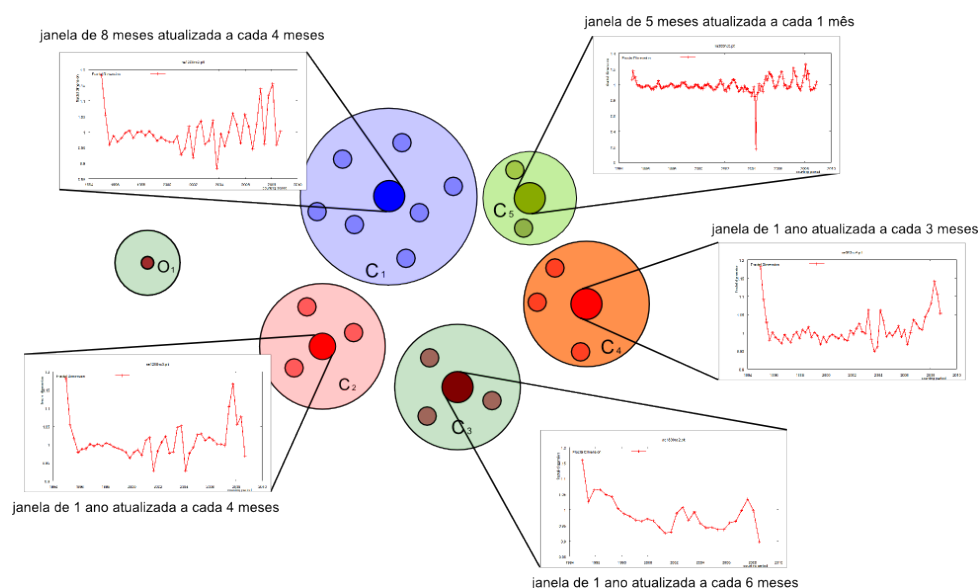


Figura 6. Grupos de gráficos D_2 para São Paulo (C - grupos e O - extremos).

5. Conclusão

A análise de séries temporais de precipitação e temperatura baseada em ferramentas da Teoria dos Fractais revela que existe um padrão rítmico no tempo entre as variáveis. Quando ocorre uma mudança de tendência, a abordagem apresentada neste trabalho é capaz de identificar a mudança e as variáveis responsáveis pela variação no padrão. Assim, é possível relacionar essas variações com as mudanças nos padrões climáticos globais, como o El Niño. Estes resultados indicam que o estudo de séries temporais com mais de uma variável pode permitir a identificação de padrões em mais detalhes do que utilizando apenas uma variável. Além disso, essa abordagem pode ajudar a compreender a interdependência entre as variáveis meteorológicas. Portanto, a análise realizada diretamente sobre dados de estações meteorológicas permite identificar o comportamento dos dados de diferentes estações, detectando eventos extremos em cada região.

Os trabalhos futuros incluem a comparação do comportamento de dados climáticos reais com o de dados simulados gerados a partir de modelos climáticos. Assim, pretende-se detectar tendências no comportamento dos dados reais e possíveis desvios nos modelos climáticos, possibilitando o refinamento desses modelos e maior compreensão dos especialistas do domínio sobre o comportamento dos dados reais.

Agradecimentos

Os autores agradecem Embrapa, Fapesp, CNPq, Capes e a Microsoft Research pelo apoio financeiro e ao Agritempo por fornecer os dados climáticos reais.

Referências

Alexander, L., Zhang, X., Peterson, T., Caesar, J., Gleason, B., Tank, A., Haylock, M., Collins, D., Trewin, B., Rahimzadech, F., Tagipour, A., Kumar, K. R., Revadekar, J., Griffiths, G., Vincent, L., Stephenson, D., Burn, J., Aguilar, E., Brunet, M. Taylor, M., New, M., Zhai, P., Rusticucci, M., and Vasquez-Aguirre, J. (2006). Global

- observed changes in daily climate extremes of temperature and precipitation. *Journal of Geophysical Research*, 111:1–22.
- Baioco, G. B., Traina, A. J. M., and Traina, C. (2007). Mamcost: Global and local estimates leading to robust cost estimation of similarity queries. In *Proceedings of SSDBM*, pages 6–16, Banff, Canada.
- Barbará, D. and Chen, P. (2000). Using the fractal dimension to cluster datasets. In *Proceedings of KDD*, pages 260–264, Boston, USA.
- Berndt, D. J. and Clifford, J. (1994). Using dynamic time warping to find patterns in time series. In *Proceedings of AAAI94 Workshop on Knowledge Discovery in Databases*, pages 359–370, Seattle, USA.
- Chakrabarti, D. and Faloutsos, C. (2002). F4: large-scale automated forecasting using fractals. In *Proceedings of CIKM*, pages 2–9, McLean, USA.
- Faloutsos, C. and Kamel, I. (1994). Beyond uniformity and independence: Analysis of r-trees using the concept of fractal dimension. In *Proceedings of PODS*, pages 4–13, Minneapolis, USA.
- Ganguly, A. R. and Steinhäuser, K. (2008). Data mining for climate change and impacts. In *Proceedings of ICDM*, pages 385–394, Pisa, Italy.
- Grimm, A. and Tedeschi, R. G. (2008). Enso and extreme rainfall events in south america. *Journal of Climate*, 22:1589–1609.
- IPCC (2007). Intergovernmental panel on climate change. <http://www.ipcc.ch/ipccreports/index.htm>. accessed: March, 2009.
- Kaufman, L. and Rousseeuw, P. J. (1990). *Finding Groups in Data: an Introduction to Cluster Analysis*. John Wiley and Sons.
- Nunes, S. A., Romani, L. A. S., Avila, A. M. H., Traina, C., Sousa, E. P. M., and Traina, A. J. M. (2010). Análise baseada em fractais para identificação de mudanças de tendências em múltiplas séries climáticas. In *Proceedings of SBB D Short Papers*, pages 65–72, Belo Horizonte, Brasil.
- Schroeder, M. (1991). *Fractals, Chaos, Power Laws*. W. H. Freeman and Company.
- Sousa, E. P. M., Traina, C., Traina, A. J. M., and Faloutsos, C. (2007a). Measuring evolving data streams' behavior through their intrinsic dimension. *New Generation Computing Journal*, 25:33–59.
- Sousa, E. P. M., Traina, C., Traina, A. J. M., Wu, L., and Faloutsos, C. (2007b). A fast and effective method to find correlations among attributes in databases. *DMKD*, 14(3):367–407.
- Traina, C., Sousa, E. P. M., and Traina, A. J. M. (2005). Using fractals in data mining. In Kantardzic, M. M. and Zurada, J., editors, *New Generation of Data Mining Applications*, volume 1, pages 599–630 (Chapter 24). Wiley/IEEE Press.
- Traina, C., Traina, A. J. M., Wu, L., and Faloutsos, C. (2000). Fast feature selection using fractal dimension. In *Proceedings of SBB D*, pages 158–171, João Pessoa, Brasil.
- Velasco, I. and Fritsch, J. (1987). Mesoscale convective complexes in the americas. *J. Geophys. Res.*, 92:9591–9613.