

Revisão

Práticas Simples em Análises Climatológicas: Uma Revisão

Rodrigo José Bombardi¹, Leila Maria Véspoli de Carvalho²¹*Department of Atmospheric, Oceanic, and Earth Sciences, George Mason University, Fairfax, VA, USA.*²*Department of Geography, University of California, Santa Barbara, CA, USA.*

Recebido em 6 de Setembro de 2016 – Aceito em 24 de Abril de 2017

Resumo

Muitos estudos meteorológicos e climatológicos utilizam metodologias que superestimam ou até subestimam a significância estatística dos resultados. Análises que subestimam o papel de tendências e dependência temporal e espacial nos dados podem levar a conclusões errôneas. Por outro lado, análises desnecessariamente rigorosas podem enfraquecer os resultados. O objetivo deste artigo é discutir algumas práticas simples, muitas vezes negligenciadas, que podem produzir resultados muito mais robustos e estatisticamente significativos. Este artigo discute alguns problemas relacionados ao cálculo do ciclo anual médio e anomalias, às análises de tendências e à dependência temporal e espacial, dando ênfase a testes de hipóteses.

Palavras-chave: ciclo anual médio, tendência, dependência temporal, significância de campo, teste de hipótese.

Simple Practices in Climatological Analyses: A Review

Abstract

Many studies in Meteorology and Climatology use methodologies that overestimate or even underestimate the statistical significance of the results. Analyses that underestimate the role of trends and temporal or spatial dependency in the data sets can lead to incorrect conclusions. On the other hand, unnecessarily rigorous analyses can undermine the conclusions. The objective of this article is to discuss some simple practices, commonly neglected, that can produce results much more robust and statistically significant. This paper discusses some problems related to the calculation of the mean annual cycle and anomalies, trend analyzes, and temporal and spatial dependency, emphasizing statistical hypothesis testing.

Keywords: mean annual cycle, trend, temporal dependency, field significance, hypothesis testing.

1. Introdução

Qualquer meteorologista ou climatologista ficaria perplexo se alguém tentasse explicar a relação entre a temperatura média global e a criminalidade em grandes centros urbanos, ou entre o aumento do buraco da camada de ozônio e o número de pessoas no Facebook. Contudo, por mais absurdas que essas relações possam parecer, correlações estatísticas entre essas variáveis podem ser altas simplesmente pelo fato que todas elas possuem tendências e essas tendências dominam o comportamento temporal dessas variáveis relativamente às flutuações de origem estocástica. Entretanto, nem tudo está perdido. Existem práticas simples em análises meteorológicas e climatológicas que podem minimizar as chances de encontrarmos relações

estatisticamente significativas entre fenômenos sem nenhuma relação. Portanto, este artigo tem como objetivo revisar algumas dessas práticas simples as quais, infelizmente, são comumente negligenciadas em análises meteorológicas e climatológicas e que podem ajudar a encontrar resultados mais robustos e auxiliar em diagnósticos e interpretações de resultados. Embora exista um número grande de práticas que poderiam ser citadas aqui, discutiremos aquelas comumente utilizadas em artigos científicos como o método mais adequado para o cálculo de anomalias do ciclo anual e a avaliação da significância estatística de tendências e de campos espaciais e temporais.

A seção 2 explora problemas comuns no cálculo do ciclo anual médio e anomalias. A seção 3 revisa problemas em análises e comparações de dados com tendências. Pro-

blemas relacionados à dependência temporal são discutidos na seção 4. A seção 5 faz uma revisão de problemas oriundos à dependência espacial de dados climatológicos. As considerações finais são apresentadas na seção 6.

2. O Cálculo do Ciclo Anual e de Anomalias

O cálculo de anomalias é provavelmente o ponto de partida da maioria das análises climatológicas. Um aspecto importante no cálculo de anomalias, que é comumente negligenciado, é a correta estimação do ciclo anual médio. Hartmann e Michelsen (1989) discutem que o ciclo anual médio não pode ser caracterizado apenas pela série temporal anual média porque essa média resulta em uma curva com oscilações de alta frequência (Fig. 1a). Parte destas oscilações de alta frequência são resíduos provenientes de imperfeições na amostragem das variáveis atmosféricas. Quando o ciclo anual médio é suavizado, estes resíduos são removidos sem que a verdadeira variância anual seja reduzida significativamente. Comumente o ciclo anual médio é suavizado filtrando-se a série temporal (Hartmann e Michelsen, 1989) ou através do ajuste dos três primeiros harmônicos do ciclo anual médio (Kikuchi *et al.*, 2012;

Kiladis *et al.*, 2014). É importante mencionar que o ciclo anual médio deve ser suavizado principalmente quando trabalhamos com dados de alta frequência (*e.g.* dados diários ou pentadais). O ciclo anual calculado a partir de dados mensais ou sazonais não precisa ser suavizado.

O problema do cálculo das anomalias discutido aqui está ilustrado no seguinte exemplo. A Fig. 1a mostra o ciclo anual médio de precipitação (CPC_UNI, Chen *et al.*, 2008; Xie *et al.*, 2007) sobre uma pequena região no Brasil central [52.5° W-47.5° W, 20° S-15° S] utilizando dados diários e considerando o período de 1979 a 2014. Estão incluídos também o ciclo anual médio suavizado utilizando-se uma média móvel de três pontos (filtro 1-2-1; aplicado 300 vezes) e os três primeiros harmônicos do ciclo anual médio. A Fig. 1b mostra anomalias de precipitação para o período de 2000 a 2010 calculadas a partir do ciclo anual médio não suavizado. Analogamente, a Fig. 1c mostra anomalias de precipitação calculadas a partir do ciclo anual médio suavizado (filtro 1-2-1). As duas séries temporais são bastante semelhantes e apresentam desvios padrões semelhantes também (Fig. 1b,c). Por outro lado, se padronizamos as séries temporais de anomalias verificamos que a suavização do ciclo anual tem um papel importante (Fig. 1d,e). A

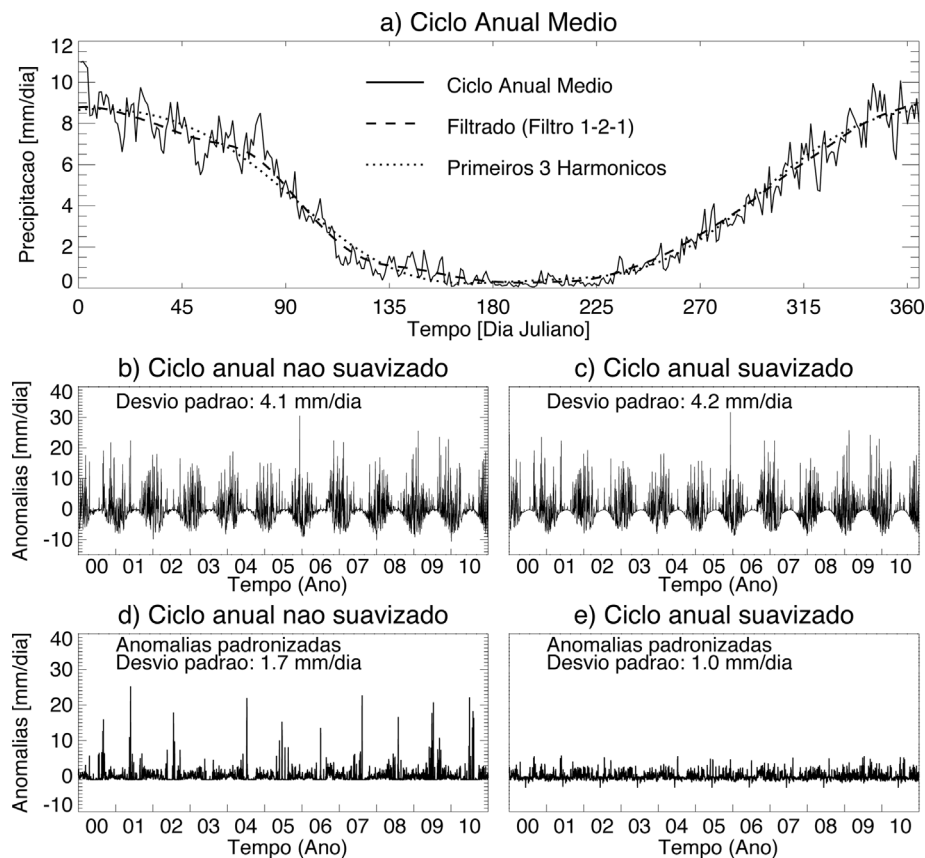


Figura 1 - a) Ciclo anual médio de precipitação não suavizado e suavizado utilizando-se um filtro 1-2-1 (passado 300 vezes) e os três primeiros harmônicos do ciclo anual médio; b) anomalias calculadas a partir do ciclo anual médio não suavizado; c) anomalias calculadas a partir do ciclo anual médio suavizado (filtro 1-2-1); d) anomalias padronizadas calculadas a partir do ciclo anual médio não suavizado; e) anomalias padronizadas calculadas a partir do ciclo anual médio suavizado (filtro 1-2-1). O valor do desvio padrão das séries de anomalias entre o período de 2000 a 2010 está mostrado nos painéis b) a e).

série temporal de anomalias padronizadas calculada a partir do ciclo anual não suavizado apresenta valores aberrantes, o que não é verificado na série de anomalias padronizadas calculadas a partir do ciclo anual suavizado (Fig. 1d,e). Ou seja, se o ciclo anual médio não é suavizado somos erroneamente levados a concluir de que há mais variância em nossos dados do que de fato existe. Isso pode ter implicações enormes em análise climáticas (por exemplo, análise de anomalias médias associadas a um determinado fenômeno) já que essa variância pode estar aleatoriamente distribuída em diversas frequências espectrais.

O cálculo de anomalias depende também do objetivo do estudo. Se o objetivo do estudo é comparar a importância relativa de anomalias em diferentes estações do ano ou comparar importância relativa de anomalias entre baixas e altas latitudes, as anomalias devem ser padronizadas. A padronização é obtida calculando-se a diferença entre o valor do dado de cada dia e o valor do mesmo dia no ciclo anual médio suavizado e dividindo-se a diferença pelo desvio padrão de cada dia (calculado com os dados de cada dia, por exemplo, todos os valores dos dias 2 de janeiro). É importante frisar que o desvio padrão de cada dia deve ser calculado em relação às anomalias obtidas com respeito à média suavizada. A anomalia padronizada é dada em número de desvios padrões para aquele dia (ou pântada, ou mês, etc.), pois a cada dia (pântada, mês, etc.) corresponde a uma variável diferente.

Se, por outro lado, o objetivo do estudo é investigar como as flutuações de duas ou mais séries temporais são coerentes, é necessário primeiro remover a tendência dos dados. Voltando ao exemplo da correlação entre a criminalidade nos centros urbanos e o aquecimento global mencionados na introdução, o único motivo pelo qual estes dois conjuntos de dados mostram uma correlação estatisticamente significativa é o fato de que ambos os dados possuem tendências temporais. Portanto, ao se correlacionar duas variáveis é preciso remover as tendências lineares de cada conjunto de dado. Só assim, estaremos realmente avaliando a covariância entre as flutuações (de origem estocástica) de cada uma das variáveis em questão. Entretanto, se o intuito do estudo é investigar tendências, o cálculo das anomalias se dá pela remoção do ciclo anual médio apenas, como descrito anteriormente.

Remover tendências é um procedimento extremamente importante não apenas no cálculo de correlações lineares de Pearson. Recomenda-se que tendências sejam removidas também quando se calculam covariâncias (ou correlações) nos procedimentos como análises de Funções Ortogonais Empíricas (EOFs). Ao se aplicar EOFs a dados com tendências, um dos modos principais (normalmente o primeiro) certamente representará a tendência dos dados. Cabe ao pesquisador decidir se a tendência deve ser incluída na análise ou não. Problemas comuns relacionados à análise de tendências são descritos na seção 3.

3. Tendências

Dois problemas comuns em análises de tendências são: 1) a escolha da metodologia apropriada para a estimativa da tendência e 2) a escolha da metodologia apropriada para a estimativa da significância estatística da tendência. Cálculos de tendências são muito comuns em análises climatológicas para identificar mudanças climáticas.

O método dos mínimos quadrados é comumente utilizado na estimativa de tendências. O problema com esse método é que este método é paramétrico, ou seja, baseado na suposição de que o conjunto de dados possui distribuição essencialmente Gaussiana. Além disso, esses métodos são apenas úteis quando se assume que a tendência monotônica é linear e pode ser representada por uma equação da reta. Métodos que não dependem dessa suposição são chamados *robustos*. O método dos mínimos quadrados também é afetado por valores aberrantes, onde alguns valores extremos podem afetar significativamente a estimativa dos parâmetros. Métodos que não são afetados por valores aberrantes são chamados *resistentes*.

O teste Mann-Kendall é uma alternativa de método mais robusto amplamente utilizado em análises de tendências (Chandler e Scott, 2011; Wilks, 2011; Zilli *et al.*, 2016). Para utilizar esse teste deve-se calcular a seguinte estatística S (e.g., Zilli *et al.*, 2016).

$$s = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{sgn}(x_j - x_i) \quad (1)$$

onde n é o número de pontos da série; x representa as medidas no tempo; i e j são índices temporais, com $i \neq j$; sgn é definido como (Wilks 2011):

$$\text{sgn} = \begin{cases} 1, & \text{se } (x_j - x_i) > 0 \\ 0, & \text{se } (x_j - x_i) = 0 \\ -1, & \text{se } (x_j - x_i) < 0 \end{cases} \quad (2)$$

O teste de Mann-Kendall deve ser aplicado para checar a hipótese nula (aquela que queremos rejeitar) de não existência de tendência contra a hipótese alternativa da presença de uma tendência. Valores positivos da estatística S indicam uma tendência positiva (de aumento da variável), enquanto que os valores negativos indicam uma tendência negativa (diminuição da variável) com o tempo. Como $n > 0$ então a estatística S segue uma distribuição Gaussiana (note que a estatística segue a distribuição Gaussiana, e não necessariamente os dados), com uma média igual a zero ($E[S] = 0$) e variância estimada por:

$$\text{var}(S) = \frac{n(n-1)(2n+5) - \sum_{i=1}^m t_i(t_i-1)(2t_i+5)}{18} \quad (3)$$

onde m representa o número de grupos de valores repetidos (se esses existirem, obviamente), e t_i é o número de valores repetidos no i -ésimo grupo (Wilks, 2011; Zilli *et al.*, 2016).

Uma vez calculada a variância, pode-se agora utilizar a estatística z (também conhecida como ‘z-scores’ em inglês), para se estimar a significância:

$$z = \begin{cases} \frac{S-1}{\sqrt{\text{var}(S)}}, S > 0 \\ \frac{S+1}{\sqrt{\text{var}(S)}}, S < 0 \end{cases} \quad (4)$$

Se o valor absoluto de z (estimado pela Eq. (4)) for maior que o valor de z ($\alpha/2$) obtido pela tabela z (assumindo-se um teste bilateral, onde α é o nível de significância desejado – por exemplo 5%) então podemos rejeitar a hipótese nula da não existência de tendência nos dados. A tabela z representa a distribuição normal $P(0 \leq Z < z_0)$. Ou seja, ela fornece a probabilidade de que a variável transformada Z esteja entre um valor z_0 e zero. A estatística Z geralmente mostra a probabilidade de apenas um lado da distribuição normal. Portanto, se o nível de significância de interesse é α , deve-se usar o valor de z_0 referente a $\alpha/2$ na tabela z . Por exemplo, se o nível de significância de interesse é $\alpha = 0.05$, o valor de z_0 de interesse é aquele cuja probabilidade é igual a 0.475 (ou 0.500 - $\alpha/2$).

Existem alguns outros problemas que precisam ser considerados. Note que o método Mann-Kendall é baseado na suposição de que a série de dados não possui dependência temporal, o que não é verdade para a maioria dos dados meteorológicos e climatológicos. Quando existe dependência temporal nos dados (quer dizer, os dados exibem um certo grau de ‘memória’ no tempo), a variância calculada pode ser subestimada e a significância estatística superestimada. A dependência temporal pode ser levada em conta no cálculo do teste adaptando-se a fórmula da variância (Eq. (3)) de modo a se considerar um número ‘efetivo’ de observações (o qual deve substituir o número total de observações) conforme será discutido na seção 4.

Entretanto, há outros meios de se reduzir a dependência temporal. Por exemplo, remover o ciclo anual dos dados (sobretudo se forem dados diários) ou agregá-los, por exemplo, como dados mensais. Para compreender este problema, considere a série temporal da precipitação discutida

acima. É fácil compreender que deve existir muito mais independência temporal entre “Janeiros” em anos consecutivos do que entre Janeiro e Fevereiro do mesmo ano. Para entender esse exemplo, considere um ano de El Niño. Se a região de estudo for significativamente afetada por esse fenômeno poderíamos esperar que as chuvas de dezembro e janeiro sejam ambas afetadas pelo fenômeno e, portanto, exibam uma maior semelhança (mais ou menos chuva, dependendo do tipo de teleconexão) do que em anos subsequentes. O fato de existir uma ‘memória’ em dados subsequentes pode resultar em uma redução da variância. Outros problemas relacionados à dependência temporal serão explorados em maior detalhe na Seção 4.

A Tabela 1 mostra valores de tendência (coeficiente angular do ajuste linear) e a significância estatística da tendência calculadas para as mesmas séries de precipitação e anomalias de precipitação utilizadas na análise da Fig. 1 e para o período entre 2000 e 2010, incluindo a série de dados brutos e as três diferentes séries de anomalias padronizadas. O coeficiente angular foi calculado utilizando o método de mínimos quadrados e o método Mann-Kendall. Note que o método dos mínimos quadrados não é a melhor opção para se estimar a tendência de dados diários de precipitação, já que os desvios em relação à média nestes dados não possuem distribuição normal. Além disso, pode haver eventos extremos de precipitação que vão afetar o ajuste de mínimos quadrados. O método dos mínimos quadrados fornece valores de coeficientes angulares muito maiores do que o método Mann-Kendall (Tabela 1).

A Tabela 1 também apresenta estimativas de significância estatística para os coeficientes angulares do ajuste de tendência nas séries temporais. Uma das vantagens do teste Mann-Kendall é o fato do método fornecer tanto o coeficiente angular do ajuste linear como a significância estatística da tendência linear. Já a significância estatística do ajuste de mínimos quadrados foi avaliada utilizando-se a técnica de re-amostragem, também conhecida como *Bootstrap* (Wilks, 2011). Esta técnica consiste em se reorganizar aleatoriamente a série temporal e calcular o ajuste de mínimos quadrados. Este procedimento é repetido várias vezes (*e.g.* 1000 vezes) e os valores dos coeficientes angulares obtidos em cada ajuste são ordenados do menor para o

Tabela 1 - Comparação entre o cálculo de tendências lineares e da significância estatística de tendências lineares para séries de precipitação e anomalias padronizadas de precipitação considerando o método de mínimos quadrados, o teste Mann-Kendall, e o teste de re-amostragem.

Série de dados	Coeficiente angular		Significância estatística	
	Mínimos Quadrados	Mann-Kendall	Re-amostragem	Mínimos Quadrados
Dados brutos	-0,000151	-0,0000001	Significativa ao nível de 5%	Não significativa
Anomalias padronizadas (ciclo anual não suavizado)	-0,000151	-0,0000001	Significativa ao nível de 5%	Não significativa
Anomalias padronizadas (ciclo anual suavizado com filtro 1-2-1)	-0,000014	-0,000017	Não significativa	Não significativa
Anomalias padronizadas (ciclo anual suavizado com harmônicos)	-0,000014	-0,000016	Não significativa	Não significativa

maior. Por fim, verifica-se a posição do ajuste original em relação à série ordenada de ajustes. O ajuste original é considerado significativo se coincide com os valores nas extremidades das caudas da distribuição de ajustes lineares das séries temporais reorganizadas (e.g. inferior a 2.5% ou superior a 97.5% para teste bilateral).

No caso da série de precipitação sobre o Brasil central, verificamos que ao se utilizar a metodologia menos apropriada, ou seja, a conjunção do teste menos apropriado (mínimos quadrados para conjunto de dados em questão) com o conjunto de dados menos apropriados (série de dados brutos ou anomalias calculadas com o ciclo anual médio não suavizado) obtêm-se erroneamente resultados estatisticamente significativos. Por outro lado, ao se utilizar a metodologia mais apropriada (teste Mann-Kendall) e o conjunto de dados mais apropriados (anomalias calculadas com o ciclo anual suavizado) a tendência temporal não é considerada significativa (Tabela 1).

Maiores informações sobre problemas relacionados a análises de tendências, sobre o cálculo de tendências e sobre métodos para se verificar a significância estatística de tendências podem ser encontrados nos sites do *Climate Data Guide*, do *Climatic Research Unit* e do *Pacific Northwest National Laboratory*.

4. Dependência Temporal

Conforme discutido anteriormente, dados meteorológicos e climatológicos possuem dependência temporal. Portanto, ao se avaliar a significância estatística de correlações entre variáveis climatológicas, é preciso levá-la em conta (Wilks, 2011). Considere, por exemplo, o caso em que queremos avaliar a significância estatística da correlação entre anomalias mensais de temperatura da superfície do mar (TSM) e o fenômeno El Niño – Oscilação Sul (ENOS). Para isso, utilizamos dados mensais de TSM (OI-SST; Reynolds *et al.*, 2007) e o índice ENOS calculado pelo *Climate Prediction Center* (CPC) para o período de 1982 - 2014. Mais informações sobre o índice ENOS pode ser encontradas no site do CPC-NOAA. O teste de significância estatística pode ser facilmente realizado calculando-se a estatística t_0 (Eq. (5)), que possui distribuição t -Student para correlação nula, com r igual ao valor do coeficiente de correlação de Pearson entre a série temporal de anomalias de TSM a cada ponto de grade e a série temporal do índice ENOS. N é o número de dados em cada série temporal.

$$t_0 = \frac{r\sqrt{N-2}}{\sqrt{1-r^2}} \quad (5)$$

Se o valor absoluto de t_0 for maior do que o valor de $t(v, \alpha)$ na tabela t de Student (e.g. Wilks, 2011) para o número de graus de liberdade v e o nível de significância de interesse α (e.g. $\alpha = 5\%$), rejeita-se a hipótese nula (na maioria das análises em climatologia e meteorologia esta-

mos interessados em testes para duas caudas, uma vez que tanto valores positivos quanto negativos são relevantes). Neste caso a hipótese nula afirma que não há correlação. O maior problema nesta análise é estimar o correto número graus de liberdade $v = N - 2$. Se considerarmos N como o número total de dados no tempo ($N = 396$ meses) estaremos superestimando o valor do número de graus de liberdade, dado que ambos os conjuntos de dados (TSM e índice ENOS) possuem alta dependência temporal. Dessa forma corremos o risco de erroneamente rejeitar a hipótese nula (erro do tipo I). Por outro lado, se considerarmos N como sendo o número de anos em nossa série temporal ($N = 33$), corremos o risco de subestimar o valor do número de graus de liberdade. Assim, estaríamos aplicando o teste estatístico com muita restrição (erro do tipo II). Uma forma simples de minimizar esse problema é simplesmente estimar o número efetivo de graus de liberdade. No caso de correlações entre conjuntos de dados diferentes, o número efetivo de graus de liberdade pode ser estimado a partir de propriedades auto-regressivas de ambos os conjuntos de dados (Livezey e Chen, 1983) (Eq. (6)).

$$\tau = \left[1 + 2 \sum_{i=1}^n C_{TSM}(i) C_{ENOS}(i) \right] \quad (6)$$

onde N é o número de dados (396 meses), i é o i -ésimo dado e C é a auto-correlação de lag i para as anomalias de TSM e o índice ENOS, respectivamente. A partir de τ podemos estimar o número de amostras independentes (N efetivo) na série temporal (Eq. (7)). O valor de N efetivo é então utilizado na Eq. (5).

$$N_{efetivo} = \frac{N}{\tau} \quad (7)$$

A Fig. 2a mostra a correlação entre anomalias de TSM e o índice ENOS incluindo a significância estatística considerando essas três formas diferentes de se estimar N . O sombreado mais claro mostra regiões de correlação significativa considerando N igual ao número total de dados (396) na série temporal (graus de liberdade superestimados). O sombreado mais escuro mostra regiões de correlação significativa considerando N igual ao número de anos (33) na série temporal (graus de liberdade subestimados). O sombreado intermediário mostra significância considerando-se o número efetivo de amostras independentes.

Outra técnica popularmente utilizada em análises exploratórias em climatologia são “composites” (ou compostos) de anomalias, em que se calcula a média de anomalias para algumas datas de interesse. Analogamente, é preciso considerar a dependência temporal destas anomalias no cálculo da significância estatística. Vamos considerar novamente o caso em que queremos avaliar a significância estatística de anomalias mensais de temperatura da superfície do mar (TSM) durante eventos El Niño. Neste caso, podemos utilizar um teste simples como o teste t para uma

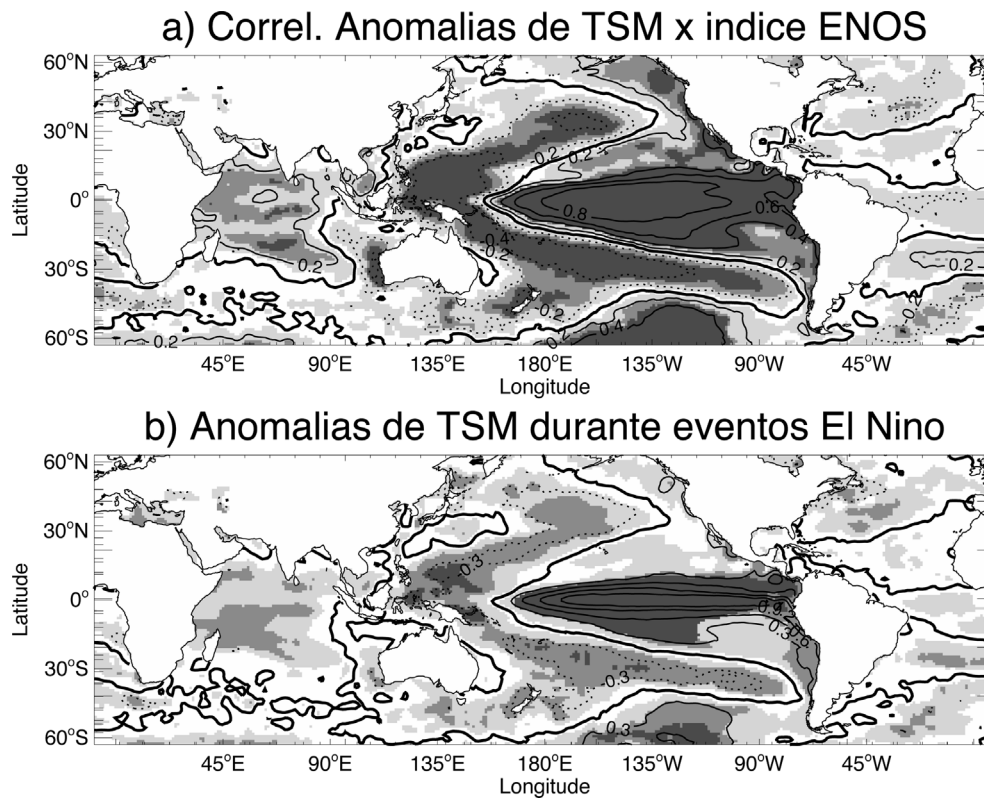


Figura 2 - a) Correlação entre anomalias de TSM a o índice ENOS (contorno) e b) composições de anomalias de TSM durante eventos El Niño (contorno). O sombreado mostra regiões estatisticamente significativas ao nível de 5%. O tom de cinza mais claro mostra regiões onde a significância estatística é estimada superestimando-se o número de graus de liberdade. O tom de cinza mais escuro mostra regiões estatisticamente significante subestimando-se o número de graus de liberdade. O tom intermediário mostra regiões estatisticamente significativas estimada com o valor apropriado para o número de graus de liberdade (veja o texto para maiores detalhes).

amostra. O teste t de uma amostra examina a hipótese nula de que a média de uma amostra retirada de uma população é centrada em um valor previamente especificado, μ_0 (no nosso caso assumimos $\mu_0 = 0$). Se o número de dados na amostra é grande o suficiente para que sua distribuição seja Gaussiana, de acordo com o Teorema do Valor Central, a estatística do teste (Eq. (8)) segue uma distribuição conhecida como t de Student, ou simplesmente distribuição t (Wilks, 2011).

$$t = \frac{\bar{x} - \mu_0}{\left[\frac{s^2}{N} \right]^{1/2}} \quad (8)$$

onde \bar{x} é a média da amostra (média das anomalias de TSM durante eventos El Niño), N é o número de dados e s^2 é a variância da amostra descrita pela Eq. (9),

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (9)$$

Na Eq. (9), x_i é o valor da i -ésima observação e n é número total de dados na amostra. Assim, o maior problema se resume a encontrar o número de graus de liberdade $v = N - 1$, onde N é o número de observações independentes. Novamente, N pode ser estimado levando em consideração

a propriedade auto-regressiva do conjunto de dados (Eq. (10)). O valor de ρ_1 é obtido da auto-correlação de lag 1 da série temporal dos dados de onde a amostra foi retirada (a série temporal completa das anomalias de TSM). O valor de N efetivo é então utilizado na Eq. (8) (Wilks, 2011).

$$N_{\text{efetivo}} = N \left(\frac{1 - \rho_1}{1 + \rho_1} \right) \quad (10)$$

A Fig. 2a mostra a composição de anomalias de TSM durante eventos El Niño incluindo a significância estatística considerando três formas diferentes de se estimar N . O sombreado mais claro mostra regiões onde as anomalias são significativas considerando N igual ao número total de dados (396) na série temporal (graus de liberdade superestimados). O sombreado mais escuro mostra regiões com anomalias significativas para N igual ao número de eventos El Niño (9) no período considerado (graus de liberdade subestimados). O sombreado intermediário mostra significância considerando-se o número efetivo de amostras independentes.

Vale lembrar que o teste t de Student é um teste paramétrico baseado na suposição de que o conjunto de dados possui distribuição essencialmente Gaussiana, o que é

verdade para anomalias de TSM. Entretanto, se estivéssemos investigando anomalias de precipitação teríamos que utilizar uma estratégia diferente. Uma alternativa seria aplicar um teste não paramétrico, que não requer que os dados pertençam a uma distribuição particular.

Outra alternativa seria transformar as anomalias de precipitação para uma distribuição normal. Vale lembrar também que o termo normalizar assume significados diferentes na literatura. Essa transformação é diferente da técnica de padronização. Existem várias técnicas para transformação como por exemplo transformação da raiz quadrada, da raiz cúbica e logarítmica, comumente aplicadas a dados de chuva. O usuário deve analisar qual a melhor transformação e se mesmo após transformar os dados eles ainda não apresentam distribuição normal, a única alternativa é o teste não paramétrico.

5. Dependência Espacial

Assim como dados meteorológicos e climatológicos apresentam dependência temporal, estes também exibem dependência espacial. O conceito de auto-correlação espacial parte do princípio de que “tudo está relacionado com todo o resto, mas coisas próximas estão mais relacionadas do que coisas distantes”, também conhecido como a primeira lei da Geografia (Tobler, 1970).

Ao se aplicar testes de significância em séries temporais obtemos algumas regiões espacialmente coerentes onde os resultados podem ser erroneamente considerados como estatisticamente significativos. Erros do tipo I em que a hipótese nula é erroneamente rejeitada tendem a se agregar no espaço, levando a conclusões errôneas de que existe uma região coerente onde os resultados são estatisticamente significativos. Este problema é conhecido na literatura como o problema da multiplicidade de testes. Portanto, sempre que se aplica múltiplos testes de hipóteses (como no caso de testes de séries temporais em pontos de grade ou mesmo em análises de perfis verticais da atmosfera) é apropriado se aplicar um teste de significância de campo (Wilks, 2006).

Os principais testes de significância de campo em meteorologia e climatologia são: 1) o teste proposto por Livezey e Chen (1983), que utiliza simulações de Monte Carlo; 2) um método de regressão proposto por Delsole e Yang (2011) e 3) um procedimento conhecido como Taxa de Falsa Detecção (*False Discovery Rate* – FDR) proposto por Ventura *et al.* (2004) e Wilks (2006).

O teste Livezey-Chen (Livezey e Chen, 1983) é baseado no método de Monte Carlo. Se tomarmos o mesmo exemplo da Fig. 2a, a ideia é calcular a correlação entre as anomalias de TSM e o índice ENOS e contar o número de pontos de grade onde a hipótese nula foi rejeitada (considerando também a dependência no tempo). Em seguida, gera-se uma série de dados aleatórios com o mesmo número de dados e as mesmas características estatísticas da série do índice ENOS original. No nosso exemplo, uma

série com distribuição Gaussiana seria suficiente. Então calcula-se a correlação entre as anomalias de TSM e a série de dados aleatórios e conta-se novamente o número de pontos onde a correlação é estatisticamente significativa. Esse processo é então repetido muitas vezes, sempre com uma série diferente de dados aleatórios. Os resultados obtidos com cada série de dados aleatórios (o número de pontos onde a correlação é estatisticamente significativa) são então ordenados do menor para o maior. O padrão de correlação original é dito ter significância de campo se o número de pontos que rejeitam a hipótese nula é superior a um limiar (*e.g.* o percentil de 95%) da cauda superior da distribuição dos números de pontos com significância estatística derivada das séries temporais aleatórias. Em resumo, 1) calcula-se várias vezes (*e.g.* 1000 vezes) o número de pontos onde a correlação é estatisticamente significativa utilizando-se diferentes séries temporais aleatórias para cada cálculo; 2) ordena-se os resultados do menor para o maior e 3) verifica-se a posição do resultado original em relação à série ordenada. A desvantagem deste teste é que ele ignora a intensidade e a localização das regiões estatisticamente significativas (Delsole e Yang, 2011; Wilks, 2016).

O teste proposto por Delsole e Yang (2011) utiliza uma regressão multivariada aplicada à análise de componentes principais das variáveis de interesse para estimar a significância de campo dos resultados. As desvantagens deste teste são que ele é mais complexo do que o testes Livezey-Chen e FDR, ignora a variabilidade ortogonal aos principais modos das componentes principais e fornece pouca informação sobre a localização das regiões estatisticamente significativas (Delsole e Yang, 2011).

O teste FDR é bastante simples de ser aplicado, tem baixo custo computacional e fornece informações sobre a localização das regiões estatisticamente significativas. Portanto, este é o método que escolhemos para demonstrar a estimativa da significância de campo dos nossos resultados (Fig. 3). Quando calculamos a significância estatística de um resultado, estamos de fato calculando a probabilidade de quanto o resultado se afasta de uma condição que satisfaz a hipótese nula (ou aquela que gostaríamos de rejeitar). Em outras palavras, calculamos a probabilidade de que o valor obtido ou a variável transformada (por exemplo, a variável z , t -student, etc.) se encontra em uma distribuição, a qual pode ser conhecida ou não. Essa probabilidade é conhecida como valor p . No caso de estudos meteorológicos ou climatológicos, a significância estatística é comumente apresentada em um campo, em um mapa. Portanto, podemos calcular o valor p para todos os pontos do nosso domínio de estudo e criar um mapa de valores p . O método FDR se baseia em ordenar os valores de um campo de valor p e encontrar o ponto em que a curva de valores p intercepta um limiar de probabilidade FDR. Esse limiar depende do nível de significância de interesse, α . Wilks (2016) sugere calcular o teste FDR utilizando um valor de α igual ao dobro do valor global de interesse ($\alpha = 2\alpha_{\text{global}}$). Ou seja,

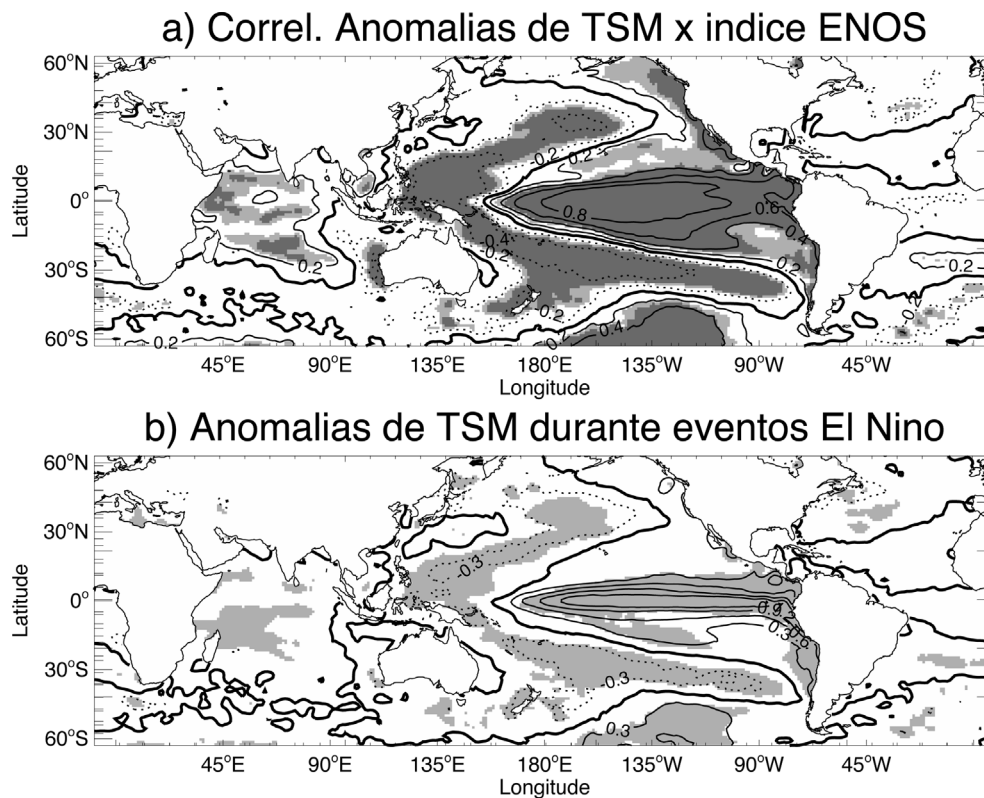


Figura 3 - a) Correlação entre anomalias de TSM a o índice ENOS (contorno) e b) composições de anomalias de TSM durante eventos El Niño (contorno). O sombreado mostra regiões estatisticamente significativas ao nível de 5%. O tom de cinza mais claro mostra regiões estatisticamente significativas estimadas com o valor apropriado para o número de graus de liberdade mas sem considerar dependência espacial. O tom de cinza mais escuro mostra regiões estatisticamente significativas estimadas com o valor apropriado para o número de graus de liberdade e considerando a dependência espacial. Note que em (b) nenhuma região é considerada estatisticamente significativa ao nível de 5% quando se aplica o teste FDR (veja o texto para maiores detalhes).

α_{global} é o nível de significância de interesse e α é o nível de significância utilizado no cálculo do teste FDR. No nosso caso o nível de significância de interesse é 5% ($\alpha_{\text{global}} = 0.05$) e, portanto, o teste FDR é calculado utilizando-se $\alpha = 0.1$. Entretanto, é importante ressaltar que as regiões significativas devem apresentar tanto valores p menores que α_{global} (significância estatística considerando-se dependência temporal) quanto significância de campo em relação a α (significância estatística considerando-se dependência espacial). Uma desvantagem do teste FDR é que o teste não é válido para o caso em que os testes de hipótese são altamente correlacionados no espaço (Delsole e Yang, 2011).

A Fig. 3 mostra o teste de significância de campo para os resultados das mesmas análises feitas na seção 2: correlação entre anomalias de TSM e o índice ENOS (Fig. 3a) e composições de anomalias de TSM durante eventos El Niño (Fig. 3b). Neste caso, primeiramente calculamos a significância estatística considerando a dependência temporal. Ou seja, considerando-se o número efetivo de amostras independentes mas sem considerar a significância de campo. Neste caso, o sombreado em cinza claro mostra as regiões estatisticamente significativas ao nível de 5%. Em

seguida, aplicamos o teste FDR e o sombreado em cinza escuro mostra as regiões estatisticamente significativas ao nível de 5% neste caso. Note que as regiões com correlações estatisticamente significativas são menores quando se considera significância de campo em comparação com o caso em que a significância de campo é ignorada (Fig. 3a).

O teste de significância de campo não é uma prática muito comum na maioria dos artigos científicos em ciências atmosféricas, apesar do fato de estes testes já existirem há bastante tempo. Uma das razões é o fato de que muitos cientistas não estão cientes do problema. Outro motivo está relacionado ao fato de que o resultado destes testes depende da escolha das fronteiras da região de estudo. Note que para o caso das anomalias de TSM durante eventos El Niño (Fig. 3b), nenhuma região foi considerada significativa quando se aplica o teste FDR. Se ao invés de calcularmos FDR sobre todo o domínio calculássemos o teste FDR apenas sobre a bacia do Oceano Pacífico obteríamos padrões espaciais muito semelhantes aos padrões espaciais obtidos considerando apenas a dependência temporal (não mostrado). Entretanto, isso seria incorreto. Escolher um domínio para incluir apenas regiões com resultados estatisticamente significantes é uma forma de manipulação de resultados, é uma prática desonesta. Portanto, métodos para

a estimativa da significância de campo devem fornecer uma explicação de como selecionar a região de interesse. Uma alternativa ao uso de testes de significância de campo seria estimar a significância estatística de um resultado científico utilizando-se a mesma análise (*i.e.* considerando apenas a dependência temporal) em diferentes fontes de dados (DelSole, comunicação pessoal e livro em preparação). Se diferentes conjuntos de dados mostram padrões espaciais de significância estatística semelhantes entre si, há menos incerteza no resultado obtido.

6. Considerações Finais

Este artigo forneceu uma breve revisão de algumas práticas simples mas muito importantes em análises climatológicas que são comumente empregadas em estudos diagnósticos e prognósticos. Essas práticas, muitas vezes ignoradas ou negligenciadas, produzem resultados robustos e estatisticamente significativos e auxiliam na interpretação de campos, diagnósticos e prognósticos.

Por exemplo, o ciclo anual não pode ser calculado como sendo simplesmente o ciclo anual médio. A melhor estimativa para o ciclo anual deve levar em consideração a suavização do ciclo anual médio. As técnicas mais comuns envolvem a utilização de uma média móvel ou dos três primeiros harmônicos do ciclo anual médio.

A estimativa de tendências em séries temporais devem ser aplicadas a séries temporais de anomalias. Deve-se, também, sempre considerar os requisitos nos quais os métodos de estimativa de tendências e estimativa de significância estatísticas são baseados. O uso de métodos inapropriados pode levar a conclusões errôneas a respeito do sinal e significância estatística de tendências.

A dependência temporal deve ser considerada na estimativa dos graus de liberdade de testes estatísticos. Erros na estimativa dos números de graus de liberdade podem levar tanto à superestimação como à subestimação de regiões com significância estatística e, portanto, a conclusões errôneas.

Como ainda não temos uma boa noção da melhor forma de se escolher as fronteiras da região de estudo em testes de significância de campo, a opinião dos autores deste artigo é que se deve aplicar testes de significância de campo com cautela na maioria das análises espaciais efetuadas. Entretanto, a recomendação é que os testes de significância de campo não sejam ignorados em caso de interpretações de significâncias em campos com estruturas fragmentadas ou baseados em número restrito de amostras independentes. Estes testes tornar-se-ão indispensáveis e mais robustos à medida que compreendamos melhor a relação entre as fronteiras do domínio de estudo e a significância estatística dos resultados.

Agradecimentos

Este artigo tem o apoio financeiro da NOAA (NA15NWS4680018). Rodrigo Bombardi gostaria de agradecer o Dr. Timothy DelSole por sua ajuda na interpretação de testes de significância de campo e por compartilhar o manuscrito do seu livro em preparação: *Statistics: An Introduction for Climate Scientists*. By Timothy DelSole and Michael Tippett. Rodrigo Bombardi gostaria de agradecer também ao Dr. Daniel Wilks por sua ajuda na implementação do teste FDR.

Referências

- CHANDLER, R.E.; SCOTT, E.M. **Statistical Methods for Trend Detection and Analysis in the Environmental Sciences**. [s.l.] John Wiley & Sons, 2011.
- CHEN, M.; SHI, W.; XIE, P.; SILVA, V.B.S.; KOUSKY, V.E. *et al.* Assessing objective techniques for gauge-based analyses of global daily precipitation. **Journal of Geophysical Research**, v. 113, n. D4, p. D04110, 29 2008.
- DELSOLE, T.; YANG, X. Field Significance of Regression Patterns. **Journal of Climate**, v. 24, n. 19, p. 5094-5107, 2011.
- HARTMANN, D.L.; MICHELSEN, M.L. Intraseasonal Peridocities in Indian Rainfall. **Journal of Atmospheric Sciences**, v. 46, n. 18, p. 2838-2862, 1989.
- KIKUCHI, K.; WANG, B.; KAJIKAWA, Y. Bimodal representation of the tropical intraseasonal oscillation. **Climate Dynamics**, v. 38, n. 9-10, p. 1989-2000, 2012.
- KILADIS, G.N.; DIAS, J.; STRAUB, K.H.; WHEELER, M.C.; TULICH, S.N. *et al.* A Comparison of OLR and Circulation-Based Indices for Tracking the MJO. **Monthly Weather Review**, v. 142, n. 5, p. 1697-1715, 2014.
- LIVEZEY, R.E.; CHEN, W.Y. Statistical Field Significance and its Determination by Monte Carlo Techniques. **Monthly Weather Review**, v. 111, n. 1, p. 46-59, 1983.
- REYNOLDS, R.W.; SMITH, T.M.; LIU, C.; CHELTON, D.B.; CASEY, K.S. *et al.* Daily High-Resolution-Blended Analyses for Sea Surface Temperature. **Journal of Climate**, v. 20, n. 22, p. 5473-5496, 2007.
- TOBLER, W.R.A Computer Movie Simulating Urban Growth in the Detroit Region. **Economic Geography**, v. 46, p. 234, 1970.
- VENTURA, V.; PACIOREK, C.J.; RISBEY, J.S. Controlling the Proportion of Falsely Rejected Hypotheses when Conducting Multiple Tests with Climatological Data. **Journal of Climate**, v. 17, n. 22, p. 4343-4356, 2004.
- WILKS, D.S. On "Field Significance" and the False Discovery Rate. **Journal of Applied Meteorology and Climatology**, v. 45, n. 9, p. 1181-1189, 2006.
- WILKS, D.S. **Statistical Methods in the Atmospheric Sciences**. Third ed. [s.l.] Academic Press, Burlington, MA 01803, USA, 2011.
- WILKS, D.S. "The stippling shows statistically significant grid-points": How Research Results are Routinely Overstated and Over-interpreted, and What to Do About It. **Bulletin of the American Meteorological Society**, p. BAMS-D-15-00267.1, 9 2016.
- XIE, P.; CHEN, M.; YANG, S.; YATAGAI, A.; HAYASAKA, T. *et al.* A Gauge-Based Analysis of Daily Precipitation over

East Asia. **Journal of Hydrometeorology**, v. 8, n. 3, p. 607-626, 2007.

ZILLI, M.T.; CARVALHO, L.M.V.; LIEBMANN, B.; SILVA, D.; MARIA, A. A comprehensive analysis of trends in extreme precipitation over southeastern coast of Brazil. **International Journal of Climatology**, v. 37, n. 5, p. 2269-2279, 2017.

Endereços de Internet

CPC-NOAA: http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears.shtml.

Climate Data Guide: <https://climatedataguide.ucar.edu/climate-data-tools-and-analysis/trend-analysis>.

Climatic Research Unit: https://crudata.uea.ac.uk/projects/stardex/Linear_regression.pdf.

Pacific Northwest National Laboratory: http://vsp.pnnl.gov/help/Vsample/Design_Trend_Mann_Kendall.htm.

Statistics: An Introduction for Climate Scientists: ftp://wxmaps.org/pub/delsole/dir_necessity/manuscript.02242017.pred.pdf.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License which permits unrestricted non-commercial use, distribution, and reproduction in any medium provided the original work is properly cited.