

## Análise de homogeneidade de temperatura mínima mensal para o Estado de São Paulo





*Empresa Brasileira de Pesquisa Agropecuária  
Embrapa Informática Agropecuária  
Ministério da Agricultura, Pecuária e Abastecimento*

# ***Boletim de Pesquisa e Desenvolvimento 41***

## **Análise de homogeneidade de temperatura mínima mensal para o Estado de São Paulo**

*José Ruy Porto de Carvalho  
Alan Massaru Nakai  
Aryeverton Fortes de Oliveira  
Eduardo Delgado Assad*

Embrapa Informática Agropecuária  
Campinas, SP  
2016

## **Embrapa Informática Agropecuária**

Av. André Tosello, 209 - Barão Geraldo  
Caixa Postal 6041 - 13083-886 - Campinas, SP  
Fone: (19) 3211-5700  
www.embrapa.br/informatica-agropecuaria  
sac: www.embrapa.br/fale-conosco/sac/

### **Comitê de Publicações**

Presidente: *Giampaolo Queiroz Pellegrino*

Secretária: *Carla Cristiane Osawa*

Membros: *Adhemar Zerlotini Neto, Stanley Robson de Medeiros Oliveira, Thiago Teixeira Santos, Maria Goretti Gurgel Praxedes, Adriana Farah Gonzalez, Neide Makiko Furukawa, Carla Cristiane Osawa*

Membros suplentes: *Felipe Rodrigues da Silva, José Ruy Porto de Carvalho, Eduardo Delgado Assad, Fábio César da Silva*

Supervisor editorial: *Stanley Robson de Medeiros Oliveira, Neide Makiko Furukawa*

Revisor de texto: *Adriana Farah Gonzalez*

Normalização bibliográfica: *Maria Goretti Gurgel Praxedes*

Editoração eletrônica/capa: *Neide Makiko Furukawa*

Imagens da capa: *Alan Massaru Nakai (mapa); <<http://mapio.net/s/29367719/>> (céu)*

### **1ª edição**

publicação digitalizada 2016

#### **Todos os direitos reservados.**

A reprodução não autorizada desta publicação, no todo ou em parte, constitui violação dos direitos autorais (Lei nº 9.610).

#### **Dados Internacionais de Catalogação na Publicação (CIP)**

Embrapa Informática Agropecuária

---

Carvalho, José Ruy Porto de.

Análise de homogeneidade de temperatura mínima mensal para o Estado de São Paulo / José Ruy Porto de Carvalho... [et al.]-  
Campinas : Embrapa Informática Agropecuária, 2016.

18 p. il. : cm. - (Boletim de pesquisa e desenvolvimento / Embrapa Informática Agropecuária, ISSN 1677-9266; 41 ).

1. Série temporal. 2. Homogeneização. 3. Climatol-R. 4. Variabilidade climatológica. I. Carvalho, José Ruy Porto de. IV. Embrapa Informática Agropecuária. IV. Título. V. Série.

---

CDD 21 ed. 519.5

© Embrapa 2016

# Sumário

<b>Resumo</b> .....	5
<b><i>Abstract</i></b> .....	7
<b>Introdução</b> .....	9
<b>Material e métodos</b> .....	10
<b>Resultados e discussão</b> .....	13
<b>Conclusões</b> .....	17
<b>Referências</b> .....	17



# Análise de homogeneidade de temperatura mínima mensal para o Estado de São Paulo

---

José Ruy Porto de Carvalho<sup>1</sup>

Alan Massaru Nakai<sup>2</sup>

Aryeverton Fortes de Oliveira<sup>3</sup>

Eduardo Delgado Assad<sup>4</sup>

## Resumo

Durante as últimas décadas, com as fortes evidências do aquecimento global, ganham importância os estudos sobre o comportamento e as mudanças dos padrões em séries climáticas. Para a coerência nas análises de dados, o processo probabilístico gerador dos valores da variável aleatória ao longo do tempo deve permanecer o mesmo. No caso de dados de estações meteorológicas, este pressuposto de homogeneidade no processo estocástico pode ser violado a partir de perturbações como mudanças de posição da estação, mudança de instrumentos ou mudança climática. Neste trabalho, o pacote Climatol-R foi aplicado a 122 séries mensais de temperatura mínima, cobrindo o interior do estado de São Paulo para os

---

<sup>1</sup> Estatístico, Ph.D. em Estatística, pesquisador da Embrapa Informática Agropecuária, Campinas, SP

<sup>2</sup> Cientista da computação, doutor em Ciência da Computação, analista da Embrapa Informática Agropecuária, Campinas, SP

<sup>3</sup> Economista, doutor em Economia Aplicada, pesquisador da Embrapa Informática Agropecuária, Campinas, SP

<sup>4</sup> Engenheiro-agrícola, Ph.D. em Sensoriamento Remoto e Geoprocessamento, pesquisador da Embrapa Informática Agropecuária, Campinas, SP

anos 1940 até 2012. Todas as estações estudadas apresentaram não homogeneidades. Uma das consequências deste trabalho de homogeneização foi a melhora da homogeneidade dos dados de temperatura mínima para a estação meteorológica IAC.00015 onde se obteve redução do desvio padrão de 0,86°C para 0,81°C. O pacote Climatol-R tende a melhorar espacialmente e temporalmente a série original, apresentando melhores resultados para a análise de variabilidade climatológica.

Palavras-chave: série temporal, homogeneização, Climatol-R, variabilidade climatológica.

# **Homogeneity analysis for monthly minimum temperature for the state of São Paulo**

---

## **Abstract**

*During the last decades, with strong evidence of global warming, become important studies on the behavior and patterns of changes in climate series. For consistency in data analysis, the probabilistic process of generating the random variable values over time should remain the same. In the case of data from meteorological station, this assumption of homogeneity in the stochastic process can be violated from disturbances such as changes in position of the station, changing instruments or climate change. In this work Climatol - R package was applied to 122 monthly minimum temperature series, covering the interior of São Paulo state for the years 1940 to 2012. All stations studied showed inhomogeneity. One consequence of this homogenization work was to improve the minimum temperature data homogeneity for the meteorological station IAC.00015 reducing the standard deviation of 0.86°C to 0.81°C. The Climatol -R package tends to improve spatially and temporally the original series showing better results for the analysis of climatological variability.*

*Index terms: time series, homogenization, Climatol -R, climatological variability.*



## Introdução

Uma série climatológica ao longo de diversos anos pode ser considerada homogênea somente quando as possíveis variações ao longo do tempo são causadas pelo clima (CONRAD; POLLAK, 1950). Entretanto, a série é usualmente afetada por fatores isolados que realmente não representam o comportamento climático. Esses fatores são inúmeros, tais como: mudança do local da estação; mudança de instrumento; mudança na forma como os dados são tratados; entre outros (HEINO, 1994). Por isso, diferentes testes estatísticos são usados para detectar mudanças artificiais ou não homogeneidades nas propriedades estatísticas das variáveis climáticas. Esses testes podem ser classificados em dois grupos: testes de homogeneidade absolutos, que consideram somente a série a ser testada, e testes relativos, aqueles que usam uma série de referência, comprovadamente homogênea, para compará-los com a série a ser testada. Homogeneização de séries temporais é amplamente reconhecida como um dos passos que devem ser tomados para a construção de conjuntos de dados confiáveis a longo prazo a partir de observações meteorológicas (SAHIN; CIGIZOGLU, 2010; TUOMENVIRTA, 2002).

O princípio de estatística mais comumente utilizado para detectar e remover os efeitos das mudanças artificiais é a homogeneização relativa, que pressupõe que estações próximas estão expostas ao mesmo sinal de clima e que, portanto, as diferenças entre as estações próximas podem ser utilizadas para detectar heterogeneidades (CONRAD; POLLACK, 1950).

Para verificar a homogeneidade de séries temporais, calcula-se uma série temporal de referência composta a partir de múltiplas estações próximas, compara-se esta série de referência à série candidata e assume-se que qualquer não homogeneidade encontrada é devida à série candidata (ALEXANDERSSON, 1986). A suposição funciona porque usando múltiplas estações como referência, a influência da falta de homogeneidade na série de referência é muito reduzida.

O objetivo deste estudo foi detectar não homogeneidades em séries temporais de temperatura mínima mensal e homogeneizá-las usando o pacote Climatol-R para o interior do estado de São Paulo.

## **Material e métodos**

Neste estudo foram utilizados dados de temperatura mínima mensal de 122 estações meteorológicas com altitude menor do que 600 metros para o período de 1940 a 2012. Esta região destaca-se no estado pela alta produção agrícola, principalmente de cana-de-açúcar. A estação IAC.00015, localizada na latitude -21,22 e longitude -48,93, foi usada como exemplo no trabalho. Esta estação apresenta uma série temporal longa, com poucas falhas nos dados e é representativa da região, portanto se apresenta como um estudo de caso pertinente.

Para que uma série de dados climáticos seja considerada homogênea em termos absolutos, as variações de observações devem apresentar propriedades estatísticas constantes ao longo do tempo. Em termos relativos, ou seja, comparando valores de séries, a homogeneidade equivale à ocorrência de um padrão climático regional que deve se repetir em todas as estações. No caso em que as observações fossem tomadas no mesmo local, dentro de um ambiente inalterado e usando instrumentos iguais e calibrados de acordo com o mesmo método, logicamente os valores deveriam ser iguais. Qualquer desvio, neste caso, ocorre em função de mudança de posicionamento de uma estação, do instrumento ou de falha de calibração, fatores esses não climáticos que devem ser isolados e corrigidos para análise do comportamento climático ao longo do tempo. As condições de existência de homogeneidade relativa entre séries de dados são favoráveis para estudos climáticos, havendo assim a probabilidade de estudar a homogeneidade absoluta das séries e atribuir estritamente ao fator climático qualquer não homogeneidade detectada. Os climatologistas recorrem às diferenças ou à relação entre as séries de estação climática e a uma série de referência criada a partir do conjunto de estações vizinhas para investigar se os valores são estatisticamente independentes e identicamente distribuídos, caracterizando assim a homogeneidade relativa de uma série de dados com o comportamento climático regional.

Peterson et al., 1998 revisaram técnicas para detectar e corrigir a não homogeneidade em séries temporais, e descrevem alguns procedimentos usados na construção de séries homogêneas. Eles ressaltam o que chamam de métodos objetivos como sendo os mais usados para a homogeneização. São os métodos de Potter, teste de homogeneidade normal

padrão, Standard Normal Homogeneity Test (SNHT), regressão linear múltipla, regressão de duas fases, teste de ponto de troca da ordem de classificação, teste de Craddock, teste T, Caussinus-Mestre e análise múltipla de séries para homogeneização. Sahin & Cigizoglu, 2010 usaram quatro métodos para detectar não homogeneidades em séries temporais: SNHT, teste de amplitude de Buishand, teste de Pettitt e teste da razão de Von Neumann.

Neste estudo foi utilizado o pacote CLIMATOL, software livre para a depuração e homogeneização de dados climatológicos desenvolvido por Guijarro Pastor (2004). É uma rotina dentro do software R que, por ser livre e multiplataforma, permite seu uso em grande variedade de situações. Além disso, o seu código é aberto e modificável, para que outros pesquisadores possam adaptá-lo as suas preferências.

Climatol é um conjunto de rotinas para aplicações climatológicas, dedicado, principalmente, à homogeneização da série mensal, porém pode ser aplicado aos dados diários. O método de homogeneização é baseado na comparação entre cada série de dados com uma série de referência construída para a mesma estação através de interpolação de relações, diferenças ou valores padronizados das estações vizinhas. A comparação da série de dados com as suas séries de referências permite a detecção de erros pontuais, possíveis mudanças e tendências por meio de testes estatísticos (MALVESTI; NERY, 2012). Os valores de referência calculados podem ser facilmente usados para preencher os dados faltantes da série.

Os dados podem ser conjuntos simples de cada estação (utilizando valores anuais), ou conjuntos de séries (valores mensais ou sazonais). O pacote Climatol foi aplicado na série diferença entre a estação a ser testada e a série de referência construída usando a média ponderada pela distância das estações próximas. A seleção destas estações próximas não se baseia somente na proximidade, mas também na correlação, pois as anomalias das séries temporais altamente correlacionadas são coincidentes no tempo. No pacote Climatol, dados originais são normalizados usando proporções (relações) ou diferenças, dependendo da variável climatológica. Proporções para valores normais climatológicos são apropriados para parâmetros meteorológicos limitados a zero com distribuição de probabilidade em forma de L (por exemplo, precipitação), enquanto que diferenças

em relação a Normal são mais adequadas para as variáveis com distribuição Normal (por exemplo, temperatura), o que equivale a aplicar um modelo estatístico de regressão linear do tipo II, em vez de conhecido tipo I. Na regressão linear ordinária (tipo I), o objetivo é o de minimizar o desvio entre as observações para a linha de regressão verticalmente. A pressuposição neste caso é que a variável independente  $X$  é controlada pelo investigador ou medida com erros insignificantes. Mas no caso do ajuste de regressão para os pares de série de uma rede climatológica, onde os erros são a priori semelhantes em todas as estações, o objetivo é o de minimizar o desvio perpendicular dos dados para a linha de regressão. Neste caso a regressão é ortogonal (tipo II).

Esta abordagem foi inspirada no método usado por Paulhus e Kohler (1952) para estimar dados faltantes na série de precipitação diária, que consiste em uma interpolação espacial da taxa de precipitação normal das estações vizinhas (GUIJARRO PASTOR, 2011; MAMARA et al., 2012). Uma característica deste tipo de regressão II é que a variância da variável estimada é a mesma que a da variável original.

Uma vez que os dados originais são normalizados, cada termo de cada série é estimado como uma média ponderada do número de dados disponíveis nas séries mais próximas. Os pesos a serem aplicados para os dados da série de referência podem ser todos iguais (média simples) ou ser calculados como uma função inversa da distância  $d$  entre os locais de observação. A função originalmente escolhida para isso foi  $(1 + d^2 / a)$ , onde o parâmetro  $a$  permite ao investigador modular o peso relativo das estações vizinhas em relação às estações mais distantes. Por conveniência, formula-se como  $(1 + d^2 / h^2)$ , pois nesta forma o parâmetro  $h$  torna-se a distância à qual o peso é a metade do peso da estação localizada no mesmo local do dado que está sendo estimado. Se séries temporais não estão completas, as médias e os desvios-padrão para todo o período de estudo não podem ser calculadas. Nesse caso, estes parâmetros são calculados a partir dos dados disponíveis, a série estimada (depois de desfazer a normalização) é usada para preencher os dados faltantes. Em seguida, as médias e os desvios-padrão são recalculados, e os dados são novamente normalizados. As estimativas da série são obtidas. Este processo é repetido até que a alteração máxima na média é menor do que (0,005 unidades).

Após ter estimado todos os dados, para cada série original é computada uma série de anomalias (diferenças entre os dados normalizados originais e estimados). É aplicado um procedimento para detectar outsiders (a série de anomalias é padronizada e anomalias  $> 5$  desvios-padrão resultará na eliminação do dado original correspondente) e mudanças na média (o teste de homogeneidade padrão - SNHT (ALEXANDERSSON, 1986) é aplicado em toda série de anomalias).

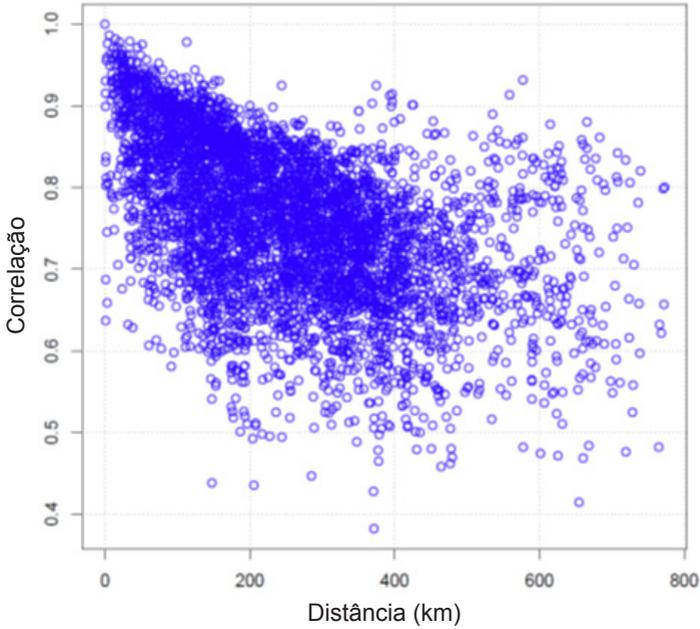
Os valores máximos do teste SNHT e suas localizações para cada conjunto de dados são retidos e a série com o maior valor de SNHT é dividida no ponto em que foi calculado o máximo. Valores após este ponto de interrupção são transferidos para uma nova série (com as mesmas coordenadas) e excluídos da série original. Quando todas as não homogeneidades detectadas pelo teste SNHT forem removidas pelo processo de separação, o SNHT é aplicado novamente para toda a série, possivelmente gerando mais quebras e correções na série.

## **Resultados e discussão**

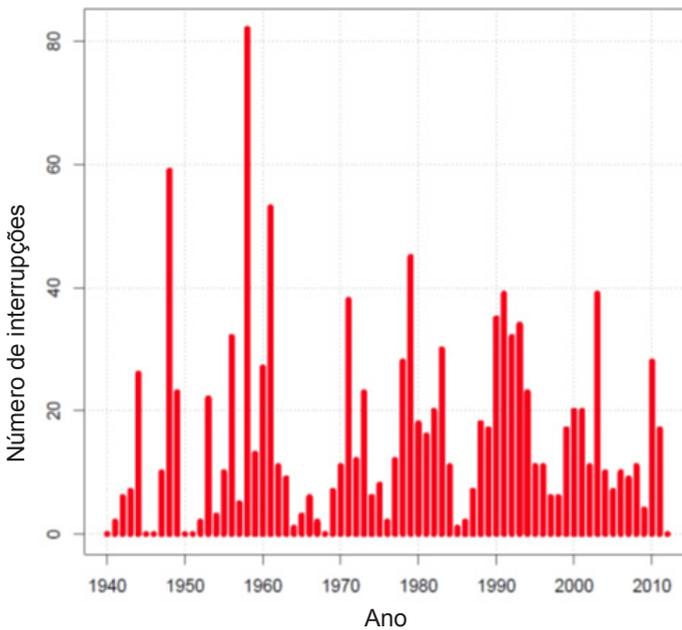
A série de referência é obtida usando a média ponderada dos valores das séries próximas com maior peso para as séries mais próximas. A proximidade geográfica tende a apresentar maior correlação com a série observada, apesar de a Figura 1 indicar que essa relação é bastante difusa.

As maiores correlações correspondem às estações mais próximas a uma distância de 50 km, entretanto para a mesma distância, verificam-se coeficientes de correlação variados, que vão de 0,65 até 0,95. Devido a esta variabilidade do coeficiente de correlação, os critérios de ponderação são decisivos. Além disso, o coeficiente de correlação apresenta falta de robustez para dados de variáveis que não são distribuídos normalmente, como os dados de precipitação, por exemplo. Cuidados com o uso do coeficiente de correlação foram observados em Freitas et al. (2013) e Venema et al. (2012).

A distribuição de não homogeneidades no período de 1940 até 2012 pode ser visualizada na Figura 2. Para cada estação, as interrupções mensais



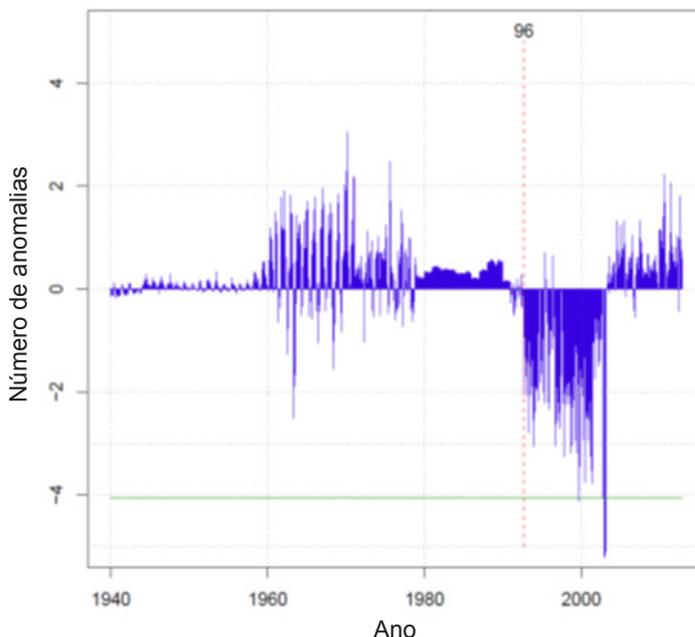
**Figura 1.** Coeficientes de correlação em função da distância.



**Figura 2.** Número de interrupções por ano para as 112 estações meteorológicas.

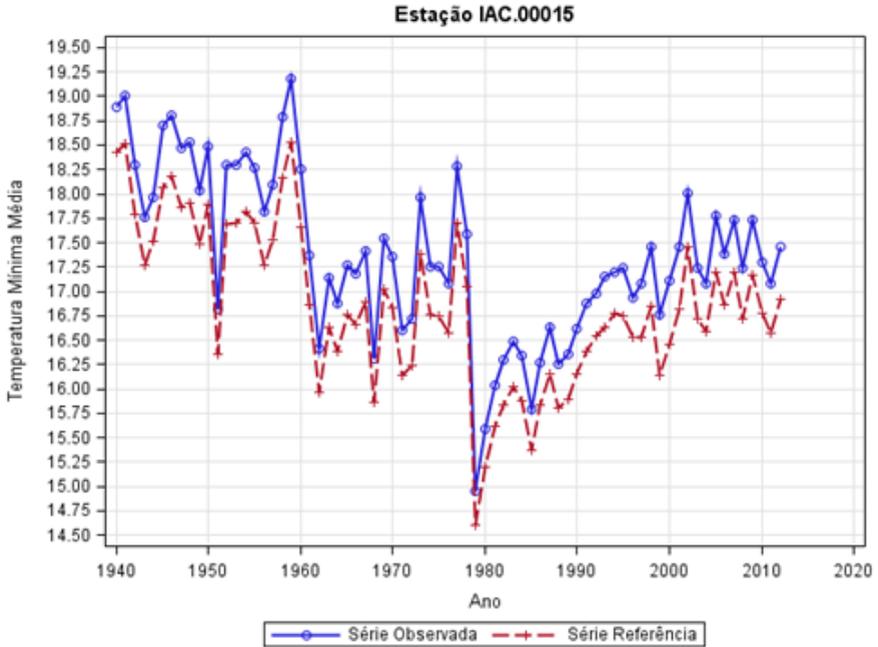
(quebra de continuidade) são combinadas para uma interrupção por ano. O maior número de interrupções na série de temperatura mínima foi em 1957, com 81 ocorrências. Grandes números de interrupções na série também podem ser encontrados em 1948 (59), 1961 (47), 1979 (42), 1992 e 2004 com 39 ocorrências. A explicação mais provável para este tipo de alteração não é climático e sim a localização das estações meteorológicas e também a possíveis mudanças em sua instrumentação.

Para a estação em estudo, verifica-se que a primeira interrupção na série observada se deu em setembro de 1996 (Figura 3) originando a primeira série de referência. Automaticamente o programa calculou as demais interrupções que aconteceram em março de 2003, abril de 1944, maio de 1948, janeiro de 1990 e julho de 2001, gerando as séries de referência 2, 3, 4, 5 e 6 respectivamente. O trabalho de Nery e Carfan, 2014 estuda detalhadamente a evolução temporal de anomalias.



**Figura 3.** Número de anomalias padronizadas (observada menos estimada) para a estação IAC.0015 no período de 1940 até 2012.

Após o cálculo da série de referência final para cada estação meteorológica, a homogeneidade da série observada pôde ser testada mediante a comparação com a série estimada, pelas diferenças entre ambas as séries que no caso ideal deveria ser semelhante a uma série aleatória de distribuição normal conforme Figura 4.



**Figura 4.** Série observada e série de referência para a estação IAC.00015.

O gráfico foi construído pelo procedimento SGPLOT (SAS INSTITUTE, 2008). A visualização da figura 4 nos revela um significativo decréscimo na temperatura mínima média em 1978. Podemos concluir que esta variação foi causada por fatores climáticos, visto que a mesma variação foi verificada na série original e na série de referência. A série original apresentou uma variação de temperatura mínima média de 4,23°C com um desvio padrão de 0,86°C. Para a série de referência, a variação de temperatura mínima média foi de 3,92°C com um desvio padrão de 0,81°C.

## Conclusões

1. Os resultados das análises permitiram concluir que todas as séries meteorológicas estudadas (122) apresentaram, pelo menos, um ponto de descontinuidade.
2. Para a estação usada como estudo de caso (IAC.00015), este ponto foi em 1996.
3. Melhoria da homogeneidade dos dados de temperatura mínima para a estação meteorológica IAC.00015 com a redução de desvios-padrão e melhora na consistência temporal das séries climatológicas.
4. O uso do pacote Climatol melhorou espacialmente e temporalmente a série original, conseqüentemente, melhorando os resultados para a análise de variabilidade climatológica.

## Referências

ALEXANDERSSON, H. A homogeneity test applied to precipitation data. **International Journal Climatology**, 6, n. 6, p. 661-675, 1986. DOI: 10.1002/joc.3370060607.

CONRAD, V.; POLLAK, C. **Methods in climatology**. Cambridge: Harvard University, 1950, 459 p. ill.

FREITAS, L.; PEREIRA, M. G.; CAMELO, L.; MENDES, M.; NUNES, L. F. Homogeneity of monthly air temperature in Portugal with Homes and Mash. **Quarterly Journal of the Hungarian Meteorological Service**, v. 117, n. 1, p. 69-90, Jan./Mar. 2013.

GUIJARRO PASTOR, J. A. Climatol: software libre para la depuración Y homogeneización de datos climatológicos. In: GARCIA CODRON, J. C.; DIEGO LIAÑO, C.; FDEZ DE ARRÓYABE HERNÁNDEZ, P. GARMENDIA PEDRAJA, C.; RASILLA ÁLVAREZ, D. (Ed.). **El clima entre el mar y la montaña**. Asociación Española de climatología y Universidad de Cantabria, 2004. 40 p. (Serie A, n. 4).

GUIJARRO, J. A. **User's guide to climatol**. 2014. 40 p. Disponível em: <<http://www.climatol.eu/climatol-guide.pdf>>. Acesso em: 20 jan. 2016.

HEINO, R. **Climate in Finland during the period of meteorological observations**. Helsinki: Finnish Meteorological Institute, 1994. 209 p. (Finnish Meteorological Institute. Contributions, n. 12).

MALVESTIO, L. M.; NERY, J. T. A importância da Amazônia para o regime pluviométrico da região sudeste do Brasil. **Revista Geonorte**, v. 1, n. 5, p. 786-795, 2012.

MAMARA, A.; ARGIROU, A. A.; ANADRANISTAKIS, M. Homogenization of mean monthly temperature time series of Greece. **International Journal Climatology**, v. 33, n. 12, p. 2649-2666, Oct. 2012. DOI: 10.1002/joc.3614.

NERY, J. T.; CARFAN, A. C. Re-analysis of pluvial precipitation in southern Brazil. **Atmosfera**, v. 27, n. 2, p. 103-115, abr. 2014.

PAULHUS, J. L. H.; KOHLER, M. A. interpolation of missing precipitation records. **Monthly Weather Review**, 80, n. 8, p. 129-133, Aug. 1952.

PETERSON, T. C.; EASTERLING, D. R.; KARL, T. R.; GROISMAN, P.; NICHOLLS, N.; PLUMMER, N.; TOROK, S.; AUER I.; BOEHM, R.; GULLETT D.; VINCENT, L.; HEINO, R.; TUOMENVIRTA, H.; MESTRE, O.; SZENTIMREY, T.; SALINGER, J.; FØRLAND, E.; HANSSEN-BAUER, I.; ALEXANDERSSON, H.; JONES, P.; PARKER, D. Homogeneity Adjustments of 'In Situ' Atmospheric Climate Data: **International Journal of Climatology**, 18, p. 1493-1518, Nov. 1998.

SAHIN, S.; CIGIZOGLU, H. K. Homogeneity analysis of Turkish meteorological data set. **Hydrological Processes**, v. 24, n. 8, p. 981-992, Apr. 2010.

SAS INSTITUTE. **SAS/Graph 9.2**: statistical graphics procedures guide. 2008. Disponível em: <<http://support.sas.com/documentation/onlinedoc/graph/index.html>>. Acesso em: 12 abr. 2012.

TUOMENVIRTA, H. Homogeneity testing and adjustment of climatic time Series in Finland, **Geophysica**, v. 38, n. 1-2, p. 15-41, 2002.

VENEMA, V.; MESTRE, O.; AGUILAR, E.; AUER, I.; GUIJARRO, J. A.; DOMONKOS, P.; VERTACNIK, G.; SZENTIMREY, T.; STEPANEK, P.; ZAHRADNICEK, P.; VIARRE, J.; MÜLLER-WESTERMEIER, G.; LAKATOS, M.; WILLIAMS, C. N.; MENNE, M.; LINDAU, R.; RASOL, D.; RUSTEMEIER, E.; KOLOKYTHAS, K.; MARINOVA, T.; ANDRESEN, L.; ACQUAOTTA, F.; FRATIANNI, S.; CHEVAL, S.; KLANCAR, M.; BRUNETTI, M.; GRUBER, C.; DURAN, M. P.; LIKSO, T.; ESTEBAN, P.; BRANDSMA, T. Benchmarking homogenization algorithms for monthly data. **Climate of the Past**, v. 8, n. 1, p. 89-115, Jan. 2012. DOI: 10.5194/cp-8-89-2012.



---

*Informática Agropecuária*

MINISTÉRIO DA  
**AGRICULTURA, PECUÁRIA  
E ABASTECIMENTO**



CGPE 12906