



ISSN:1984-2295

Revista Brasileira de Geografia Física

Homepage: www.ufpe.br/rbgfe



Sistema de Controle de Qualidade para Dados Diários de Variáveis Meteorológicas

Fabrcio Daniel dos Santos Silva¹, Rafael Marconi Ramos², Rafaela Lisboa Costa³, Pedro Vieira de Azevedo⁴

¹Meteorologista do Instituto Nacional de Meteorologia e Doutorando em Meteorologia pela Universidade Federal de Campina Grande, PB, Brasil, fabrcio.silva@inmet.gov.br.

²Analista de Sistemas do Instituto Nacional de Meteorologia, Brasília, DF, Brasil, rafael.ramos@inmet.gov.br.

³Meteorologista, Doutoranda em Geoprocessamento e Análise Ambiental, Universidade de Brasília, DF, Brasil, rafaelalisboac@gmail.com.

⁴Unidade Acadêmica de Ciências Atmosféricas, Universidade Federal de Campina Grande, PB, pvieira@dca.ufcg.edu.br.

Artigo recebido em 09/09/2014 e aceito em 27/12/2014

RESUMO

Neste trabalho utilizou-se a técnica *MICE* (do inglês "*Multivariate Imputation by Chained Equations*") para imputação de dados diários de variáveis meteorológicas, para séries temporais provenientes de 96 estações convencionais do INMET, entre período de 1961 e 2012, para em seguida aplicar os testes de um sistema de controle de qualidade desenvolvido para detecção, correção e possível substituição de dados suspeitos. Tanto as técnicas de preenchimento de falhas utilizadas, quanto do sistema de controle de qualidade, evidenciaram ser possível solucionar dois dos maiores problemas encontrados nas séries temporais de dados diários medidos em estações meteorológicas: a geração de valores plausíveis, para cada variável de interesse, a fim de sanar a ausência de observações, como detectar e permitir a devida correção, a valores suspeitos provenientes das observações.

Palavras-chave: dados imputados, controle de qualidade, séries históricas.

Quality Control System for Data Daily Weather Variables.

ABSTRACT

In this work we used the technique *MICE* ("Multivariate imputation by Chained Equations") for allocating daily data of meteorological variables for time series from 96 conventional stations of INMET, between 1961 and 2012, and then to apply the tests for a system of quality control developed for detection, correction and possible replacement of suspect data. Both gap filling technique used, as the system of quality control, showed to be possible to solve two major problems encountered in time series of daily data measured at weather stations: the generation of plausible values for each variable of interest, the order to remedy the lack of observations, how to detect and allow correction due to suspicious values from the observations.

Keywords: imputed data, quality control, time series.

* E-mail para correspondência:
fabrcio.silva@inmet.gov.br (Silva, F. D. S.).

Introdução

O Brasil é um país de dimensões continentais e, conseqüentemente, com alta variabilidade espacial e temporal da distribuição das variáveis meteorológicas observadas (Gregorczyk e Cena, 1967). Este é um dos principais fatores que exigem, para qualquer tipo de estudo de natureza climatológica, séries temporais de qualquer variável com boa qualidade de dados, e pequeno número de falhas, assim como é igualmente importante que os pontos de medição sejam bem distribuídos espacialmente.

Todos os sistemas de monitoramento do clima e conjuntos de dados climáticos exigem a melhoria da qualidade dos dados, continuidade e homogeneidade. Isso vale para praticamente todas as aplicações climatológicas, mas torna-se uma necessidade essencial em termos da capacidade de medir, detectar ou atribuir mudanças climáticas. Para dados diários, onde a acessibilidade ainda é muito restrita, Sistemas de Controle de Qualidade (SCQ), particularmente rigorosos, são essenciais (WMO, 2003; Camargo e Hubbard, 1999). Os dados não devem ser considerados como satisfatórios para arquivamento permanente até que tenham sido submetidos a um nível adequado de controle de qualidade. Um bom SCQ deve combinar dois fatores: pessoas e máquinas. Um software de qualidade de dados deve fornecer uma lista de dados suspeitos, mas as decisões finais sobre a correção ou atualização do arquivo digital devem ser do pessoal competente que estará analisando tais relatórios. No entanto, a massa de dados disponíveis nos Serviços Meteorológicos Nacionais (particularmente no Instituto Nacional de Meteorologia – INMET) só aumenta, por exemplo, com a inserção de dados provenientes de estações meteorológicas automáticas de alta resolução. No entanto, a contrapartida geralmente está associada a restrições de pessoal e na qualidade da verificação automática.

Um SCQ não deve ser uma ferramenta perfeita, que acusa um possível erro e automaticamente o dado é retirado da série. Algumas técnicas de SCQ com forte componente estocástica podem levar a uma alta probabilidade de rejeição de boas observações (Guttman, 1988; Meek, 1994). Um SCQ deve disponibilizar ao tomador de decisão, saídas gráficas facilmente compreensíveis e resumos contendo a lista de dados suspeitos, que são excelentes ferramentas para exames visuais que podem ser de grande ajuda, uma vez que a mente humana é especialmente hábil em identificar padrões espaciais.

Alguns testes de consistência são um importante conjunto de verificações de possíveis erros, pois devem explorar a inter-relação temporal e espacial dos dados climatológicos. Os três principais meios de verificações de consistência são os internos, os temporais, e os espaciais:

- os testes de consistência interna exprimem as relações físicas entre os diferentes elementos climatológicos. Em alguns casos são testes lógicos do tipo: se determinado elemento encontra-se em algum intervalo,

outro deverá estar contido, também, em outro determinado intervalo (Gandin, 1988).

- testes de consistência temporal são baseados na persistência no tempo de elementos climatológicos. Certos limiares de mudanças selecionadas dependem da variável em questão, do período do ano e da região climática a que pertence os elementos da série temporal (Eischeid et al., 1995).

- testes de consistência espacial exploram a suave variação no espaço das variáveis climatológicas. Geralmente este tipo de teste envolve a estimativa de um determinado elemento com base nas observações vizinhas em uma mesma região climática (Hubbard et al., 2005). O limite de diferenças aceitas dependerá do tipo de variável, da região climática e da distancia entre as estações. Logo, a eficácia deste tipo de teste vai depender da disponibilidade de estações vizinhas (You et al., 2007).

Devido aos problemas enfrentados ao trabalhar com dados observados, a maioria dos estudos referentes a tendências climáticas, por exemplo, utiliza séries temporais sem o devido tratamento de erros e falhas, e muitas vezes os estudos recorrem a dados provenientes de análises gradeadas, disponibilizadas por diversas instituições internacionais, podendo-se citar as reanálises (Kalnay, 1993).

Assim, este estudo tem como objetivo avaliar um SCQ para dados diários de variáveis meteorológicas medidas em estações convencionais, após um processo de preenchimento de falhas, e, desta forma, facilitar ao futuro usuário dos dados a resolução de dois grandes problemas: as muitas falhas presentes nas séries de dados e, após o preenchimento das falhas, a detecção eficaz dos dados suspeitos, possibilitando-o a se tornar o tomador da decisão de aceitar, corrigir ou excluir esta informação. Todos os procedimentos que serão descritos a seguir foram elaborados usando a linguagem computacional R (2011).

Material e Métodos

Dados meteorológicos observados

Os dados meteorológicos usados são relativos a variáveis medidas em estações meteorológicas convencionais do Instituto Nacional de Meteorologia (INMET), para o período de 1961 até 2012. Para este estudo, as variáveis analisadas são a Temperatura Máxima (TMax), Temperatura Média (TMed), Temperatura Mínima (TMin), Umidade Relativa (UR), Velocidade Média do vento (VMed), Insolação (Insol) e Precipitação (Prec), a níveis diários.

Preenchimento de dados faltantes

Uma etapa anterior à avaliação do SCQ consistiu em aplicar uma metodologia para preenchimento de falhas em dados diários das variáveis meteorológicas. A grande quantidade de falhas observadas foi preenchida usando a técnica *MICE* (do inglês “*Multivariate Imputation by Chained Equations*”).

Essa técnica é utilizada para preenchimento de dados faltantes em séries históricas, faltas essas cau-

sadas por motivos diversos, como falha no instrumento ou problemas nas estações meteorológicas. Os dados faltantes que se deseja completar são caracterizados por “NA”, e foi usado o numero de imputações múltiplas padrão do pacote *MICE* versão 2.12 (m=5 iterações) do software estatístico livre R, para versões ≥ 2.10 (Van Buuren, 2011). As imputações são geradas de acordo com o método padrão, que é, para dados numéricos, o método PMM (*Predictive Mean Matching*). Como exemplo, a Tabela 1 mostra a série de precipitação original com falhas da estação meteorológica, colocada lado a lado dos quatro pontos de grade mais próximos da localidade fornecidas pelo Climate Prediction Center (CPC), em uma matriz.

Como a falta de dados pode ocorrer em qualquer segmento de uma série temporal, o algoritmo atribui uma coluna de dados incompletos onde gera valores sintéticos “plausíveis” de acordo com as demais colunas, onde cada dado desta coluna quando coincide com a falha a ser tratada, tendo seu próprio conjunto de equações baseadas em indicadores específicos. Desta forma, o método “conserva” os

dados originais no arquivo de saída mesmo levando-os em consideração na geração de dados sintéticos.

As séries temporais das diversas variáveis, com exceção da insolação diária, tiveram as falhas preenchidas separadamente com auxílio de séries oriundas de pontos de grade próximos à estação selecionada, fornecidas pelo CPC para precipitação, e do *National Center for Atmospheric Research* (NCAR) para temperaturas máximas, mínimas e médias, umidade relativa, pressão atmosférica e velocidade média diária do vento. Após estas variáveis terem suas falhas preenchidas, as falhas nos dados de insolação foram preenchidas levando-se em conta as relações desta variável com as demais. A grande vantagem desta técnica é preservar os dados originais observados. Com a finalidade de validação do método foram geradas falhas em dados observados para determinados anos – princípio da validação cruzada. A metodologia apresentou resultados promissores. Foram obtidos altos valores de correlações tanto entre os dados diários observados e imputados, assim como quando estes foram acumulados mensalmente (Costa et al., 2012).

Tabela 1. Ilustração de dados de precipitação a serem imputados, representados por NA, referentes à estação de Remanso (a), e dados com as falhas preenchidas (b). Os dados originais (coluna Orig) estão ao lado das séries gradeadas mais próximas (Gr-01; Gr-02; Gr-03 e Gr-04). As falhas preenchidas aparecem destacadas em vermelho.

(a)								(b)							
Ano	Mês	Dia	Orig	Gr-01	Gr-02	Gr-03	Gr-04	Ano	Mês	Dia	Orig	Gr-01	Gr-02	Gr-03	Gr-04
1995	12	29	8.6	5.7	4.8	27	26.3	1995	12	29	8.6	5.7	4.8	27	26.3
1995	12	30	0	0.1	1.2	0.7	0.9	1995	12	30	0.0	0.1	1.2	0.7	0.9
1995	12	31	12.4	0.5	7	1.9	6.9	1995	12	31	12.4	0.5	7	1.9	6.9
1996	1	1	NA	19.1	10	5.9	4.2	1996	1	1	30.8	19.1	10	5.9	4.2
1996	1	2	NA	2.2	1.6	9.8	9.1	1996	1	2	0.0	2.2	1.6	9.8	9.1
1996	1	3	0	0	0	0	0	1996	1	3	0.0	0	0	0	0

Sistema de controle de qualidade para dados diários

Todos os sistemas de monitoramento do clima e conjuntos de dados climáticos exigem a melhoria da qualidade dos dados, continuidade e homogeneidade. Isso vale para praticamente todas as aplicações climatológicas, mas torna-se uma necessidade essencial em termos da capacidade de medir, detectar ou atribuir mudanças climáticas. Para dados diários, onde a acessibilidade ainda é muito restrita, Sistemas de Controle de Qualidade (SCQ), particularmente rigorosos, são essenciais (WMO, 2003; Camargo e Hubbard, 1999). Os dados não devem ser considerados como satisfatórios para arquivamento permanente até que tenham sido submetidos a um nível adequado de controle de qualidade. Um bom SCQ deve combinar dois fatores: pessoas e máquinas. Um *software* de qualidade de dados deve fornecer uma lista de dados suspeitos, mas as decisões finais sobre a correção ou atualização do arquivo digital devem ser do pessoal competente que estará analisando tais relatórios. No entanto, a massa de dados disponíveis nos Serviços Meteorológicos Nacionais (particularmente no Instituto Nacional de Meteorologia - INMET) só aumenta, por exemplo, com a inserção de

dados provenientes de estações meteorológicas automáticas de alta resolução. No entanto, a contrapartida geralmente está associada a restrições de pessoal e na qualidade da verificação automática.

Um SCQ não deve ser uma ferramenta perfeita, que acusa um possível erro e automaticamente o dado é retirado da série. Algumas técnicas de SCQ com forte componente estocástica podem levar a uma alta probabilidade de rejeição de boas observações (Guttman, 1990; Meek, 1994). Um SCQ deve disponibilizar ao tomador de decisão, saídas gráficas facilmente compreensíveis e resumos contendo a lista de dados suspeitos, que são excelentes ferramentas para exames visuais que podem ser de grande ajuda, uma vez que a mente humana é especialmente hábil em identificar padrões espaciais.

Alguns testes de consistência são um importante conjunto de verificações de possíveis erros, pois devem explorar a inter-relação temporal e espacial dos dados climatológicos. Os três principais meios de verificações de consistência são os internos, os temporais, e os espaciais:

- os testes de consistência interna exprimem as relações físicas entre os diferentes elementos climatológicos.

Em alguns casos são testes lógicos do tipo: se determinado elemento encontra-se em algum intervalo, outro deverá estar contido, também, em outro determinado intervalo (Gandin, 1988).

- testes de consistência temporal são baseados na persistência no tempo de elementos climatológicos. Certos limiares de mudanças selecionadas dependem da variável em questão, do período do ano e da região climática a que pertence os elementos da série temporal (Eischeid et al., 1995).
- testes de consistência espacial exploram a suave variação no espaço das variáveis climatológicas. Geralmente este tipo de teste envolve a estimativa de um determinado elemento com base nas observações vizinhas em uma mesma região climática (Hubbard et al., 2005). O limite de diferenças aceitas dependerá do tipo de variável, da região climática e da distancia entre as estações. Logo, a eficácia deste tipo de teste vai depender da disponibilidade de estações vizinhas (You et al., 2007).

O SCQ usado aplicado neste estudo é baseado em uma série de testes, denominados Grupos de Testes. O fluxograma mostrado na Figura 1 detalha o passo a passo de cada etapa à qual todos os registros de dados são analisados. Todas as colunas com informações são cuidadosamente estudadas, desde a coluna que contém o código de identificação da OMM da estação, a coluna que contém as datas da coleta dos dados, e os respectivos valores das variáveis meteorológicas. Uma etapa importante é a confecção de um arquivo chamado metadados, que contém as informações básicas da estação: o código de identificação da OMM, o nome da estação, a longitude e a latitude em graus, minutos e segundos, assim como em décimos de graus, a altitude, o País e Estado a qual pertence, a data de início e fim de suas operações, a Instituição a qual pertence e o tipo da estação, se convencional ou automática. No pré-processamento do SCQ, estas informações contidas nos metadados são lidas e servem de base para alguns dos testes gerais, mostrados na Figura 1.

Na etapa final de verificação e correção dos dados duvidosos, a fim de minimizar o processo de verificação, desenvolver-se-á uma rotina que identificará os casos de dados duvidosos ocorridos em mais de um teste. Se determinado dado for apontado como suspeito em mais de um dos testes do SCQ, então ele já será considerado como um dado incorreto, e não precisa ser avaliado pelo especialista, indo diretamente para um processo de correção.

Resultados

Para todas as áreas de pesquisa envolvendo dados observados, seria fundamental poder contar com series de observações sem dois grandes problemas, as falhas existentes e os valores suspeitos. Aqui, após demonstração do desempenho da uma técnica de preenchimento de falhas aplicada com sucesso a variáveis meteorológicas, passasse a uma etapa subsequente de igual importância, avaliar um SQC para estes dados diários livres de falhas, e, desta forma, facilitar ao usuário dos dados a detecção eficaz dos dados suspeitos, possibilitando-o a se tornar o tomador

da decisão de aceitar, corrigir ou excluir esta informação, e, finalmente, ter em mãos, além de series completas para determinado período de interesse, series com alta confiabilidade.

O SCQ usado é baseado em uma série de testes, denominados Grupos de Testes, os quais obedecem a um fluxograma, como foi mostrado na Figura 1. Para exemplificar os resultados de cada teste, o SCQ foi aplicado à estação de Remanso do INMET (código OMM 82979), município localizado no Vale do Rio São Francisco, no sertão da Bahia. Esta estação teve os dados diários analisados de 1961 a 2012, onde a porcentagem de falhas preenchidas pelo método citado no item 2.2 foi de aproximadamente 30% do total de dados diários.

Com relação ao grupo de testes inicial, denominados testes gerais, o primeiro teste deste grupo verifica se os dados estão relacionados a um único código de identificação (número da OMM), ou seja, se todos os valores listados no arquivo são referentes à mesma estação. Este teste é muito simples, mas importante para verificar se, por algum erro, a seqüência de dados de uma estação não foi contaminada com dados de outra estação, com outro código de identificação. O resultado deste teste é mostrado em um gráfico de barras mostrando os valores do código de identificação encontrados. O ideal é que o gráfico tenha uma só barra, o que significa que os dados pertencem a uma só estação, com um único código de identificação.

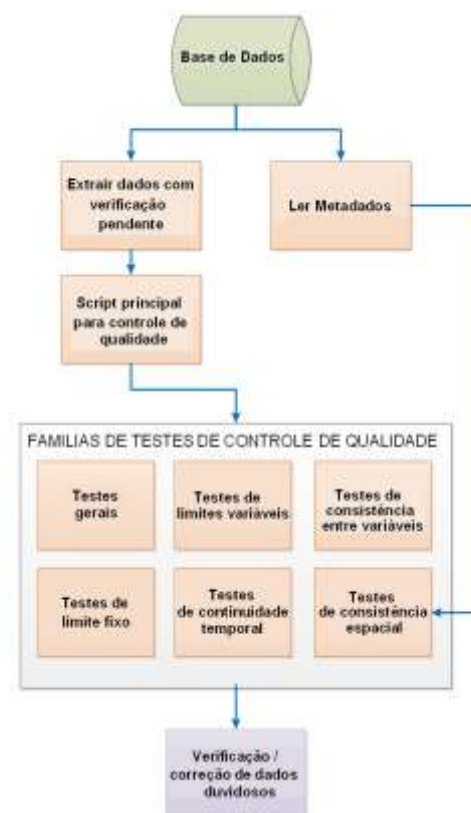


Figura 1: Descrição da rotina de avaliação do SCQ, desde a extração de uma série do banco de dados, a aplicação dos testes, até a verificação ou correção dos dados duvidosos.

O segundo grupo de testes, denominados testes de limites fixos, é utilizado para as temperaturas máximas, mínimas e médias, para a precipitação, a umidade relativa, e a insolação. O teste consiste em estabelecer limites fixos para cada variável, um limite inferior e superior para valores aceitáveis. O resultado deste teste dependerá do rigor imposto por quem está aplicando o SCQ, assim como pelo grau de conhecimento sobre a estação que está sendo trabalhada. No caso da estação de Remanso, a publicação ampliada e revisada das Normais Climatológicas do Brasil para o período 1961-1990 (Ramos et al., 2009), foi usada como referência para o

limite inferior e superior de temperaturas máximas e mínimas, que foram, para Remanso, 10,4°C e 41°C. Logo, para este teste foram escolhidos os limites de 10 e 41°C para verificar se havia nos registros valores inferiores ou superiores a estes para as temperaturas mínimas, máximas e médias. O resultado impresso mostrou não haver problemas de valores além destes limites, em todo o período 1961-2012 da estação. Para as demais variáveis foram detectados os seguintes erros mostrados na Tabela 2, para a pressão atmosférica, com limites fixos entre 955 e 975 hPa, e a insolação, com limites fixos entre 0 e 12 horas.

Tabela 2. Registros de valores suspeitos encontrados pelo SCQ relativo ao Teste de Limites Fixos, para pressão atmosférica (em hPa) e insolação (em horas/dia).

Estação: Remanso (82979)			
Ano	Mes	Dia	Pressao
1996	1	28	953,8
1996	8	7	977,7
1997	7	27	997,7
1999	3	20	1002,1
1999	3	21	982,5
1999	6	4	322,8
1999	6	5	0
1999	6	6	646,1
1999	6	14	975,8
2002	3	4	953,7
Ano	Mes	Dia	Insolação
1998	8	25	25

Os testes de limites variáveis, pertencentes ao terceiro grupo de testes, identificam resíduos “extremos” em relação a um ciclo sazonal ajustado para a variável considerada. Como limites para identificar resíduos extremos se definem dois limiares de percentis (um inferior e um superior), por exemplo, 0,01 (ou seja, o percentil 1%) e 0,99 (ou seja, o percentil 99%). Os percentis extremos podem ser calculados de duas formas, para cada mês, ou com todos os dados. Resíduos menores que o percentil inferior e maiores que o percentil superior são considerados extremos. Tomando como exemplo os dados de temperaturas máximas da estação de Remanso, temos, por exemplo, que para o mês de janeiro, os percentis de 1% e 99% são respectivamente os valores 25,7°C e 36,6°C. Se tomarmos todos os valores da série, sem distinção mensal, os valores dos percentis de 1% e 99% são respectivamente os valores

26,3°C e 37,2°C. Logo, valores que ultrapassem estes limiares serão considerados extremos, e suspeitos, e serão reportados em uma saída impressa para que estes valores sejam analisados, como exemplifica a Tabela 3.

O resultado deste teste também é fornecido de forma gráfica, em um gráfico do tipo box-plot. A Figura 2 mostra estes resultados para as variáveis: umidade relativa, temperatura máxima, temperatura média e temperatura mínima, onde os pontos em vermelho fazem referência aos valores que ultrapasaram estes limites variáveis referentes aos percentis de 1% e 99%. Para a precipitação, estima-se que para cada mês os percentis de 95%, 97,5% e 99%, serão usados como limites para identificar potenciais valores extremos de chuvas diárias. Estes percentis são calculados usando os parâmetros da distribuição gama ajustados para cada mês do ano.

Tabela 3. Registros de valores suspeitos encontrados pelo SCQ relativo ao teste de limites variáveis, para a temperatura máxima, destacados em vermelho (°C), para um trecho da série no mês de abril de 1961. Para abril, os valores dos percentis de 1% e 99% são respectivamente 26,6°C e 36,4°C.

Estação: Remanso(82979)			
Ano	Mes	Dia	Tmax (°C)
1961	4	1	34,4
1961	4	2	30,6
1961	4	3	36,4
1961	4	4	37,0
1961	4	5	34,8
1961	4	6	37,2
1961	4	7	36,6
1961	4	8	36,0
1961	4	9	35,8
1961	4	10	33,6

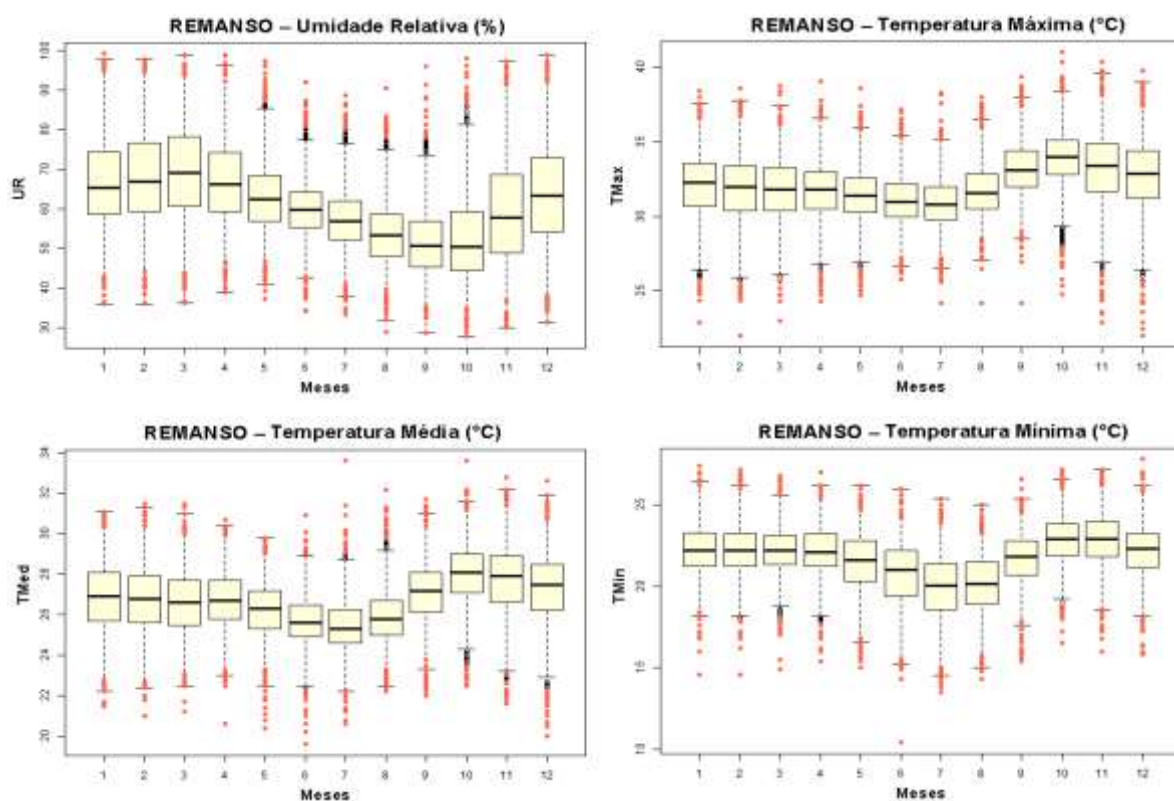


Figura 2: Resultado do teste de limites variáveis para os percentis de 1% e 99%, para as variáveis: umidade relativa, temperatura máxima, média e mínima. Os pontos em vermelho indicam os valores que extrapolaram estes limiares.

O quarto grupo de testes, chamados testes de continuidade temporal, investiga dois problemas muito comuns em séries de dados diários de estações meteorológicas convencionais: as seqüências de valores repetidos e os *shifts*, ou saltos, que ocorrem nos dados. A função para detectar seqüências de valores repetidos é determinada para encontrar seqüências a partir de um limite definido, por exemplo, a partir de 3 dias. Para a detecção de saltos ou discontinuidades extremas entre valores de uma variável para dias consecutivos, monta-se uma série com todas as diferenças diárias: a diferença do dia 02 de janeiro de

1961 com o dia 01 de janeiro de 1961, em seguida a diferença do dia 03 de janeiro de 1961 com o dia 02 de janeiro de 1961, e assim por diante. Com estas séries de diferenças, calculam-se os percentis (por exemplo, 95%, 99% e 99,5%) dos valores absolutos das diferenças, que serão usados como limites para se definir saltos extremos. Se o valor absoluto (ou seja, uma diferença positiva ou negativa) é maior que o percentil usado como limite, este salto é identificado como extremo. Para este estudo foi usado o percentil de 99,5% como limite.

A saída deste teste é impresso e gráfico, contendo o dia, a variável, e o valor suspeito. A Figura 3 mostra estes resultados para as variáveis: umidade relativa, temperatura máxima, temperatura média e temperatura mínima, onde são mostradas duas formas de apresentar o mesmo resultado, onde os pontos e/ou

linhas vermelhas fazem referencia aos valores que não respeitaram o limite do percentil de 99,5 das diferenças, e foram caracterizadas como dados suspeitos.

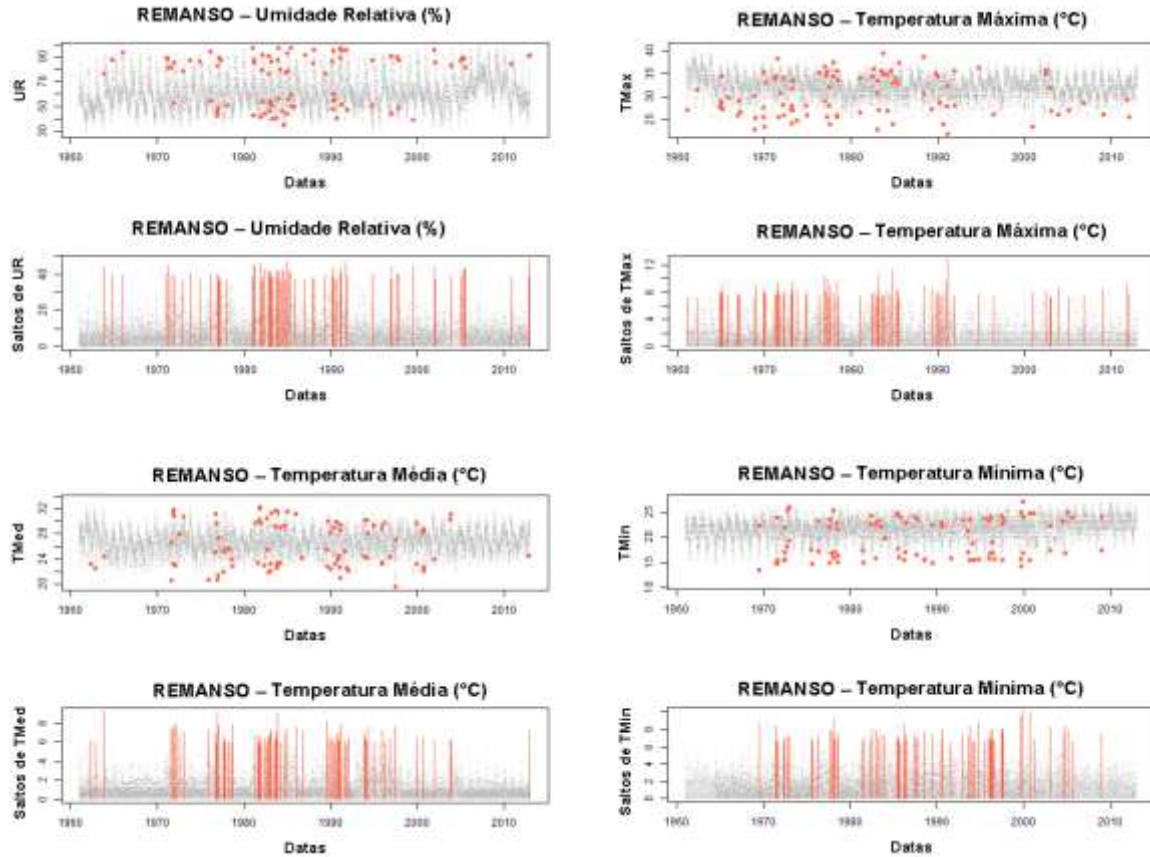


Figura 3: Resultado do teste de limites móveis para os percentis de 1% e 99%, para as variáveis: umidade relativa, temperatura máxima, média e mínima. Os pontos em vermelho indicam os valores que extrapolaram estes limiares.

No quinto grupo de testes, ou testes de consistência entre variáveis, é verificada a consistência entre variáveis relaciona as temperaturas mínimas, médias e máximas. Este teste parte da condição de que as temperaturas mínimas não devem ser maiores que as temperaturas máximas, e também de que as temperaturas médias devem estar entre as temperaturas mínimas e máximas diárias.

A saída destes testes é impressa e gráfica. A Figura 4 mostra a condição ideal esperada, à esquerda a figura mostra que todos os valores de temperaturas máximas foram maiores que os da temperatura média, e a da direita mostram que todos os valores de temperatura média da série foram maiores que os valores de temperaturas mínimas.

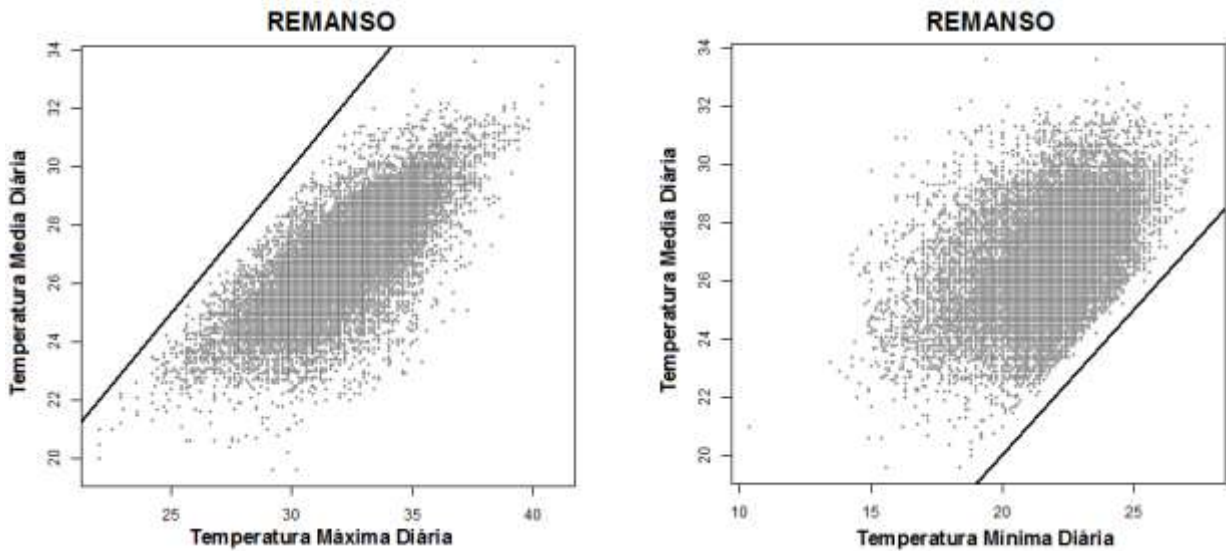


Figura 4: Resultado do primeiro teste de consistência entre variáveis. A figura da esquerda mostra que nenhum valor de temperatura média foi maior que a temperatura máxima diária, e a figura da esquerda mostram que nenhum valor de temperatura média foi menor que a temperatura mínima diária. Caso alguma das condições falhasse devido a dados suspeitos, estes valores seriam identificados como pontos em vermelho, à esquerda ou a direita, da linha diagonal dos gráficos.

A temperatura média diária obtida para uma estação meteorológica e chamada de temperatura média compensada, pois é estimada pela seguinte equação:

$$T = [T(12h) + 2 \times T(24h) + T_{Max} + T_{Min}] \div 5$$

onde T = temperatura média compensada diária;
 T(12h) = temperatura observada às 12 horas UTC;
 T(24h) = temperatura observada às 24 horas UTC;
 TMax = temperatura máxima diária e TMin = temperatura mínima diária.

Logo, após a verificação do primeiro teste, outra forma de identificar eventuais problemas nos

dados diários de temperaturas médias, é comparar o seu valor diário diretamente com a média diária entre a temperatura máxima e a temperatura mínima. Para a execução deste teste, é montada uma série das diferenças entre a temperatura média diária da estação e as médias obtidas entre as temperaturas máximas e mínimas diárias. A partir desta série de diferenças, é calculado o percentil de 99%, sendo este o limiar de tolerância máxima para a diferença entre a temperatura média diária e a média diária das temperaturas máximas e mínimas. A saída deste teste é impressa e gráfica. A Figura 5 mostra o resultado deste teste.

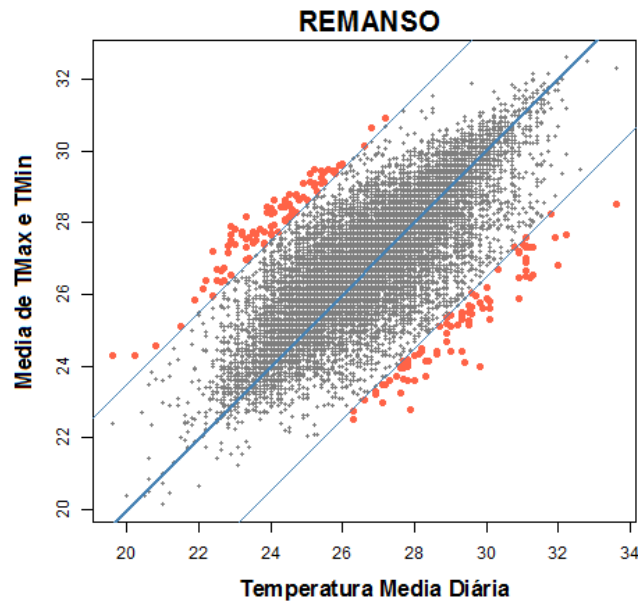


Figura 5: Resultado do teste de consistência para a temperatura média diária. Os pontos em vermelho representam valores onde a diferença entre a temperatura média diária e a média da temperatura máxima e mínima excedeu o percentil de 99%.

Com a preocupação de permitir que o SCQ possa ser utilizado como rotina de trabalho para um constante acompanhamento da qualidade dos dados, será formulada uma rotina para minimizar o número de casos a serem avaliados. No caso da estação analisada, verificou-se que o número de dados suspeitos apontados pelos testes do SCQ foi significativo em sua base histórica. Com isso o processo de avaliação tornou-se oneroso. Com a utilização dessa rotina será possível minimizar o número de casos a serem avaliados pelo especialista. Diversos casos foram sinalizados por essa rotina como incorretos, pois foram identificados em mais de um dos testes do SCQ. Os dados apontados deverão passar diretamente para o processo de definição como NA, e posteriormente para a rotina de preenchimento de falhas ao final do processo como descrito em 2.2, cabendo ao especialista verificar a veracidade de um número bem menor de dados considerados suspeitos, por terem sido indicados por apenas um dos testes do SCQ.

Conclusões

Neste estudo foram apresentadas, de forma resumida, as potencialidades de um SCQ desenvolvido para avaliar possíveis erros em dados diários de estações meteorológicas, convencionais e automáticas. Este SCQ está em estágio avançado, faltando apenas à finalização dos testes de consistência espacial. Estes testes serão particularmente úteis para a tomada de decisão em aceitar ou excluir valores extremos em dados diários de precipitação. No entanto, o conjunto de testes apresentados já é suficiente para analisar e dar ao pesquisador o poder de tomar a decisão de aceitar, corrigir, ou eliminar os dados suspeitos, em praticamente sua totalidade. Todas as saídas após a aplicação de cada teste são na forma impressa e gráfica, algumas delas aqui apresentadas quando aplicadas a series de uma estação meteorológica do INMET, o que facilita ainda mais o processo de tomada de decisão.

O potencial desta ferramenta aqui exibida é ainda maior do que os exemplos apresentados, haja vista que resultados de outros testes, para estas e outras variáveis, não foram sequer citados, como exemplo, as aplicações para as variáveis: temperatura do ponto de orvalho, rajada máxima diária do vento e direção do vento. Além disso, existe a possibilidade de melhorar o modelo de automatização do processo de avaliação dos dados suspeitos, onde se poderá trabalhar futuramente com modelos que combinem os resultados dos testes de forma ponderada com pesos específicos por tipo de teste.

Diante das inúmeras pesquisas, principalmente de análises de mudanças climáticas desenvolvidas hoje no Brasil que precisam ter em sua base dados observados, um cuidadoso SCQ deve ser o primeiro procedimento a ser tomado, que consiste em analisar rigorosamente a qualidade dos dados que serão trabalhados, a fim de gerar maior confiança nos resultados obtidos das pesquisas.

Agradecimentos

O primeiro autor agradece a Universidade de Miami (Rosenstiel School of Marine and Atmospheric Science, Miami, Florida-USA), na pessoa do professor e pesquisador Guillermo Podestá, e a equipe de meteorologistas do Serviço Meteorológico Nacional da Argentina (SMN), pelo compartilhamento dos programas para testes. Agradece ainda o apoio do Inter-American Institute, (IAI) projeto CRN303, financiado pelo US National Science Foundation (Grant GEO-1128040.)

Referências

- Camargo, M. B. P., Hubbard, K. G. 1999. Spatial and temporal variability of daily weather variables in sub-humid and semi-arid areas of the U.S. High Plains. *Agricultural and Forest Meteorology.*, 93, 141-148.
- Costa, R. L., Silva, F. D. S., Sarmento, G., Lúcio, P. S. 2012. Imputação Multivariada de Dados Diários de Precipitação e Análise de Índices de Extremos Climáticos. *Revista Brasileira de Geografia Física* 03, 661-675.
- Eischeid, J. K., Baker, C. B., Karl, T., Diaz, H. F. 1995. The quality control of long-term climatological data using objective data analysis. *Journal of Applied Meteorology*, 34, 2787-2795.
- Gandin, L. S. 1988. Complex quality control of meteorological observations. *Monthly Weather Review*, 116, 1137-1156.
- Gregorczyk, M., Cena, K. 1967. Distribution of effective temperature over the surface of the Earth. *International Journal of Biometeorology.*, 11 (2): 145-149.
- Guttman, N. V., Quayle, R. G. 1990. A review of cooperative temperature data validation. *Journal of Atmospheric and Oceanic Technology.*, 7, 334-339.
- Hubbard, K. G., Goddard, S., Sorensen, W. D., Wells, N., Osugi, T. T. 2005. Performance of Quality Assurance Procedure for an Applied Climate Information System., *Journal of Atmospheric and Oceanic Technology.*, 22, 105-112.
- Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, M. Chelliah, W. Ebisuzaki, J. Janowiak, K. C. Mo, C. Ropelewski, J. Wang, A. Leetmaa, R. Reynolds, E. Kung and D. Salstein, 1993: The NCEP/NCAR CDAS/Reanalysis Project. NMC Office Note 401. Available from NOAA/NCEP Development Division, Washington DC 20233.
- Meek, D. W., Hatfield, J. L. 1994. Data quality checking for single station meteorological databases. *Agricultural and Forest Meteorology.*, 69, 85-109.

- Ramos, A. M., Santos, L. A. R., Fortes, L. T. 2009. Normais Climatológicas do Brasil 1961-1990. Brasília, DF, INMET, 465p.
- R Development Core Team, 2011. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN3-900051-07-0, URL: <http://www.R-project.org/>.
- Van Buuren, S., Groothuis-Oudshoorn, K. 2011. MICE: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software, 45(3), 1-67.
- WMO, 2003. Guidelines on Climate Observation Networks and Systems. WMO Technical Document, Geneva, Switzerland.
- You, J. K., Hubbard, G., Goddard, S. 2007. Comparison of methods for spatially estimating station temperatures in a quality control system. International Journal of Climatology, 28, 777-787, DOI: [joc.1571](https://doi.org/10.1002/joc.1571).