

**UNIVERSIDADE FEDERAL DE SANTA MARIA
CENTRO DE CIÊNCIAS NATURAIS E EXATAS
CURSO DE GRADUAÇÃO EM METEOROLOGIA**

Jonas Barboza Corrêa

**CONTROLE DE QUALIDADE DOS DADOS DA TEMPERATURA
DO AR PARA AS ESTAÇÕES METEOROLÓGICAS AUTOMÁTICAS
DO INMET NA REGIÃO SUL DO BRASIL**

Santa Maria, RS
2018

Jonas Barboza Corrêa

**CONTROLE DE QUALIDADE DOS DADOS DA TEMPERATURA
DO AR PARA AS ESTAÇÕES METEOROLÓGICAS AUTOMÁTICAS
DO INMET NA REGIÃO SUL DO BRASIL**

**Relatório final de
iniciação científica**

ORIENTADOR: Prof. Jônatan Dupont Tatsch

**Santa Maria, RS
2018**

SUMÁRIO

- 1. INTRODUÇÃO**
- 2. METODOLOGIA**
 - 2.1. SELEÇÃO DO PERÍODO DE DADOS
 - 2.2. TESTE DE CONTROLE DE QUALIDADE
 - 2.2.1. Intervalo de Variação (QC1)
 - 2.2.2. Persistência Temporal (QC2)
 - 2.2.3. Consistência Interna (QC3)
 - 2.2.4. Consistência Temporal (QC4)
- 3. RESULTADOS**
- 4. CONCLUSÃO**
- 5. REFERÊNCIAS BIBLIOGRÁFICAS**

1. INTRODUÇÃO

As observações meteorológicas (e relacionadas ambientais e geofísicas) são feitas por uma variedade de razões. Eles são usados para a preparação em tempo real de análises meteorológicas, previsões e advertências meteorológicas severas, para o estudo do clima, para operações locais dependentes do clima (por exemplo, operações locais de vôo de aeródromo, trabalhos de construção em terra e no mar), para hidrologia e meteorologia agrícola, e para pesquisa em meteorologia e climatologia (WMO, 2012).

Nos últimos anos, a automação da estação e o aumento das velocidades de transmissão de dados estão em progresso constante. É necessário um controle de qualidade rápido e efetivo para identificação e sinalização de erros ou observações suspeitas para proporcionar acesso rápido à informação e disseminação de observações confiáveis quanto possível aos usuários. Geralmente, os objetivos para o desenvolvimento de sistemas de controle de qualidade são os seguintes: tornar o controle de qualidade mais eficaz e mais próximo do tempo real; identificar erros de calibração, medição e comunicação tão próximos da fonte de observação quanto possível; focar no desenvolvimento automático de algoritmos de controle de qualidade; desenvolver um sistema abrangente de sinalização para indicar o nível de qualidade dos dados; para facilitar aos usuários de dados identificar dados suspeitos e errados e destacar os valores corrigidos (VEJEN et al., 2002). Entre os principais tipos de erros, encontram-se os erros aleatórios, os erros sistemáticos, erros grandes e os erros micrometeorológicos.

Os erros aleatórios são distribuídos de forma mais ou menos simétrica em torno de zero e não dependem do valor medido. Erros aleatórios, por vezes, resultam em superestimação e às vezes em subestimação do valor real. Em média, os erros se cancelam mutuamente. Os erros sistemáticos, por outro lado, são distribuídos de forma assimétrica em torno de zero. Em média, esses erros tendem a polarizar o valor medido acima ou abaixo do valor real. Uma razão de erros aleatórios é uma deriva a longo prazo de sensores. Erros grandes (ásperos) são causados por mau funcionamento de dispositivos de medição ou por erros cometidos durante o processamento de dados; Os erros são facilmente detectados por cheques. Os erros micrometeorológicos (representatividade) são o resultado de perturbações em pequena escala ou sistemas meteorológicos que afetam a observação do tempo. Estes sistemas não são completamente observáveis pelo sistema de observação devido à resolução temporal ou espacial do sistema de observação. No entanto, quando tal fenômeno ocorre durante uma observação de rotina, os resultados podem parecer estranhos em comparação com as observações circundantes que ocorrem ao mesmo tempo (ZAHUMENSKÝ, 2004).

No Brasil, os dados fornecidos, em sua maioria, encontram-se em sua forma bruta, sem que a qualidade dos mesmos seja verificada. Este trabalho consiste em, através da aplicação de certos métodos de controle de qualidade (QC), verificar e rotular tais dados como suspeitos ou não, para que futuramente possam ser melhores aproveitados em seu uso, com um maior grau de confiabilidade.

2. METODOLOGIA

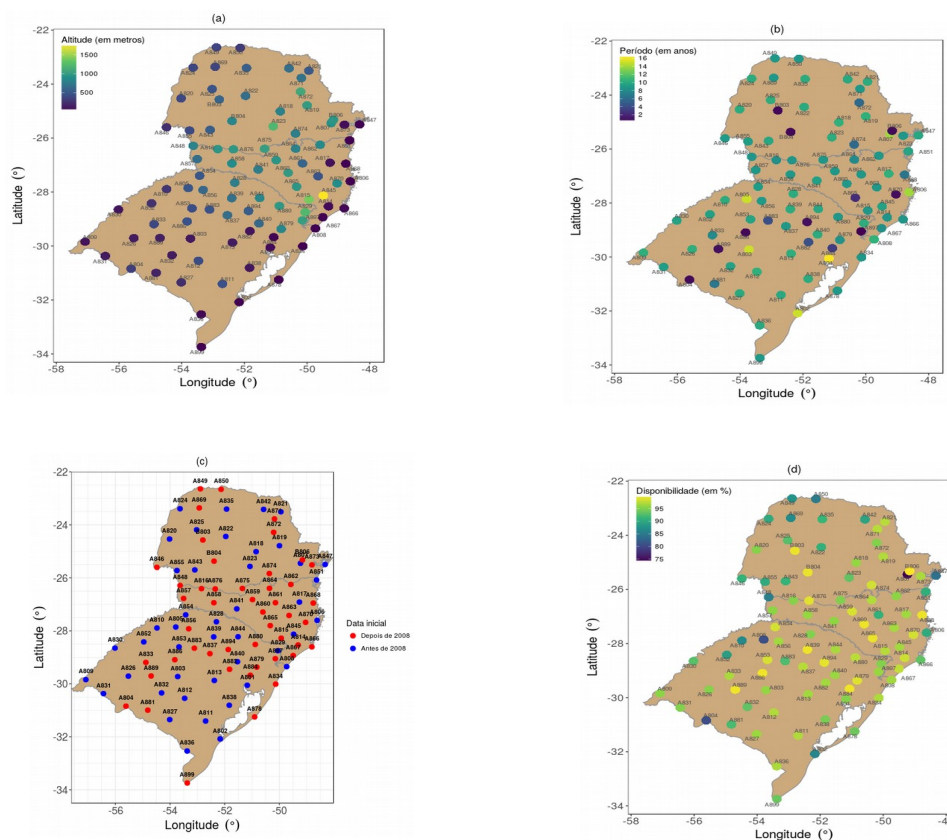
A metodologia adotada consiste em duas partes, inicialmente os dados brutos passarão por uma seleção inicial baseada em certos fatores, em seguida, eles serão submetidos à uma série de testes de controle de qualidade, que os qualificarão como suspeitos ou não, além, de obter-se certas informações relevantes a respeito deles.

2.1 SELEÇÃO DO PERÍODO DE DADOS

Neste trabalho, serão usados dados horários da Temperatura do Ar (Tar) de 91 Estações Meteorológicas Automáticas (EMAs) do Instituto Nacional de Meteorologia (INMET), localizadas na região sul do Brasil. Ao todo, os estados do Paraná (PR), Rio Grande do Sul (RS) e Santa Catarina (SC), possuem, respectivamente, 27, 42 e 22 EMAs em funcionamento. Os dados horários de Tar utilizados estão no padrão do Tempo Universal Coordenado (UTC).

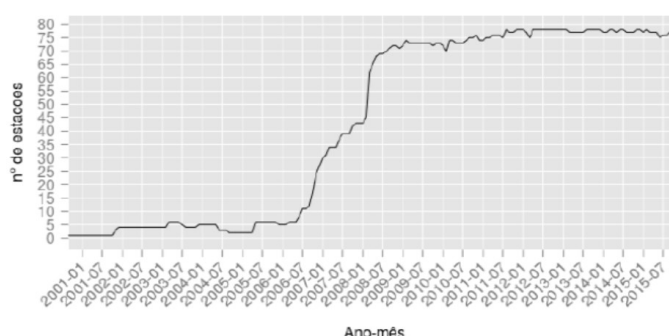
O modelo dos sensores de temperatura usados é o QMH102 da fabricante Vaisala, estes possuem acurácia de $\pm 0,2^{\circ}\text{C}$, e são instalados a uma altura de 2 metros acima do solo. As EMAs devem estar em uma área livre de obstruções (como árvores e prédios), em solo gramado, tendo área mínima de 14mx18m, e estando cercada por tela metálica.

Figura 2.1 – (a) Localização das EMAs do INMET no sul do Brasil, (b) Período de dados das EMAs em anos, (c) EMAs com início de funcionamento pré-pós 2008, (d) Disponibilidade dos dados das EMAs em %.



Dois critérios serão usados para a seleção dos dados que serão utilizados, (i) as estações meteorológicas automáticas devem ter no mínimo quatro anos de dados (podendo serem descontínuos), para haver um tempo mínimo para análise ser consistente, e (ii) o período de análise terá início em 01/01/2008, tendo como data final 31/12/2016. Este ano inicial foi escolhido, pois a partir dele notou-se um bom acréscimo na quantidade de EMAs existente, além de que, a distribuição espacial delas tornou-se mais homogênea.

Figura 2.2 – Evolução temporal mensal do número de EMAs.



Fonte: KANNENBERG, 2016

Para que a seleção fosse feita, antes, foi necessário uma regularização das séries temporais de cada EMA, para assim garantir que todas tenham 24 horas em cada dia, e 365 (ou 366 dias, se ano bissexto) em cada ano.

Após a seleção, restaram ao todo 80 EMAs para análise, estando 24 localizadas no estado do Paraná (PR), 36 no estado do Rio Grande do Sul (RS) e 20 no estado de Santa Catarina (SC). No quadro 2.1 é mostrado as EMAs que apresentaram o maior período de dados.

Quadro 2.1 – EMAs com maior período de dados.

| Código | Estado | Nome | Período (em anos) |
|--------|--------|------------------------|-------------------|
| A801 | RS | Porto Alegre | 16,3 |
| A802 | RS | Rio Grande | 15,1 |
| A803 | RS | Santa Maria | 15,1 |
| A805 | RS | Santo Augusto | 15,1 |
| A806 | SC | Florianópolis-São José | 14 |
| A807 | PR | Curitiba | 13,9 |
| A808 | RS | Torres | 10,6 |
| A812 | RS | Caçapava Do Sul | 10,5 |
| A817 | SC | Indaial | 10,5 |
| A818 | PR | Ivaí | 10,5 |
| A819 | PR | Castro | 10,5 |

Na realização deste trabalho, foi utilizado a linguagem de programação R (Version 1.1.383 – © 2009-2017 RStudio, Inc.) no sistema operacional Linux, com o software livre com ambiente de desenvolvimento integrado, Rstudio (<https://www.rstudio.com>). Os seguintes pacotes foram utilizados: dplyr (WICKHAM, 2017), ggplot2 (WICKHAM, 2016), kableExtra (ZHU, 2017), knitr (XIE, 2017), lubridate (GROLEMUND, 2016), magrittr (BACHE; WICKHAM, 2014), openair (CARSLAW; ROPKINS, 2017), padr (THOEN, 2017), plyr (WICKHAM, 2016), raster (HIJMANS, 2016), scales (WICKHAM, 2017), stringr (WICKHAM, 2017), tidyverse (WICKHAM, 2017).

2.2 TESTES DE CONTROLE DE QUALIDADE

Os testes aplicados aos dados da temperatura do ar, serão divididos em categorias, sendo elas: Limites do Intervalo de Variação, Persistência Temporal, Consistência Interna e Consistência Temporal. Os dados serão considerados suspeitos quando os testes abaixo forem válidos.

2.2.1 Limites do Intervalo de Variação (QC1)

$$(a) \quad T_{avg}(h) < -50^{\circ}\text{C} \quad \text{ou} \quad T_{avg}(h) > 60^{\circ}\text{C} \quad (\text{ESTÉVEZ et al., 2011})$$

Onde:

- $T_{avg}(h)$ é a média aritmética simples de $T_{min}(h)$ (temperatura mínima do ar) e $T_{max}(h)$ (temperatura máxima do ar);
- -50°C e 60°C são, respectivamente, os limites mínimo e máximo instrumental;

$$(b) \quad T_{avg}(h) < T_{low} \quad \text{ou} \quad T_{avg}(h) > T_{high} \quad (\text{ESTÉVEZ et al., 2011})$$

Onde:

- $T_{low}(h)$ e $T_{high}(h)$ são, respectivamente, os valores mínimo e máximo já registrados pela estação climatológica mais próxima de cada EMA;

2.2.2 Persistência Temporal (QC2)

$$(a) \quad T_{avg}(h) = T_{avg}(h-1) = T_{avg}(h-2) = \dots = T_{avg}(h-n) \quad (\text{MEEK; HATFIELD, 1994})$$

Onde:

- n é um número natural, correspondente à(s) hora(s) anteriores à $T_{avg}(h)$;

2.2.3 Consistência Interna (QC3)

$$(a) \quad T_{min}(h) \geq T_{max}(h) \quad (\text{ESTÉVEZ et al., 2011})$$

Onde:

- $T_{min}(h)$ e $T_{max}(h)$ são, respectivamente, os valores mínimo e máximo horário;

$$(b) \quad T_{inst}(h) < T_{min}(h) \quad \text{ou} \quad T_{inst}(h) > T_{max}(h)$$

Onde:

- T_{inst} é a temperatura instantânea do ar, que corresponde à média de um minuto de 12 valores de amostragens medidos a cada cinco segundos

$$(c) \quad T_{avg}(h) < T_{davg,f}(h)$$

Onde:

- $T_{davg,f}(h)$ é a temperatura média do ponto de orvalho filtrada apenas com dados que passaram nos testes anteriores de QC3, aplicados para a variável $T_{davg}(h)$ (temperatura média do ponto de orvalho horária);

$$(d) \quad T_{avg}(d) < T_{min}(d) \quad \text{ou} \quad T_{avg}(d) > T_{max}(d) \quad (\text{MEEK; HATFIELD, 1994})$$

Onde:

- $T_{avg}(d)$ é média aritmética simples das $T_{avg}(h)$;
- $T_{min}(d)$ é o valor menor de temperatura registrada no dia das $T_{min}(h)$;
- $T_{max}(d)$ é o valor maior de temperatura registrada no dia das $T_{max}(h)$;

$$(e) \quad T_{max}(d) < T_{min}(d-1) \quad (\text{ESTÉVEZ et al., 2011})$$

Onde:

- $T_{min}(d-1)$ é a temperatura mínima do dia anterior;

$$(f) \quad T_{min}(d) \geq T_{max}(d-1) \quad (\text{ESTÉVEZ et al., 2011})$$

Onde:

- $T_{max}(d-1)$ é a temperatura máxima do dia anterior;

2.2.4 Consistência Temporal (QC4)

$$(a) \quad (T_{avg}(h) - T_{avg}(h-dt)) > T_{tol}(dt) \quad (\text{WMO, 1993})$$

Onde:

- dt são valores correspondentes a uma determinada hora, sendo eles: $1h$, $2h$, $3h$, $6h$ e $12h$;
- $T_{tol}(dt)$ é um valor tolerável de temperatura, que varia com dt , recebendo os seguintes valores respectivos: 4°C ($1h$), 7°C ($2h$), 9°C ($3h$), 15°C ($6h$) e 25°C ($12h$);

3. RESULTADOS

3.1 Limites do Intervalo de Variação (QC1)

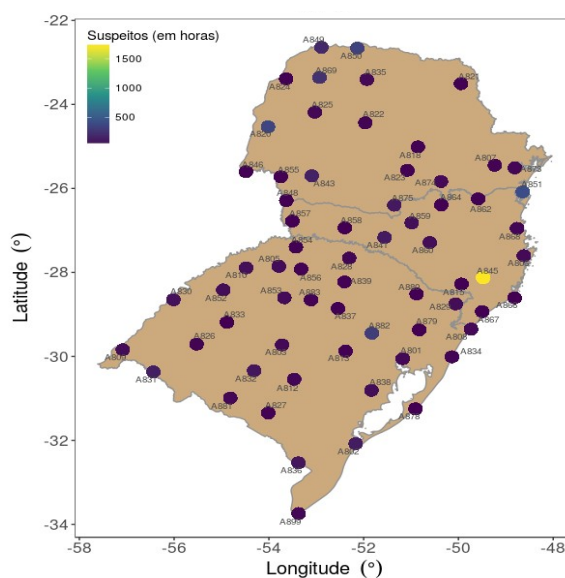
(a) Das 80 EMAs , nenhuma apresentou dados suspeitos.

(b) Das 80 EMAs, 66 apresentaram dados suspeitos.

Quadro 3.1 – EMAs com maior número de dados considerados suspeitos no teste QC1b.

| Código | Nome | Dados Suspeitos (em horas) |
|--------|---------------------------------------|----------------------------|
| A845 | Morro Da Igreja (Bom Jardim Da Serra) | 1726 |
| A851 | Itapoá | 402 |
| A820 | Marechal Cândido Rondon | 354 |
| A850 | Paranapoema | 337 |
| A882 | Teutônia | 306 |
| A869 | Cidade Gaúcha | 229 |
| A849 | Diamante Do Norte | 183 |
| A843 | Dois Vizinhos | 136 |
| A831 | Quaraí | 122 |
| A802 | Rio Grande | 109 |

Figura 3.1 – Distribuição espacial das EMAs com dados considerados suspeitos no teste QC1b.



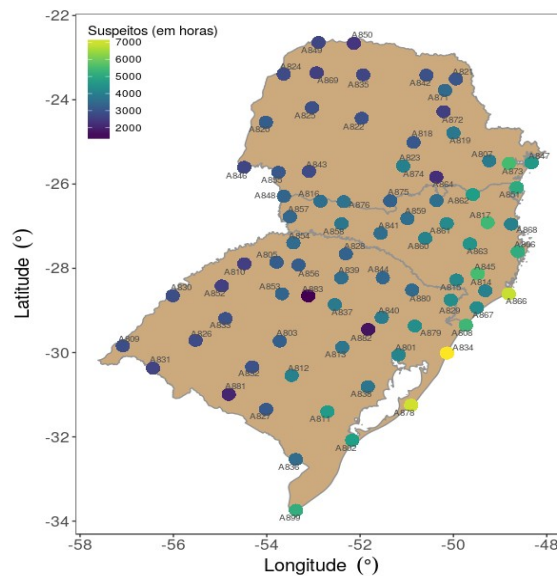
3.2 Persistência Temporal (QC2)

(a) Das 80 EMAs, 80 apresentaram dados suspeitos.

Quadro 3.2 – EMAs com maior número de dados considerados suspeitos no teste QC2a.

| Código | Nome | Dados Suspeitos (em horas) |
|--------|-------------------|----------------------------|
| A831 | Quaraí | 2585 |
| A852 | São Luiz Gonzaga | 2524 |
| A869 | Cidade Gaúcha | 2489 |
| A872 | Ventania | 2486 |
| A810 | Santa Rosa | 2464 |
| A850 | Paranapoema | 2289 |
| A874 | São Mateus Do Sul | 2149 |
| A881 | Dom Pedrito | 1962 |
| A882 | Teutônia | 1629 |
| A883 | Ibirubá | 1360 |

Figura 3.2 – Distribuição espacial das EMAs com dados considerados suspeitos no teste QC2a.



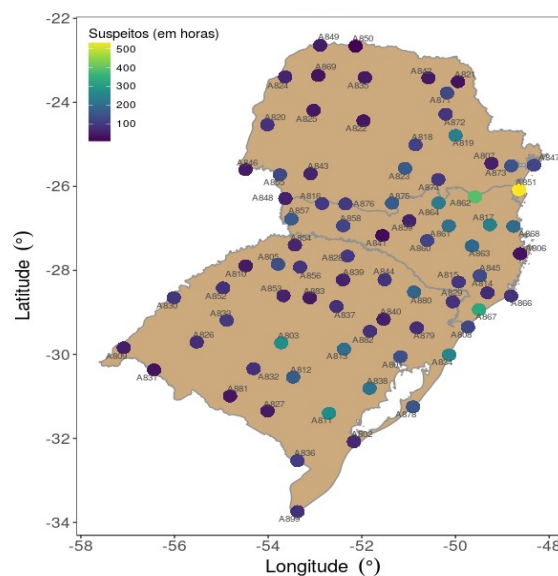
3.3 Consistência Interna (QC3)

(a) Das 80 EMAs, 80 apresentaram dados suspeitos.

Quadro 3.3 – EMAs com maior número de dados considerados suspeitos no teste QC3a.

| Código | Nome | Dados Suspeitos (em horas) |
|--------|------------------------|----------------------------|
| A825 | Goioere | 44 |
| A883 | Ibirubá | 44 |
| A822 | Nova Tebas | 42 |
| A881 | Dom Pedrito | 39 |
| A806 | Florianópolis-São José | 37 |
| A821 | Joaquim Távora | 36 |
| A831 | Quaraí | 36 |
| A869 | Cidade Gaúcha | 36 |
| A841 | Joaçaba3a25 | 25 |
| A850 | Paranapoema | 15 |

Figura 3.3 – Distribuição espacial das EMAs com dados considerados suspeitos no teste QC3a.

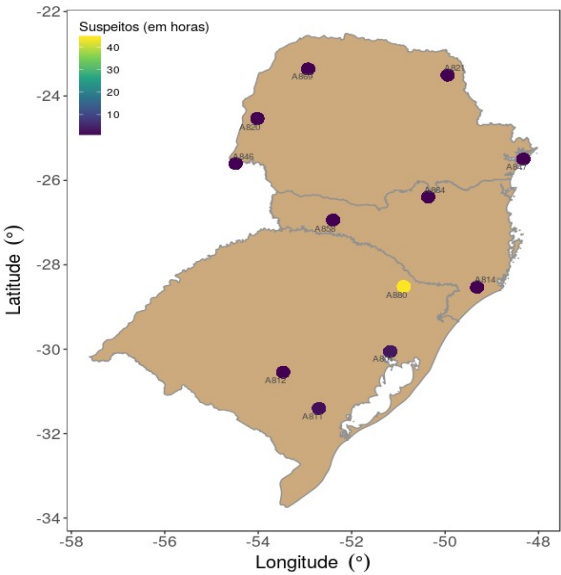


(b) Das 80 EMAs, 12 apresentaram dados suspeitos.

Quadro 3.4 – EMAs com maior número de dados considerados suspeitos no teste QC3b.

| Código | Nome | Dados Suspeitos (em horas) |
|--------|-------------------------|----------------------------|
| A880 | Vacaria | 44 |
| A801 | Porto Alegre | 3 |
| A811 | Canguçu | 2 |
| A812 | Caçapava Do Sul | 1 |
| A814 | Urussanga | 1 |
| A820 | Marechal Cândido Rondon | 1 |
| A821 | Joaquim Távora | 1 |
| A846 | Foz Do Iguaçu | 1 |
| A847 | Ilha Do Mel | 1 |
| A858 | Xanxerê | 1 |

Figura 3.4 – Distribuição espacial das EMAs com dados considerados suspeitos no teste QC3b.

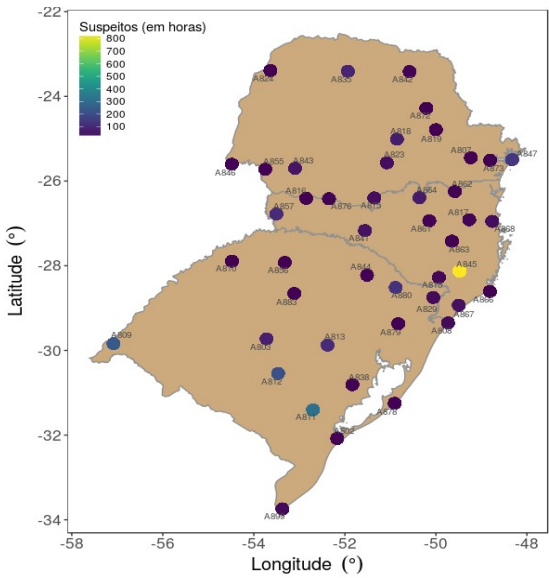


(c) Das 80 EMAs, 45 apresentaram dados suspeitos.

Quadro 3.5 – EMAs com maior número de dados considerados suspeitos no teste QC3c.

| Código | Nome | Dados Suspeitos (em horas) |
|--------|---------------------------------------|----------------------------|
| A845 | Morro Da Igreja (Bom Jardim Da Serra) | 841 |
| A811 | Canguçu | 309 |
| A809 | Uruguaiana | 241 |
| A812 | Caçapava Do Sul | 208 |
| A847 | Ilha Do Mel3c137 | 137 |
| A880 | Vacaria | 122 |
| A813 | Rio Pardo | 104 |
| A857 | São Miguel Do Oeste | 102 |
| A835 | Maringá3 | 87 |
| A803 | Santa Maria | 71 |

Figura 3.5 – Distribuição espacial das EMAs com dados considerados suspeitos no teste QC3c.



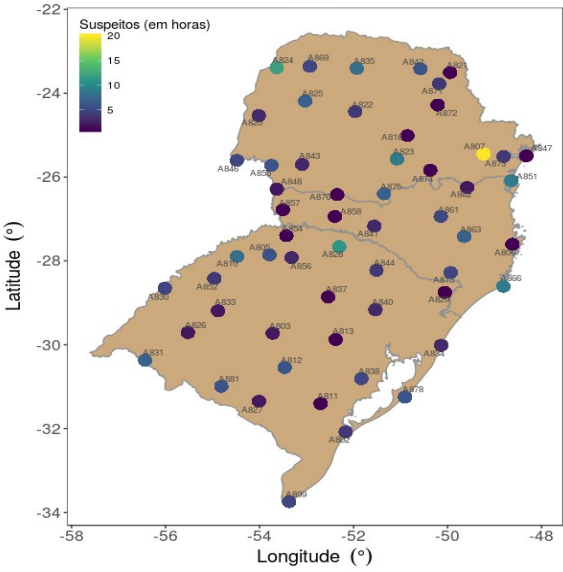
(d) Das 80 EMAs, nenhuma apresentou dados suspeitos.

(e) Das 80 EMAs, 57 apresentaram dados suspeitos.

Quadro 3.7 – EMAs com maior número de dados considerados suspeitos no teste QC3e.

| Código | Nome | Dados Suspeitos (em horas) |
|--------|----------------------------|----------------------------|
| A807 | Curitiba | 20 |
| A824 | Icaraíma | 12 |
| A828 | Erechim | 11 |
| A823 | Inácio Martins | 9 |
| A851 | Itapoá | 9 |
| A866 | Laguna (Farol Santa Marta) | 9 |
| A810 | Santa Rosa | 8 |
| A835 | Maringá | 8 |
| A825 | Goioere | 7 |
| A831 | Quaraí | 7 |

Figura 3.7 – Distribuição espacial das EMAs com dados considerados suspeitos no teste QC3e.

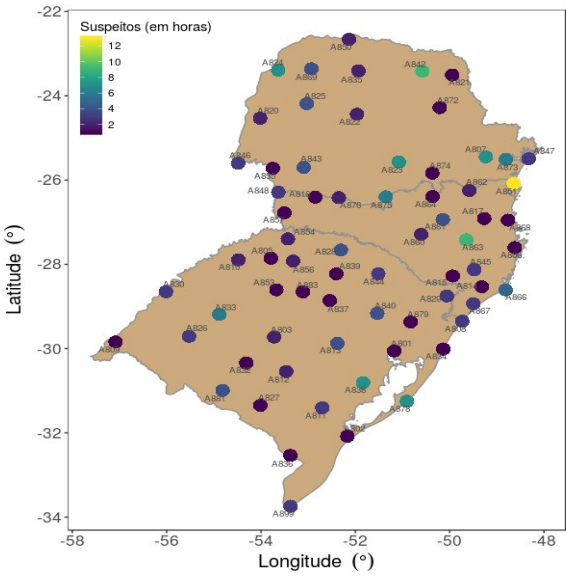


(f) Das 80 EMAs, 69 apresentaram dados suspeitos.

Quadro 3.8 – EMAs com maior número de dados considerados suspeitos no teste QC3f.

| Código | Nome | Dados Suspeitos (em horas) |
|--------|----------------|----------------------------|
| A851 | Itapoá | 13 |
| A842 | Nova Fátima | 9 |
| A863 | Ituporanga | 9 |
| A807 | Curitiba | 7 |
| A823 | Inácio Martins | 7 |
| A824 | Icaraíma | 7 |
| A838 | Camaquã | 7 |
| A878 | Mostardas | 7 |
| A833 | Santiago | 6 |
| A873 | Morretes | 6 |

Figura 3.8 – Distribuição espacial das EMAs com dados considerados suspeitos no teste QC3f.



3.4 Consistência Temporal (QC4)

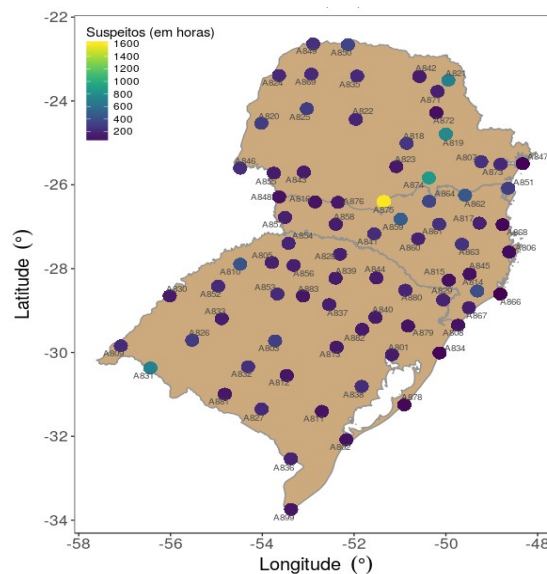
(a) Este controle de qualidade é dividido em partes, de acordo com a variação horária da temperatura.

- Para 1 hora: Das 80 EMAs, 80 apresentaram dados suspeitos.

Quadro 3.9 – EMAs com maior número de dados considerados suspeitos no teste QC4a (1h).

| Código | Nome | Dados Suspeitos (em horas) |
|--------|-------------------|----------------------------|
| A875 | General Carneiro | 1606 |
| A874 | São Mateus Do Sul | 869 |
| A819 | Castro | 772 |
| A831 | Quaraí | 715 |
| A821 | Joaquim Távora | 696 |
| A859 | Caçador | 487 |
| A862 | Rio Negrinho | 445 |
| A810 | Santa Rosa | 420 |
| A814 | Urussanga | 395 |
| A850 | Paranapoema | 340 |

Figura 3.9 – Distribuição espacial das EMAs com dados considerados suspeitos no teste QC4a (1h).

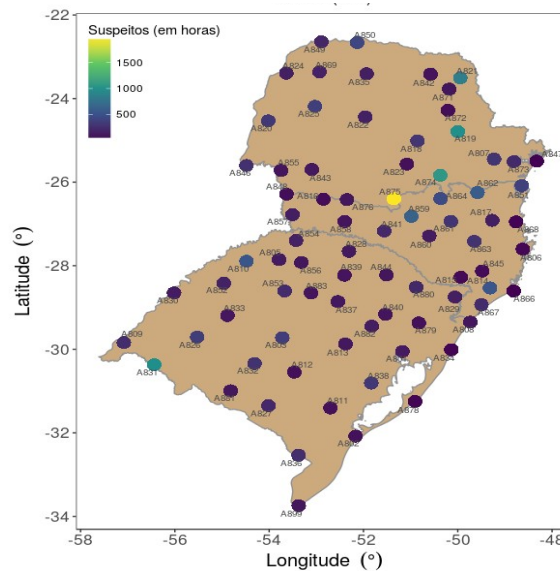


- Para 2 horas: Das 80 EMAs, 80 apresentaram dados suspeitos.

Quadro 3.10 – EMAs com maior número de dados considerados suspeitos no teste QC4a (2h).

| Código | Nome | Dados Suspeitos (em horas) |
|--------|-------------------|----------------------------|
| A875 | General Carneiro | 1970 |
| A874 | São Mateus Do Sul | 1062 |
| A819 | Castro | 990 |
| A831 | Quaraí | 988 |
| A821 | Joaquim Távora | 844 |
| A859 | Caçador | 630 |
| A862 | Rio Negrinho | 587 |
| A814 | Urussanga | 538 |
| A810 | Santa Rosa | 524 |
| A864 | Major Vieira | 447 |

Figura 3.10 – Distribuição espacial das EMAs com dados considerados suspeitos no teste QC4a (2h).

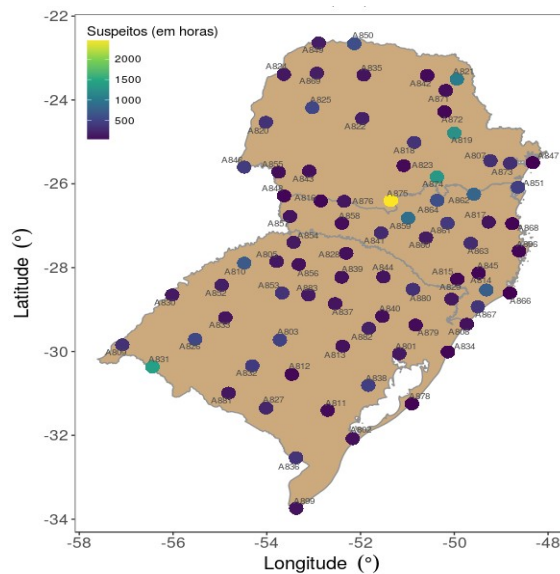


- Para 3 horas: Das 80 EMAs, 80 apresentaram dados suspeitos.

Quadro 3.11 – EMAs com maior número de dados considerados suspeitos no teste QC4a (3h).

| Código | Nome | Dados Suspeitos (em horas) |
|--------|-------------------|----------------------------|
| A875 | General Carneiro | 2476 |
| A831 | Quaraí | 1405 |
| A874 | São Mateus Do Sul | 1320 |
| A819 | Castro | 1250 |
| A821 | Joaquim Távora | 1047 |
| A859 | Caçador | 904 |
| A862 | Rio Negrinho | 807 |
| A814 | Urussanga | 755 |
| A810 | Santa Rosa | 711 |
| A850 | Paranapoema | 604 |

Figura 3.11 – Distribuição espacial das EMAs com dados considerados suspeitos no teste QC4a (3h).

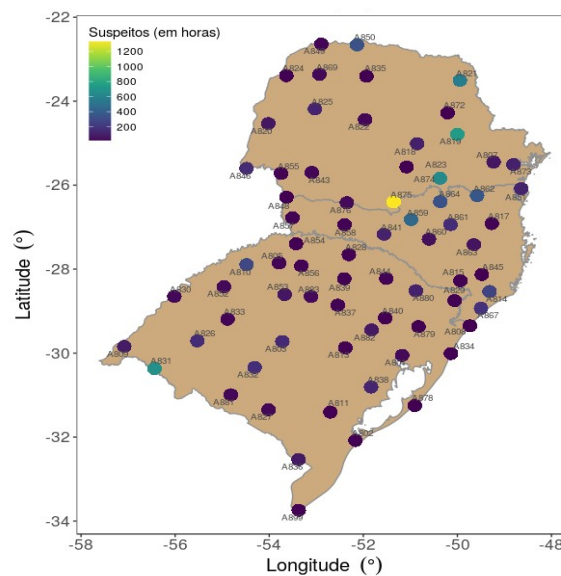


- Para 6 horas: Das 80 EMAs, 72 apresentaram dados suspeitos.

Quadro 3.12 – EMAs com maior número de dados considerados suspeitos no teste QC4a (6h).

| Código | Nome | Dados Suspeitos (em horas) |
|--------|-------------------|----------------------------|
| A875 | General Carneiro | 1336 |
| A819 | Castro | 731 |
| A831 | Quaraí | 668 |
| A874 | São Mateus Do Sul | 638 |
| A821 | Joaquim Távora | 567 |
| A859 | Caçador | 446 |
| A862 | Rio Negrinho | 365 |
| A850 | Paranapoema | 340 |
| A864 | Major Vieira | 340 |
| A810 | Santa Rosa | 288 |

Figura 3.12 – Distribuição espacial das EMAs com dados considerados suspeitos no teste QC4a (6h).



- Para 12 horas: Das 80 EMAs, 1 apresentou dados suspeitos.

A única EMA identificada com dados suspeitos foi a A875, General Carneiro, ela apresentou 1h de dado suspeito.

4. CONCLUSÃO

Somando os resultados dos testes, as estações que apresentaram o maior número de dados suspeitos foram A875 - General Carneiro com 10943 horas; A845 – Morro Da Igreja (Bom Jardim Da Serra) com 8465 horas; A834 – Tramandaí com 7825; A819 – Castro com 7778 horas; A862 – Rio Negrinho com 7382 horas; A851 – Itapoá; A878 – Mostardas com 7289; A866 – Laguna (Farol Santa Marta) com 7055; A873 – Morretes com 6595 horas e A831 – Quaraí com 6526 horas.

No geral, embora a maioria das estações meteorológicas automáticas tenham apresentado ao mínimo um teste onde foram detectados dados suspeitos, a porcentagem de dados suspeitos detectada em cada estação, em relação ao número total de horas de dados na mesma, é suficientemente pequeno.

5. REFERÊNCIAS BIBLIOGRÁFICAS

Carslaw, D. C. and K. Ropkins, (2012) **openair --- an R package for air quality data analysis**. Environmental Modelling & Software. Volume 27-28, 52-61.

Edwin Thoen (2017). **padr: Quickly Get Datetime Data Ready for Analysis**. R package version 0.3.0. <https://CRAN.R-project.org/package=padr>

ESTÉVEZ, J. et al. **Guidelines on validation procedures for meteorological data from automatic weather stations**. Journal of Hydrology, p. 147, 2011.

Garrett Golemund, Hadley Wickham (2011). **Dates and Times Made Easy with lubridate**. Journal of Statistical Software, 40(3), 1-25. URL <http://www.jstatsoft.org/v40/i03/>.

Hadley Wickham, Romain Francois, Lionel Henry and Kirill Müller (2017). **dplyr: A Grammar of Data Manipulation**. R package version 0.7.4. <https://CRAN.R-project.org/package=dplyr>

H. Wickham. **ggplot2: Elegant Graphics for Data Analysis**. Springer-Verlag New York, 2009

Hadley Wickham (2017). **scales: Scale Functions for Visualization**. R package version 0.5.0. <https://CRAN.R-project.org/package=scales>

Hadley Wickham (2017). **stringr: Simple, Consistent Wrappers for Common String Operations**. R package version 1.2.0. <https://CRAN.R-project.org/package=stringr>

Hadley Wickham (2017). **tidyverse: Easily Install and Load 'Tidyverse' Packages**. R package version 1.1.1. <https://CRAN.R-project.org/package=tidyverse>

Hadley Wickham (2011). **The Split-Apply-Combine Strategy for Data Analysis**. Journal of Statistical Software, 40(1), 1-29. URL <http://www.jstatsoft.org/v40/i01/>.

Hao Zhu (2017). **kableExtra: Construct Complex Table with 'kable' and Pipe Syntax**. R package version 0.5.2. <https://CRAN.R-project.org/package=kableExtra>

MEEK, D. W.; HATFIELD, J. L. **Data quality checking for single station meteorological databases**. Agricultural and Forest Meteorology, p. 90-91, 1994.

KANNENBERG, C. **DESAGREGAÇÃO DA TEMPERATURA DO AR DA ESCALA DIÁRIA PARA HORÁRIA PARA O SUL DO BRASIL**, 2016, p. 24. Tese (TCC para obtenção do grau de Bacharel em Meteorologia) – Universidade Federal de Santa Maria, RS, 2016.

Robert J. Hijmans (2016). **raster: Geographic Data Analysis and Modeling**. R package version 2.5-8. <https://CRAN.R-project.org/package=raster>

Stefan Milton Bache and Hadley Wickham (2014). **magrittr: A Forward-Pipe Operator for R**. R package version 1.5. <https://CRAN.R-project.org/package=magrittr>

VEJEN et al. **Quality control of meteorological observations**: Automatic Methods Used in the Nordic Countries. Norwegian Meteorological Institute, p. 11, 2002.

WMO. **GUIDE ON THE GLOBAL DATA-PROCESSING SYSTEM**, p. VI.21, 1993.

WMO. **Guide to Meteorological Instruments and Methods of Observation**, p. 1, 2012.

Yihui Xie (2017). **knitr: A General-Purpose Package for Dynamic Report Generation in R**. R package version 1.17.

ZAHUMENSKÝ, I. **Guidelines on Quality Control Procedures for Data from Automatic Weather Stations**. World Meteorological Organization, p. 3-4, 2004.