

## **XXVI SIMPÓSIO BRASILEIRO DE RECURSOS HIDRÍCOS**

### **Metodologia de Controle e Garantia de Qualidade de Dados Meteorológicos Medidos Observados com Foco no Cálculo da Evapotranspiração de Referência**

*Bruno Comini de Andrade<sup>1</sup>; Pedro Miranda<sup>1</sup>; Vinicius Siqueira<sup>1</sup>; Fernando Fan<sup>1</sup>;*

*Rodrigo Paiva<sup>1</sup>; Anderson Ruhoff<sup>1</sup> & Morris Warren<sup>2</sup>*

**Abstract:** Meteorological data are observed at weather stations, which capture atmospheric behavior near the Earth's surface. This data is extremely useful for climate monitoring and is used as input in a multitude of hydrometeorological models. However, both conventional and automatic stations suffer from recording errors, which must be identified and addressed to ensure confidence in their application. The objective of this work is to develop an automated procedure for quality assurance and quality control (QAQC) of meteorological data, focusing on the input variables for reference evapotranspiration ( $ET_r$ ). The procedure includes a series of steps to check for suspicious data in both hourly and daily datasets, generating a post-processed database that includes  $ET_r$  data. The checks involve filters applied to individual weather stations, identifying errors like duplicate data, extreme values, repeating value streaks, temporal inconsistencies, and statistical outliers. Filters are also applied considering the regional context, such as regression and corroboration of anomalies across neighboring stations. The procedure was applied to INMET's automatic monitoring network, consisting of 616 stations with data available from 2000 onward, totaling over 300 million available hourly data points. The automated procedure yielded satisfactory results, identifying errors in about 3% of INMET's data. However, a final stage of visual data analysis is still necessary.

**Resumo:** Dados meteorológicos são observados em estações meteorológicas, que capturam o comportamento da atmosfera próximo à superfície terrestre. Estes dados são extremamente úteis para o monitoramento climático e são utilizados como entrada em uma infinidade de modelos hidrometeorológicos. Entretanto, tanto estações convencionais quanto automáticas sofrem com erros nos registros, que devem ser identificados e tratados para que haja confiança em sua aplicação. O objetivo deste trabalho consiste na concepção de um procedimento automático de controle e garantia de qualidade (CGQ) de dados meteorológicos, com foco nas variáveis de entrada da evapotranspiração de referência ( $ET_r$ ). O procedimento inclui uma série de etapas de checagem de dados suspeitos tanto em bases de dados horárias quanto diárias, com a geração de uma base pós-processada incluindo dados de  $ET_r$ . As etapas de checagem incluem filtros aplicados a uma única estação meteorológica, como a identificação de períodos com dados duplicados, valores extremos, séries de valores repetidos, incoerências temporais e outliers estatísticos; e filtros aplicados avaliando-se o contexto regional, como filtros de regressão e corroboração de anomalias entre estações. O CGQ foi aplicado à rede de monitoramento automática do INMET, composta de 616 estações e com dados disponíveis de 2000 em diante, totalizando mais de 300 milhões de dados horários disponíveis. O procedimento automático apresenta resultados satisfatórios, identificando erros em cerca de 3% dos dados INMET. Entretanto, uma última etapa de análise visual dos dados ainda se faz necessária.

**Palavras-Chave** – meteorologia; clima; processamento de dados

---

1) Universidade Federal do Rio Grande do Sul, Instituto de Pesquisas Hidráulicas, Avenida Bento Gonçalves, 9500, Porto Alegre, RS, 91501-970.

2) Agência Nacional de Águas e Saneamento Básico, Setor Policial Sul, Área 5, Quadra 3, Blocos B, L e M, Brasília, DF, 70610-200

## 1. INTRODUÇÃO

A crescente demanda por informações confiáveis sobre o clima e os processos hidrológicos tem impulsionado o uso intensivo de dados meteorológicos em grade obtidos de modelos de reanálise, como o ERA5, GLDAS e o MERRA2, entre outros (GELARO *et al.*, 2017; HERSBACH *et al.*, 2020; RODELL *et al.*, 2004). Contudo para a construção dessas bases, assim como para a verificação de sua acurácia, são necessários dados observados em estações de monitoramento in situ. Estas estações, sejam convencionais ou automáticas, registram o comportamento da atmosfera próximo à superfície terrestre. Essas informações são amplamente utilizadas em estudos climáticos, diagnósticos ambientais, modelagem hidrológica e previsão do tempo, além de serem cruciais para o manejo racional dos recursos hídricos, especialmente em cenários de variabilidade e mudança climática.

Apesar da ampla utilização, a qualidade dos dados registrados nessas estações pode ser comprometida por uma série de fatores, como falhas técnicas nos sensores, problemas de comunicação, interferências externas e erros humanos (ALLEN, 1996). Esses erros introduzem incertezas nas análises e modelos que utilizam esses dados como entrada, podendo comprometer a confiabilidade dos resultados. Devido a magnitude das séries de dados automáticos, torna-se essencial dispor de mecanismos automáticos e sistemáticos que avaliem a consistência e a plausibilidade dos dados meteorológicos, identificando e corrigindo erros de forma eficiente.

O presente trabalho propõe a concepção de um procedimento automático de controle de qualidade de dados meteorológicos, com especial atenção às variáveis utilizadas no cálculo da evapotranspiração de referência ( $ET_r$ ), como radiação solar, temperatura e umidade do ar, e velocidade do vento. O procedimento contempla uma sequência estruturada de testes de checagem de incoerências, aplicáveis tanto a séries horárias quanto diárias, visando à geração de uma base de dados confiável e pronta para aplicação em estudos que demandem estimativas precisas da  $ET_r$ . Para avaliar a efetividade do procedimento de controle de qualidade, o mesmo foi aplicado a dados automáticos observados em estações meteorológicas do Instituto Nacional de Meteorologia (INMET) e os resultados foram inspecionados visualmente.

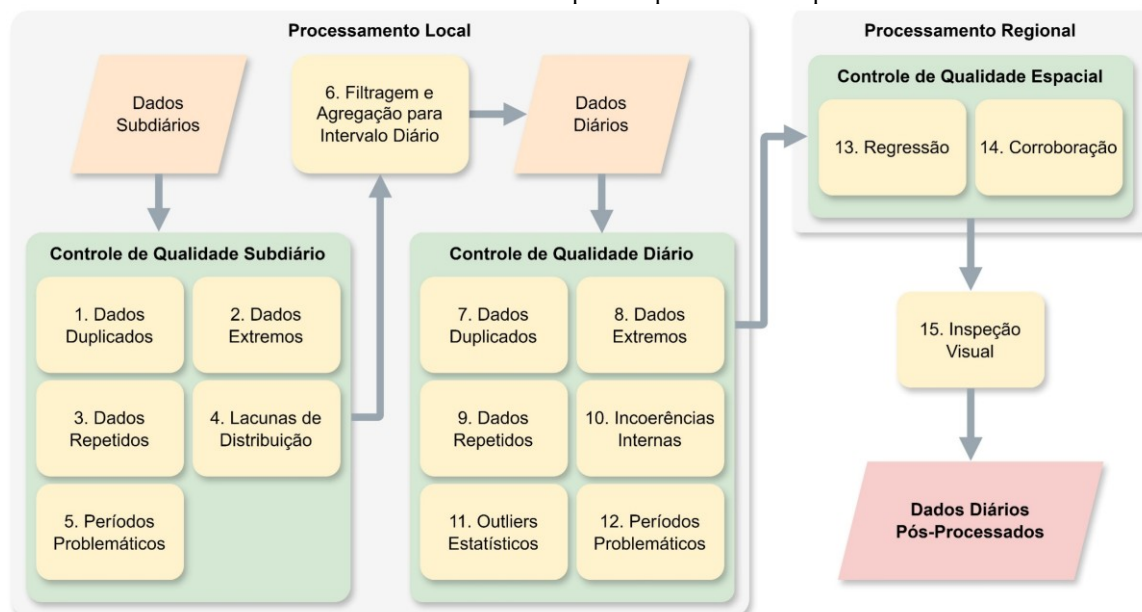
## 2. METODOLOGIA

A Figura 1 apresenta as etapas de identificação de potenciais erros de monitoramento dos dados meteorológicos. O processo de controle de qualidade pode ser aplicado tanto a dados subdiários (intervalos de meia hora, uma hora etc.) quanto a dados diários. As etapas são aplicadas sequencialmente, com a filtragem dos dados antes da aplicação da etapa seguinte. Para bases subdiárias, estão concebidas etapas adicionais (1 a 6) que resultam em uma base diária, enquanto para dados diários a priori, são aplicadas as etapas 7 em diante. A seguir, são detalhadas cada uma das etapas do processo de CC.

### 1. Identificação de dados duplicados

Devido a erros de compilação dos dados, pode haver períodos com séries duplicadas. Por isso é feita uma checagem para identificar meses consecutivos, ou um mesmo mês, em anos consecutivos que apresentem valores idênticos (DURRE *et al.*, 2010). Para cada mês (e. g. abril de 2022) da série de dados de uma estação, são avaliados o mês seguinte (e. g. maio de 2022) e o mesmo mês do ano seguinte (e. g. abril de 2023). Caso dois meses apresentem valores idênticos, os dois são marcados como erros.

Figura 1 – Fluxograma de procedimento de controle de qualidade de dados meteorológicos subdiários e diários para geração de base de dados pós-processada. As etapas 1 a 14 são automáticas, enquanto a etapa 15. Inspeção Visual é manual e deve ser executada por um profissional experiente.



## 2. Identificação de dados extremos

Essa etapa avalia se os dados medidos extrapolam valores considerados implausíveis para as condições em que são medidas as variáveis meteorológicas. Para a radiação solar global ( $R_g$ ), é utilizado um modelo matemático para determinar os limites teóricos para os dados observados. Já para a temperatura do ar ( $T_a$ ), a temperatura do ponto de orvalho ( $T_d$ ) e a velocidade do vento a 2 m de altura da superfície ( $u_2$ ), são usados recordes globais segundo a *World Meteorological Organization* (WMO, 2024). A Tabela 1 apresenta os limiares utilizados para a classificação de dados a serem marcados como potenciais erros.  $R_{so}$  representa a radiação solar incidente sobre a superfície terrestre sob condições de céu limpo (ALLEN *et al.*, 1998) e calculada a partir da Equação (1):

Tabela 1 – Limites de filtragem de valores subdiários extremos.

Variável	Mínimo	Máximo
$R_g$ ( $W\ m^{-2}$ )	0	$1.1\ R_{so}$
$T_a$ ( $^{\circ}C$ )	-90	67
$T_d$ ( $^{\circ}C$ )	-90	67
$u_2$ ( $m\ s^{-1}$ )	0	50

$$R_{so} = (0.75 + 2 \times 10^{-5} z) R_a \quad (1)$$

em que  $R_a$  é a radiação solar no topo da atmosfera, em  $W\ m^{-2}$ . Dados de  $R_g$  fora do intervalo de 0 a 110% de  $R_{so}$  foram marcados como erros. O valor de 0 foi definido para considerar a cobertura de nuvens e a opacidade da atmosfera devido a umidade e presença de sólidos em suspensão, que reduzem a radiação que atinge a superfície. O limite de 110% foi definido para incorporar efeitos de radiação incidente difusa e incertezas no cálculo de  $R_{so}$ .

## 3. Identificação de dados repetidos

Este processo identifica períodos em que um valor se repete por uma duração maior que um limite determinado, conforme apresentado na Tabela 2. O número de repetições foi determinado com base em inspeções visuais, que indicam que períodos de até 12 horas podem apresentar valores

repetidos de  $R_g$  e  $T_a$ , enquanto  $T_d$  e  $u_2$  podem se repetir por períodos maiores em condições de baixa turbulência e alta umidade. Logo, fixou-se o limite de 24 horas para  $T_d$  e  $u_2$ .

Tabela 2 – Limites para a filtragem de valores subdiários repetidos.

Variável	Número de repetições (horas)
$R_g$	12
$T_a$	12
$T_d$	24
$u_2$	24

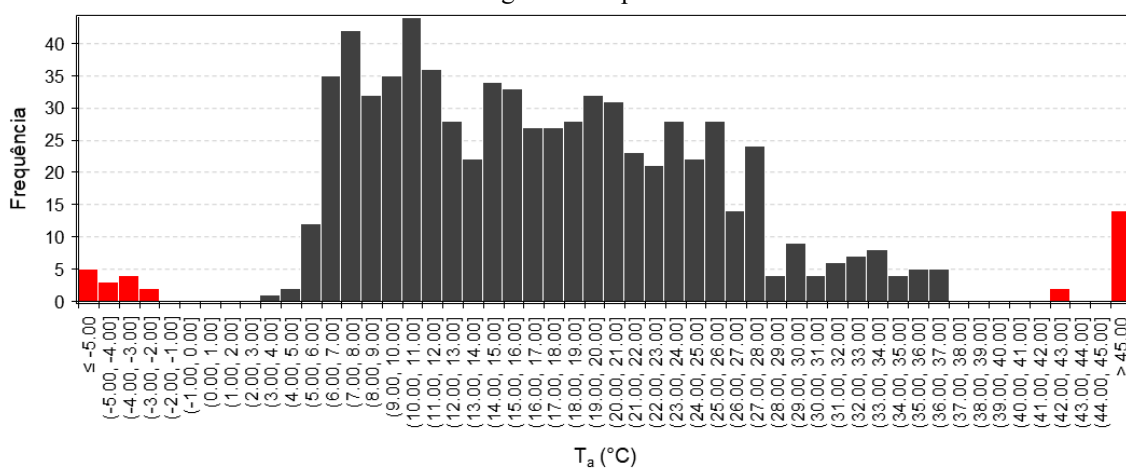
#### 4. Identificação de lacunas de distribuição

Para cada mês da série de dados da estação, os dados são distribuídos em ordem crescente e a diferença entre valores adjacentes é calculada. Diferenças que superam um limiar (Tabela 3) são identificadas. Para variáveis com distribuição assimétrica,  $R_g$  e  $u_2$ , as caudas superiores da distribuição são marcadas caso haja diferenças maiores que o limiar. Já para as outras variáveis ( $T_a$  e  $T_d$ ), ambas as caudas inferior e superior da distribuição são marcadas. Este filtro é capaz identificar grandes anomalias na série temporal que não são identificadas no filtro de valores extremos, mas que são improváveis para o período em que se encontram (DURRE *et al.*, 2010). A Figura 2 apresenta um exemplo para valores de  $T_a$  identificados neste filtro

Tabela 3 – Limites para a filtragem de valores repetidos.

Variável	Diferença máxima
$R_g$	250 W m <sup>-2</sup>
$T_a$	5 °C
$T_d$	5 °C
$u_2$	5 m s <sup>-1</sup>

Figura 2 – Exemplo de histograma de um mês de dados subdiários de  $T_a$ . As colunas em vermelho identificam as caudas com lacunas iguais ou superiores a 5 °C.



#### 5. Marcação de períodos problemáticos

No último filtro para dados subdiários, são identificados períodos que apresentam muitos dados marcados nas etapas anteriores como erros. O objetivo aqui é minimizar o número de possíveis falsos negativos, ou seja, dados que são problemáticos, mas não foram filtrados. Para isso, foram determinadas diferentes janelas móveis para cada filtro, como apresentado na Tabela 4. Para cada dado na série temporal, é extraída uma janela centralizada nesse dado e, caso o número de dados filtrados seja superior a um limiar, todos os dados da janela são marcados.

Tabela 4 – Tamanhos das janelas analisadas e limites respectivos de erros para a identificação de janelas problemáticas em dados subdiários.

Filtro	Tamanho da janela	Número máximo de erros
Dados extremos	24, 72	4, 8
Dados repetidos	24	4
Lacunas na distribuição	24, 72	4, 8
Qualquer filtro acima	24, 72	4, 8

## 6. Filtragem e agregação para intervalo diário

Os dados subdiários são agregados para intervalo diário como uma média de todos os valores de cada dia. Para a  $T_a$ , além da média ( $T_{a, med}$ ), são calculados ainda a  $T_a$  mínima e máxima diária,  $T_{a, min}$  e  $T_{a, max}$ , respectivamente. Primeiramente, são definidos os dados diurnos e os noturnos, a partir de dados de  $R_{so}$ . Instantes em que a  $R_{so}$  é positiva são considerados diurnos, enquanto se  $R_{so}$  é igual ou menor a zero, são noturnos. Falhas de  $R_g$  noturnas são preenchidos com  $R_g = 0$ , assumindo-se que não há radiação durante a noite. Para todas as variáveis, sequências de falhas consecutivas de até 4 horas para o período noturno e de até 2 horas para o período diurno são preenchidas por meio de interpolação linear com os dados adjacentes (VOLK *et al.*, 2023). Dias com falhas consecutivas maiores que os limites mencionados não são agregados e são considerados falhas. Todos os dias restantes são agregados para o intervalo diário.

## 7. Identificação de dados duplicados

Esta etapa segue os mesmos passos da etapa 1, mas aplicada a dados diários, assim como todas as etapas a seguir.

## 8. Identificação de dados extremos

Esta etapa é idêntica à etapa 2, mas aplicada a dados diários.

## 9. Identificação de dados repetidos

Esta etapa é idêntica à etapa 3, mas aplicada a dados diários. Contudo, o número máximo de valores repetidos adotado aqui é de 5 dias para todas as variáveis.

## 10. Identificação de incoerências internas

Essa etapa compara incoerências temporais na série de uma mesma variável e na sua relação com outras variáveis (DURRE *et al.*, 2010). Inicialmente, identifica-se grandes variações entre dois dias consecutivos, marcando os dois se a diferença seja maior que os limiares apresentados na Tabela 5.

Tabela 5 – Diferença máxima aceita entre dias consecutivos.

Variável	Diferença máxima
$R_g$	400 W m <sup>-2</sup>
$T_a$	25 °C
$T_d$	25 °C
$u_2$	50 m s <sup>-1</sup>

Em seguida são marcados os valores em que são verificadas as incoerências apresentadas na Tabela 6. São verificadas incoerências na temperatura do ar que sejam matematicamente impossíveis, como  $T_{a, max}$  menor que  $T_{a, min}$  em um mesmo dia, ou improváveis, como  $T_{a, max}$  de um dia menor que  $T_{a, min}$  do dia adjacente. Também é calculada a umidade relativa (UR) e os dados de temperatura são marcados se UR for maior que 100%, limite matemático para a UR, ou menor que 0.5%, próximo ao



menor registro histórico global (MASTERS, 2017). Esse processo de identificação de incoerências é repetido até que elas não sejam mais encontradas na série temporal da estação.

Tabela 6 – verificações de incoerências entre variáveis meteorológicas observadas na mesma estação. Os valores 0 e 1 entre parêntesis denotam o dia avaliado.

Incoerência	Variáveis marcadas
$T_{a, \max}(0) < T_{a, \min}(0)$	$T_{a, \max}(0)$ e $T_{a, \min}(0)$
$T_{a, \text{med}}(0) < T_{a, \min}(0)$	$T_{a, \text{med}}(0)$ e $T_{a, \min}(0)$
$T_{a, \text{med}}(0) > T_{a, \max}(0)$	$T_{a, \text{med}}(0)$ e $T_{a, \max}(0)$
$T_{a, \max}(0) < T_{a, \min}(1) - 1 \text{ }^{\circ}\text{C}$	$T_{a, \max}(0)$ e $T_{a, \min}(1)$
$T_{a, \min}(0) > T_{a, \max}(1) + 1 \text{ }^{\circ}\text{C}$	$T_{a, \min}(0)$ e $T_{a, \max}(1)$
$UR < 0.5\%$ ou $UR > 100\%$	$T_{a, \min}$ , $T_{a, \text{med}}$ , $T_{a, \max}$ e $T_d$

### 11. Identificação de outliers estatísticos

Esta etapa consiste no cálculo da estatística z-score modificado (CROSBY; IGLEWICZ; HOAGLIN, 1994), descrito a seguir:

$$z_{\text{score}, \text{mod}}(i) = \frac{0.6745}{\text{MAD}(J)} (x(i) - \tilde{x}(J)) \quad (2)$$

em que  $z_{\text{score}, \text{mod}}(i)$  é o z-score modificado no dia  $i$ , adimensional;  $x(i)$  e  $\tilde{x}$  são a observação no dia  $i$  e a mediana de uma variável  $x$  no dia juliano  $J$ ;  $\text{MAD}(J)$  é a mediana dos desvios absolutos (*Mean Average Deviation*, em inglês) da variável  $x$ , calculada para o dia juliano ( $J$ ). Os valores de  $\tilde{x}$  e  $\text{MAD}$  são calculados com base no dia juliano da observação, que corresponde ao dia do ano, entre 1 e 365, ou 366 para anos bissextos. Para cada dia juliano, são agregados todos os dados observados da série em uma janela de 31 dias centralizada em  $J$ , ou seja, para o dia 15 de janeiro ( $J = 15$ ), os dados de 1 a 31 de janeiro de todos os anos da série são considerados para o cálculo de  $\tilde{x}(J)$  e  $\text{MAD}(J)$ . Finalmente, são marcados os dados da série cujo  $z_{\text{score}, \text{mod}}$  seja maior que 6.0 ou menor que -6.0.

Para algumas condições, esse filtro não é aplicado, como para  $T_d$ , que experimenta variações bruscas com grandes anomalias, especialmente durante períodos de transição entre as estações seca e úmida. O filtro também não é aplicado aos limites inferiores de  $R_g$  e  $u_2$ , para as quais o valor de zero é considerado o limite inferior. Por fim, o processo é repetido iterativamente até que outliers não sejam mais identificados.

### 12. Identificação de períodos problemáticos

Esta etapa se assemelha à etapa 5, porém com parâmetros apropriados a séries de intervalo diário, apresentados na Tabela 7.

Tabela 7 – Tamanhos das janelas analisadas e limites respectivos de erros para a identificação de janelas problemáticas em dados diários.

Filtro	Tamanho da janela	Número máximo de erros
Dados repetidos	3	2
Dados extremos	3, 7, 30	2, 3, 10
Incoerências internas	3, 7, 30	2, 3, 10
Outliers estatísticos	3, 7, 30	2, 3, 10
Qualquer filtro acima	3, 7, 30	2, 3, 10
Dados ausentes	7	3

### 13. Avaliação de regressão espacial

A etapa de checagem da regressão espacial (DURRE *et al.*, 2010) permite identificar discrepâncias nos valores medidos por uma estação em relação às suas vizinhas. Primeiramente, são

identificadas as estações vizinhas da estação a ser filtrada, dentro de um raio de 75 km. Para garantir a robustez da análise, apenas estações que possuam pelo menos 3 vizinhas são filtradas. Em seguida é realizado o processo de filtragem, descrito a seguir.

Para cada mês da série temporal, é extraída uma janela temporal estendida por 15 dias antes e depois do mês analisado, totalizando cerca de 60 dias. Apenas meses cujas janelas apresentam pelo menos 40 dias com dados na estação de interesse e nas vizinhas são considerados para a filtragem. Se menos de três estações apresentarem dados suficientes na janela, o mês respectivo não é filtrado. Para aumentar a confiabilidade da regressão, as estações vizinhas são ordenadas de forma decrescente em relação ao índice de concordância ( $d$ ) entre os seus dados e o da estação filtrada, e as sete primeiras são selecionadas. O cálculo de  $d$  é descrito pela equação (3):

$$d = 1 - \frac{\sum_1^n |y_i - x_i|}{\sum_1^n (|y_i - x_i| + |x_i - \bar{x}|)} \quad (3)$$

em que  $d$  é o índice de concordância, adimensional;  $n$  é o número de dados disponíveis na janela temporal;  $y_i$  e  $x_i$  são os valores da estação vizinha e da estação de interesse, respectivamente, para o dia  $i$ ; e  $\bar{x}$  é a média dos valores da estação de interesse na janela temporal.

Selecionadas as estações preditoras, é construída a função a seguir:

$$\widehat{y(i)} = \frac{\sum_{k=1}^m [a(k) x'(i, k) + b(k)] d(k)}{\sum_{k=1}^m d(k)} \quad (4)$$

em que  $\widehat{y(i)}$  é a estimativa no dia  $i$ , para a estação analisada;  $m$  é o número de estações vizinhas selecionadas para a construção da função na janela;  $a(k)$  e  $b(k)$  são o coeficiente angular e o intercepto da regressão linear calculada para a estação vizinha  $k$ ;  $x'(i, k)$  é o valor medido na estação vizinha  $k$ , dentro de uma janela de 3 dias centralizada no dia  $i$ , que mais se aproxima do valor da estação de interesse no dia  $i$ ; e  $d(k)$  é o índice de concordância da estação  $k$  com a estação de interesse. Com o modelo de regressão construído, é calculada a correlação entre ele e os dados da estação de interesse, e apenas meses cuja correlação apresentar valor maior que 0.80 são filtrados. Por fim, se todos os critérios forem cumpridos, os dados do mês que apresentem diferença maior que os limiares apresentados na Tabela 8, ou diferença normalizada maior que 4.0, são marcados como erros.

Tabela 8 – Diferença máxima aceita entre os valores da estação e o valor estimado a partir do modelo de regressão linear múltipla dos dados de estações vizinhas.

Variável	Diferença máxima
$R_g$	80 W m <sup>-2</sup>
$T_a$	8 °C
$T_d$	8 °C
$u_2$	4 m s <sup>-1</sup>

#### 14. Avaliação de corroboração espacial das anomalias

Esta etapa busca avaliar se desvios do comportamento médio observado para a variável medida são observados em estações suficientemente próximas. Novamente, são selecionadas estações vizinhas num raio de 75 km. O filtro é aplicado considerando uma janela de três dias em torno do dia juliano. Para cada dia da série, é computada a janela de três dias centralizada no dia avaliado. Nela, o dado medido na estação é comparado com os dados da janela das vizinhas e o valor mais próximo de cada vizinha é utilizado para o cálculo de comparação das anomalias. Essa avaliação em janela mitiga os efeitos da passagem de frentes atmosféricas, que podem afetar as estações meteorológicas em diferentes tempos. As anomalias são computadas como a diferença da média de longo termo, aplicada

a uma janela temporal 31 dias em torno do dia juliano da data analisada. Para cada janela de três dias, o filtro só é aplicado se, para cada um dos dias, há pelo menos três estações com medições. As sete estações vizinhas mais próximas são selecionadas para a aplicação do filtro. Por fim, são filtrados os valores cuja anomalia difere das anomalias das estações vizinhas além de um limiar, apresentado na Tabela 9. Enquanto para as outras variáveis esse filtro é aplicado para anomalias negativas e positivas, para  $R_g$  e  $u_2$  ele é aplicado somente a anomalias positivas, devido à distribuição assimétrica dessas grandezas.

Tabela 9 - Diferença máxima aceita entre as anomalias da estação e a anomalia mais próxima entre as estações vizinhas.

Variável	Diferença máxima
$R_g$	$100 \text{ W m}^{-2}$
$T_a$	$10 \text{ }^{\circ}\text{C}$
$T_d$	$10 \text{ }^{\circ}\text{C}$
$u_2$	$5 \text{ m s}^{-1}$

Essa abordagem é complementar ao filtro de regressão espacial, pois, por não precisar de boa concordância entre os dados da estação de interesse e as vizinhas, pode ser aplicada a locais e períodos em que a regressão não é aplicável. Por outro lado, quando aplicada, a regressão espacial permite identificar inconsistências espaciais de menor magnitude que no processo de corroboração (DURRE *et al.*, 2010).

### 15. Inspeção visual

A última etapa do processo de controle de qualidade consiste na inspeção visual de das séries temporais de todas as variáveis das estações meteorológicas, para identificação de valores que não foram capturados pelos filtros automáticos.

## 3. BASE DE DADOS

O controle de qualidade foi aplicado à base de dados meteorológicos de estações automáticas disponibilizada pelo Instituto Nacional de Meteorologia (INMET, 2025). Foram consideradas 616 estações distribuídas pelo Brasil, com dados disponíveis de 2000 a 2024.

## 4. EVAPOTRANSPIRAÇÃO DE REFERÊNCIA

A  $ET_r$  foi calculada a partir de dados diários, por meio da equação de Penman-Monteith, aplicada a uma superfície de referência padrão (ALLEN *et al.*, 1998; ASCE-EWRI, 2005), equivalente a uma grama manejada e irrigada.

## 5. RESULTADOS

A Tabela 10 apresenta a contagem de dados filtrados nas 616 estações automáticas do INMET. Como os dados passaram pelo processo completo, a base horária apresenta uma razão maior de dados filtrados do que a base diária. Entre os dados horários,  $R_g$  e  $u_2$  apresentam o maior número de erros, do tipo extremo e repetido, respectivamente. No total, menos de 3% dos dados horários foram filtrados. Quanto à base diária, grande parte dos dados foram filtrados por estarem em períodos problemáticos. Entre os dados de  $ET_r$ , 1,63% foram classificados como errôneos, com mais de 2 milhões de dados (66%) aceitos.

Para exemplificar a etapa de inspeção visual, a Figura 3 apresenta o resultado do CGQ na estação meteorológica A521, em Belo Horizonte, MG. Entre os dados horários, se verifica a presença de erros extremos, nos dados de  $R_g$ , lacunas em  $T_d$ , e repetições em  $u_2$  que, devido a sua alta frequência, acarreta em períodos identificados como problemáticos. Após a agregação para intervalos diários, há uma drástica redução na quantidade de erros detectados automaticamente. Visualmente,

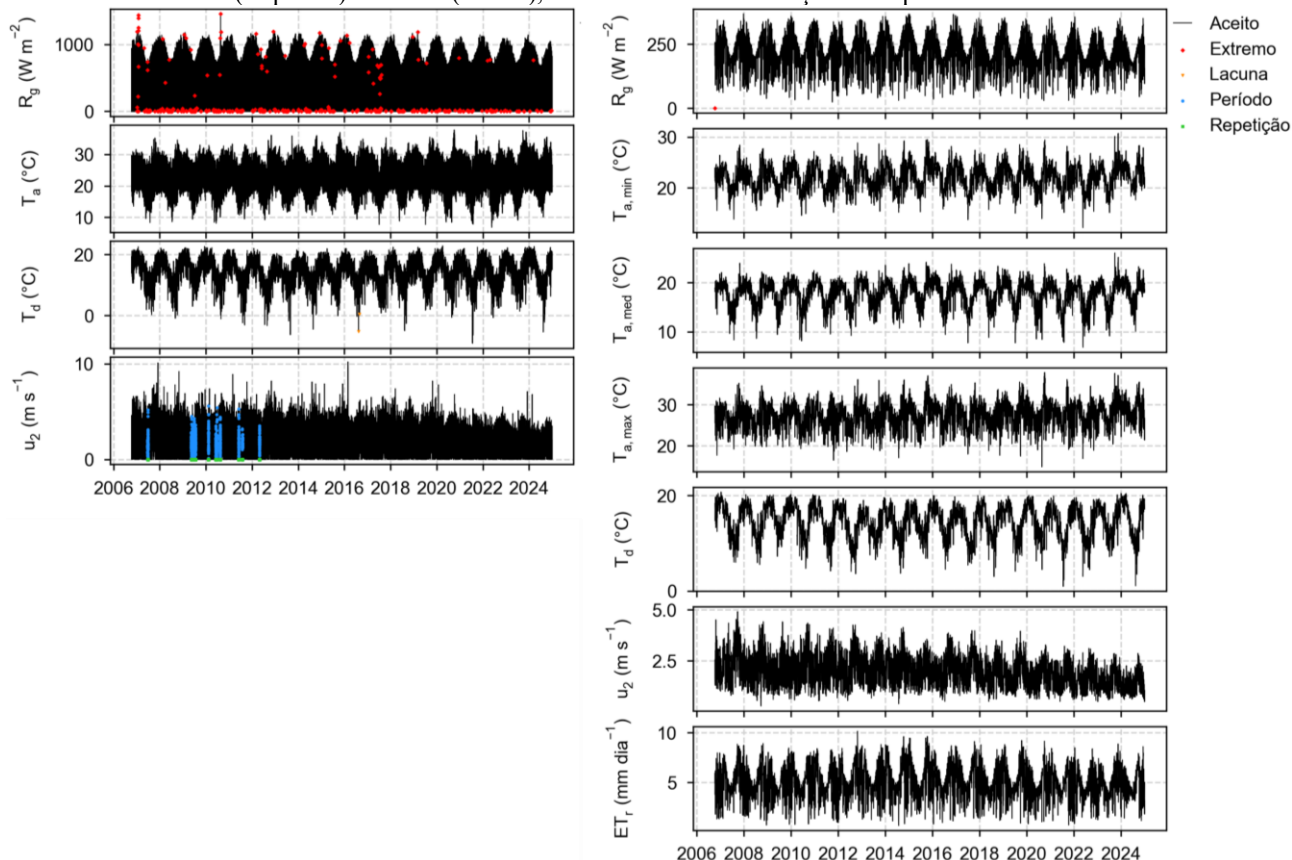


não se observa grandes problemas nas séries temporais. Entretanto, há uma tendência de redução nos dados de  $u_2$ , que pode estar associada a mudanças ambientais no entorno da estação, à degradação do instrumento medidor, ou ainda à variabilidade ou mudança climática de maior escala.

Tabela 10 – Contagem de dados aceitos ou identificados como erros no procedimento de CGQ, tanto para intervalo horário quanto diário. A razão entre os dados marcados como erro e o total de dados também é apresentada.

Base horária											
Variável	Aceitos	Falhas	Duplicados	Extremos	Repetidos	Lacunas	Períodos	Razão			
R <sub>g</sub>	71.333.139	6.072.194	24.456	1.255.423	728	1.056	1.356.844	3,57%			
T <sub>a</sub>	68.551.944	11.491.140	0	0	0	556	200	0,00%			
T <sub>d</sub>	66.807.909	13.231.752	0	0	0	2.960	1.219	0,01%			
u <sub>2</sub>	61.241.030	13.581.264	175.488	0	2.526.523	38	2.519.497	7,86%			
Total	267.934.022	44.376.350	199.944	1.255.423	2.527.251	4.610	3.877.760	2,85%			
Base diária											
Variável	Aceitos	Falhas	Dup.	Ext.	Rep.	Incoer.	Out.	Per.	Reg.	Corrob.	Razão
R <sub>g</sub>	2.684.530	620.716	8.560	8.170	0	0	0	15.991	0	0	1,20%
T <sub>a. min</sub>	2.777.251	551.656	0	0	0	0	0	9.060	0	0	0,33%
T <sub>a. med</sub>	2.777.250	551.656	0	0	0	2	0	9.059	0	0	0,33%
T <sub>a. max</sub>	2.777.251	551.656	0	0	0	0	0	9.060	0	0	0,33%
T <sub>d</sub>	2.674.733	647.000	0	0	0	0	0	16.234	0	0	0,60%
u <sub>2</sub>	2.435.739	880.365	0	139	0	0	0	21.724	0	0	0,89%
ET <sub>r</sub>	2.208.272	1.093.077	151	5.568	0	0	0	30.899	0	0	1,63%
Total	18.335.026	4.896.126	8.711	13.877	0	2	0	112.027	0	0	0,73%

Figura 3 – Exemplo visualização do resultado do procedimento automático de CGQ na estação A521, MG, para dados horários (esquerda) e diários (direita), com a devida identificação dos tipos de erros detectados.



## 6. CONCLUSÃO

A metodologia proposta de controle e garantia de qualidade utiliza procedimentos atuais de reconhecimento de falhas em dados meteorológicos. A aplicação à rede de monitoramento do INMET mostra grande potencial para sua aplicação de forma automática e operacional. Não obstante, a identificação de erros residuais mostra que a inspeção visual não deve ser descartada por enquanto. Essa limitação pode ser melhorada com o refinamento dos parâmetros de filtragem e com a incorporação de técnicas de decomposição de séries temporais.

## AGRADECIMENTOS

Os autores gostariam de agradecer o apoio da Agência Nacional de Águas e Saneamento Básico (ANA) para o desenvolvimento do projeto Desenvolvimento de Tecnologias Inovadoras Baseadas em Modelagem Hidrológica e Sensoriamento Remoto para Monitoramento da Agricultura Irrigada no Brasil, através de um Termo de Execução Descentralizada (TED) firmado entre o Instituto de Pesquisas Hidráulicas da Universidade Federal do Rio Grande do Sul (IPH/UFRGS) e a Superintendência de Fiscalização da ANA (TED Nº 03/2023/ANA). Também expressamos nossa gratidão ao Instituto Nacional de Meteorologia (INMET) pela disponibilização pública dos dados meteorológicos observados.

## REFERÊNCIAS

- ALLEN, Richard G. Assessing Integrity of Weather Data for Reference Evapotranspiration Estimation. *Journal of Irrigation and Drainage Engineering*, v. 122, n. 2, 1996.
- ALLEN, Richard G. *et al.* *FAO Irrigation and Drainage Paper No. 56 - Crop Evapotranspiration (guidelines for computing crop water requirements)*. Rome: Food and Agriculture Organization of the United Nations, 1998.
- ASCE-EWRI. The ASCE standardized reference evapotranspiration equation: ASCE-EWRI Standardization of Reference Evapotranspiration Task Committee Report. *American Society of Civil Engineers*, 2005.
- CROSBY, Teri; IGLEWICZ, Boris; HOAGLIN, David C. How to Detect and Handle Outliers. *Technometrics*, v. 36, n. 3, 1994.
- DURRE, Imke *et al.* Comprehensive automated quality assurance of daily surface observations. *Journal of Applied Meteorology and Climatology*, v. 49, n. 8, 2010.
- GELARO, Ronald *et al.* The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *Journal of Climate*, v. 30, n. 14, 2017.
- HERSBACH, Hans *et al.* The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, v. 146, n. 730, 2020.
- Instituto Nacional de Meteorologia - INMET. Disponível em: <<https://portal.inmet.gov.br/>>. Acesso em: 16 jun. 2025.
- MASTERS, Jeff. *A World Record Low Humidity? 116°F With a 0.36% Humidity in Iran | Weather Underground*. Disponível em: <<https://www.wunderground.com/cat6/world-record-low-humidity-116f-036-humidity-iran>>. Acesso em: 29 maio 2025.
- RODELL, Matthew *et al.* The Global Land Data Assimilation System. *Bulletin of the American Meteorological Society*, v. 85, n. 3, 2004.
- VOLK, John *et al.* Development of a Benchmark Eddy Flux Evapotranspiration Dataset for Evaluation of Satellite-Driven Evapotranspiration Models Over the CONUS. *Agricultural and Forest Meteorology*, v. 331, p. 109307, 15 mar. 2023. Acesso em: 26 fev. 2023.
- WMO. *World Weather and Climate Extremes Archive*. Disponível em: <<https://wmo.int/files/records-of-weather-and-climate-extremes-table?book=21486>>. Acesso em: 28 maio 2025.