

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

GABRIEL NEVES DA SILVA.

**ESTUDO A RESPEITO DA AQUISIÇÃO DE OBSERVAÇÕES METEOROLÓGICAS
NO CPTEC/INPE.**

CACHOEIRA PAULISTA,

2019.

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

GABRIEL NEVES DA SILVA.

**ESTUDO A RESPEITO DA AQUISIÇÃO DE OBSERVAÇÕES METEOROLÓGICAS
NO CPTEC/INPE.**

Relatório entregue para prestação de atividades realizadas no CPTEC/INPE, Cachoeira Paulista, sob orientação de Eduardo B. M. Barbosa.

CACHOEIRA PAULISTA,

2019.

RESUMO

O Centro de Previsão do Tempo e Estudos Climáticos de Cachoeira Paulista – CPTEC/INPE – recebe diariamente um grande volume de dados meteorológicos por meio do GTS – Global Telecommunications System –, e para manter a qualidade e precisão desses dados se faz necessário aplicar métodos estatísticos como o Kolmogorov-Smirnov, que foi aplicado aos dados meteorológicos de 2017, e desta forma foi comprovada a normalidade dos dados mês a mês, de janeiro a setembro, fazendo uso também de gráficos de Controle Estatístico de Processos – CEP –, desenvolvidos no software Rstudio. Além disso, falhas no sistema de recepção e transmissão de mensagens puderam ser apontadas mediante a aplicação de ferramentas da qualidade como o PDCA e o diagrama de Ishikawa.

Palavras-chave: dados, normalidade e qualidade.

Abstract: The Center for Weather and Climate Studies of Cachoeira Paulista – CPTEC / INPE – receives daily a large volume of meteorological data through the GTS – Global Telecommunications System –, and to maintain the quality and accuracy of these data it is necessary to apply statistical methods such as Kolmogorov-Smirnov, which was applied to the meteorological data of 2017, and this way it was proved the normality of the data from month to month, from January to September, also using Statistical Process Control (SPC) charts developed in the software Rstudio. In addition, failures in the system of reception and transmission of messages could be pointed out by the application of quality tools like the PDCA and the Ishikawa diagram.

Keywords: data, normality and quality.

1. INTRODUÇÃO

A estatística é de suma importância para compreender e até mesmo apontar falhas dentro de um conjunto de dados, assim, é possível verificar sua qualidade e precisão. Por meio de gráficos de Controle de Processo – CEP, pode-se encontrar onde está ocorrendo variações para serem corrigidas, através de um auxílio mútuo dos testes de normalidade.

O Centro de Previsão de Tempo e Estudos Climáticos (CPTEC) adquire um vasto conjunto de dados meteorológicos por meio do GTS – *Global Telecommunications System*, em formatos variados, que são utilizadas como insumo nas rotinas operacionais, no desenvolvimento de trabalhos científicos, execuções de modelos numéricos de previsão de tempo e monitoramento. Os dados do GTS são recebidos em dois formatos estabelecidos pela OMM - Organização Meteorológica Mundial, são eles BUFR e TAC.

- BUFR, *Binary Universal Form for the Representation of meteorological data*, basicamente é um tipo de dado de formatação binária que permite trabalhar de forma mais organizada, sendo um tipo tratado em primeira instância.

A Sigla BUFR segue que este é um formato padrão e extremamente flexível(universal). Evitando falsas expectativas com relação a isso, prefere-se dizer que o BUFR não é, de fato, um formato padrão, mas sim um padrão de formatação. (FERREIRA, S.H.S; ALMEIDA, W.G., 2006)

- TAC, *Traditinal Alphanumeric Codes*, combinação de caracteres alfanuméricos utilizados em transmissões de bandas, com baixa precisão, e que foram trabalhados posteriormente.

Além dos formatos, existem classificações de dados por cada método de extração, alguns deles:

- METAR – Informe meteorológico regular de aeródromo.
- SYNOP – Informe sinóticas à superfície, procedentes de estações terrestre fixa, dotada de pessoal ou automática.
- PILOT – Informe de vento em altitude, procedente de estações terrestre fixa.
- EMA – Informe de instrumentos meteorológicos (Sensores) das estações meteorológica automáticas.

- TEMP – Informe de pressão, temperatura, umidade e vento em altitude, procedente de estação terrestre.
- BUOY – Informe de boia de observação.
- SHIP – Informe de observação à superfície, de estação marítima

Cada respectivo método de extração é tratado individualmente e em junção dos formatos BUFR e TAC.

1.1 OBJETIVOS

O presente estudo tem por objetivo avaliar aquisição de dados do GTS e analisar a sua qualidade.

2. MATERIAIS E MÉTODOS

2.1 OBSERVAÇÕES METEOROLÓGICAS

2.1.1 GTS – GLOBAL TELECOMMUNICATION SYSTEM

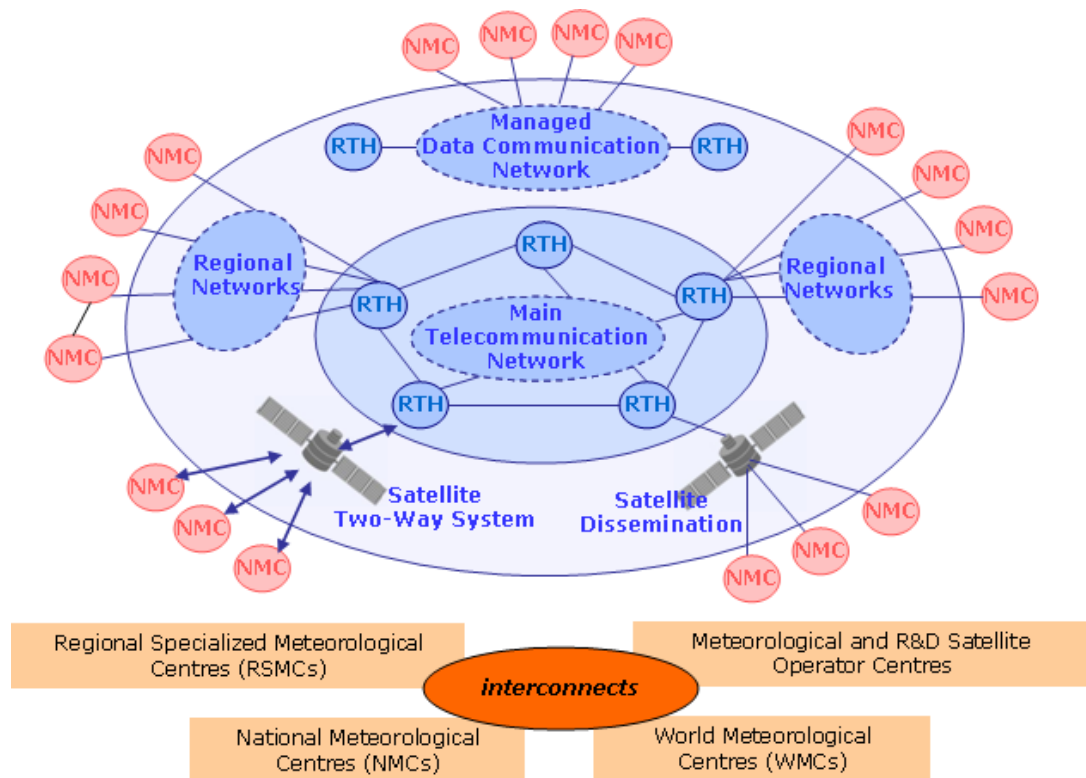
O sistema global de telecomunicações tem como um dos princípios: Ser uma “rede integrada para a coleta, intercâmbio e distribuição de informações em todo o mundo, com vistas a atender, de maneira eficiente e eficaz, os requisitos de todos os Serviços Meteorológicos Nacionais e também os requisitos da Organização Mundial e Regional.”(manual GTS)

O GTS da OMM - Organização Meteorológica Mundial, é um componente de comunicação e gestão de dados que permite a Vigilância meteorológica mundial (VMM ou WWW) operar através da coleta e distribuição de informações críticas para seus processos. Ele é implementado e operado pelos serviços meteorológicas nacionais dos países membros da OMM e de organizações internacionais como o ECMWF e EUMETSAT.

A WWW - *World Weather Watch* é o programa responsável pela troca instantânea de dados entre os países do globo. Para prever o tempo, a moderna meteorologia depende fundamentalmente dessa troca de dados.

Um exemplo da estrutura do GTS é mostrado na figura 1.

Figura 1: Estrutura do GTS



Fonte: Disponível em: <http://www.wmo.int/pages/prog/www/TEM/GTS/index_en.html>

A NMC – *National Meteorological Centres* tem como função coletar dados observacionais, receber e distribuir informações meteorológicas em nível nacional.

O RTH - *Regional Telecommunication Hubs* da MTN – *Main Telecommunication Network* é uma rede central que fornece um serviço de comunicação eficiente, rápido e confiável entre os MTCs - *Meteorological Telecommunication Centres* .

2.2 CONTROLE ESTATÍSTICO DE PROCESSOS

O CEP – Controle Estatístico de processo, uma das 7 ferramentas de qualidade, é um método de coleta de amostras, e permite a capacidade de monitorar as características de qualidade de um processo, assegurando o monitoramento e manutenção dentro dos limites preestabelecido: LSC: Limite superior de controle e o LIC: limite inferior de controle, que são limites permitidos de variação nos resultados do processo, e entre eles, é traçado à LM: linha média, e assim indicar quando adotar ações de correção e melhoria.

Segundo Ribeiro (1998) a variabilidade está sempre presente em qualquer processo produtivo, independente de quão bem ele seja projetado e operado. Há dois tipos de variabilidade, causa comuns e causa especiais:

- Causas de variação comuns: são consideradas aleatórias e inevitáveis, as variáveis do processo seguem uma distribuição normal, são variações que estejam dentro dos limites de controle.
- Causas de variação especiais: ocorrem por motivos identificáveis. As causas especiais alteram o parâmetro do processo, média e desvio padrão, pois estão fora dos limites de controle.

O CEP constituem ferramentas, onde seu objetivo é analisar o processo e identificar a variabilidade. Algumas ferramentas são: lista de verificação, uso de histograma, diagrama de dispersão, gráficos de controle.

Existem dois tipos básicos de gráficos de Controle:

- Gráficos por variáveis:
 - Gráficos \bar{X} e R (média e amplitude)
 - Gráficos \bar{X} e S (média e desvio padrão)
 - Gráficos \tilde{X} e R (mediana e amplitude)
 - Gráficos para *IMR* ou *XMR* Valores Individuais (X ou I) e Amplitude Móvel (*MR*)
- Gráficos por atributos:
 - Gráfico p (proporções não conforme)
 - Gráfico np (unidades não conforme)
 - Gráfico c (número de não conformidade por unidade)
 - Gráfico u (taxa de não conformidade por unidade)

O tipo gráfico a ser selecionado depende da característica da qualidade a ser controlada. Sendo esta uma característica contínua, os gráficos para valores individuais poderão ser utilizados. Muitas vezes instrumentos passa/não passa são utilizados, o que produz uma discretização das características contínuas, permitindo assim a utilização de gráficos por atributos.

2.3 TESTE DE NORMALIDADE

Os testes são importantes para encontrar probabilidades da distribuição de um conjunto de dados a serem normais. Existem diversos teste de normalidade, como os tipos Shapiro-Wilk e KS – Kolmogorov-Smirnov, e para selecionar um teste adequado para um conjunto de dados, é importante avaliar seu tipo, um dos mais eficientes para quando existem diversas amostragens é a do tipo KS.

2.3.1 KOLMOGOROV-SMIRNOV

KS é um teste não paramétrico sobre a igualdade de distribuições de probabilidade contínuas e unidimensionais, utilizadas para comprar amostras com distribuição de probabilidade de referência. No caso do teste da normalidade, as amostras são padronizadas e comparadas com uma distribuição normal padrão.

2.4 IDE RSTUDIO

Software livre e um ambiente computacional onde vem se aprimorando no desenvolvimento e manipulação de dados estatístico, análises e visualizações gráficas.

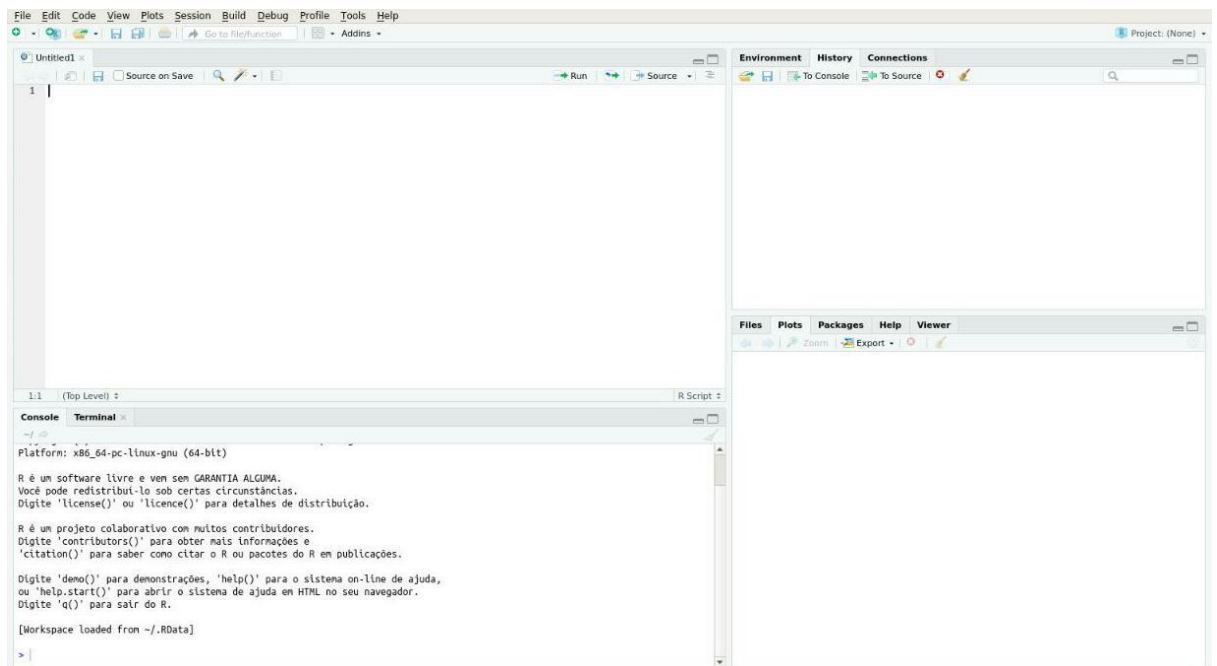
Figura 2: Logo RStudio



Fonte: Disponível em: <<https://www.rstudio.com/>>

O ambiente utiliza a linguagem R, que na atualmente é muito utilizada entre estatísticos e analista de dados para desenvolverem seus gráficos simples à gráficos mais complexos. O RStudio inclui um console, um terminal, editor de realce de sintaxe que suporta execução direta de código, bem como ferramentas para plotagem, histórico, depuração e gerenciamento de espaço de trabalho dentre muitas outras funcionalidades.

Figura 3: Ambiente de trabalho no RStudio



Fonte: Disponível na IDE RStudio

A IDE é facilmente encontrada no google, e se encontra disponível para os sistemas operacionais: Windows, Mac e Linux.

3. APRESENTAÇÃO E ANÁLISE DE RESULTADOS

3.1 ESTATÍSTICAS DA AQUISIÇÃO DE OBSERVAÇÕES METEOROLÓGICAS

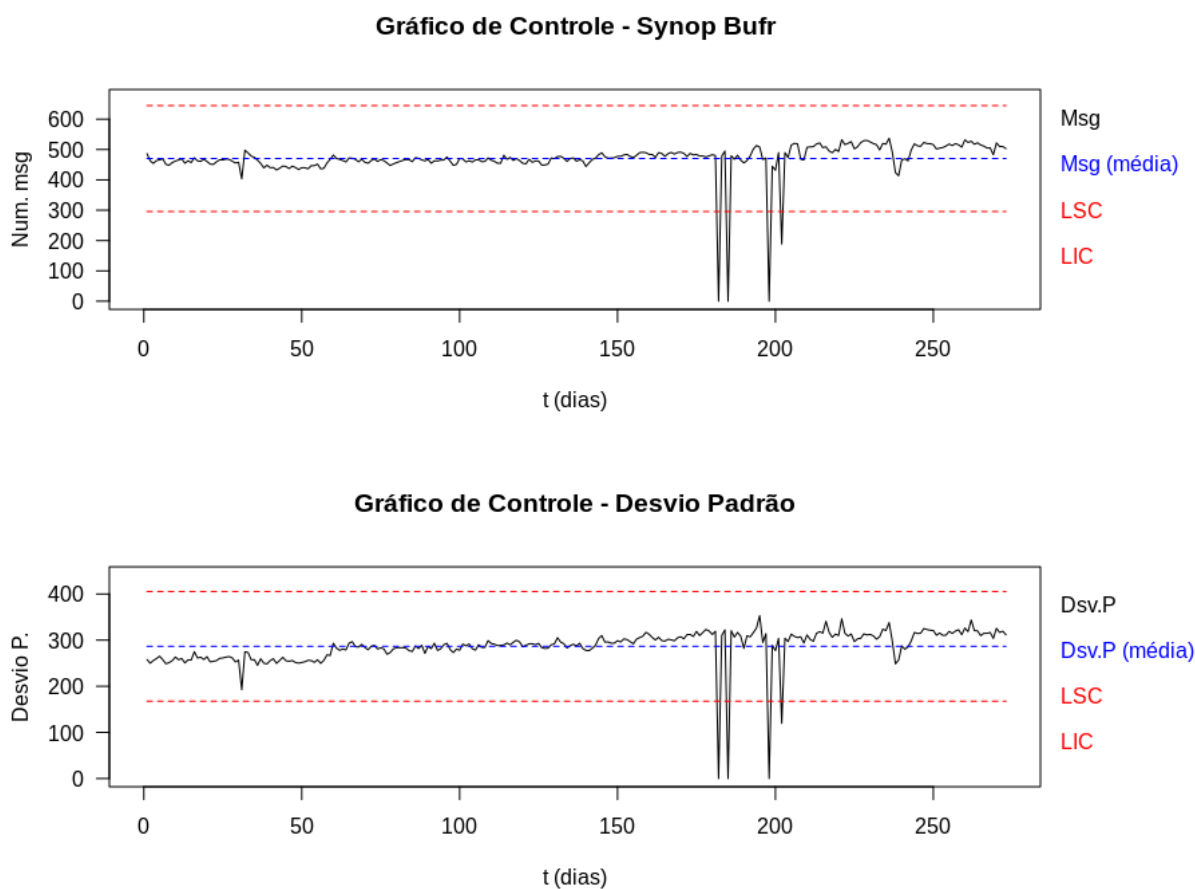
Há diversas maneiras de fazer comparação de gráficos, uma forma utilizada foi feita olhando para o gráfico de controle contendo X-Barra, pois é importante olhar para um gráfico, e obter análises e assim tirar uma solução daquilo que está tendo uma variação. Olhando junto do gráfico de controle do desvio padrão, nele poderemos visualizar diferenças especiais que não são encontradas no gráfico do X-Barra, isso irá mostrar se o gráfico está dentro ou fora dos parâmetros.

Foram verificados e analisados os formatos SYNOP e TEMP do tipo BUFR e TAC, e METAR e EMA do tipo TAC.

3.1.1 GRÁFICO DOS DADOS SYNOP

Analisando a Figura 4 SYNOP-BUFT, podemos perceber que de acordo com que o tempo vai passando a tendência é receber uma maior quantidade de mensagens ultrapassando um pouco a média, mas nunca chegar perto do limite superior, e mesmo com essa alteração do número das mensagens, ele tende a ficar dentro do controle estatístico. Percebe-se que há variações brutas mesmo havendo uma tendência de crescimento, entre o tempo 175 a 210, percebe-se que três mensagens chegam a zerar, isso quer dizer que naquele dia mensagens do formato BUFT não foram recebidas, enquanto uma chega a ficar um pouco abaixo do limite inferior. Observando o seu desvio para verificar essas quatro quedas e nos dados, foi confirmado que nas três primeiras quedas o número de mensagens zerou e no quarto o número chegou a diminuir pela metade.

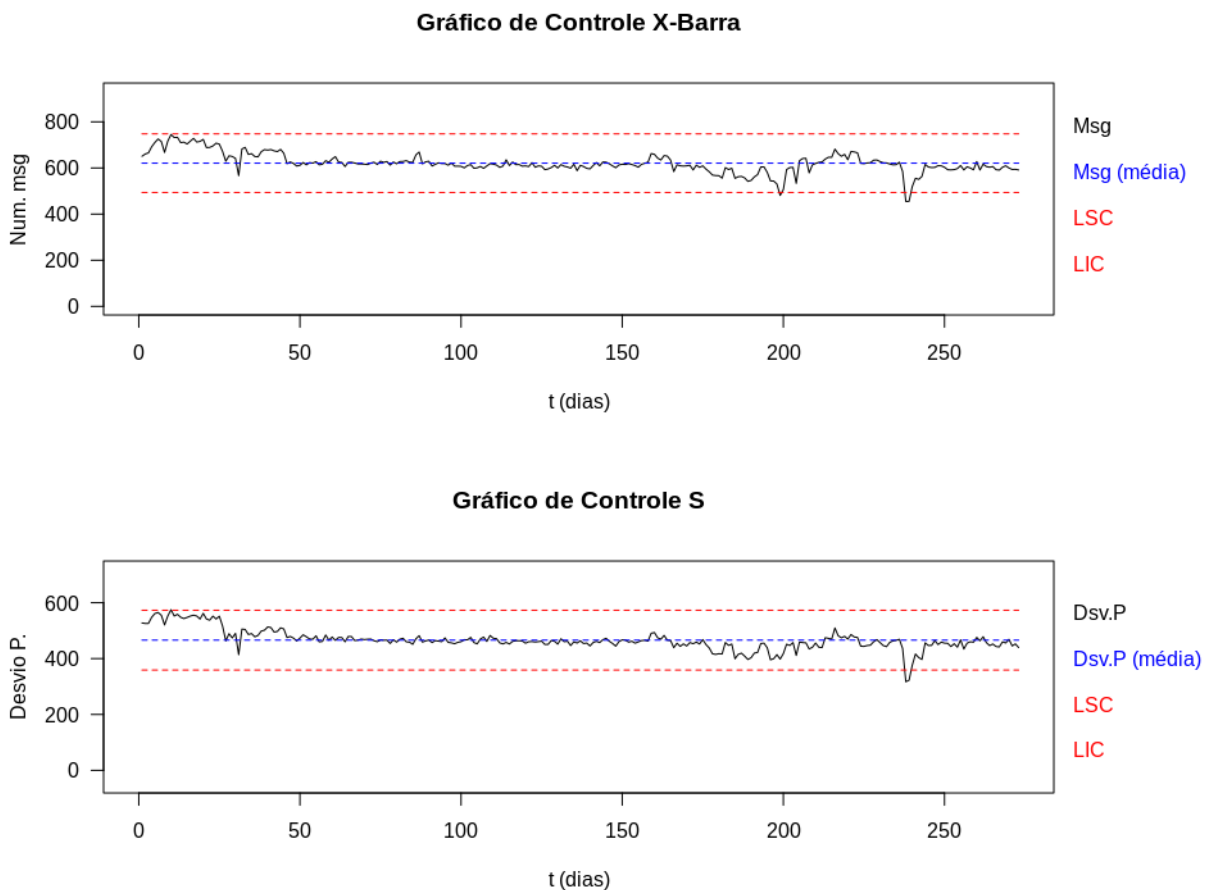
Figura 4: Synop-Bufr



Fonte: Autor(2019). Elaborado no RStudio.

Analisando a figura 5 SYNOP-TAC, podemos perceber no tipo TAC ela inicia com altas quantidades de mensagens ficando acima da média, e assim há uma caída no número das mensagens a medida que o tempo vai passando chegando a se manter na média, mantendo o controle dentro dos parâmetros aceitáveis, havendo poucas variações para baixo. Percebe-se que no tipo TAC não há variações bruscas como no tipo BUFR, apenas dois dados chegam a ultrapassar o limite inferior, mas nada comparado ao BUFR que conteve dias que chegou a zerar o número de mensagens, compara-se também olhando para o gráfico em S as mesmas variações ocorrem, foram verificadas apenas três pequenas alterações abaixo do limite inferior.

Figura 5: SYNOP-TAC

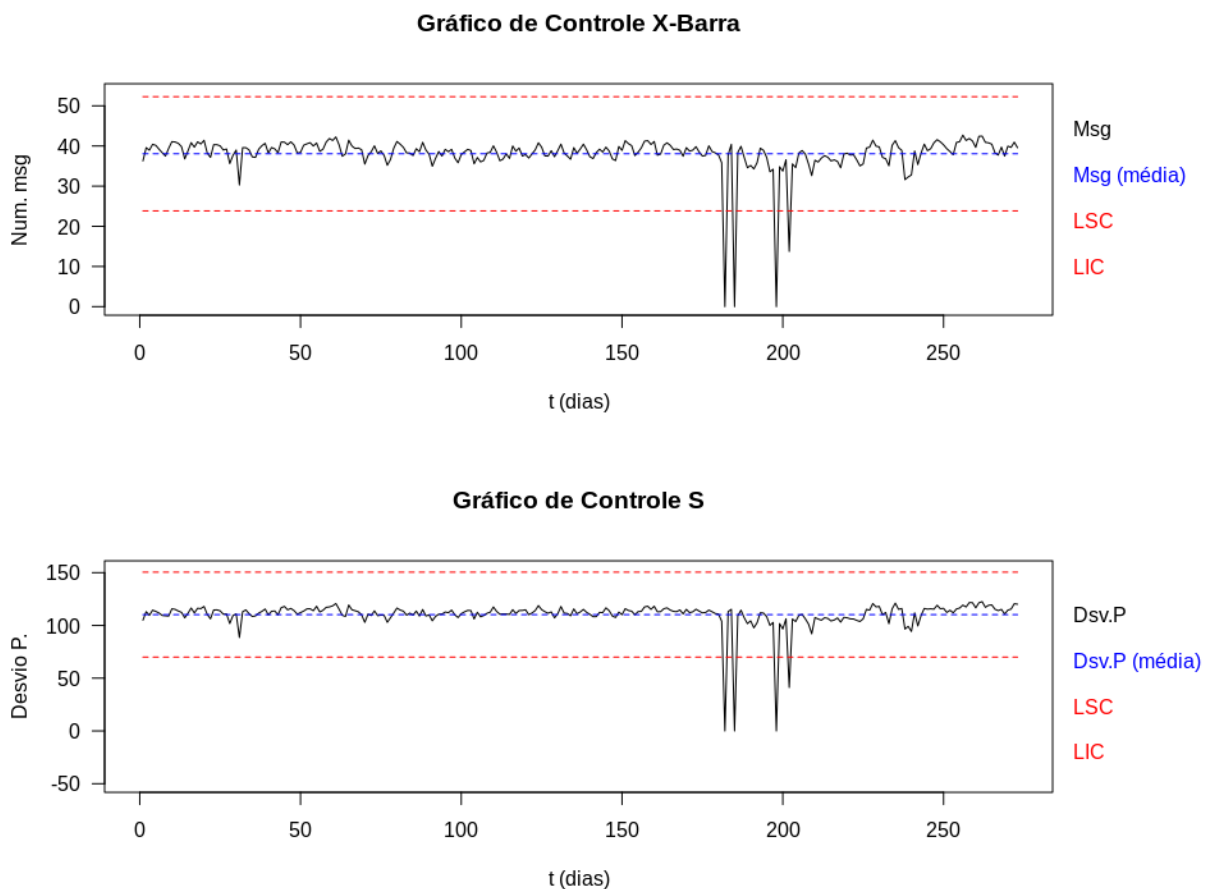


Fonte: Autor(2019). Elaborado no RStudio.

3.1.2 GRÁFICO DOS DADOS TEMP

Análise da figura 6: As mensagens tendem a se manterem ao nível da média, assim como a amplitude da variação tende a se manter parecidas até o final do período. Nota-se que à medida que o tempo vai chegando a 175 começa a ter um declínio do número de mensagens e também ocorrendo às mesmas quatro quedas que os outros tipos de dados BUFR. Após a queda, o número de mensagens tende a voltar a linha da média. Concluiu-se comparando com o gráfico em S que o tipo TEMP-BUFR tende a ficar dentro do controle estatístico, pois não houve variações significantes a serem questionáveis, exceto as mesmas quedas que os outros tipos BUFR também tiveram.

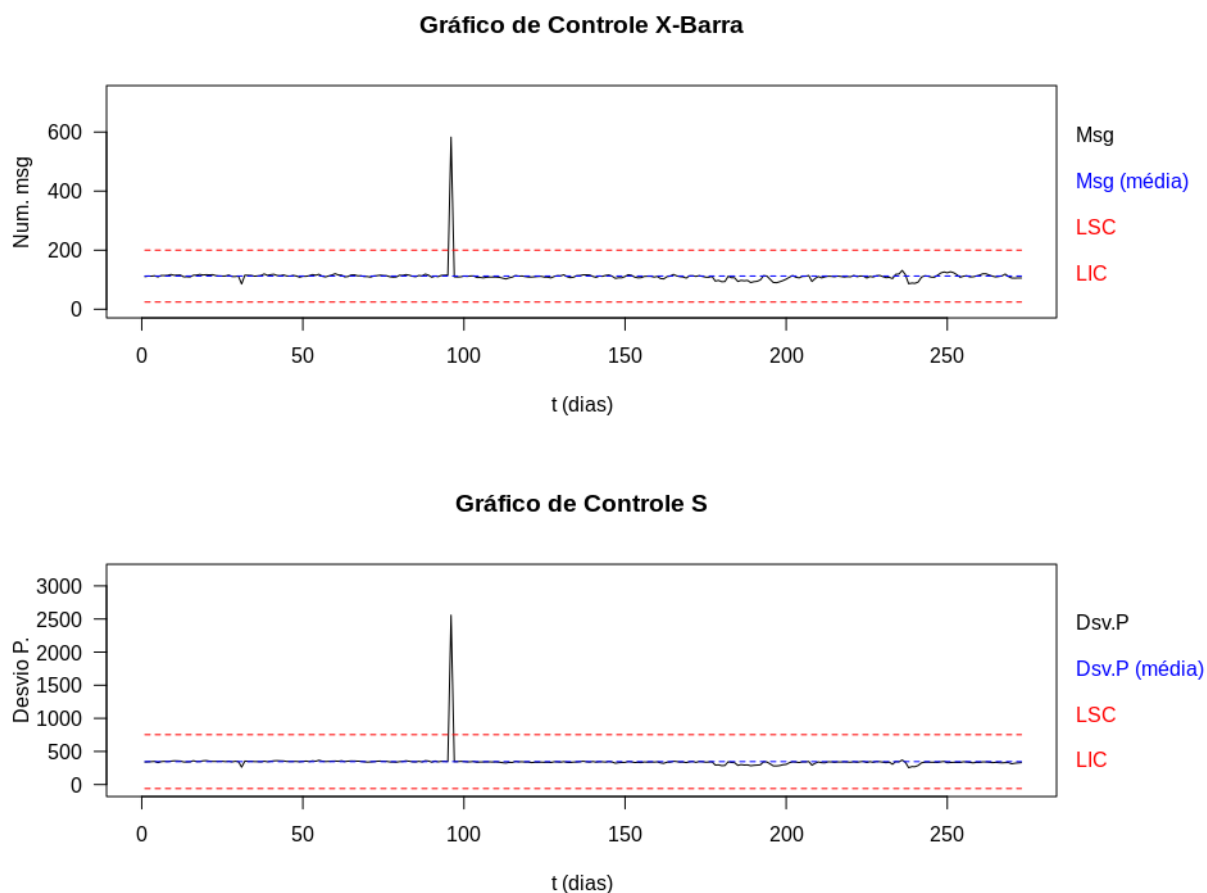
Figura 6: TEMP-BUFR



Fonte: Autor(2019). Elaborado no RStudio.

Análise da Figura 7: O Comportamento do tipo TEMP TAC tende a se manter na linha da média, mas com pequenas oscilações depois do tempo cento e cinquenta. Nota-se que próximo ao tempo cem, tem um aumento brusco de aproximadamente mais de quinhentas mensagens, isso pode-se considerar um erro da parte de recepção do dado. Esse mesmo erro foi confirmado comparando com o gráfico em S, portanto conclui-se que o tempo TEMP TAC manteve dentro do controle estatístico, havendo pequenas alterações onde foram possíveis visualizar essa alteração no número de mensagens.

Figura 7: TEMP-TAC

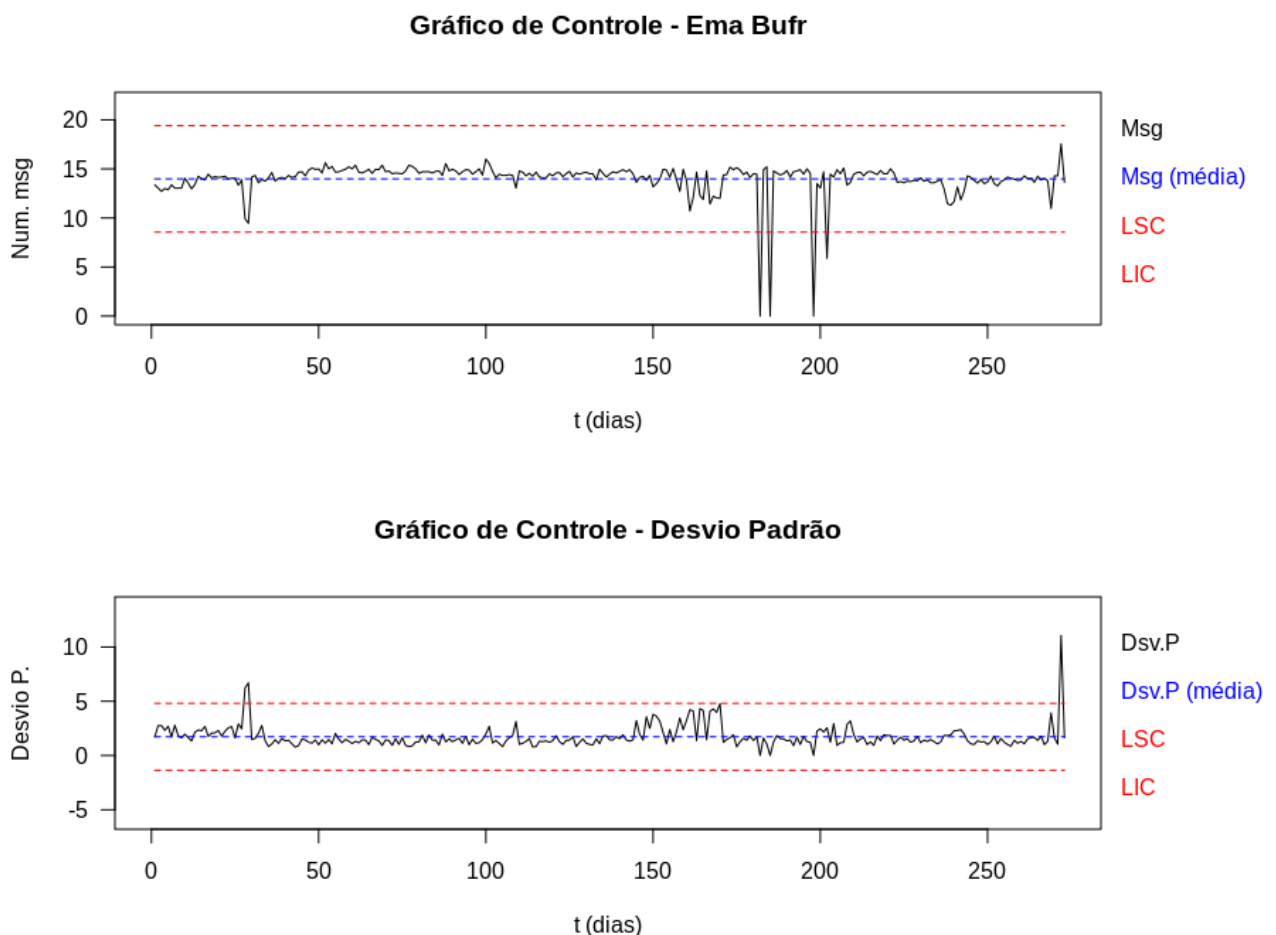


Fonte: Autor(2019). Elaborado no RStudio.

3.1.5 GRÁFICOS DOS DADOS EMA

Analisando a Figura 10 SYNOP-BUFR, percebe-se que o gráfico X-Barra tende a permanecer em torno da média, assim como o gráfico S permanece na média, mas mesmo ocorrendo algumas variações a ser questionáveis próximas do dia cento e cinquenta, o número de mensagens ainda estão dentro do controle estatístico. O tipo ema BUFR também encontra as mesmas três quedas que chegam a zerar o número de mensagens. Observa-se que no início do gráfico e no final no gráfico a picos tanto de diminuição e aumento do número de mensagens, portanto confirmou-se essa variação utilizando o gráfico em S, onde mesmo no X-Barra ele se encontra dentro dos parâmetros, o S mostra uma alteração. Foi confirmado que nesses dias o número de mensagens chegou a cair pela metade ou mais do que a metade.

Figura 11: EMA-Bufr

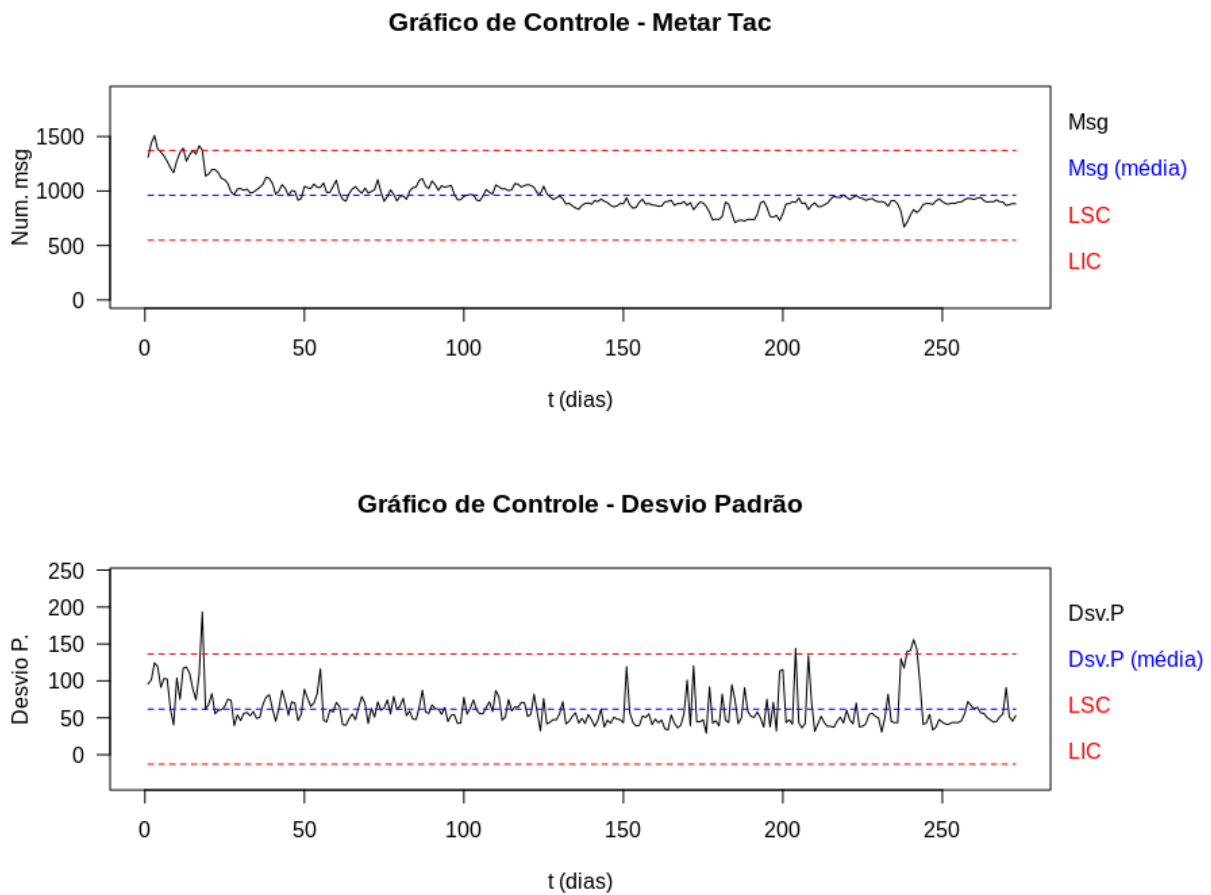


Fonte: Autor(2019). Elaborado no RStudio.

3.1.6 GRÁFICOS DOS DADOS METAR

O número de mensagens de Metar Tac observado na figura 12, inicia-se em uma quantidade considerável, mas ao decorrer do tempo o mesmo vai diminuindo e se estabilizando abaixo da linha da média, durante essa queda e as pequenas variações no número de mensagens, percebe-se que no desvio ocorre uma grande variação, todo momento onde a variação da amplitude é alta, o número de mensagens teve ou uma queda repentina ou uma subida repentina. Conclui-se que o tipo metar Tac tende a ficar dentro do controle estatístico, e há pontos que foram questionados onde mesmo estando dentro do controle no gráfico X-Barra, o gráfico em S mostra uma alteração.

Figura 12: Metar-Tac



Fonte: Autor(2019). Elaborado no RStudio.

CONSIDERAÇÕES FINAIS

Com o término do projeto foi possível analisar por meio dos gráficos e dos testes de normalidade, a qualidade dos dados recebidos no período de nove meses - janeiro a setembro - do ano de 2017.

Observando os gráficos foi possível visualizar datas onde a recepção de mensagens decaiu e onde a recepção teve um aumento ultrapassando os limites, é possível também comparar com seu desvio para averiguar se a sua qualidade ainda está dentro dos parâmetros aceitáveis mesmo que o número de mensagem tenha excedido ou diminuído, e assim questionar os motivos, sendo esses motivos problemas externos de estrutura ou momentos onde não foram recebidas as mensagens. Essas quedas de recepção retornavam em previsões não tão precisas. Os tipos de dados como o BUFR se assemelham por ocorrerem o mesmo tipo de erro, no presente estudo - o número de mensagem chegar a zero. Já os tipos TAC, ele tem uma recepção independente da do BUFR, não havendo nenhuma alteração mesmo que o número de mensagens tenha caído no BUFR, observou-se casos onde o número de mensagens pode ser maior em um tipo do que em outro.

Portanto, por meio dessa análise, pode-se definir soluções por meio do diagrama de Ishikawa e o PDCA, e assim, melhorar a qualidade dos dados meteorológicos.

Conclui-se também, que a utilização do CEP foi de suma importância para a realização da verificação da qualidade das amostras obtidas nesses nove meses.

REFERÊNCIAS

AERONÁUTICA. **Divulgação de informações Meteorológicas.** Meteorologia, ICA 105-1, 2016.

FERREIRA, S. H. S., CARVALHO, L. S. M., O’FILHO, E. F. **Banco de Dados Meteorológicos para Previsão de Tempo e Estudos Climáticos.** Cachoeira Paulista, SP, INPE.

FERREIRA, S.H.S; ALMEIDA, W.G. **Módulo para Codificação e Decodificação de Dados Meteorológicos em Formato BUFR: Guia Básico de Utilização.** São José dos Campos: Instituto Nacional de Meteorologia, 2006.

FERREIRA, S. H. S., ALMEIDA, W. G., FERREIRA, A. L. T. **Utilização do formato Bufr no CPTEC/INPE e Ferramental Desenvolvido.** Cachoeira Paulista, SP, INPE.

RIBEIRO, J.L. & CATEN, C.T. (1998) – **Controle Estatístico do Processo.** Apostila do Curso de Pós –Graduação em Engenharia de Produção. Editora da UFRGS. Porto Alegre

ROSÁRIO, K. P.,DANTAS, L. M., OEIRAS, E. N. **Aplicação de Controle Estatístico de Processo no Monitoramento de Peso médio de polpas de frutas: um estudo realizado numa empresa de médio porte.** Fortaleza, CE, Brasil, 2015.

WERKEMA, M.C.C. **Ferramentas estatísticas básicas para o gerenciamento de processos.** Belo Horizonte: Fundação Christiano Ottoni, 2006.

RIBEIRO, J.L. & CATEN, C.T. (1998) – **Controle Estatístico do Processo.** Apostila do Curso de Pós –Graduação em Engenharia de Produção. Editora da UFRGS. Porto Alegre