

CONTROLE DE QUALIDADE DOS DADOS

1. INTRODUÇÃO

A complexidade de cada algoritmo está relacionada com o tipo de dado a ser qualificado. Existe uma tendência natural dos testes e algoritmos se tornarem mais robustos e confiáveis ao longo do tempo. Desta forma, o *feedback* dos usuários é essencial para progressivamente aprimorar a qualidade dos dados disponibilizados.

A sistemática da qualificação automática de dados se baseia em comparar as medições com limites ou padrões previamente estabelecidos. Caso o valor medido exceda o limite ou não se encaixe no padrão, recebe um *soft* ou *hard flag*.

Um *hard flag* é aplicado a um dado quando não há dúvidas quanto à sua má qualidade. Após ser marcado com um *hard flag*, identificação de número 4, o dado não é disponibilizado ao público, exceto se sua marcação for retirada através de uma intervenção manual de um analista qualificado. O *soft flag* é atribuído aos dados que possuem indícios que questionam sua qualidade. Ao ser marcado com um *soft flag*, identificações de números 2 e 3, o dado é disponibilizado ao público e periodicamente avaliado pelo analista para definir a sua veracidade. Tais marcações seguem a seguinte estrutura:

Tabela 1. Relação dos diferentes tipos de *flags* e suas interpretações.

Flag	Interpretação
0	Controle de qualidade não realizado
1	Dados Bons
2	Dados Provavelmente Bons
3	Dados Provavelmente Ruins
4	Dados Ruins

Alguns testes são interdependentes, ou seja, alguns testes-base são pré-requisitos para a execução dos demais testes. Exemplo, caso o dado seja marcado como espúrio durante o teste de verificação do tempo, ele não prossegue pelas demais rotinas, sendo descartado. Outro fator importante a ser considerado é a íntima relação de algumas variáveis. Por exemplo, caso a

intensidade do vento seja identificada com algum problema é muito provável que os dados de rajada também estejam degradados.

Os testes de Controle de Qualidade aplicados nos dados do PNBOIA são referenciados na Documentação Técnica nº 09-02 do *National Data Buoy Center* (NDBC). Nesta seção será descrito cada teste aplicado aos dados bem como proceder para sua correta interpretação.

2. IDENTIFICAÇÃO DOS FLAGS

Para facilitar a interpretação dos dados, nos subitens a seguir encontram-se sumarizados os *hard* e *soft flags* utilizados na qualificação dos dados do PNBOIA, juntamente com o seu identificador.

Um dado pode ser identificado durante um ou mais testes, o que significa que o mesmo foi rejeitado de acordo com um ou mais critérios preestabelecidos.

2.1. *Hard flags*

A tabela a seguir apresenta a lista de *hard flags* utilizados durante o controle de qualidade dos dados do PNBOIA. Em adição a referida tabela, ainda há um *hard flag* identificado com o número 0, que se refere a “Correlação de dados”.

Nome do Teste	Identificador
Verificação de Valores Inexistentes	1
Verificação dos Intervalos de Medição	2
Verificação dos Intervalos de Climatológicos	9
Verificação da Altura Significativa de Onda x Altura Máxima	4
Verificação da Altura Significativa de Onda x Período Médio	7
Verificação da Velocidade Máxima do Vento x Rajada	3
Verificação de Voltagem da Bateria	5
Verificação de Dados Repetidos	6
Verificação da Continuidade no Tempo	8

2.2 *Soft flags*

A tabela a seguir apresenta a lista de *soft flags* utilizados durante o controle de qualidade dos dados do PNBOIA.

Tabela 3. Relação dos testes que podem resultar em *soft flags* e seus identificadores

Nome do Teste	Identificador
Verificação da Temperatura do Ponto de Orvalho x Temperatura do ar	51
Exceção 1 da Entrada de Frente Fria	53
Exceção 2 da Entrada de Frente Fria	54
Exceção 3 da Entrada de Frente Fria	55
Exceção 4 da Entrada de Frente Fria	56
Exceção 5 da Entrada de Frente Fria	57
Exceção 6 da Entrada de Frente Fria	58

3. ROTINAS DE QUALIFICAÇÃO

3.1. Verificação do Tempo de Coleta (Time Check)

Essa é a primeira fase de teste ao qual o dado é submetido. Essa rotina tem o objetivo de verificar se o dado possui um valor de tempo (data e hora) que não se refere ao futuro.

Código:

```
T_now = time.time()  
timeflag=[0]*len(Epoch)  
for i in xrange(len(Epoch)):
```

Ressalta-se que é utilizado um fator de segurança de 10 minutos para o futuro, para que assim se previna a inserção errônea de um *flag* a um dado que esteja somente desalinhado por alguns minutos.

```
if Epoch[i] > T_now+600:  
timeflag[i]=4  
else:
```

continue

Se o dado for rejeitado neste teste, ele não prossegue na rotina de qualificação, sendo descartado.

3.2. Verificação de dados Inexistentes (*Missing Value Check*)

Durante essa etapa são filtrados os dados inexistentes, que normalmente são marcados com um valor indicativo (ex.: 999, *NaN*, *misvalue*, *None*, etc) que porventura possam ter sido transmitidos. Caso o dado não passe neste teste, ele recebe um *flag* 4 em relação ao identificador

Código:

```
if dado(i) == 999 or dado(i) == NaN or dado(i) == "misvalue"
    flag(i) = 4;
    idf(i) = 'I'
else:
    continue
```

3.3. Verificação do Intervalo dos Dados (*Range Check*)

3.3.1. Verificação dos Limites de Aquisição

Esse teste visa identificar possíveis dados espúrios que estão fora dos limites mínimos e máximos de aquisição pré-definidos pelo fabricante dos sensores (tabela a seguir). Caso o dado esteja fora dos limites ele recebe um *flag* 4, para o identificador 2 e não é armazenado no banco.

Parâmetro	Limite Inferior	Limite Superior
Altura de Onda (m)	0	19.9
Período Médio de Onda (s)	1.7	30
Direção Média de Onda (°)	0	360
Velocidade do Vento (m/s)	0	59
Direção do Vento (°)	0	360
Rajada (m/s)	0	59
Temperatura do Ar (°C)	-39	59
Pressão Atmosférica (dbar)	501	1099
Temperatura do Ponto de Orvalho (°C)	-29	39
Temperatura da Superfície do Mar (°C)	-3	29
Umidade Relativa do ar (%)	25	102

Intensidade de Corrente (cm/s)	-499	499
Direção de Corrente (°)	0	360

Código:

```

if dado(i) > limite_superior | dado(i) < limite_inferior
    flag(i) = 4;
    idf(i) = '2'
else:
    continue

```

3.3.2. Verificação dos Limites Climatológicos

Esta verificação permite que determinados dados sejam comparados à climatologia regional, possibilitando a identificação de dados espúrios. Dentre os parâmetros adquiridos pelas boias, os que recebem tal tratamento são: Altura de onda, período de pico, velocidade do vento, rajada, temperatura do ar, pressão atmosférica, temperatura do ponto de orvalho, temperatura da superfície do mar e intensidade de corrente.

Inicialmente os dados climatológicos são provenientes de bases mundiais (ex. *National Oceanographic Data Center*, NODC) ou de fontes da literatura. A tendência é que, com o tempo, a própria rede de boias gere uma base de dados robusta com uma climatologia mais precisa e específica para cada ponto de coleta.

Por exemplo, o intervalo climatológico dos dados de TSM foi obtido a partir da média anual do NODC para a área de atuação da rede, com resolução de 1/4 de grau, mais ou menos três vezes o valor do desvio padrão. A partir deste cálculo obteve-se o intervalo de 9,8°C-37,16°C. São atribuídos *flags* do tipo 4, para o identificador '9', aos dados cujos valores são comprovadamente fora dos intervalos climatológicos.

Código:

```

if dado(i) > clima_superior | dado(i) < clima_inferior
    flag(i) = 4;
    idf(i) = '9'
else:

```

continue

3.4. Verificação da Altura Significativa de Onda x Altura Máxima

A Altura Significativa de Onda (Hs) representa o valor médio da altura do 1/3 das maiores ondas registradas durante o período de amostragem, enquanto que a Altura Máxima (Hmax) representa a altura da maior onda registrada nesse mesmo período. Logo, por definição, o valor da Hs não pode ser superior ao valor da Hmax.

Nessa etapa da qualificação, os dados de Hs que não respeitam a premissa descrita acima são marcados com o *hard flag*, para o identificador 4, assim como os dados Hmax.

Código:

```
if Hs (i) > Hmax (i)
    flag Hs (i) = 4;
    idf Hs (i) = '4'
    flag Hmax (i) = 4;
    idf Hmax (i) = '4'
else:
    continue
```

3.5. Verificação da Velocidade Máxima do Vento x Rajada

Seguindo a mesma linha teórica apresentada no item 3.4, o presente teste verifica se os valores de intensidade média dos ventos são superiores aos dados de rajada, assim como confere se os dados de rajada são inferiores a 0.5 m/s. Se a primeira verificação for satisfeita, os dois parâmetros são marcados com *flags* do tipo 4, para o marcador 3. Da mesma forma, se o teste aplicado exclusivamente aos dados de rajada for positivo, este será identificado pelo mesmo *flag* apresentado anteriormente, para o marcador 3.

Código:

a) <i>if wind_speed (i) > gust (i)</i>	b) <i>if gust (i) < 0.5</i>
<i>flag w (i) = 4;</i>	<i>flag g (i) = 4;</i>

<i>idf w (i) = '3';</i>	<i>idf g (i) = '3';</i>
<i>flag g (i) = 4;</i>	<i>else:</i>
<i>idf g (i) = '3';</i>	<i>continue</i>
<i>else:</i>	
<i>continue</i>	

3.6. Verificação da Temperatura do Ponto de Orvalho x Temperatura do Ar

Durante esse teste, os dados de Temperatura do Ponto de Orvalho são comparados com a Temperatura do ar para verificar se os mesmos estão de acordo com a teoria apresentada no item 2.1.5 deste manual. Se a Temperatura do Ponto de Orvalho for maior que a Temperatura do Ar, a mesma receberá um *soft flag* tipo 51 e seu valor será igualado ao da Temperatura do Ar.

Código:

```

if dewp (i) > atmp (i)
    dewp (i) = atmp (i);
    flag d (i) = 3;
    idf d (i) = '51';
else:
    continue

```

3.7. Verificação da Redundância e Hierarquia dos sensores

Idealmente, um sistema de coleta de dados deve possuir todos os dados coletados em redundância, para em caso de falha do sensor primário, o dado continuar sendo provido pelo sistema secundário. Entretanto, por questões logísticas e financeiras nem sempre esta configuração é possível.

Devido à grande importância para a previsão meteorológica, as boias do PNBOIA coletam em redundância os dados de vento (i.e. direção, intensidade media e rajada). A fase de qualificação descrita abaixo é aplicada nestes dados em específico, mas ele pode ser estendido para os demais dados coletados em redundância em outras plataformas meteoceanográficas.

Os dois anemômetros são organizados em uma hierarquia (primário e secundário). Assume-se que os dados do sensor primário são mais confiáveis e precisos. Entretanto, durante a

sua utilização podem ocorrer situações que sua qualidade é degradada (ex. desgaste natural, abalroamento, danos causados por aves marinhas, etc). Esse teste visa garantir que o dado a ser disponibilizado corresponde ao equipamento em melhor estado. Para tal, após os testes iniciais os dados dos dois anemômetros são comparados.

Se ambos os equipamentos estiverem bons, os dados do sensor primário são disponibilizados. Caso o dado do sensor primário tenha recebido um *flag* durante a qualificação, são disponibilizados os dados do sensor secundário.

Para que se possa realizar a comparação entre os dados de vento com alturas diferentes, os mesmos são corrigidos para a altura de 10 metros, utilizando a metodologia descrita por LIU *et al.* (1979): *Bulk parameterization of air-sea exchanges in heat and water vapor including molecular constraints at the interface, Journal of Atmospheric Science*, 36, pp 1722-1735.

3.8. Verificação de Voltagem da Bateria x Pressão Atmosférica

Para seu correto funcionamento, o sensor de pressão atmosférica das boias requer da bateria que alimenta o sistema da boia uma carga mínima de 10.5 V. Caso a voltagem seja inferior a este valor a qualidade dos dados de pressão fica comprometida, sendo assim, eles recebem um *hard flag* para o marcador 5.

Código:

```
if battery (i) < 10.5
    flag pressure (i) = 4;
    idf pressure (i) = '5'
else:
    continue
```

3.9. Verificação de Dados Repetidos

Durante esta etapa da qualificação é verificado se os sensores não estão congelados no tempo, ou seja, se há uma variação nos valores dos dados entre as amostragens. Os últimos dados recebidos são comparados e caso não haja variação entre eles, recebem um *hard flag* para o marcador 6. Atualmente, considera-se, para fins de verificação, a repetição de seis dados

contínuos (ou 6 horas). Entretanto, a experiência da aplicação deste teste mostrará qual será a quantidade de dados a ser utilizada.

Código:

```
if dado [t(i)] = dado [t(i-1)] = dado [t(i-2)...]  
    flag (i) = 4;  
    idf (i) = '6';  
else:  
    continue
```

3.10. Verificação da Continuidade no Tempo

O Teste da Continuidade visa verificar a consistência dos dados no tempo. Ou seja, ele verifica se a variação temporal dos dados está dentro de um intervalo aceitável, de acordo com a seguinte fórmula:

$$\sigma_t = 0.58 \sigma \sqrt{T}$$

onde:

σ_t é o desvio padrão aceitável dentro do intervalo de tempo T, e;

σ é o desvio padrão estimado a partir de um conjunto de dados pretéritos.

Os dados são comparados com os dados das últimas 3 horas. Caso a variação exceda o desvio padrão máximo estipulado, o dado recebe um *hard flag* para o marcador 8. Se não houver dados válidos nas ultimas 3 horas, o teste não é aplicado.

3.11. Verificação das Exceções pelos Sistemas Frontais

Esta é uma etapa de suma importância durante a qualificação de dados meteorológicos, pois ela visa retirar possíveis *flags* errôneos aplicados aos dados associados a sistemas frontais (i.e. frentes frias, cristas, cavados, etc.). Por uma questão óbvia, dados obtidos durante a passagem de eventos meteorológicos (tempo ruim/instável) possuem um maior valor agregado do que dados relativos a períodos de tempo bom.

Foram definidos seis tipos de exceções ao teste de continuidade no tempo. Estas exceções estão comumente relacionadas ao comportamento característico das variáveis ambientais quando da passagem de sistemas frontais:

Exceção nº1: Temperatura do Ar x Direção do Vento

Se o dado da temperatura do ar e direção do vento receberem um *flag* durante o teste de continuidade, mas a mudança na direção do vento entre as duas últimas medidas for superior a 40°, o *flag* da temperatura do ar é retirado.

Exceção nº2: Direção do Vento x Temperatura do Ar

Se o dado da direção do vento e temperatura do ar receberem um *flag* durante o teste de continuidade, mas a mudança na direção do vento entre as duas últimas medidas for superior a 40°, o *flag* da direção do vento é retirado.

Exceção nº3: Temperatura do Ar x Velocidade do Vento

Se o dado da temperatura do ar receber um *flag* durante o teste de continuidade, mas o valor da velocidade do vento for superior a 7 m/s, o *flag* da temperatura do ar é retirado.

Exceção nº4: Intensidade do Vento x Pressão Atmosférica

Se o dado de velocidade do vento for identificado com um *flag* no teste de continuidade, mas a pressão atmosférica correspondente e a imediatamente anterior forem menores que 995 hPa, o *flag* da velocidade é retirado.

Exceção nº5: Pressão Atmosférica x Pressão Atmosférica

Se o dado da pressão atmosférica for filtrado durante o teste de continuidade, mas o valor obtido previamente for menor que 1000 hPa, o *flag* é retirado.

Exceção nº6: Altura de Onda x Intensidade do Vento

Os dados de Altura de Onda que foram excluídos durante o Teste de Continuidade são reaceitos e disponibilizados se a intensidade do vento for igual ou superior a 15 m/s.

3.12. Correlação de dados

Caso alguns dados sejam considerados espúrios (*hard flag*), automaticamente os dados correlatos também ser considerados espúrios. Desta maneira, caso um dado medido pelo correntômetro obtenha um *hard flag*, todos os outros dados coletados pelo equipamento receberão um indicador de *hard flag* '0'. A mesma relação ocorre para os dados coletados pelo ondógrafo e os dados relativos ao vento.