

Análise de Consistência e Exploratória de Dados Meteorológicos

Elaborado por:



EÓLICA
engenharia

Miguel Hiroo Hirata
Regina Araújo
Cecilia Araújo
Renata Machado

Para:

Agência de Cooperação
Técnica Alemã – GIZ
Deutsche Gesellschaft für
Internationale Zusammenarbeit (GIZ) GmbH

Outubro 2011

giz



Programa Energia
Brasil-Alemanha



Análise de Consistência e Exploratória de Dados Meteorológicos

Elaborado por: Eólica Engenharia Ltda.

Autores: Miguel Hiroo Hirata
Regina Araújo
Cecilia Araújo
Renata Machado

Para: Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH

Programa: Programa Energia Brasileiro-Alemão

No do Programa: 2007.2189.4-001.00

Coordenação: Torsten Schwab (GIZ),
Juarez Lopes (EPE)

Outubro 2011

Informações Legais

1. Todas as indicações, dados e resultados deste estudo foram compilados e cuidadosamente revisados pelo(s) autor (es). No entanto, erros com relação ao conteúdo não podem ser evitados. Consequentemente, nem a GTZ ou o(s) autor (es) podem ser responsabilizados por qualquer reivindicação, perda ou prejuízo direto ou indireto resultante do uso ou confiança depositada sobre as informações contidas neste estudo, ou direta ou indiretamente resultante dos erros, imprecisões ou omissões de informações neste estudo.
2. A duplicação ou reprodução de todo ou partes do estudo (incluindo a transferência de dados para sistemas de armazenamento de mídia) e distribuição para fins não comerciais é permitida, desde que a GTZ seja citada como fonte da informação. Para outros usos comerciais, incluindo duplicação, reprodução ou distribuição de todo ou partes deste estudo, é necessário o consentimento escrito da GTZ.

Introdução	2
1. Características do AMA	4
2. Análise de consistência de dados	6
2.1 Premissas adotadas na análise de consistência de dados	8
3. Procedimentos e metodologias para identificação de erros	10
3.1 Erros grosseiros	10
3.2 Erros sistemáticos ou sistêmicos	11
3.3 Erros aleatórios ou valores atípicos	15
4. Aplicativo computacional para análise de consistência e exploratória de dados	21
Anexo I: Códigos de erros para os arquivos AMA	32
Anexo II: Arquivo de dados consistidos	34
II.1. Cabeçalho	34
II.2. Registro	36

Introdução

A Empresa de Pesquisas Energéticas (EPE) desenvolveu e está processando o aplicativo computacional “Sistema de Acompanhamento de Medições Anemométricas – AMA” com o objetivo de formar um Banco de Dados Anemométricos e Climatológicos a partir dos registros de medições efetuadas em parques eólicos vencedores dos leilões de energia de reserva e dos leilões de fontes alternativas.

Com o AMA espera-se constituir um Banco de Dados Anemométricos que retrate com alta confiabilidade o regime eólico dos locais de medições e que sirva de referência para estudos prospectivos de locais potencialmente candidatos para futuros aproveitamentos de geração de energia elétrica, além de propiciar um avanço nacional no desenvolvimento de técnicas e metodologias de planejamento, operação e integração de parques eólicos ao sistema elétrico brasileiro.

Para que o objetivo do AMA seja atingido é necessário que o Banco de Dados seja abastecido com dados coerentes e de qualidade confiável, o que requer uma análise prévia da consistência das medições antes que façam parte do banco.

Este relatório apresenta metodologias e procedimentos a serem seguidos para a análise de consistência dos dados de campo enviados pelos empreendedores para a EPE. A análise se dá em duas etapas. Na primeira é verificada se o arquivo de dados está de acordo com as características do AMA. Se não estiver, o arquivo é automaticamente devolvido ao responsável pela operação da estação meteorológica para a devida compatibilização. Para os arquivos enviados na forma correta, é feita para cada variável do arquivo a análise de consistência das medições considerando seus limites físicos, comportamento temporal em relação ao histórico das medições, comparação das estatísticas no intervalo de medição e cruzamento com os dados medidos em outros sensores. Se nas proximidades da estação meteorológica existir outra estação, a consistência pode abordar o cruzamento dos dados das duas estações, mas que não é objeto deste estudo.

Estão apresentadas no relatório as funcionalidades de um aplicativo computacional para auxiliar o usuário do Banco de Dados na tomada de decisão de usar ou não um dado considerado com suspeita de erro. O aplicativo deverá ainda possuir ferramentas estatísticas para uma análise exploratória dos dados, com os resultados apresentados em tabelas e gráficos.

Este relatório está organizado em quatro capítulos e dois anexos. O capítulo 1 descreve as características do sistema AMA a serem observadas na metodologia da análise de consistência de seus dados. O capítulo 2 trata do processo de análise de consistência de dados, apresenta o fluxograma desse processo e contém as premissas utilizadas. O capítulo 3 detalha os

procedimentos e metodologias para identificação de erros, sendo dividido em três seções de acordo com o tipo de erro. Finalmente, o capítulo 4 descreve as funcionalidades e as ferramentas de interação com o usuário de um aplicativo computacional para a análise de consistência e exploratória dos dados do AMA e apresenta sugestões de saídas em tela de vídeo para tais ferramentas. O Anexo I descreve a metodologia para código de erros elaborada pela EPE a ser utilizada neste aplicativo. O Anexo II detalha a estrutura do arquivo de dados consistidos citado no capítulo 2 e sugere como seria a sua escrita em uma linguagem computacional.

1. Características do AMA

Este item descreve as características do aplicativo AMA que devem ser consideradas na metodologia para a análise de consistência objetivada.

- a) Cada parque eólico tem somente um ponto de coleta de dados meteorológicos, onde os sensores para as medições estão instalados em uma única torre, conjunto chamado de estação meteorológica.
- b) Na estação meteorológica estão instalados: 1 termômetro, 1 higrômetro, 1 barômetro e pelo menos 2 anemômetros e 2 *wind vane*. A quantidade de anemômetros e *wind vane* e as alturas em que estão instalados são indicadas na ficha de cadastro da estação no AMA.
- c) Os dados são medidos continuamente, com taxa de amostragem de 1 segundo, e a cada 10 minutos são gravados no sistema de aquisição as seguintes variáveis:
 - média da temperatura do ar, °C;
 - média da umidade relativa do ar, [%];
 - média da pressão atmosférica, [hPa];
 - média, máxima, mínima e desvio padrão da velocidade do vento, [m/s], para cada anemômetro;
 - média e desvio padrão da direção do vento em relação ao norte verdadeiro (geográfico), [°], para cada *wind vane*.
- d) A estação tem um cadastro com a sua localização geográfica, altitude, altura da torre de medição e as datas de calibração ou substituição dos sensores.
- e) A cada 15 dias um arquivo com os dados é enviado para a EPE através de sua página na internet, onde as três primeiras linhas contêm a identificação da estação, o calendário e horário de início, o calendário e horário do fim das medições, e em seguida os dados meteorológicos. O nome do arquivo segue uma dada especificação, iniciando com o código numérico da estação com 6 algarismos, seguido dos calendários de início e fim no formato (AAAAMMDD – ano,mês e dia).

Para o arquivo 000002_20101216_20101230.TXT, apresentado na “Nota Técnica DEA 14/09. Leilão de Reserva 2009: Instruções para as medições anemométricas e climatológicas”, tem-se os seguintes dados::

```
Estação | 00002
Início | 201012160500 (formato AAAAMMDDhhmm)
Fim | 201012300900 (formato AAAAMMDDhhmm)
20101216 | 050000| 000| 1009| 32,9| 56,2| 8,65| 9,50| 7,80| ...
...
...
20101230 | 090000| 002| 998| 26,3| 66,9| 9,75|10,50| 6,60| ...
```

A ordem das variáveis nas linhas com os registros dos dados está apresentada na Tabela 1.

Tabela 1 – Ordem das variáveis nas linhas com os registros de dados

CH	Variável	formato
01	calendário do início do intervalo	AAAAMMDD – ano, mês, dia
02	horário início	hhmmss – hora, minuto e segundo
03	código de erro	Anexo I - Tabela I.1 elaborada pela EPE

CH	instrumento	Estatística em 10 minutos	Variável [unidade]	Posição na torre
04	barômetro	média	pressão atmosférica [hPa]	altura de 10 m
05	termômetro	média	temperatura do ar [°C]	altura de 10 m
06	higrômetro	média	umidade relativa do ar [%]	altura de 10 m
07	anemômetro 1	média	velocidade do vento [m/s]	topo da torre
08		máxima		
09		mínima		
10		desvio padrão		
11	<i>wind vane 1</i>	média	direção do vento [°]	1,5m abaixo anemômetro 1
12		desvio		
13	anemômetro 2	média	velocidade do vento [m/s]	20 metros abaixo do topo
14		máxima		
15		mínima		
16		desvio padrão		
17	<i>wind vane 2</i>	média	direção do vento [°]	1,5m abaixo anemômetro 2
18		desvio padrão		
...			
...	anemômetro n			

- f) O arquivo de dados é formatado em ASCII e na falta de um dos dados no registro, o campo correspondente será preenchido com hífen “-”.
- g) É possível que nas proximidades do parque em consideração existam outros parques. Neste caso é possível realizar uma análise de consistência através do cruzamento dos dados de duas estações meteorológicas considerando a variabilidade espacial do fenômeno meteorológico entre as mesmas, esta análise não é objeto deste estudo.

2. Análise de consistência de dados

Uma série temporal de dados meteorológicos resulta de medidas efetuadas no campo ao longo do tempo através de instrumentação apropriada. Ainda que as medições sejam feitas com os modernos sistemas informatizados e automáticos de aquisição de dados, que são mais precisos e confiáveis que os instrumentos convencionais, não é incomum, ao utilizar as séries de dados, deparar-se com os mais variados tipos de erros de medição. Alguns são de fácil detecção, tais como os grosseiros e os sistemáticos. Todavia, existem os erros aleatórios que são mais difíceis de serem detectados, pois muitas vezes eles são comparáveis com valores atípicos ou extremos da série, ainda que apresentem um grande afastamento dos demais valores da série.

A tarefa de verificar a existência em um conjunto de dados brutos de possíveis erros de medições é chamada de análise de consistência de dados.

Na Figura 1 estão apresentados em forma de fluxograma os procedimentos a serem adotados para a análise de consistência e qualidade dos dados do AMA. Os itens que seguem detalham os procedimentos.

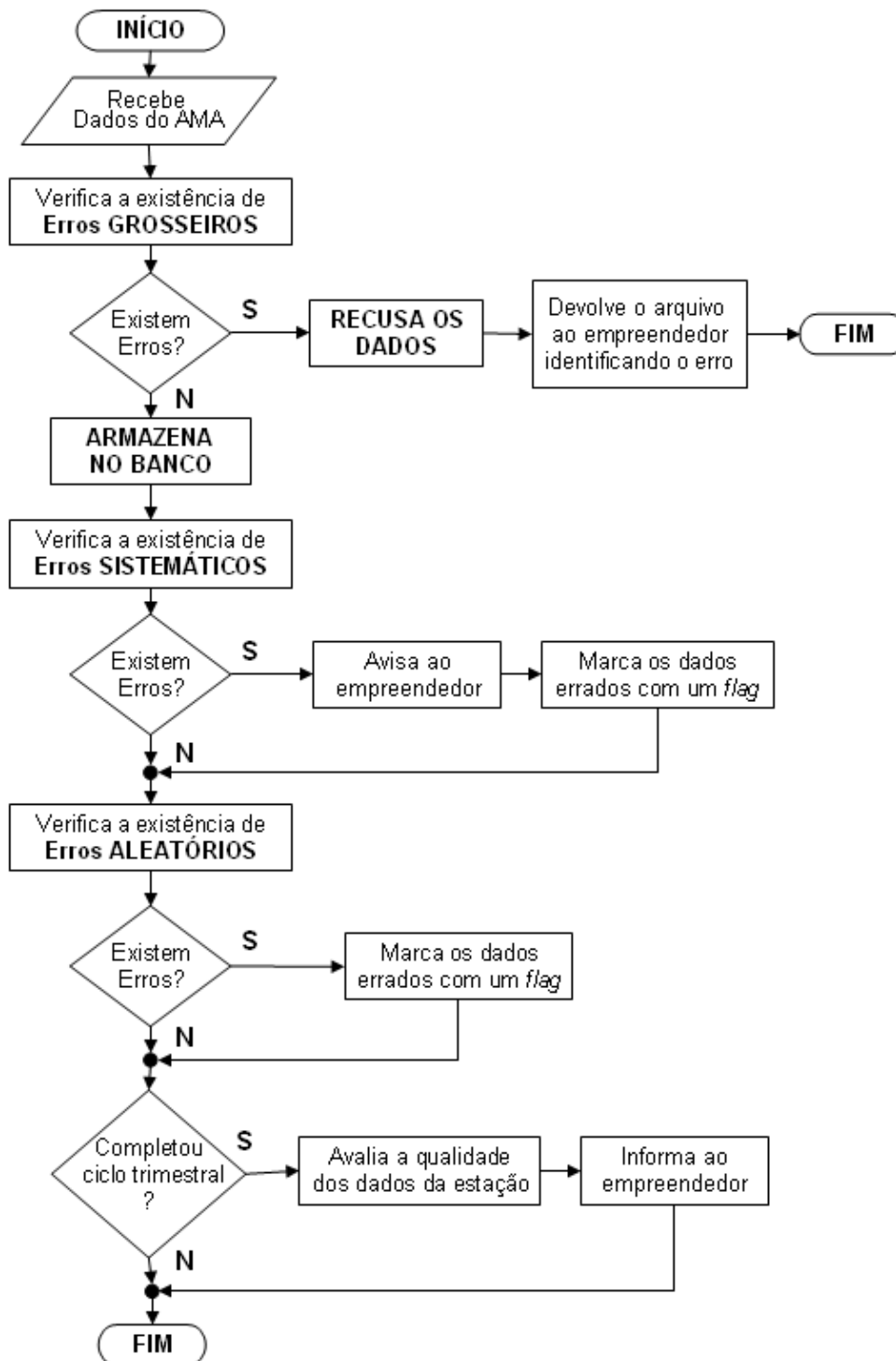


Figura 1 - Procedimentos para análise de consistência e qualidade dos dados do AMA

2.1 Premissas adotadas na análise de consistência de dados

- a) Os dados originais, medidas de campo, são preservados sem nenhuma modificação, ou seja, não é feita nenhuma correção nos dados considerados com suspeita de erro.
- b) Os arquivos quinzenais sem erros grosseiros são copiados para um arquivo específico da estação que contém todos os dados enviados pelo empreendedor, acrescidos de um campo por variável e por registro que contém um código indicando a qualidade de cada um dos dados. Para os dados considerados com suspeitas de erros, um código sinalizando o tipo da inconsistência verificada é gravado neste arquivo copiado.
- c) Neste arquivo são gravadas outras informações específicas da estação, tais como: número e altura dos sensores, limites de valores para consistência e estatísticas da série.
- d) O arquivo servirá de cópia de segurança dos dados enviados e deve ser escrito em formato binário. Mesmo com a inclusão de um campo a mais por variável, o arquivo em formato binário fica menor que o arquivo original. Além do mais, a leitura pelo aplicativo do arquivo binário é mais rápida.
- e) O arquivo é de livre consulta pelo usuário, ficando a critério do mesmo a utilização ou não do dado considerado suspeito.

Como este arquivo, específico de cada estação meteorológica, será citado outras vezes no relatório, o mesmo será aqui nomeado de ARQCONSISTIDO.BIN. No Anexo II – “Arquivos de dados consistidos” encontra-se descrita uma proposta de estrutura para o arquivo.

A análise de consistência dos dados, aqui proposta, é feita em três etapas. Se forem encontrados erros, os mesmos são classificados em três tipos: grosseiros, sistemáticos ou sistêmicos e aleatórios ou atípicos. Os arquivos com dados grosseiros não serão incorporados no Banco de Dados, mas sim devolvidos para o empreendedor que deverá processar as devidas correções ou adequações ao formato do arquivo AMA. Os arquivos livres de erro grosseiro são incorporados ao Banco de Dados e submetidos às demais etapas de análise de consistência.

A metodologia aqui proposta considera cada variável de forma isolada e também o seu cruzamento com outras variáveis correlatas, medidas simultaneamente em uma mesma altura e em alturas diferentes. Em linhas gerais a metodologia contempla as seguintes situações:

- identificação de erros grosseiros, identificando se o arquivo de dados corresponde à estação e valores fora dos limites esperados para a região da estação,
- confronto com as estatísticas do histórico da série. Para isto, a série histórica é dividida em períodos sazonais, para os quais são calculados as médias e os desvios padrão e verificado

se o dado em análise está dentro de um intervalo onde os valores máximos e mínimos são obtidos pela média e mais ou menos um número de desvios padrão,

Para os dados medidos em uma mesma altura é verificado se:

- a velocidade máxima instantânea é menor que a velocidade média de 10 minutos,
- a relação entre as velocidades máxima instantânea e a média de 10 minutos está fora do intervalo especificado,
- a velocidade máxima instantânea é maior que a velocidade média de 10 minutos acrescida de um certo número de desvios padrão. O número de desvios é dado em função da rugosidade do terreno,
- o desvio padrão da velocidade do vento é nulo durante longo tempo contínuo. O fato é forte indicação de defeito no anemômetro,
- o desvio padrão da direção do vento é nulo durante longo tempo contínuo. O fato é forte indicação de defeito no *wind vane*.

Para os dados medidos em alturas diferentes, a metodologia leva em consideração que a velocidade média do vento aumenta com a altura e o seu desvio padrão diminui, já que o atrito com obstáculos que provocam as variações decresce na sua intensidade. Nestes casos, são considerados se:

- a relação entre as velocidades médias nos diferentes níveis está fora de um intervalo especificado,
- a relação entre as velocidades máximas nos diferentes níveis está fora de um intervalo especificado,
- a defasagem entre as direções nos diferentes níveis está fora de um intervalo especificado.

3. Procedimentos e metodologias para identificação de erros

3.1 Erros grosseiros

Esta etapa é executada no momento do recebimento do arquivo de dados. Se nesta primeira análise for identificado algum erro o arquivo será recusado e devolvido ao empreendedor juntamente com um relatório descrevendo os tipos e a localização dos erros no arquivo.

1. Identificação da estação.

A identificação da estação é um código numérico de 6 dígitos e deverá constar no nome e na primeira linha do arquivo com os dados de medição. Portanto a primeira tarefa é verificar a coerência entre o nome do arquivo, o cabeçalho e o cadastro da estação no AMA. Exemplo:

Nome do arquivo: 000002_20101216_20101230.TXT

Primeira linha do arquivo: 000002

2. Continuidade no tempo com o último arquivo recebido.

O calendário e o horário do primeiro registro do arquivo recebido devem corresponder ao do último arquivo recebido acrescido do intervalo de 10 minutos. O calendário e horário de início estão informados na segunda linha do arquivo de dados. O calendário de início consta, também, no nome do arquivo e deve ser comparado com a segunda linha do arquivo. Exemplo:

Nome do arquivo: 000002_20101216_20101230.TXT

Segunda linha do arquivo: 201012160500

3. Intervalo de 10 minutos entre os registros.

É verificado se os registros estão ordenados sequencialmente no tempo, em intervalos de 10 minutos.

4. Número de variáveis em cada registro.

Cada linha, ou registro, do arquivo deve conter o número de variáveis descrito na formatação dos dados do AMA.

5. Valores numéricos nos registros.

Todos os dados dos registros devem ser numéricos. É ainda verificado se todos os valores são positivos, com exceção da temperatura do ar.

6. Final inesperado do arquivo.

É verificado se o calendário e horário do último registro do arquivo coincidem com os valores definidos a priori para o final do arquivo. O calendário final consta, também, no nome do arquivo e deve ser comparado com a terceira linha do arquivo. Exemplo:

Nome do arquivo: 000002_20101216_20101230.TXT

Terceira linha do arquivo: 201012300900

De acordo com as regras do AMA, se durante a operação da estação meteorológica ocorrer uma falha no sistema de aquisição de dados ocasionando a perda de um ou mais intervalos de medição, as linhas correspondentes a esses horários devem ser inseridas pelo empreendedor no arquivo. Neste caso a linha do arquivo deve conter o calendário, o horário, o código de erro correspondente à falha de todas as variáveis e para cada uma das variáveis o símbolo “-“ (hífen).

A Tabela 2 contém uma descrição dos erros grosseiros com as correspondentes mensagens abreviadas e códigos numéricos, para serem usados no aplicativo de análise de consistência.

Tabela 2 - Erro grosseiro

Descrição do erro	Código erro	
	Mensagem no aplicativo	numérico
identificação da estação	ID não coincide	1
continuidade no tempo com o último arquivo recebido	data início	2
intervalo de 10 minutos entre os registros	pulou registro	3
número de variáveis não esperado	número variáveis	4
valor não numérico	não numérico	5
valor negativo	negativo	6
final inesperado do arquivo	data fim	7

3.2 Erros sistemáticos ou sistêmicos

Esta etapa é executada posteriormente à inclusão do arquivo de dados no AMA. Como os erros sistemáticos ou sistêmicos são atribuídos a falhas ou defeitos no sistema automático de aquisição de dados, tão logo seja constatada a sua ocorrência, o empreendedor deve ser informado para que os mesmos sejam sanados.

1. Valores fora do limite físico

Os dados de cada variável são comparados com os valores extremos esperados para a região onde a estação meteorológica está instalada. Os limites serão descritos em um arquivo específico da estação (ARQCONSISTIDO.BIN) e poderão ser alterados a qualquer momento. Inicialmente esses limites podem ser dilatados e definidos em função das medições de outras estações meteorológicas existentes na região, com o passar dos anos os valores podem ser revistos em função dos dados medidos na própria estação. Os limites iniciais para a temperatura do ar, umidade relativa do ar e pressão atmosférica podem ser

estabelecidos, por exemplo, com base nas estações do INMET – Instituto Nacional de Meteorologia. Os limites da direção do vento são 0° e 360°. Para as velocidades mínimas, máximas e médias do vento os limites dependem da altura do anemômetro e os limites iniciais podem ser estabelecidos com base nos dados medidos e utilizados no projeto do parque eólico.

2. Sensor de velocidade do vento travado

Se todas as estatísticas média, máxima, mínima e desvio padrão da velocidade do vento em 10 minutos forem iguais a zero, podem ter ocorridas duas situações: vento calmo ou sensor travado. Se for devida a vento calmo, após um período a situação se normalizará; caso contrário ocorreu um defeito no sensor. Para classificar como defeito as estatísticas devem permanecer igual a zero em seis intervalos consecutivos de 10 minutos, ou seja, 1 hora. Lembrar que essas medidas são feitas em parques eólicos e a alturas elevadas

3. Sensor de direção do vento travado

Quando ocorre um mau funcionamento do sensor de direção do vento devido a problemas mecânicos como travamento dos rolamentos, a direção fica persistente. Este problema pode ser detectado analisando uma sequência de intervalos contínuos, por exemplo, 6 intervalos que correspondem a 1 hora. Se nesta sequência as velocidades médias de 10 minutos forem maiores que 3 m/s e as direções constantes deve ter ocorrido um travamento. Outro indicativo do travamento do sensor é a ocorrência do desvio padrão da direção igual a zero por mais de uma hora.

4. Valores discrepantes entre as velocidades do vento medidas em duas alturas

Fisicamente, o perfil vertical da velocidade do vento aumenta com a altura. Nos modelos de cálculos este aumento segue uma função logarítmica ou uma potência da relação entre as alturas, isto é:

$$\frac{V_1}{V_2} = \frac{\ln\left(\frac{h_1}{z_0}\right)}{\ln\left(\frac{h_2}{z_0}\right)} \quad \text{ou} \quad \frac{V_1}{V_2} = \left(\frac{h_1}{h_2}\right)^{\frac{1}{n}}$$

onde:

V_1 e V_2 são as velocidades do vento nas alturas h_1 e h_2 , respectivamente, e

z_0 e n são parâmetros que dependem da rugosidade do solo. Dependendo do local da estação a rugosidade pode variar com a direção do vento.

A consistência das velocidades medidas em duas alturas pode ser verificada através da formulação:

$$k_{min} \leq \frac{V_2}{V_1} \leq k_{max}$$

Onde:

V_2 e V_1 são as velocidades do vento nos anemômetros mais alto e mais baixo, respectivamente;

k_{min} e k_{max} são os limites adotados na análise e dependem da relação entre as alturas dos dois anemômetros, que nem sempre é a mesma. Por isso, não há como estabelecer um par de valores únicos para todas as estações e anemômetros. Nos casos em que o período de medições de dados é curto, menos de três meses, pode-se adotar, como primeira aproximação, $k_{min} = 0,9$ e $k_{max} = 2$. Com o aumento do período de medições estes limites podem ser reavaliados e se necessário modificados no arquivo ARQCONSISTIDO.BIN.

Teoricamente o limite mínimo deveria ser maior que 1. Entretanto existem situações climáticas onde ocorre uma inversão do perfil da velocidade do vento com a altura, o que justifica o uso de $k_{min} = 0,9$. Se na análise quinzenal dos ventos com velocidades maiores que 3 m/s forem encontrados um número excessivo de $k_{min} < 1$, digamos 5% do tempo, é provável que esteja ocorrendo um mau funcionamento dos anemômetros. Outra causa pode ser creditada a efeitos orográficos devido à posição da torre no parque.

5. Defasagem nas direções do vento medidas em duas alturas

Tem-se observado em medições realizadas no país para fins de estudos de aproveitamento eólico que não ocorrem variações significativas da direção do vento com a altura na faixa entre 30 e 80 metros. Isto pode ser atribuído ao fato de que nestas alturas há pouca influência dos obstáculos no solo que poderiam modificar a direção do vento com velocidades maiores que de 3 a 4 m/s. Com base nesta observação, caso ocorram sistematicamente diferenças entre as direções medidas em duas alturas for maior que 30 graus é um indicativo da incorreta orientação de um ou dos dois sensores em relação ao norte verdadeiro. O limite aceitável de defasagem entre as direções está gravado no arquivo ARQCONSISTIDO.BIN. A Figura 2 ilustra este tipo de erro.

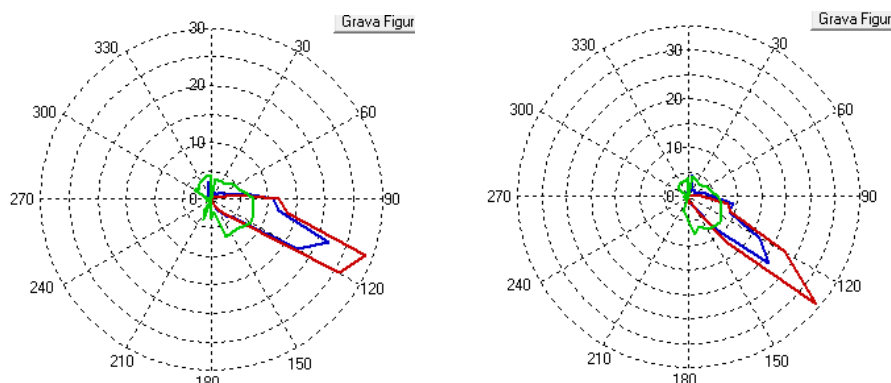


Figura 2 Wind vane instalados em alturas diferentes, com defasagem na orientação

6. Inversão de valores estatísticos da velocidade do vento

As estatísticas da velocidade do vento no intervalo de tempo de 10 minutos devem rigorosamente seguir a seguinte ordem de valores iguais ou decrescentes: máxima, média e mínima. Se ocorrer uma inversão da ordem é provável que a origem do erro esteja no sistema computacional do *data logger*.

7. Troca ou manutenção de sensor

Se ocorrer a substituição ou a manutenção de um sensor, os dados dos próximos 30 dias após a intervenção devem ser comparados com os dados dos últimos 30 dias anteriores. Esta comparação pode ser realizada através de gráficos com a evolução temporal dos dois períodos e de testes estatísticos de hipóteses da igualdade da média e do desvio padrão.

As datas de substituição e manutenção dos sensores estão anotadas na ficha de identificação da estação.

Na Tabela 3 estão apresentadas uma descrição dos erros sistemáticos ou sistêmicos com a sugestão das mensagens abreviadas a serem apresentadas no aplicativo de análise de dados e o código numérico correspondente que será gravado no ARQCONSISTIDO.BIN ao lado da variável com o erro encontrado. Observar que os códigos de erros não são todos sequenciais; propositalmente os códigos estão ordenados por tipo de instrumento e existem lacunas para inserção de outros tipos de erros que porventura não tenham sido aqui considerados.

Tabela 3 - Erros sistemáticos ou sistêmicos

Sensor	Descrição do erro	Código erro	
		Mensagem	num
Anemômetro N	mínima fora da faixa de valores previstos	Vmin N fora faixa	101
	média fora da faixa de valores previstos	Vmed N fora faixa	102
	máxima fora da faixa de valores previstos	Vmax N fora faixa	103
	mínima persistente	Vmin N persistente	104
	média persistente	Vmed N persistente	105
	máxima persistente	Vmax N persistente	106
	desvio padrão igual a zero	Vdp N = 0	107
	média maior que a máxima	Vmed N > Vmax N	108
média menor que a mínima	Vmed N < Vmin N	109	
Anemômetro N Anemômetro K	mínima anemômetro inferior maior que a mínima do anemômetro superior	Vmin N >> Vmin, K	120
	média anemômetro inferior maior que a média do anemômetro superior	Vmed N >> Vmed, K	121
	máxima anemômetro inferior maior que a máxima do anemômetro superior	Vmax N >> Vmax K	122
wind vane N	direção menor que 0° ou maior que 360°	Dmed N <> [0, 360]	130
	desvio padrão da direção igual a zero	Ddp N = 0	131
	Direção persistente	Dmed N persistente	132
wind vane N wind vane K	Defasagem nas direções entre duas alturas	Dmed N – Dmed k > 40	140
Termômetro	temperatura do ar fora da faixa de valores previstos	Tar fora da faixa	150
Higrômetro	umidade relativa fora da faixa de valores previstos	Urel fora da faixa	151
Barômetro	pressão atmosférica fora da faixa de valores previstos	Pat fora da faixa	152

3.3 Erros aleatórios ou valores atípicos

Esta etapa é executada posteriormente à inclusão do arquivo de dados no AMA. Como esse tipo de erro é identificado com base em estatísticas dos históricos dos dados, à medida em que o período de medições aumentar, os valores limites utilizados na análise de consistência podem ser alterados. Estes limites ficam armazenados no arquivo ARQCONSISTIDO.BIN.

Classificar se um dado, com características fora dos padrões de seu histórico, é um erro aleatório ou um valor atípico é uma tarefa um tanto subjetiva. Por isso, uma boa prática ao se deparar com esse tipo de dado não é removê-lo da série, mas sim marcá-lo como sendo um dado com suspeita de erro, deixando que a experiência do usuário defina o seu uso. Algumas vezes um dado atípico medido no início da operação da estação pode, com o aumento da massa de dados, tornar-se um dado dentro dos padrões de seu histórico. Portanto, é recomendado que as séries sejam re-analisadas uma vez por ano.

1. Valores suspeitos de não pertencer à população dos dados – Metodologia 1

A existência de erros aleatórios pode ser verificada através da variabilidade diária e mensal dos fenômenos meteorológicos. Neste caso, a série histórica é dividida em períodos sazonais, para quais são calculados as médias e os desvios padrão.

Isto feito verifica-se se o dado em análise está dentro de um intervalo onde os valores máximos e mínimos são obtidos pela média e mais ou menos um dado número de desvios padrão. Neste procedimento está-se admitindo que as medições naquele horário e mês tenham uma distribuição normal. As Figuras 3 e 4 exemplificam essa metodologia.

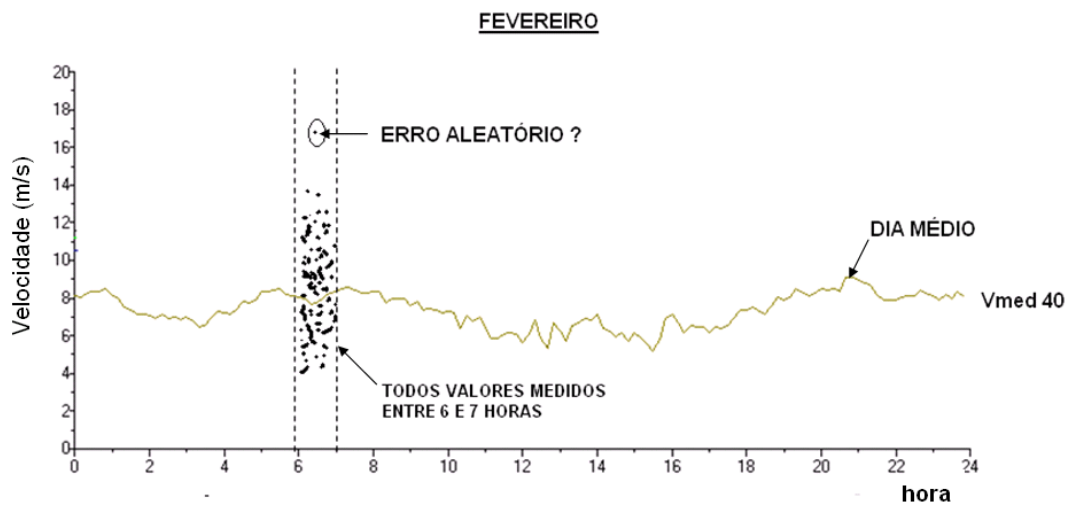


Figura 3 – Divisão da série para verificação de erros aleatórios

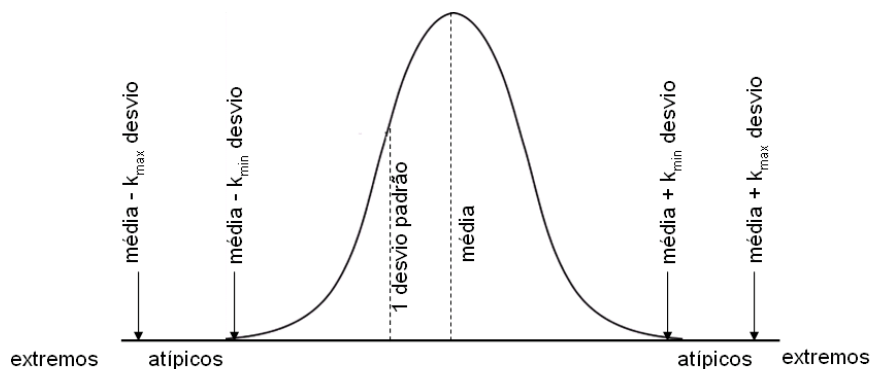


Figura 4 – Distribuição normal usada para identificar dados atípicos e extremos – Metodologia

2. Valores suspeitos de não pertencer à população dos dados – Metodologia 2

Outro procedimento semelhante ao anterior, mas que não considera os dados do mês e da hora segundo uma distribuição normal, está baseado no gráfico de Box-Plot. Este procedimento é oriundo da metodologia conhecida como análise exploratória de dados. Na Figura 5 está apresentado o gráfico que é construído da seguinte forma:

- calcula-se a mediana, o quartil inferior (Q1) e o quartil superior (Q3);
- subtrai-se o quartil superior do quartil inferior ($L = Q3 - Q1$);
- os valores que estiverem no intervalo de $Q3 + k_{\min} L$ e $Q3 + k_{\max} L$ e no intervalo $Q1 - k_{\min} L$ e $Q1 - k_{\max} L$, serão considerados atípicos, podendo portanto ser aceitos na população com alguma suspeita;
- os valores que forem maiores que $Q3 + k_{\max} L$ e menores que $Q1 - k_{\min} L$ devem ser considerados suspeitos de pertencer à população, devendo ser investigada a origem da dispersão. Esses pontos podem ser valores extremos ou erros aleatórios.
- Os valores de k_{\min} e k_{\max} dependem da variável e da localização da estação. Como valores iniciais pode-se usar os valores de 1,5 e 3, respectivamente, que serão paulatinamente melhor estimados e modificados com o aumento do período de coleta de dados.

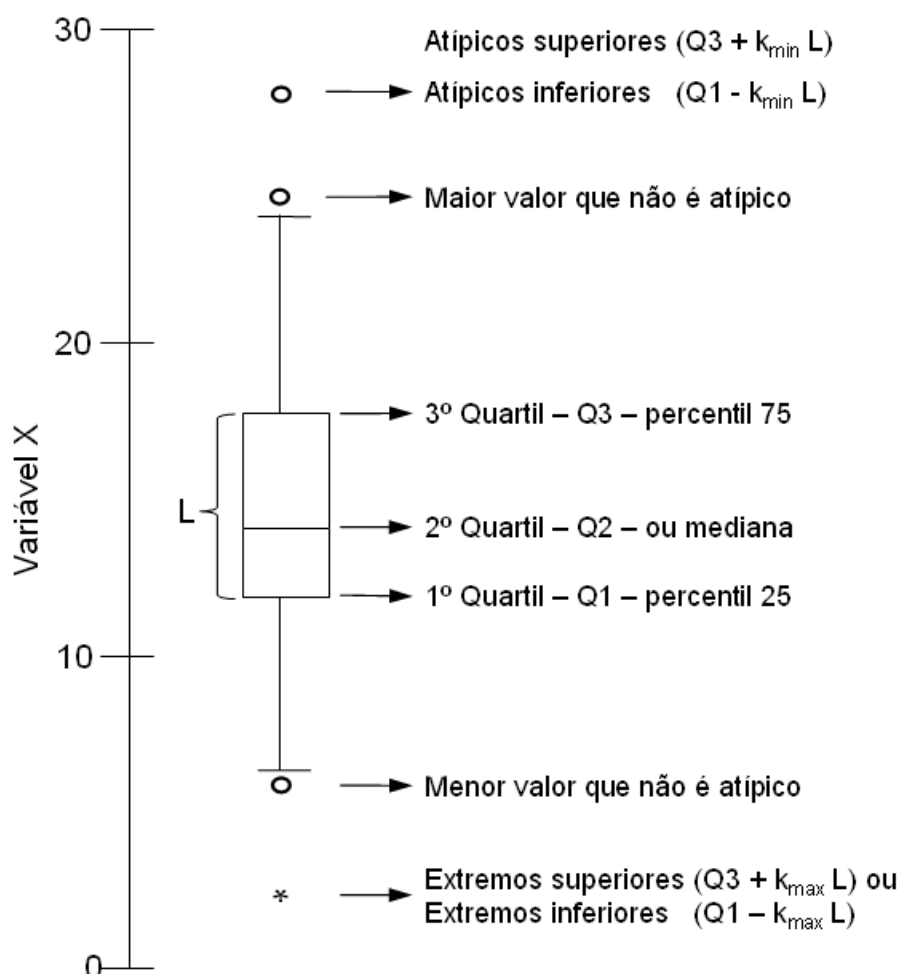


Figura 5 – Gráfico de Box-plot usado para identificar dados atípicos e extremos - Metodologia 2

3. Relação entre a média, máxima e desvio padrão

Teoricamente, no período de 10 minutos a velocidade máxima do vento pode ser expressa pela velocidade média acrescida de k desvios padrão, ou seja:

$$V_{\max} = V_{\text{med}} + k V_{\text{dp}}$$

O valor de k depende da rugosidade do solo e da velocidade média do vento. São estabelecidos dois valores para k , aqui chamados de k_{\min} e k_{\max} , e são usados para classificar se a suspeita de erro está relacionada a um valor atípico ou extremo. Para fins de análise de consistência, para velocidades médias acima de 4 m/s são sugeridos como valor iniciais de $k_{\min} = 3$ e $k_{\max} = 6$. Após um período de tempo de medições estes valores são reavaliados e se for o caso modificados para melhor se adequarem às características do local da estação meteorológica.

4. Desvio padrão da velocidade do vento maior que a velocidade média

O coeficiente de variação, expresso pela relação entre o desvio padrão e a média da velocidade em 10 minutos, define a intensidade de turbulência do vento. Como a turbulência é influenciada pela rugosidade do solo, ela decresce com o aumento da altura de medição do vento e com a intensidade do vento. Valores típicos da intensidade de turbulência se encontram no intervalo de 0,5 a 0,1. Portanto, um dos testes adotados na análise de consistência é verificar se a velocidade média de 10 minutos é maior que o desvio padrão ($V_{dp} > V_{med}$).

Tabela 3 - Erros aleatórios

sensor	Descrição do erro	Código erro	
		Mensagem	num
Anemômetro N	valor atípico da velocidade mínima	Vmin N atípica	160
	valor atípico da velocidade média	Vmed N atípica	161
	valor atípico da velocidade máxima	Vmax N atípica	162
	valor atípico do desvio padrão da velocidade	Vdp N atípica	163
	valor extremo da velocidade mínima	Vmin N extrema	164
	valor extremo da velocidade média	Vmed N extrema	165
	valor extremo da velocidade máxima	Vmax N extrema	166
	valor extremo do desvio padrão da velocidade	Vdp N extrema	167
	velocidade máxima atípica	Vmax N > Vmed+kmin Vdp	168
	velocidade máxima extrema	Vmax N > Vmed+kmax Vdp	169
	Velocidade média maior que desvio padrão	Vmed N > Vdp N	170
wind vane N	valor atípico da média da direção	Dmed N atípica	180
	valor atípico do desvio padrão da direção	Ddp N atípica	181
	valor extremo da média da direção	Dmed N extrema	182
	valor extremo do desvio padrão da direção	Ddp N extrema	183
Termômetro	Valor atípico da temperatura do ar	Tar atípica	200
Higrômetro	Valor atípico da umidade relativa	Urel atípica	201
Barômetro	Valor atípico da pressão atmosférica	Pat atípica	202
Termômetro	Valor extremo da temperatura do ar	Tar atípica	203
Higrômetro	Valor extremo da umidade relativa	Urel atípica	204
Barômetro	Valor extremo da pressão atmosférica	Pat atípica	205

4. Aplicativo computacional para análise de consistência e exploratória de dados

Este item descreve as funcionalidades e as ferramentas de interação com o usuário de um aplicativo computacional para a análise de consistência e a análise exploratória de dados meteorológicos integrantes do “Sistema de Acompanhamento de Medições Anemométricas – AMA”.

As funcionalidades do aplicativo serão atendidas com as seguintes tarefas:

- Verificar a adequação do arquivo de dados enviado pelo empreendedor ao padrão de formatação do AMA.
- Analisar o arquivo de dados para identificar erros grosseiros.
- Emitir relatório endereçado ao empreendedor com a aceitação ou a rejeição do arquivo de dados. Neste último caso deve ser relatado o motivo da rejeição, o tipo e a posição do erro no arquivo.
- Copiar o arquivo com os dados medidos no campo para um arquivo espelho, codificado em binário.
- Analisar a consistência das séries temporais dos dados meteorológicos para identificar erros sistêmicos ou sistemáticos e aleatórios ou atípicos.
- Gravar os erros no arquivo espelho.
- Apresentar em forma de gráficos e tabelas os erros encontrados.
- Apresentar em forma de gráficos e tabelas os resultados estatísticos da análise exploratória dos dados

Para facilitar a interação do usuário com o Banco de Dados, o aplicativo deve:

- Apresentar em uma tabela os valores máximos e mínimos dos limites físicos das variáveis e dos demais limites utilizados na análise de consistência. Esta tabela pode ser editada e seus dados modificados;
- Apresentar em tabelas os erros encontrados, classificados por tipo. A tabela deve ter o número do registro da variável com erro ou suspeita de erro e o tipo de erro. Ao clicar na tabela, todo o registro da linha do arquivo deve ser apresentado em uma outra tabela com a variável com erro realçada.
- Gráficos com as series temporais discretizadas por dia. Podem ser apresentados os gráficos de diversos dias sequenciais em uma mesma saída. Os gráficos de dias com erros ou suspeitas de erros devem ter fundos com cores diferentes, para facilitar a identificação desses dias pelo usuário. Esses gráficos podem ser ampliados e apresentados em uma tela única e nos dias com erros os mesmos são apresentados em tabelas. Esses gráficos e tabelas servem para auxiliar o usuário na decisão de utilizar ou não um dado suspeito. As variáveis a serem apresentadas nos gráficos podem ser selecionadas pelo usuário.
- Visualização em gráficos e de tabelas com os resultados da análise exploratória dos dados, contendo:

- Distribuições estatísticas ajustadas aos histogramas de frequências das velocidades, com a opção de escolha pelo usuário das distribuições de Rayleigh, Weibull e log-normal:
 - Rosa dos ventos com as frequências da direção do vento
 - Dia típico, com a média em cada hora ao longo do dia
- Cálculo prospectivo da energia a ser gerada, com os resultados apresentados em gráficos e tabelas:
 - Por mês a mês, por estação do ano e global
 - Discretizada por intervalo da direção do vento

A seguir são apresentadas sugestões de saídas em telas de vídeo para o desenvolvimento de um aplicativo computacional.

Tela 1 – Tabela com valores máximos e mínimos de limites de intervalos

Valores limites para análise da consistência dos dados

Consistência univariada			
No.	Variável	Adverti	Erro
3	Velocidade máxima a 0 metros	5	6
4	Direção da velocidade máxima a 0 metros	5	6
5	Velocidade média a 0 metros	5	6
6	Desvio padrão da velocidade média a 0 metros	6	8
7	Média vetorial da direção a 0 metros	5	6
8	Desvio padrão da direção a 0 metros	8	11
9	Média do cubo da velocidade a 0 metros	8	10
10	Velocidade máxima a 0 metros	5	6
11	Direção da velocidade máxima a 0 metros	5	6
12	Velocidade média a 0 metros	5	6
13	Desvio padrão da velocidade média a 0 metros	6	8
14	Média vetorial da direção a 0 metros	5	6
15	Desvio padrão da direção a 0 metros	8	11
16	Média do cubo da velocidade a 0 metros	8	10
17	Velocidade máxima a 0 metros	5	6
18	Velocidade média a 0 metros	5	6
19	Desvio padrão da velocidade média a 0 metros	6	8
20	Média do cubo da velocidade a 0 metros	8	10
21	Temperatura do ar (C) - média	4	5
22	Umidade relativa (%) - média	6	8
23	Pressão atmosférica (mb) - média	4	5
24	Tensão da bateria (V)	4	5

Consistência multivariada - mesmo nível		
Variável	Mínimo	Máximo
Vmax/Vmed	1	3
Vmax/Vmed+K*Vdp	4	7
Vcub/Vmed	0,8	2
Tensão mínima bateria	12	11

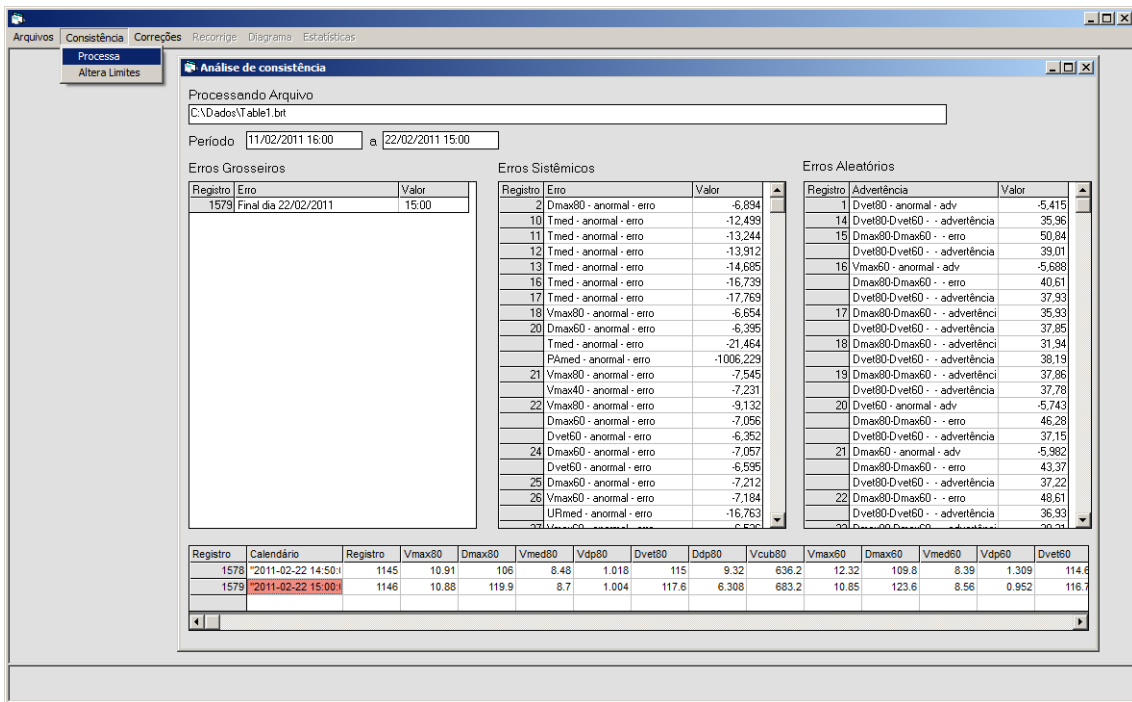
Consistência multivariada-níveis diferentes		
Variável	Mínimo	Máximo
Vmax80/Vmax60	0,8	2
Vmed80/Vmed60	0,8	2
Vcub80/Vcub60	0,7	2,5
Vmax60/Vmax40	0,8	2
Vmed60/Vmed40	0,8	2
Vcub60/Vcub40	0,7	2,5
Dmax80-Dmax60	30	40
Dvet80-Dvet60	30	40

Velocidade mínima para cruzamento mesmo nível

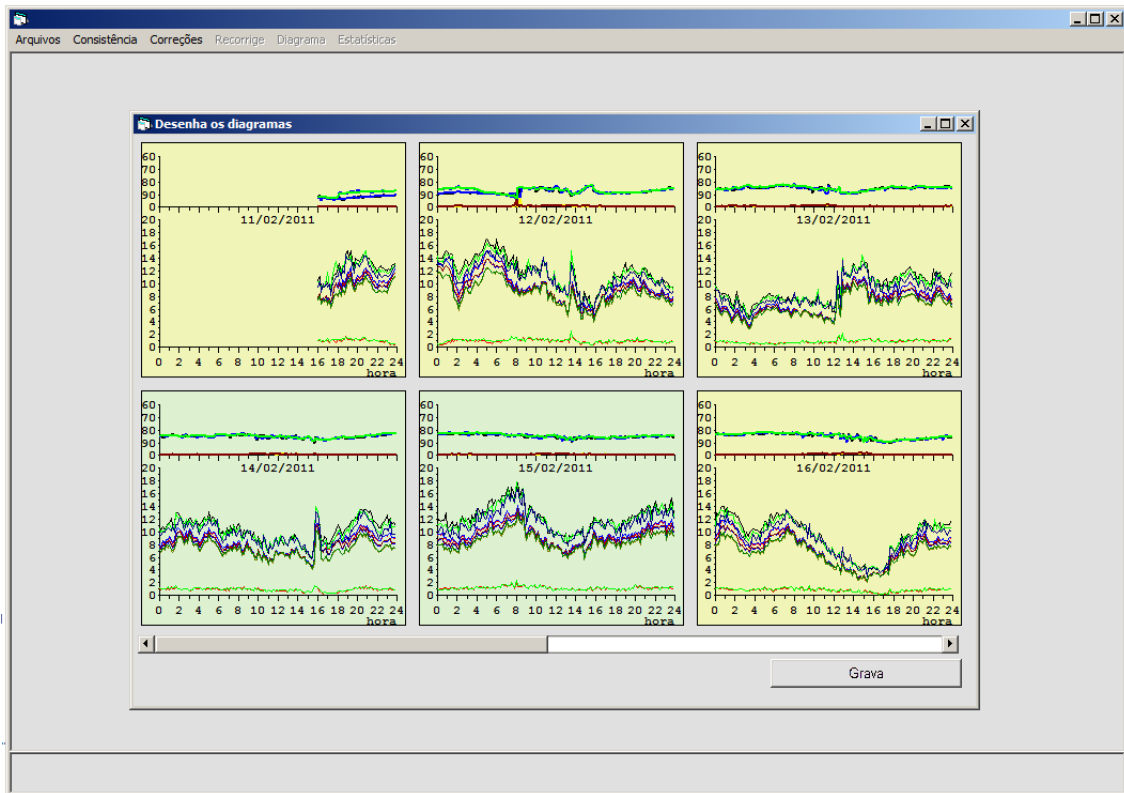
Velocidade mínima para cruzamento níveis diferentes

Valores padrão Grava Cancela

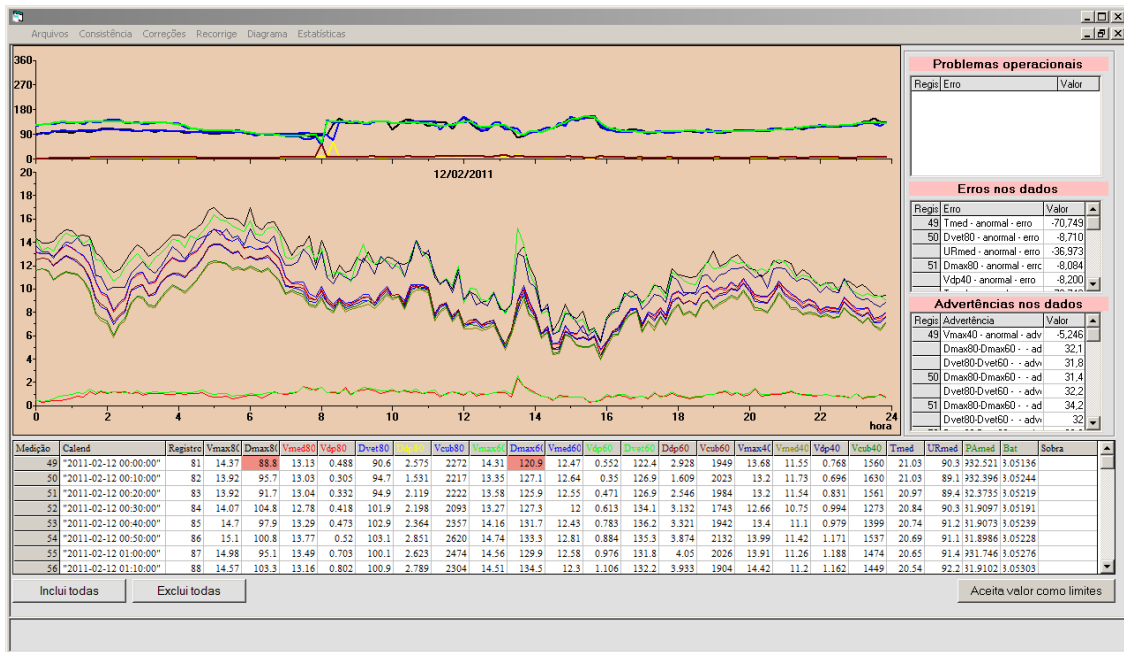
Tela 2 – Tabela com erros classificados por tipo



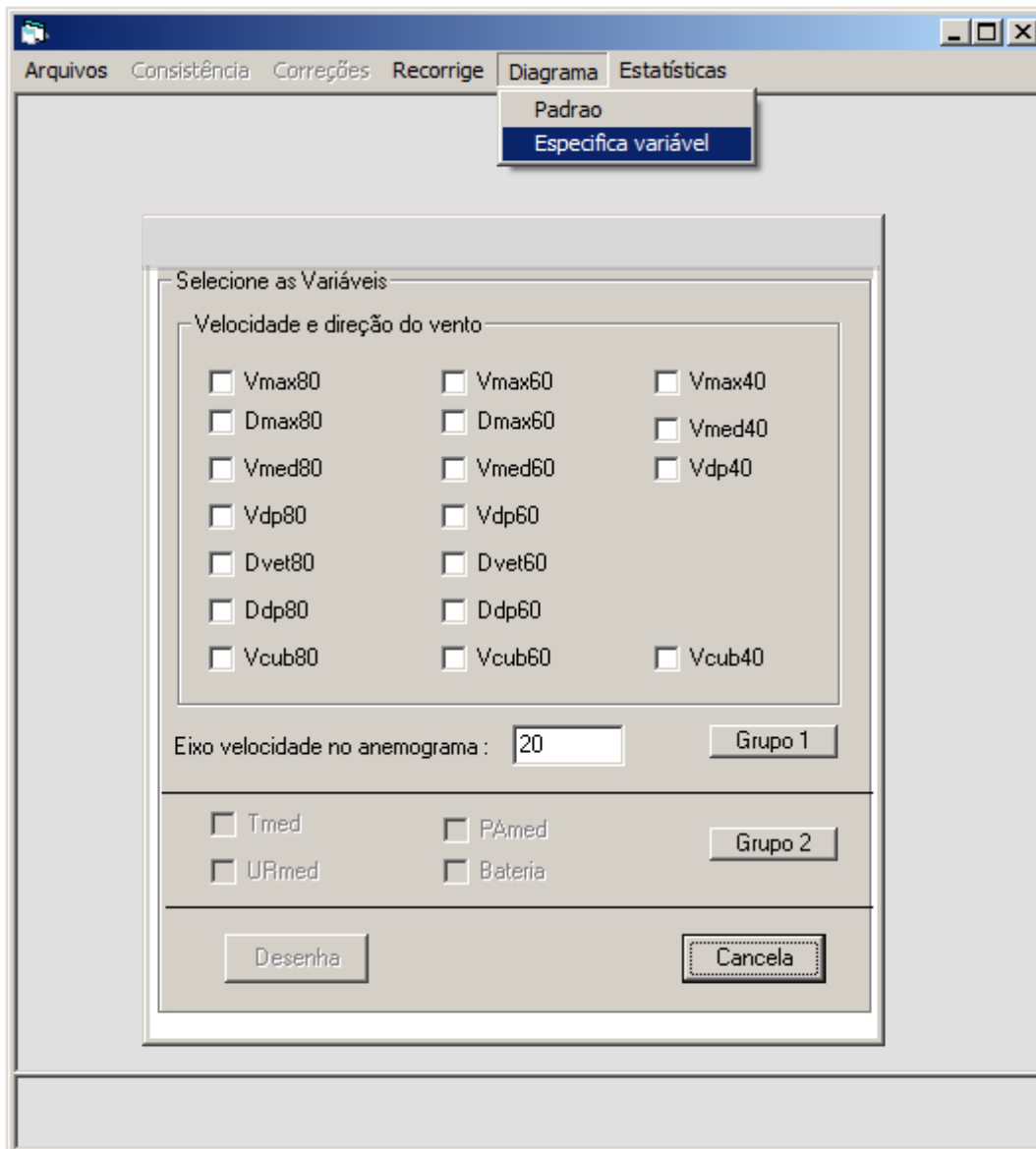
Tela 3 – Gráficos com a variabilidade diária dos dados em vários dias sequenciais



Tela 4 – Gráficos com a variabilidade dos dados em 1 dia e respectiva tabela de erros



Tela 5 – Quadro para a seleção das variáveis a serem consideradas



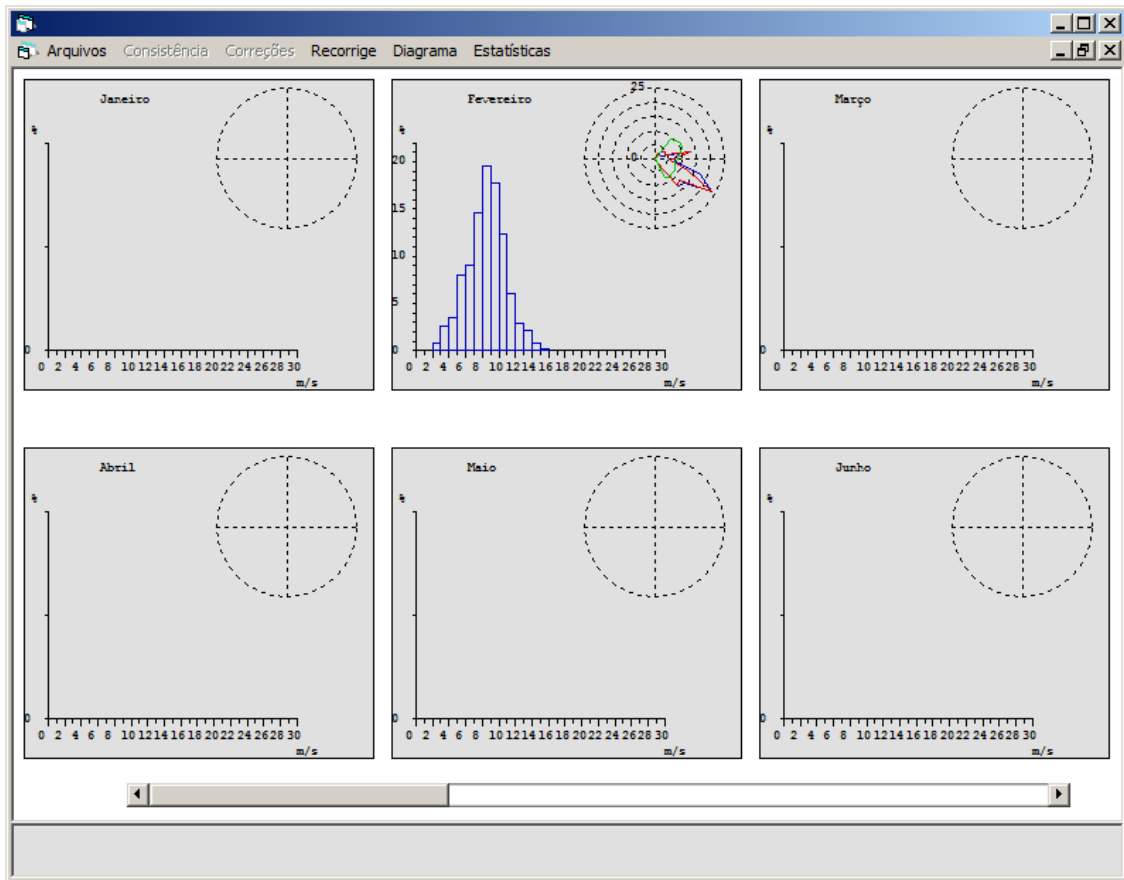
Tela 6 – Tela para seleção da distribuição estatística

The image shows a software window titled "Eólica Engenharia" with a menu bar containing "Arquivos", "Consistência", "Correções", "Recorre", "Diagrama", and "Estatísticas". The "Estatísticas" menu is open, showing "Histogramas" and "Dia médio". The main window has a title bar "Selecione as variáveis e o período" and contains several sections:

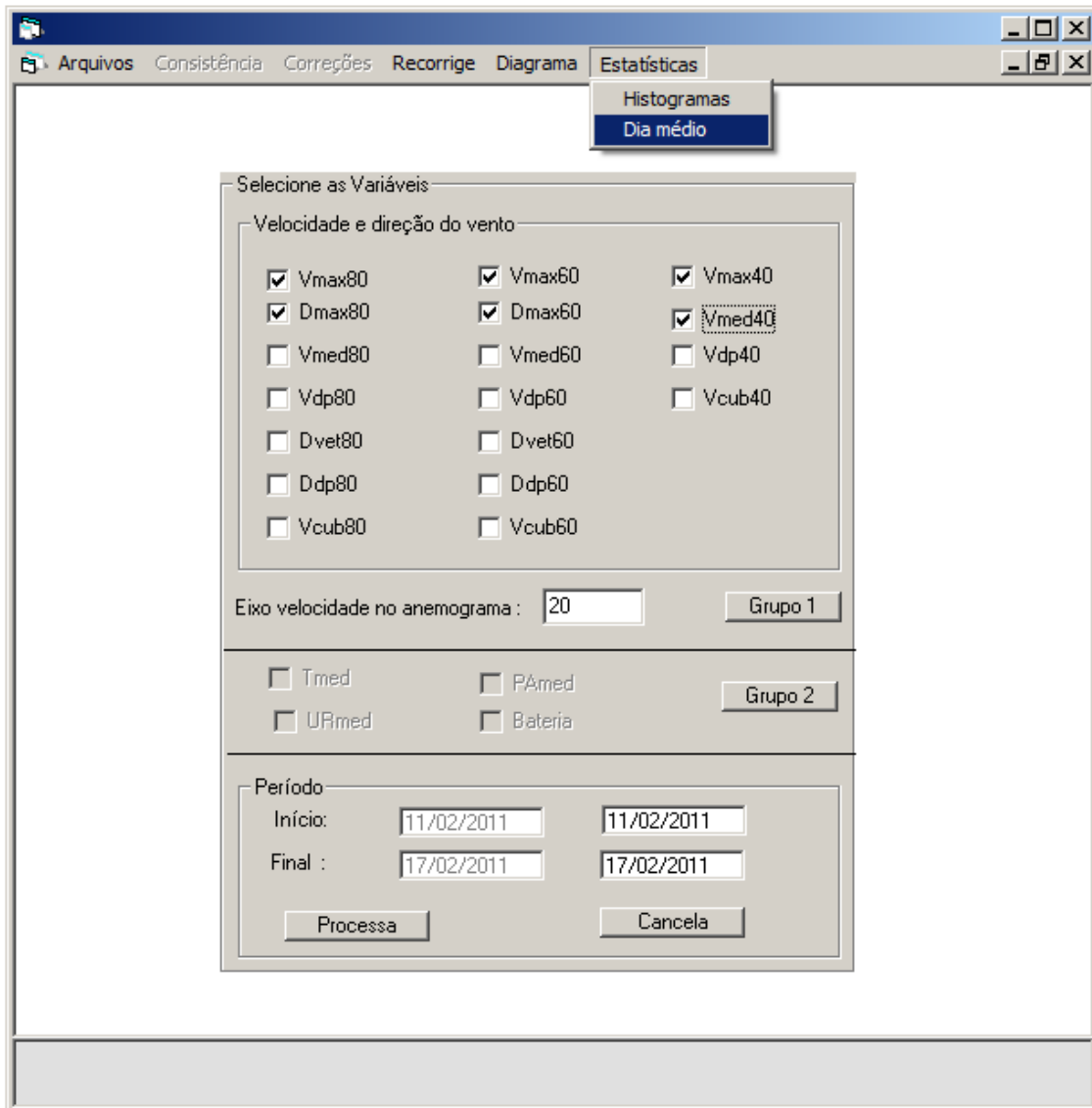
- Altura de medição:** Radio buttons for "80 metros" (selected), "60 metros", and "40 metros".
- Variável:** Radio buttons for "Vmed / Dvet" (selected), "Vcub", and "Vmax / Dmax".
- Declinação magnética:** A text box containing "-23,95".
- Período:** Two pairs of date pickers. "Início:" shows "11/02/2011" and "11/02/2011". "Final:" shows "17/02/2011" and "17/02/2011".
- Histograma velocidade do vento:** "Número de classes" is a text box with "30". "Intervalo das classes" is a text box with "1" followed by "m/s".
- Rosa dos ventos da direção:** "Intervalo das classes" is a text box with "10" followed by "graus".
- Distribuição estatística:** Three checkboxes: "Weibull", "Rayleigh", and "Log-Normal", all of which are currently unchecked.

At the bottom of the window are two buttons: "Processa" and "Cancela".

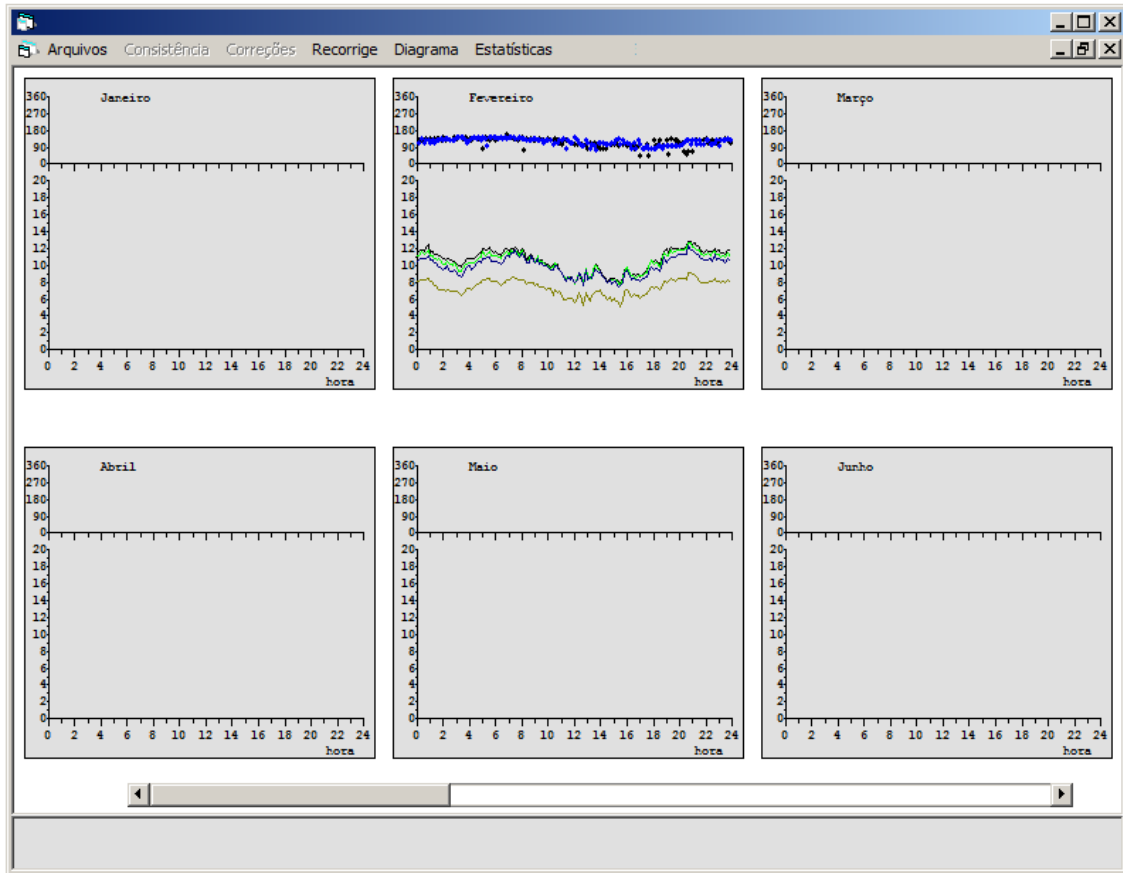
Tela 7 – Histogramas e rosas dos ventos por mes



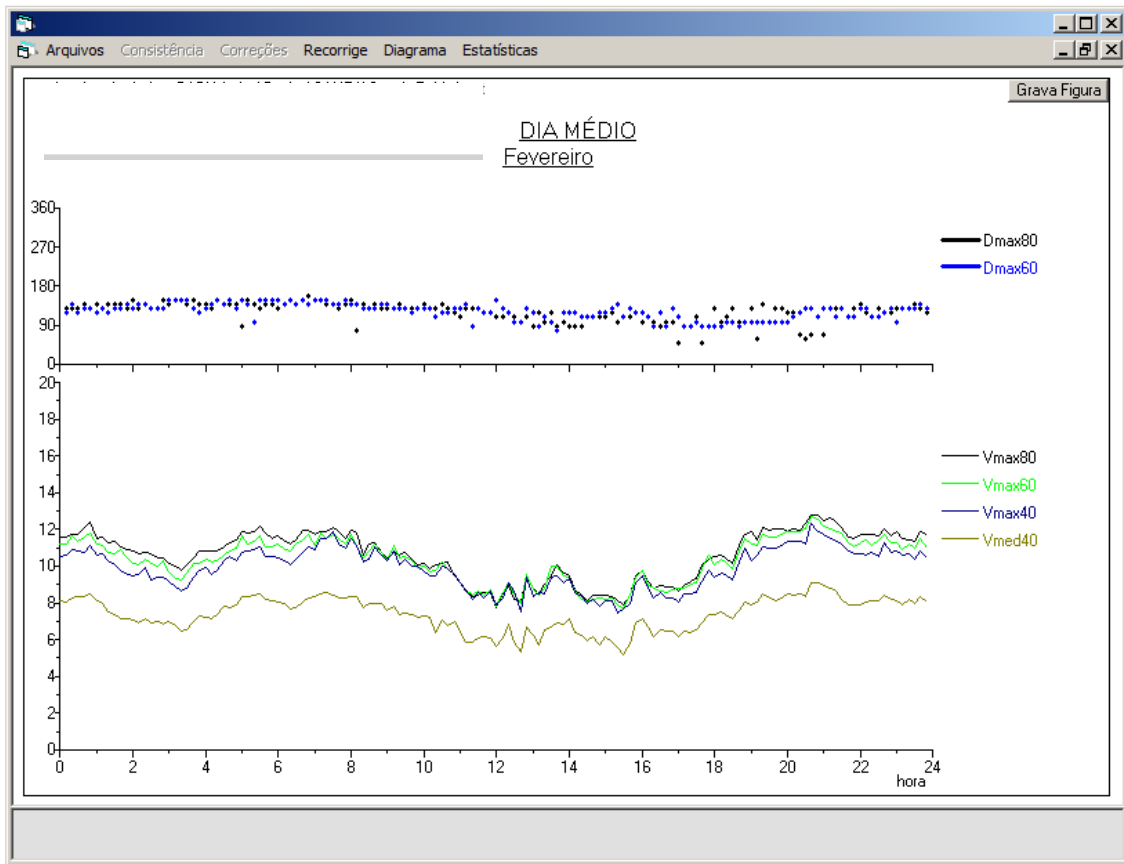
Tela 7 - Quadro para selecionar as variáveis para compor o dia médio



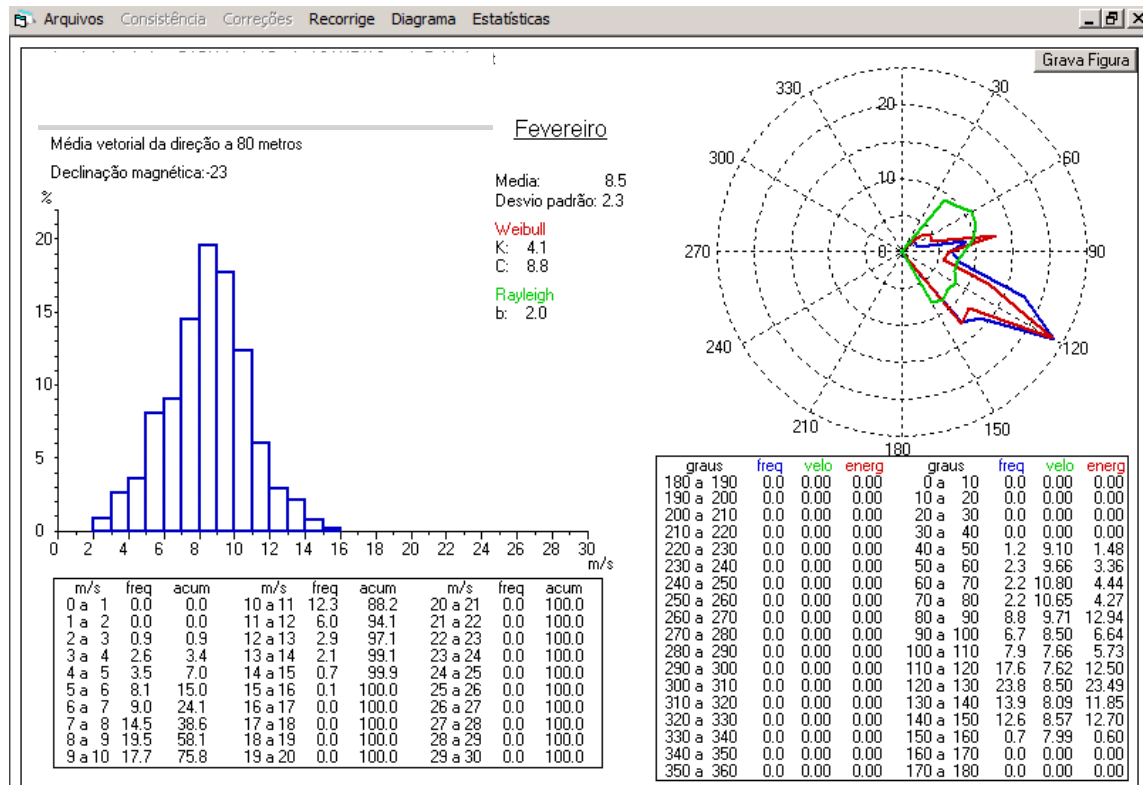
Tela 8 – Dia médio de cada mês



Tela 9 – Dia médio de um mês - Ampliada



Tela 10 – Energia prospectiva a ser gerada



Anexo I: Códigos de erros para os arquivos AMA

Neste anexo está descrita a formatação do código elaborado pela EPE utilizado pelo AMA para identificar se a medição de uma dada variável está com erro.

No canal 03 de cada linha de arquivos com o formato do AMA é informado um código para indicar se ocorreu erro na medição em cada um dos sensores.

No cálculo do potencial eólico os dados da temperatura do ar, da umidade relativa do ar e da pressão atmosférica são usados simultaneamente. A ocorrência de erro na medição de qualquer uma dessas medições inviabiliza que o calculo seja processado. Portanto, o erro nessas três variáveis é considerado como sendo 1.

Código de Erros na Base de Dados

	Cod Erro
P, T ou U	1
Anem S	2
Wins S	4
Anem I	8
Wind I	16
Anem 3	32
Linha	999

Código de Erros Possíveis

{1+1}	P, T ou U	Anem S	Wind S	Anem I	Wind I	Anem 3
P, T, ou U	1	3	5	9	17	33
Anem S	3	-	6	10	18	34
Wind S	5	6	-	12	20	36
Anem I	9	10	12	-	24	40
Wind I	17	18	20	24	-	48
Anem 3	33	34	36	40	48	-
{2+1}	P, T ou U	Anem S	Wind S	Anem I	Wind I	Anem 3
PTU + An S	3		7	11	19	35
An S + WV S	7	-		14	22	38
An I + WV I	25	26	28	-		56

Sendo:

P – Pressão

T – Temperatura

U – Umidade

Anem ou An – Anemômetro

WV – Wind Vane

I – Inferior

S - Superior

Determinação dos Instrumentos com Defeito

a) Se CodErro é ímpar P, T, U e pelo menos mais um instrumento

Calcular $\text{CodErro}-1$ Exemplo: $\text{CodErro}=29$ Erro = 1 + 28 1 Erro em P, T ou U

Subtrair o maior código tal que resulte em no. ≥ 0 CodErro	$28-16=12$	16 Erro no WV I
Subtrair o maior código tal que resulte em no. ≥ 0 CodErro	$12-6=4$	8 Erro no Anem I
Subtrair o maior código tal que resulte em no. ≥ 0 CodErro	$4-4=0$	4 Erro no WV S

b) Se CodErro é par P, T, U estão corretos

Exemplo: $\text{CodErro}=46$

Subtrair o maior código tal que resulte em no. ≥ 0 CodErro	$46-32=14$	32 Erro no Anem 3
Subtrair o maior código tal que resulte em no. ≥ 0 CodErro	$14-8=6$	8 Erro no Anem I
Subtrair o maior código tal que resulte em no. ≥ 0 CodErro	$6-4=2$	4 Erro no WV S
Subtrair o maior código tal que resulte em no. ≥ 0 CodErro	$2-2=0$	2 Erro no Anem S

Exemplo: $\text{CodErro}=56$

Subtrair o maior código tal que resulte em no. ≥ 0 CodErro	$56-32=24$	32 Erro no Anem 3
Subtrair o maior código tal que resulte em no. ≥ 0 CodErro	$24-16=8$	16 Erro no WV I
Subtrair o maior código tal que resulte em no. ≥ 0 CodErro	$8-8=0$	8 Erro no Anem I

Anexo II: Arquivo de dados consistidos

O arquivo com os dados consistidos é escrito em formato binário, com uma estrutura formada por dois blocos.

No primeiro, chamado de Cabeçalho, estão as informações gerais da estação e os limites dos intervalos dos testes de consistência para cada variável meteorológica.

O segundo bloco contém, em cada linha ou registro do arquivo, os dados medidos e os códigos indicando o resultado da análise de consistência de cada uma das variáveis. Se não foi encontrado erro, o código é igual a zero, caso contrário é gravado o código numérico do erro encontrado.

Os itens que seguem exemplificam a estrutura do arquivo e ao mesmo tempo sugerem como seria a sua escrita em uma linguagem computacional, no caso Visual Basic.

II.1. Cabeçalho

Neste bloco são descritos:

- o numérico da estação, com 6 algarismos;
- os calendários do início e do final do arquivo com o formato AAAAMMDD e hhmms;
- o código numérico que identifica a falha ou falta de dado de uma variável. Sugere-se o valor de 6999, pois fisicamente nenhuma das variáveis medidas pode ter esse valor;
- o número de anemômetros e as suas alturas na torre, lidos do arquivo de cadastro da estação no AMA;
- o número de *wind vane* e as suas alturas na torre, lidos do AMA;
- os valores dos limites físicos, máximos e mínimos, das variáveis e dos parâmetros medidos;
- os valores máximos e mínimos das relações entre os parâmetros estatísticos de cada variável e de variáveis correlatas medidas em cada uma das alturas e entre as diferentes alturas.

O aplicativo computacional deve ter a opção de sugerir ao usuário os valores iniciais para os máximos e mínimos. Com o aumento do período de coleta estes valores podem ser modificados pelo usuário, para melhor adequação às características do histórico dos dados medidos

Na Figura II.1 está apresentada uma sugestão da estrutura deste bloco.

```

Public Type TipoCalendario
  AAAAMMDD      As Long
  hhmmss       As Long
End Type
Public Type TipoLimiteFisico
  minimo      As Single
  maximo      As Single
End Type
Public Type TipoLimiteRelacao
  minimo      As Single
  maximo      As Single
End Type
Public Type TipoCabecalho
  numeroEstacao      As Single
  calendInicio       As Single
  calendFinal        As Single
  codFalha           As Single      ' código de falta de dado 6999
  nAnemo             As Byte       ' número de anemómetros
  nVane              As Byte       ' número de wind vane
  ' ERROS SISTEMICOS ou SISTEMÁTICOS
  termometro         As TipoLimiteFisico
  higrometro         As TipoLimiteFisico
  barometro          As TipoLimiteFisico
  anemometro(1 To nAnemo) As TipoLimiteFisico
  windvane(1 To nAnemo) As TipoLimiteFisico
  VminN_VminK       As TipoLimiteRelacao ' relação entre velocidades mínimas
  ' anemometro superior e inferior
  VmedN_VmedK       As TipoLimiteRelacao ' idem velocidades médias
  VmaxN_VmaxK       As TipoLimiteRelacao ' idem velocidades máximas
  DmedN_DmedK       As Byte          ' relação entre direções médias
  ' wind vane superior e inferior
  ' ERROS ALEATÓRIOS
  atipico(1 To nVariaveis) As TipoLimiteRelacao ' número de desvios padrão do intervalo
  ' de valores atipicos
  ' se for utilizado o Box-Plot são os múltiplicadores
  ' da diferença (L = Q3 - Q1)
  ' nVariaveis é o número de variáveis de cada registro
  ' valor do coeficiente k da relação Vmax = Vmed + k Vdp
  Vmed_k_Vdp        As TipoLimiteRelacao
End Type

```

Figura II.1 Estrutura do cabeçalho de dados no arquivo consistido

II.2. Registro

Este bloco é repetido tantas vezes quantas forem os intervalos de 10 minutos entre os calendários do início e do final do arquivo., que constam do bloco Cabeçalho. Como todos os valores do registro são numéricos, o código “-“ (hifen) que indica falta de dado deverá ser substituído pelo código numérico especificado no bloco Cabeçalho.

Uma sugestão da sua estrutura na linguagem Visual Basic está apresentada na Figura II.2.

```

Public Type TipoVelocidade
    maxima           As Single
    media            As Single
    minima           As Single
    dPadrao          As Single
    codErro(1 To 4) As Byte
End Type

Public Type TipoDirecao
    media            As Integer
    dPadrao          As Integer
    codErro(1 To 2) As Byte
End Type

Public Type TipoMeteoro
    termometro       As Single
    higrometro       As Single
    barometro        As Single
    codErro(1 To 3) As Byte
End Type

Public Type TipoCalendario
    AAAAMMDD        As Long
    hhmmss          As Long
End Type

Public Type TipoRegistroArquivo
    calendario      As TipoCalendario
    codErroAma      As Integer
    meteoro         As TipoMeteoro
    anemometro(1 To nAnemo) As TipoVelocidade
    windvane(1 To nVane) As TipoDirecao
End Type

Global registro As TipoRegistroArquivo
    
```

Figura II.2 Estrutura do registro de dados no arquivo consistido