# **Verificação do Sistema de Previsão por Conjunto Global (oensMB09)**

Grupo de Avaliação de Modelos - GAM

Divisão de Modelagem Numérica do Sistema Terrestre - DIMNT

Carlos Frederico Bastarz

## **Atributos das Previsões Numéricas**

Previsões numéricas possuem atributos importantes que devem ser verificados para mensurar a sua habilidade, acurácia e precisão e corrigir os seus erros sistemáticos. Dentre estes atributos há aqueles que estão relacionados com o aspecto determinístico das previsões e outros que estão relacionados com o aspecto probabilístico de um conjunto de previsões. Segundo Hudson (2017), a verificação das previsões numéricas pode ter diferentes motivações (e.g., aplicações operacionais e desenvolvimentos) e as previsões numéricas por conjunto devem ser avaliadas de forma que sejam considerados os seguintes atributos quanto à sua qualidade e valor:

* **Acurácia** (*accuracy*): a acurácia é a correspondência ou nível de concordância/aderência entre as previsões e as observações (i.e., referência);
* **Habilidade** (*skill*): a habilidade é uma quantidade que nos permite afirmar que uma previsão numérica pode ser mais habilidosa ("ter mais skill") do que uma referência (e.g., persistência, climatologia). Para esse tipo de verificação, pode-se calcular um *score*, e.g., ;
* **Confiabilidade** (*reliability*): pode-se confiar na chance de chuva calculada (e.g., 80%, i.e., probabilidade = 0,8)? A confiabilidade é um atributo relacionado com estimativas dicotômicas (e.g., determinação de eventos de chuva/não-chuva) das previsões numéricas e que pode ser melhorado por meio de calibração;
* **Discriminação e resolução** (*discrimination and resolution*): a resolução é um atributo que está relacionado com a sensibilidade com a qual um resultado da previsão muda (e.g., devido à uma condição inicial diferente) quando a previsão numérica muda (e.g., considerando um conjunto de previsões numéricas, a chance de precipitação sobre uma área é de 80%; mas se estas previsões numéricas mudarem, qual é o impacto na chance de precipitação?). Por outro lado, a discriminação trata da verificação individual dos membros dentro do mesmo contexto: considerando um conjunto de previsões numéricas, a chance de precipitação sobre uma área é de 80%, mas se uma destas previsões numéricas muda, qual é o impacto na chance de precipitação devido à esta previsão numérica?). Estes dois atributos indicam o uso potencial das previsões e não podem ser melhorados por calibração. De outra forma, pode-se entender a discriminação como sendo o discernimento entre as médias dos eventos observados e não observados (i.e., a previsão de chuva que não foi observada), considerando-se desvios-padrão semelhantes. Se estas médias forem muito próximas, então este pode ser um indicativo de que o conjunto de previsões tem a mesma capacidade de prever eventos observados e não observados.
* **Nitidez** (*sharpness*): é a tendência para a previsão de valores extremos (valores de probabilidades próximos a 0 ou 1, ou chances próximas a 0 ou 100%), ao invés da previsão de valores ao redor da média (a previsão de valores climatológicos, não possui *sharpness*). Previsões "nítidas" são úteis, mas quando não são confiáveis, implicam em confiança não realística.

## **Acurácia e Precisão**

Muitos são os termos utilizados para referenciar determinados atributos da verificação de previsões numéricas. Por vezes, torna-se confuso o uso dos termos **acurácia** e **precisão**. A acurácia, como fora mencionado, é um atributo que permite verificar o grau de concordância (ou discrepância) entre os valores representados pelas previsões numéricas e um valor de referência. Nesse sentido, um conjunto de previsões numéricas é mais (menos) acurado quando suas previsões forem mais (menos) próximas dos valores de referência. Por outro lado, um conjunto de previsões numéricas pode ser mais (menos) preciso quando suas previsões possuírem menor (maior) dispersão em relação à sua média. De forma análoga, pode-se também verificar a exatidão de um conjunto de previsões numéricas: estas podem ser mais (menos) exatas quando suas previsões estão menos (mais) distantes de um valor de referência. Para exemplificar estes conceitos, considere a Figura 1 a seguir:

|  |  |  |  |
| --- | --- | --- | --- |
| a) | b) | c) | d) |

Figura 1 - Precisão e Acurácia como medidas dos erros aleatório e sistemático, respectivamente. Adaptado de <https://blog.idwall.co/o-que-e-acuracia/>.

Na Figura 1, considere um arqueiro que dispara suas flechas contra um alvo. Na primeira tentativa (Figura 1a), o arqueiro acerta com acurácia e precisão o centro do alvo (denotado pelo círculo vermelho e as flechas pelos pontos em verde). Na segunda tentativa (Figura 1b), o disparo das flechas ainda é acurado, porém menos preciso: o arqueiro ainda consegue acertar o centro do alvo, mas o desvio-padrão dos seus disparos é maior (por isso ele é menos acurado do que a primeira tentativa). Na terceira tentativa (Figura 1c), a sua precisão é muito menor em relação às duas primeiras tentativas: os seus disparos estão longe do centro do alvo, mas a sua acurácia é maior do que a segunda tentativa, pois o desvio-padrão dos seus disparos é menor. Na última tentativa (Figura 1d), os seus disparos não são nem acurados e nem precisos: ele erra o centro do alvo e os seus disparos possuem maior desvio-padrão.

Analogamente a um modelo numérico, podemos verificar que as suas previsões podem conter "bias" (i.e., viés) e erros sistemáticos. No contexto do exemplo apresentado, o viés é uma medida do erro aleatório do modelo e está relacionado com a precisão com a qual as previsões numéricas representam uma determinada quantidade física (e.g., a temperatura, umidade, pressão, precipitação etc). O erro sistemático, por outro lado, está relacionado com a acurácia com a qual um modelo numérico representa estas quantidades físicas. Nesse exemplo, o viés de um modelo pode ser corrigido depois que as previsões são pós-processadas, mas o seu erro sistemático não - nesse caso, é necessário corrigir o modelo para melhorar a sua acurácia.

Pode-se entender também a exatidão como uma medida a ser verificada quando é possível mensurar a sua distância ou discrepância em relação ao valor observado (ou referência). Essa medida é diferente dos dois primeiros momentos estatísticos (média e desvio-padrão) de uma distribuição de probabilidades. Para um conjunto de previsões numéricas, a exatidão pode ser calculada como a discrepância entre a média do conjunto de previsões e o valor de referência. Em geral, na verificação de previsões numéricas, utiliza-se esta diferença para o cálculo de outras quantidades, como a Viés (*Bias*), a Raiz do Erro Quadrático Médio (*Root Mean Square Error* - RMSE), o Coeficiente de Correlação de Anomalias (*Anomaly Correlation Coefficient* - ACC) entre outros.

## Método de Perturbação baseado em EOFs (oensMB09)

O Sistema de Previsão por Conjunto (SPCON) Global do CPTEC é um sistema composto pelo modelo de circulação geral da atmosfera desenvolvido pelo CPTEC e um módulo de perturbação da condição inicial baseado em Funções Ortogonais Empíricas (EOF, em inglês). O SPCON global tem sido realizado operacionalmente no CPTEC desde 1999, e seu objetivo principal são as previsões numéricas de tempo para até 15 dias. Para isso, o SPCON global é realizado duas vezes ao dia (às 00 e 12Z), na resolução TQ0126L028 (~100km de resolução horizontal e 28 níveis verticais em coordenada sigma) com 15 membros, sendo 1 membro controle (realizado a partir de uma análise determinística) e mais 14 membros (realizados a partir de perturbações somadas e subtraídas à análise controle). O limite da previsão útil do SPCON global do CPTEC é de 14 (11) dias para a previsão da temperatura do ar ao nível de 850 hPa às 00Z (12Z) e 12 (14) dias para a previsão da pressão em superfície às 00Z (12Z), sobre o Hemisfério Sul (Cunningham et al., 2015).

Nas seções a seguir, são apresentadas as ferramentas disponíveis no CPTEC para a avaliação do oensMB09.

## Ferramentas de Verificação para o oensMB09

No CPTEC, as ferramentas de avaliação para o oensMB09 são baseadas em ferramentas desenvolvidas pelo NCEP. No repositório do GAM no GitHub, no *branch* CRPS (disponível no endereço <https://github.com/GAM-DIMNT-CPTEC/oensMB09-AVAL>) está disponível o código para o cálculo do *Continuous Ranked Probability Score* (CRPS). Apesar disso, nem todas as ferramentas que estão disponíveis estão habilitadas para uso. Até 2017, apenas o CRPS estava sendo utilizado como métrica para a verificação do oensMB09 dentro do período de uma estação.

### Continuous Ranked Probability Score (CRPS)

Considere as Funções Densidade de Probabilidade (*Probability Density Function* - PDF) representadas na Figura 2 a seguir. Estas PDFs representam as distribuições de probabilidades de uma determinada quantidade física (e.g., a temperatura) em termos dos seus dois primeiros momentos estatísticos (i.e., a média e o desvio-padrão). As suas respectivas Distribuições Cumulativas de Probabilidades (*Cumulative Distribution Funcion* – CDF, e.g., a curva azul da Figura 3), por outro lado, medem a probabilidade de ocorrência dessa quantidade (e.g., temperatura de 30o C). Matematicamente, a PDF representa a curva e a CDF a área sob a curva, sendo 1 a sua área total. Para uma PDF contínua, a probabilidade é a área sob a curva à esquerda do ponto de interesse.

Na Figura 2, as PDFs representando a representação da frequência de previsões de uma quantidade física. Nas figuras, a área em cinza representa apenas a interseção entre as previsões de eventos observados e não observados de acordo com a proximidade da média das distribuições.

|  |  |  |
| --- | --- | --- |
| a) | b) | c) |

Figura 2 – Em “a”, as duas PDFs representam distribuições distintas quanto à previsão de eventos observados e não observados, as previsões são nítidas e é fácil distinguir entre esses os eventos. Em “b”, as PDFs representam distribuições de previsões pouco nítidas e os eventos observados e não observados nas previsões tem a mesma chance de ocorrer. Em “c”, as PDFs representam distribuições de previsões nítidas. Em termos de acurácia e precisão, se o valor de referência estivesse no meio, entre as duas curvas, em “a”, as distribuições seriam nem acuradas e nem precisas; em “b”, as distribuições seriam acuradas, mas não seriam precisas e em “c”, seriam acuradas e precisas.

O CRPS é a medida da discrepância quadrática (uma distância) entre duas CDFs (resolução de um conjunto de previsões numéricas). Para o caso de um conjunto de previsões, o CRPS equivale à Raiz do Erro Quadrático Médio (*Root Mean Square Error* - RMSE) entre a CDF prevista (i.e., do conjunto de previsões) e a CDF de referência (idealmente, das observações) sendo, portanto, uma medida de acurácia (ou incerteza) de um conjunto de previsões. Devido ao fato de que o CRPS é calculado em termos das CDFs das distribuições de probabilidades do conjunto de previsões e das observações, a característica determinística das observações e probabilística do conjunto de previsões é abstraído e, nesse sentido, o CRPS é uma generalização do Erro Médio Absoluto (*Mean Absolute Error* - MAE).

onde,

* é a CFD predita (i.e., do conjunto de previsões);
* é a função Heaviside (função degrau, i.e., 0 ou 1 em diferentes partes do domínio);
* é o valor observado.

|  |
| --- |
|  |

Figura 3 - Interpretação do CRPS: a área em cinza, é a quantificação do CRPS (calculado como a área entre as duas curvas).

Em termos do valor do CRPS, quanto mais próximo de zero, mais próximo o conjunto de previsões verificado está da referência (basta lembrar que o CRPS mede a discrepância quadrática entre duas CDFs).

#### Cálculo do CRPS para o oensMB09

Para o cálculo do CRPS para o oensMB09, considerando o período de uma estação (i.e., 3 meses), são realizadas as seguintes etapas:

1. Realização da suíte de previsões por conjunto global (método de perturbação oensMB09 e modelo global) para o período completo, considerando as análises das 00 e 12Z, com previsões para 360 horas (15 dias) a cada 6 horas (nesta etapa, já se considera também o pós-processamento das previsões);
2. Verificação por falhas nos arquivos pós-processados (falhas durante as submissões podem ocorrer - na situação operacional, esta verificação é feita ao final de cada submissão da suíte);
3. Recorte dos dados a serem avaliados. Nesta etapa, são selecionadas as variáveis que serão utilizadas na verificação: temperatura do ar em 850 hPa, altura geopotencial em 500 hPa e pressão reduzida ao nível médio do mar;
4. Submissão do cálculo do CRPS:
5. Leitura de todos os arquivos de previsões para os domínios de interesse (Hemisférios Norte e Sul e região Tropical);
6. Leitura de todos os arquivos de referência: conforme mostrado em Cunningham et al. (2015), os dados de referência são representados pelos dados de reanálise do ERA Interim;
7. Cálculo do CRPS e escrita dos resultados em arquivos binários e “txt” com o nome, e.g., “CRPS4CPTECEPS.24ForecastFor2020083100.txt”. Ambos os arquivos, i.e., binário e “txt” possuem as mesmas informações e o arquivo binário (acompanhado do seu respectivo arquivo `ctl`, é utilizado para a confecção de figuras com o GrADS). O nome dos arquivos com os valores do CRPS, trazem as seguintes informações:
   * Prefixo: “CRPS4CPTECEPS.”;
   * Tempo da previsão: “24” para 24 horas (são consideradas todas as previsões no intervalo entre 24 e 360 horas);
   * Data de referência (“ForecastFor”): “2020083100”, ou seja, a previsão de 24 horas válida para a data “2020083100”.

O conteúdo do arquivo apresenta dois valores (exemplo válido para a altura geopotencial em 500 hPa): “5.0877 57.2291”

Neste exemplo, o primeiro valor representa o CRPS calculado da previsão em relação à referência (i.e., a discrepância entre a CDF da previsão e a CDF da referência) e o segundo valor representa o CRPS calculado da climatologia em relação à referência. Os dois CRPSs são calculados porque os resultados são depois apresentados como um *score* na forma do *Continuous Probability Skill Score* (CRPSS), da seguinte forma:

Produção das figuras a partir dos cálculos realizados:

|  |  |
| --- | --- |
| a) | b) |

Figura 4 - Exemplo do CRPSS calculado para dois conjuntos de previsões (XC50 e Tupã). Válido para a pressão reduzida ao nível médio do mar sobre o hemisfério norte (a) e sobre o hemisfério sul (b), para o período entre 2020061500 e 2020083100.

Na Figura 4, CRPSS da pressão reduzida ao nível médio do mar sobre o hemisfério norte, verificado durante o período de 2020061500 a 2020083100, indica que a validade das previsões do experimento Tupã é de aproximadamente 6 dias, enquanto que a validade das previsões do experimento XC50 é de aproximadamente 5 dias. Na Figura 4, a linha sólida preta em indica o limite útil da previsão, no sentido do *score* calculado, i.e., à medida em que o tempo de previsão aumenta (previsões de 24, 48, 72 até 360 horas), o valor do CRPS da previsão e da climatologia tendem a diminuir pois a distância entre as CDFs da previsão e da referência (reanálise do ERA Interim) e da climatologia e da referência, tendem a aumentar. Como o CRPS da previsão é maior do que o CRPS da climatologia, logo a curva representada pelo CRPSS tende a ser negativa (no sentido de diminuição dos valores), pois a razão tende a aumentar e se torna maior do que a unidade.

### Rank Histogram (RH)

O *Rank Histogram* (RH), também conhecido como Histograma de Talagrand, é um tipo de histograma de categorias (*ranks*) que permite verificar a acurácia de um conjunto de previsões (confiabilidade de um conjunto de previsões numéricas). Na prática, a depender da forma característica do histograma calculado, é possível verificar se o espalhamento do conjunto de previsões é superestimado ou subestimado: um histograma em forma de "n", indica que o conjunto de previsões numéricas é superdispersivo (i.e., o espalhamento superestimado onde há excesso variabilidade no conjunto, com maior frequência nas categorias intermediárias), enquanto que um histograma em forma de "u" indicam que o conjunto de previsões numéricas é subdispersivo (i.e., o espalhamento é subestimado onde há falta de variabilidade no conjunto, com maior frequência nas categorias das extremidades). Por outro lado, um histograma "reto", em que a altura das barras está alinhada, indica um conjunto de previsões numéricas acurado. Segundo Hamil (2001), esta interpretação está sujeita à confiabilidade da referência utilizada para o cálculo do RH, i.e., se elas são realmente confiáveis e independentes (e.g., não são contaminadas por erros).

O cálculo do RH para o oensMB09, segue os mesmos procedimentos apresentados para o cálculo do CRPS em termos da organização dos dados, com a diferença de que uma outra rotina é chamada a partir do programa principal do CRPS, além do que outros *scripts* de plotagem são utilizados.

Para a construção do RH, faz-se a suposição de que a observação (i.e., a referência) é estatisticamente indistinguível dos membros do conjunto de previsões numéricas. Levando-se em consideração esta suposição, os valores das previsões dos membros do conjunto são ordenados para formar as categorias, e então, os valores da referência são testados em cada categoria. Como são consideradas previsões e observações colhidas ao longo de um período, então, idealmente, a probabilidade de que as observações possam estar em qualquer uma das categorias, é equiprovável. Com isso, para cada observação testada dentro de cada categoria, atribui-se o valor 0 ou 1, para o caso de a referência estar de dentro ou foram da categoria definida de acordo com as previsões numéricas do conjunto.

Cada intervalo do RH representa uma categoria (de tamanho ) em que cada membro do conjunto (de tamanho ) é testado contra a referência. Nesse sentido, os membros são antes ordenados de forma crescente dentro da categoria, de forma que o número de intervalos corresponde ao tamanho do conjunto de previsões numéricas mais um (). Se um conjunto de previsões numéricas é acurado e preciso, então uma referência é indistinguível dos membros do conjunto e tem a mesma probabilidade de estar em qualquer categoria (equiprovável). Dessa forma, calcular o RH implica em contar quantas vezes uma referência está dentro de cada intervalo do histograma. No final cálculo, uma com colunas (16 colunas, para o caso do oensMB09 com 15 membros) será escrita e, para cada coluna, haverá 11.520 valores com os quais será calculada a frequência relativa. O que é plotado no RH, são as frequências relativas em cada coluna.

|  |  |
| --- | --- |
| a) | b) |
| c) | d) |

Figura 5 - Exemplos de RH para as previsões de 24, 120, 240 e 230 horas (1, 5, 10 e 15 dias).

Na Figura 5, são apresentados RHs calculados a partir dos conjuntos de previsões de 24 horas dos experimentos oensMCGA, oensMB09 e oensMB09\_mcgav4.0 (oensMB09 realizado com diferentes versões do modelo global do CPTEC). As figuras referem-se à variável Temperatura do Ar em 850 hPa, calculada para o mês de janeiro de 2015 e são válidas para o horário das 00Z. Além dos histogramas, são apresentados também os *boxplots* associados a estes histogramas com o objetivo de auxiliar no diagnóstico do espalhamento dos conjuntos de previsões de cada experimento. Os RHs são calculados para cada uma das datas consideradas no período de avaliação. Portanto, se um período de 3 meses (como na avaliação dos experimentos oensMCGA, oensMB09 e oensMB09\_mcgav4.0) é considerado e sendo dois horários por dia (às 00Z e 12Z), são produzidos 240 histogramas para cada variável considerada. Para facilitar a avaliação foram então produzidos histogramas médios nos quais barras foram adicionadas indicando as variações nas frequências dos histogramas. Nesse sentido, os *boxplots* auxiliam na compreensão da distribuição dos erros da probabilidade do posicionamento das observações em cada um dos intervalos calculados. Além disso, a barra horizontal apresentada nos histogramas e nos *boxplots* indica a média das frequências relativas de cada intervalo de forma que, quando o histograma apresenta uma distribuição de frequência o mais alinhada possível com a barra, o diagnóstico é o melhor possível, i.e., indicando que o espalhamento do conjunto de previsões é adequado e possivelmente sem viés.

|  |  |  |
| --- | --- | --- |
| a) | b) | c) |
| d) | e) | f) |

Figura 6 - Histogramas de Talagrand das previsões de 24 horas da Temperatura do Ar em 850 hPa dos experimentos oensMCGA ("a" e "d"), oensMB09 ("b" e "e"), oensMB09\_mcgav4.0 ("c" e "f"), da esquerda para a direita, respectivamente. Válido para o mês de Janeiro de 2015, às 00Z. Adaptado de Bastarz et al. (2017).

### Outras Métricas

Dentre as ferramentas de verificação disponíveis no pacote do CRPS do oensMB09, encontram-se (vide rotina “VFPROB\_pub.f” (disponível no endereço <https://github.com/GAM-DIMNT-CPTEC/oensMB09-AVAL/blob/CRPS/source/VFPROB_pub.f>):

1. *Reliability* (*diagram*) para probabilidades ( membros);

2. *Ranked probability score* e *skill scores*;

3. *Brier score* e *Brier skill score*;

4. *Hit rate* e *false alarm rate*;

5. *Resolution* e *reliability*;

6. *Economic values*;

7. *Information content*.

Para a utilização destas ferramentas, é necessário estudar melhor o programa do CRPS e habilitar a escrita dos arquivos com os resultados e estudar a forma de plotar as figuras.

## Outros Tipos de Verificação

No CPTEC, outras ferramentas de verificação foram utilizadas para a avaliação da média das análises contra as previsões controle oensMB09, além da verificação do RMSE da média do conjunto contra o espalhamento do conjunto.

### Scorecards

Por meio SCANPLOT, pode-se plotar as tabelas com as estatísticas do SCANTEC em forma de um *scorecard* que resume as avaliações realizadas pelo SCANTEC. Na Figura 7 a seguir, são apresentados 2 exemplos de *scorecards* que apresentam um *score* calculado com base no Coeficiente de Correlação de Anomalia (CCA), para os membros médio e controle de um conjunto de previsões. Nesse caso, o CCA foi calculado para o período de 2020061500 a 2020083100 para as variáveis indicadas nas figuras.

|  |  |
| --- | --- |
| a) | b) |

Figura 7 - *Scorecards* dos experimentos apresentando o resumo das estatísticas calculadas pelo SCANTEC. Nas figuras, são comparados os membros médios TENM e XENM contra os seus respectivos membros controle T126 e X126, respectivamente.

Nesse tipo de verificação, embora meramente determinístico, pode-se avaliar outro aspecto importante de um conjunto de previsões, que é a melhoria do *skill* da média do conjunto em relação ao membro controle.

### RMSE da Média do Conjunto X Espalhamento

Outra medida importante sobre a verificação de um conjunto de previsões, é a comparação entre o RMSE da média do conjunto e o seu espalhamento (i.e., o desvio-padrão dos membros do conjunto em relação à sua média). Nesse tipo de verificação, o ideal é que as curvas que representam estas quantidades para diferentes tempos de previsão, sejam próximas entre si. Isso significa que, para um conjunto de previsões bem calibrado, a acurácia da média dos membros deve ser comparável ao desvio-padrão do conjunto.

|  |  |
| --- | --- |
| a) | b) |

Figura 8 - Exemplos da comparação entre o RMSE da média do conjunto e o espalhamento do conjunto, válido para a altura geopotencial em 500 hPa, válido para janeiro de 2013. A diferença entre as figuras está na consideração da ponderação do cosseno das latitudes da grade Gaussiana (aproximação dos pontos de grade em direção aos polos), sendo que a figura b) é a correta.

## Dificuldades Relacionadas com a Aplicação dos Métodos de Verificação

* A aplicação dos métodos de verificação de previsões com conjunto está intrinsecamente relacionada com o tamanho dos conjuntos de análises e previsões, o que se traduz e uma grande quantidade de arquivos a serem tratados e armazenados. Para a verificação com o CRPS, considerando uma realização típica do SPCON global para uma data qualquer, tem-se: 15 análises x 60 previsões (15 dias de previsões a cada 6 horas) = 900 arquivos no formato “grib”. Contabilizando os arquivos “idx” e “ctl”, tem-se ao todo 2.700 arquivos a serem manipuladas para apenas uma única data. Se a verificação a ser realizada for para um período de 3 meses, serão tratados 2.700 arquivos x 30 dias x 2 horários (00 e 12Z) x 3 meses = 486.000 arquivos.
* A falta de documentação dos programas dificulta o trabalho e nos força a encontrar outras soluções;
* Equipe reduzida;
* Treinamento.

## Referências

Bastarz, C. F. et al.: Sistema de Modelagem por Conjunto Global. Relatório Técnico INPE, 2016. – Disponível em <http://mtc-m21b.sid.inpe.br/col/sid.inpe.br/mtc-m21b/2016/08.17.14.20/doc/publicacao.pdf>

Cunningham, C. et al.: Assessing improved CPTEC probabilistic forecasts on medium-range timescale. RMets, 2014. – Disponível em <https://rmets.onlinelibrary.wiley.com/doi/full/10.1002/met.1464>.

Hamill, T. M. (2001). Interpretation of Rank Histograms for Verifying Ensemble Forecasts, Monthly Weather Review, 129(3), 550-560. Retrieved Jun 3, 2022, from <https://journals.ametsoc.org/view/journals/mwre/129/3/1520-0493_2001_129_0550_iorhfv_2.0.co_2.xml>

Hudson, D.: Ensemble Verification Metrics. ECMWF Annual Seminar, 2017. – Disponível em <https://www.ecmwf.int/sites/default/files/elibrary/2017/17626-ensemble-verification-metrics.pdf>

Portal da Física: O que é Acurácia, Precisão e Exatidão. Disponível em <https://portalfisica.com/2018/08/24/acuracia-precisao-e-exatidao/>

Stephanie Glen. "Rank Histogram / Talagrand Diagram" From StatisticsHowTo.com: Elementary Statistics for the rest of us! – Disponível em <https://www.statisticshowto.com/rank-histogram/>