

VERIFICAÇÃO DA PREVISÃO OPERACIONAL DE CHUVA SOBRE BACIAS HIDROGRÁFICAS DO ESTADO DO CEARÁ

Dirceu Silveira Reis Junior¹; Sérgio S. Sombra², Alexandre Araújo Costa³; Otacílio Neto⁴; Antonio Carlos Santana dos Santos⁵; Mosefran Barbosa Macedo Firmino⁶

RESUMO --- Este artigo apresenta um estudo de verificação das previsões de chuva emitidas pela Fundação Cearense de Meteorologia e Recursos Hídricos (FUNCEME) no período 2008/09, obtidas pelo modelo numérico de tempo RAMS, alimentado pelo modelo global do CPTEC. A análise foca nas previsões de chuva com horizontes de 24, 48 e 72 horas sobre 39 bacias e sub-bacias hidrográficas do Estado do Ceará, muitas delas contribuintes aos mais importantes reservatórios do estado. Este estudo de verificação foca numa análise categórica de chuva, baseada em tabelas de contingência, procurando descrever da forma mais completa possível a distribuição conjunta de probabilidades entre as previsões e as observações. O artigo apresenta os resultados de maneira agregada, como forma de avaliar a qualidade da previsão no estado. O artigo também apresenta e discute os resultados da previsão de chuva após passar por um processo de correção estatística, que reduz o *bias* de forma bastante eficiente, promovendo uma melhoria significativa da qualidade da previsão. Os resultados mostram que o sistema de previsão de chuva, quando corrigido de forma adequada, é melhor do que um sistema de previsão baseado na persistência, porém sua qualidade diminui com a intensidade da chuva que se quer prever.

ABSTRACT --- This paper presents a precipitation forecast verification study based upon the forecasts issued by Fundação Cearense de Meteorologia e Recursos Hídricos (FUNCEME) during the 2008/09 period. These operational forecasts are based on the results of a regional atmospheric model. The analysis focuses on precipitation forecasts for 24, 48 and 72 hours lead times over 39 carefully selected watersheds over the state. The verification methods employed here are based on categorical precipitation described by contingency tables. Several verification metrics are employed in order to completely describe the characteristics of the joint probability distribution of forecasts and observed precipitations. The paper presents the results with pooled data so as to have a description of the forecast quality over the entire state. This paper also presents and discusses the results obtained by the forecast system when forecasts outputs are corrected by a bias-correction procedure. Results show that the forecasts, when properly corrected by a statistical model, performs better than a persistent model, however, its quality deteriorates as the precipitation threshold increases.

Palavras-chave: Previsão, modelos numéricos de tempo, verificação

¹ Pesquisador da Fundação Cearense de Meteorologia e Recursos Hídricos, Av. Rui Barbosa, 1246, Aldeota, Fortaleza, CE, CEP: 60.115-221, e-mail: dirceu.reis@gmail.com

² Pesquisador do Departamento de Meteorologia da FUNCEME, e-mail: sergiossombra@gmail.com

³ Professor Titular do Curso de Mestrado em Ciências Físicas Aplicadas da Universidade Estadual do Ceará. Av. Paranjana, 1700 – Itaperi, Fortaleza - CE - Brasil - CEP: 60740-903. alexandrearaujoc@gmail.com.

⁴ Pesquisador do Departamento de Meteorologia da FUNCEME, e-mail: otacilioneto@gmail.com

⁵ Professor Assistente do Curso de Mestrado em Ciências Físicas Aplicadas da Universidade Estadual do Ceará. Av. Paranjana, 1700 – Itaperi, Fortaleza - CE - Brasil - CEP: 60740-903. dr.acss@gmail.com.

⁶ Pesquisador do Departamento de Recursos Hídricos da FUNCEME, e-mail: mosefran@gmail.com

1- INTRODUÇÃO

A previsão de curto prazo de chuva, baseada em modelos numéricos de tempo, fornece informação importante para diversos setores da sociedade, principalmente para aqueles setores sensíveis às variáveis atmosféricas, como a agricultura, recursos hídricos e transporte. Além disso, a possibilidade de prever eventos extremos com antecedência permite que a sociedade tome decisões que ao menos minimizem os impactos de tais eventos, reduzindo assim o grau de vulnerabilidade de uma determinada comunidade.

Nos últimos 30 anos tem-se observado uma melhora na qualidade das previsões de curto e médio prazo das variáveis meteorológicas. Isto se deve, em grande parte, principalmente no caso do hemisfério sul, à disponibilidade de satélites meteorológicos e ao uso de sofisticadas ferramentas de assimilação de dados, que permitem uma caracterização mais acurada das condições iniciais da atmosfera (Kerr, 2004). Além disso, o desenvolvimento da capacidade de processamento e armazenamento dos computadores possibilitou o uso de modelos cada vez mais realistas, baseados em grades cada vez mais densas, o que também contribuiu para a melhora da qualidade da informação.

Uma atividade extremamente importante para o desenvolvimento da modelagem numérica de tempo, e das previsões de maneira geral, é entender as qualidades e deficiências dos sistemas de previsão utilizados em uma dada região. Esta análise recebe o nome de estudo de verificação da previsão, que fornece não apenas subsídios para o aprimoramento destes sistemas, mas também uma idéia aos usuários da informação sobre a qualidade das mesmas.

A literatura mostra que o desenvolvimento de estratégias de verificação de previsão de variáveis meteorológicas vem de longa data, ainda no século XIX, com a publicação de Finley (1884) sobre a verificação de um sistema de previsão de tornados. Este artigo gerou tanta controvérsia, que acabou incentivando o desenvolvimento desta área da ciência com a introdução de diversas métricas de verificação, muitas delas usadas até hoje (Murphy, 1997).

No Estado do Ceará, a Fundação Cearense de Meteorologia e Recursos Hídricos vem utilizando modelos numéricos de tempo como ferramenta para fornecer uma previsão de curto prazo de chuva. Esta previsão é utilizada por diversos setores da sociedade. Não é simples avaliar se as informações contidas nestas previsões são de fato úteis, ou se possuem valor econômico, o que depende não apenas da qualidade da previsão propriamente dita, mas também das características da tomada de decisão de cada caso específico. De todo, é natural imaginar que quanto melhor o sistema de previsão, maior é a probabilidade da informação contida na previsão ser útil. Diversos artigos discutem como identificar o nível de qualidade de um sistema de previsão [Murphy, 1988; Murphy & Epstein, 1989; Murphy, 1991; Gandin & Murphy, 1992; Gerrit, 1992; Krzysztofowicz,

1992; Murphy, 1995; Brooks & Doswell, 1996; Murphy, 1996; Murphy, 1997; Krzysztofowicz & Sigrest, 1999; Zhang & Casey, 2000; Bradley et al., 2003; Thompson & Brier, 1955; Wilks, 2001]. A literatura sobre o valor da informação meteorológica, principalmente do ponto de vista econômico, não é tão extensa quanto àquela sobre a qualidade da previsão, embora não seja desprezível [Anderson, 1973; Murphy, 1977; Murphy et al., 1985; Murphy & Ye, 1990; Buizza, 2001].

Este artigo limita-se a analisar a qualidade da previsão sob o enfoque de bacias hidrográficas, continuando o estudo inicial de verificação apresentado em Reis *et al.* (2008), onde o sistema de previsão de chuva adotado pela FUNCEME foi avaliado para as chuvas que ocorreram no ano de 2008 sobre as bacias contribuintes aos reservatórios de Orós e Castanhão, os maiores reservatórios do estado.

O estudo apresentado aqui foca nas previsões de chuva média sobre 39 bacias hidrográficas localizadas no estado do Ceará, para horizontes de previsão de 24, 48 e 72 horas. Este estudo específico limita-se a uma análise agregada sobre todo o estado, embora uma análise estratificada esteja em andamento, que revelará diferenças em termos de qualidade da previsão para diferentes regiões. O artigo segue com a metodologia empregada neste estudo de verificação, incluindo uma descrição de todas as métricas empregadas. Os dados utilizados no estudo, assim como uma descrição sucinta do modelo numérico de previsão utilizado nestas previsões são apresentados na Seção 3. A Seção 4 apresenta uma discussão dos resultados encontrados, enquanto que a seção 5 descreve as conclusões do artigo.

2- METODOLOGIA PARA VERIFICAÇÃO

Um estudo de verificação tem o objetivo de avaliar a qualidade de um sistema de previsão. O estudo de verificação apresentado aqui segue a estratégia sugerida por Murphy & Winkler (1987), cuja idéia é descrever da melhor forma possível as propriedades da distribuição conjunta de probabilidades entre as previsões de chuva e suas respectivas observações.

Neste caso específico, assim como na grande maioria de estudos de verificação apresentados na literatura, a distribuição conjunta de probabilidades entre as previsões e as observações de chuva é descrita por frequências relativas das diferentes combinações possíveis entre previsão e observação. Neste artigo, embora a precipitação seja uma variável contínua, o estudo de verificação baseia-se na previsão de variáveis discretas. Isto é feito dividindo a faixa de valores possíveis de precipitação em duas categorias, abaixo e acima de um valor limite pré-estabelecido. Portanto, a verificação é baseada em tabelas de contingência 2 x 2. Na verdade, o estudo avalia o sistema de previsão para diferentes limites de precipitação, fornecendo uma idéia da capacidade do sistema em prever eventos de diferentes intensidades.

A distribuição conjunta de probabilidades entre as previsões e as observações, função primordial para uma análise de verificação, é dada por,

$$p(f_i, o_j) = P\{f_i, o_j\} = P\{f_i \cap o_j\} \quad (1)$$

onde f_i indica a previsão de chuva na faixa i e o_j indica a observação de chuva na faixa j . Para fins de verificação de previsão, esta distribuição conjunta é usualmente reescrita de duas formas distintas. A primeira é dada por,

$$p(f_i, o_j) = p(o_j | f_i) p(f_i) \quad (2)$$

onde $p(o_j | f_i)$ é a probabilidade condicional de observar chuva na faixa j , dado que o sistema de previsão previu chuva na faixa i , enquanto que $p(f_i)$ é a probabilidade marginal de prever chuva na faixa i . Esta expressão é chamada de fatoração de calibração da distribuição conjunta. A segunda forma de reescrever a distribuição conjunta, denominada na língua inglesa de fatoração *likelihood-base rate*, é dada por,

$$p(f_i, o_j) = p(f_i | o_j) p(o_j) \quad (3)$$

onde $p(f_i | o_j)$ é a probabilidade condicional de ter previsto chuva na faixa i , dado que a chuva observada foi na faixa j , enquanto que $p(o_j)$ é a probabilidade marginal de observar chuva na faixa j .

Estas duas formas de fatoração da distribuição conjunta são de fato complementares, e ajudam a entender, de forma conceitual, as propriedades da distribuição conjunta de probabilidades entre as previsões e as observações (Murphy, 1997).

Diversos livros e artigos descrevem detalhadamente os atributos de um bom sistema de previsão, e as métricas que devem ser utilizadas num estudo de verificação [Wilks (2006), Murphy (1997), Murphy & Wrinkler (1997)]. Estas métricas seguem a estratégia de tentar descrever características da distribuição conjunta de probabilidades entre as previsões e as observações, e geralmente são termos, ou função de termos, que aparecem nas duas equações de fatoração da distribuição conjunta descritas acima. Não existe uma métrica única que consiga caracterizar de forma completa a distribuição conjunta entre previsão e observação. Daí a necessidade de utilização de diversas delas para avaliar a qualidade de um sistema de previsão.

Nurmi (2003) e WWRP/WGNE (2004) apresentam diversas recomendações sobre como realizar estudos de verificação de modelos numéricos atmosféricos, de modo que os resultados possam ser comparáveis com os resultados de outros modelos. Os leitores interessados nestes detalhes devem procurar estas publicações.

Por falta de espaço, o artigo continua com uma descrição da tabela de contingência 2 x 2, que é a base para o cálculo das métricas utilizadas aqui para o estudo de verificação. Vale notar que boa

parte das recomendações feitas por Nurmi (2003) e WWRP/WGNE (2004) foram seguidas neste trabalho.

A Tabela 1 apresenta uma tabela de contingência 2 x 2, que expressa o número de vezes que ocorreram as diversas combinações entre previsão e observação. No caso específico de uma tabela 2 x 2, os valores possíveis de precipitação são divididos em duas categorias, abaixo de um limite pré-estabelecido, e acima deste limite. Os valores de a, b, c e d são utilizados para estimar as frequências relativas de cada uma destas combinações. A descrição das métricas apresentadas a seguir faz uso destes valores.

Tabela 1: Tabela de contingência 2 x 2

Previsão	Observação	
	Sim	Não
Sim	a	b
Não	c	d

2.1 Métricas de verificação

2.1.1 Razão de bias (B)

A razão de *bias* (B) indica uma possível tendência do sistema de previsão em superestimar ou subestimar os valores de precipitação. Tendo como base uma tabela de contingência 2 x 2, a razão de bias é calculada tomando como base a faixa acima do limite pré-estabelecido, ou seja,

$$B = \frac{a+b}{a+c} \quad (4)$$

de modo que B = 1 significa que o sistema de previsão não é tendencioso, ou seja, não possui tendência a subestimar ou superestimar os valores de precipitação. Isto acontece quando o número de dias que o modelo previu chuva acima do limite pré-estabelecido é igual ao número de dias onde a chuva observada foi maior do que este mesmo limite. Isto não significa dizer que todo o dia onde se observou chuva acima do limite, o sistema também previu chuva acima deste limite. A razão de bias é apenas um indicador do comportamento médio do sistema de previsão. De todo modo, é importante que um sistema de previsão seja não tendencioso, ou seja, que sua razão de bias seja próxima da unidade.

2.1.2 Proporção de acertos (PC)

Como o próprio nome diz, PC expressa, em termos percentuais, a fração do tempo na qual o sistema de previsão acertou a faixa, ou categoria da chuva, independente se a chuva foi abaixo ou

acima do limite pré-estabelecido para definir a tabela de contingência. PC é calculado da seguinte forma:

$$PC = \frac{a + d}{n} \quad (5)$$

onde n é o número de dias do período de verificação. Se o sistema de previsão for perfeito, $PC = 1$. Vale notar que a PC não faz qualquer distinção entre acertar o evento em questão (precipitação acima de um limite pré-estabelecido) e acertar a não ocorrência do evento (precipitação abaixo do limite pré-estabelecido). Murphy (1996) mostra que esta métrica pode ser extremamente enganosa, principalmente quando a probabilidade de ocorrência do evento que se deseja prever for pequena. Neste artigo, Murphy discute o famoso artigo de Finley (1884) sobre previsão de tornados, onde é apresentado os resultados de um estudo de verificação do sistema de previsão com um PC de 96,6%, um valor extremamente alto. Este resultado chamou a atenção de muitos, e levou à publicação, logo em seguida, de diversos artigos, que sugeriram diversas métricas de verificação para casos onde o evento de interesse tem baixa probabilidade de ocorrência. Neste tipo de situação, é possível que prever sempre a não ocorrência do evento pode resultar numa PC maior do que prever co base num sistema de previsão. Portanto, em casos como este, obter uma PC maior não diz muito sobre o sistema de previsão, já que não há qualquer utilidade prática em prever sempre a não ocorrência de um. Em casos como este, a métrica mais recomendada é o *Threat Score (TS)*, descrita na sequência.

2.1.3 Threat Score (TS)

O *Threat Score (TS)* é uma métrica recomendada para substituir a PC quando o evento a ser previsto possui uma baixa probabilidade de ocorrência. O TS pode ser considerado como uma proporção de acertos, porém calculada após a remoção dos dias nos quais o sistema de previsão corretamente previu a não ocorrência do evento de interesse. O TS é calculado da seguinte forma:

$$TS = \frac{a}{a + b + c} \quad (6)$$

2.1.4 Taxa de acerto (Hit Rate – H)

A taxa de acerto (H) é a representação empírica da probabilidade condicional do sistema de previsão ter previsto a ocorrência do evento, dado que o evento de fato ocorreu, $p(f_i|o_i)$. Esta probabilidade condicional faz parte da eq. (3), que é uma das formas de representação da distribuição conjunta de probabilidades entre previsão e observação. A taxa de acerto (H) é também chamada de probabilidade de detecção (POD).

$$H = \frac{a}{a + c} \quad (7)$$

Um sistema de previsão perfeito possui $H = 1$, ou seja, o evento nunca acontece sem que o sistema de previsão o tenha previsto. Obviamente que sistemas reais de previsão não são perfeitos, mas quanto maior o valor de H , menor é chance da sociedade ser surpreendida por um evento que não tenha sido previsto.

2.1.5 Taxa de alarme falso (*False Alarm Rate – F*)

A taxa de alarme falso (F) representa a probabilidade condicional do sistema de previsão ter previsto chuva acima de um determinado limite pré-estabelecido, dado que a observação foi abaixo deste limite, $p(f_1|o_2)$, onde 1 representa chuva acima do limite pré-estabelecido. A taxa de alarme falso é calculada por,

$$F = \frac{b}{b + d} \quad (8)$$

Na prática, valores altos de F podem comprometer a confiança que a sociedade deposita sobre o sistema de previsão.

2.1.6 Razão de alarme falso (*False Alarm Ratio – FAR*)

A razão de alarme falso (FAR) não deve ser confundida com a taxa de alarme falso (F), embora os nomes sejam muito similares. A FAR também representa uma probabilidade condicional, porém ela expressa a probabilidade de não observar chuva acima de um determinado limite, dado que o sistema de previsão previu chuva acima deste limite, $p(o_2|f_1)$, sendo 1 a faixa acima do limite pré-estabelecido. A FAR é calculada da seguinte maneira,

$$FAR = \frac{b}{a + b} \quad (9)$$

Assim como no caso de F , valores elevados de FAR podem comprometer a confiança da sociedade no sistema de previsão.

2.1.7 Heidke Skill Score (*HSS*)

É bastante comum em estudos de verificação a utilização de *skill scores* para resumir a qualidade do sistema de previsão. *Skill scores* são estatísticas que expressam a qualidade relativa de um sistema de previsão em relação a um sistema de previsão de referência. Geralmente, o sistema de previsão de referência empregado para a determinação da qualidade relativa do sistema de previsão em questão é a própria climatologia ou um sistema de previsão baseado na persistência.

O *Heidke Skill Score* (HSS) é uma das estatísticas mais utilizadas para fornecer uma idéia da qualidade relativa de um sistema de previsão. O HSS utiliza a proporção de acertos (PC) como métrica para calcular a qualidade relativa do sistema de previsão. O sistema de referência neste caso consiste de uma previsão aleatória, independente das observações, que conserva as probabilidades

marginais de previsão ($p(f_i)$) do sistema que está sendo analisado. Deste modo, o HSS é calculado da seguinte forma,

$$\begin{aligned} HSS &= \frac{PC_{prev} - PC_{ref}}{1 - PC_{ref}} = \frac{(a + d)/n - [p(f_1)p(o_1) + p(f_2)p(o_2)]}{1 - [p(f_1)p(o_1) + p(f_2)p(o_2)]} \\ &= \frac{2(ad - bc)}{(a + c)(c + d) + (a + b)(b + d)} \end{aligned} \quad (10)$$

onde PC_{prev} é a proporção de acertos obtida pelo sistema de previsão em questão.

2.1.8 Peirce Skill Score (PSS)

O *Peirce Skill Score (PSS)* segue a mesma lógica do *HSS*. A única diferença está no denominador, onde a probabilidade marginal da previsão, $p(f_i)$, é igualada à probabilidade marginal da observação, $p(o_i)$, ou seja, o sistema de previsão de referência, utilizado para avaliar a qualidade relativa do sistema de previsão em questão, é um sistema não tendencioso ($B = 1$). O *PSS* é calculado pela expressão abaixo,

$$PSS = \frac{ad - bc}{(a + c)(b + d)} \quad (11)$$

Vale notar que se o sistema de previsão em questão for não tendencioso, ou seja, se $B = 1$, o *PSS* será igual ao *HSS*.

2.2 Incerteza amostral das métricas de verificação

Métricas de verificação são estimadas em amostras de pares de previsão e observação, portanto, estão sujeitas à variabilidade amostral. Sendo assim, sempre que possível, é importante que um estudo de verificação apresente, além dos valores estimados das métricas, uma descrição das incertezas associadas, geralmente apresentada na forma de intervalos de confiança. Estes intervalos de confiança permitem que se realizem testes de hipóteses para definir se um sistema de previsão é melhor ou pior do que um sistema alternativo.

Muitas das métricas apresentadas na seção 2.1, como por exemplo, H, F e FAR, podem ser vistas como proporções que representam probabilidades condicionais. Deste modo, intervalos de confiança podem ser estimados com base em variáveis binomiais (Wilks, 2006).

O intervalo de confiança $(1 - \alpha)$ para probabilidades condicionais é calculado da seguinte forma,

$$p = \frac{\hat{p} + \frac{z_{1-\alpha/2}^2}{2N} \pm z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{N} + \frac{z_{1-\alpha/2}^2}{4N^2}}}{1 + \frac{z_{1-\alpha/2}^2}{N}} \quad (12)$$

onde \hat{p} é a probabilidade condicional estimada com base na amostra, N é o número de dias associado com a observação do evento ter acontecido ou não, o que depende de que tipo de proporção esteja sendo calculada, e $z_{1-\alpha/2}$ é o percentil da variável normal padrão associado a uma probabilidade de excedência $\alpha/2$.

3- DADOS UTILIZADOS

3.1 Bacias hidrográficas e período de verificação

O estudo de verificação apresentado aqui utilizou dados de previsão e observação de chuva média sobre 39 bacias e sub-bacias hidrográficas do Estado do Ceará, muitas delas contribuintes aos mais importantes reservatórios do estado. A Figura 1 apresenta as 39 bacias empregadas no estudo, cujas áreas de drenagem variam entre 1.000 e 16.800 km².

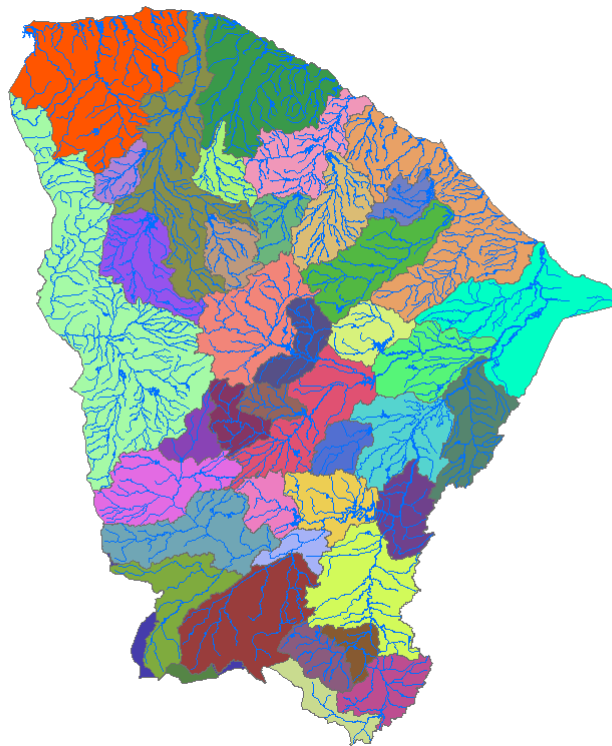


Figura 1: Mapa do Estado do Ceará com as 39 bacias e sub-bacias utilizadas no estudo de verificação.

Devido à impossibilidade de realizar previsões de anos históricos, o estudo apresentado aqui foca no período chuvoso (primeiro semestre) dos anos de 2008 e 2009, quando foram observadas cheias importantes no estado. O período em que há disponibilidade de pares de valores de previsão e observação de chuva vai de 09 de janeiro a 28 de maio de 2008, e de 10 de janeiro a 30 de maio de 2009, totalizando 252 pares para cada bacia hidrográfica. Entretanto, durante este período, houve

25 dias em que não foi possível realizar a previsão, de modo que estes dias não foram empregados na análise de verificação.

3.2 Previsão com modelo numérico de tempo

As previsões de chuva utilizadas neste estudo de verificação foram aquelas geradas operacionalmente pela FUNCEME, obtidas com o *Regional Atmospheric Modeling System* (RAMS) 6.0. O RAMS 6.0 é um modelo de área limitada desenvolvido a partir do conjunto completo de equações não-hidrostáticas compressíveis da atmosfera, além de equações de conservação para grandezas escalares. Conta ainda com diferentes opções de parametrização que permitem a reprodução de grande parte dos fenômenos atmosféricos mais relevantes (Cotton *et al.*, 2003; Pielke *et al.*, 1992).

A cada dia são efetuadas duas simulações cobrindo um prazo de 84 horas com uma grade horizontal de 115 por 115 pontos, espaçados em 25 km sobre a região Nordeste. Na determinação das condições de contorno são usados resultados da previsão do modelo global T126L28 mantido pelo CPTEC/INPE.

A caracterização da superfície é feita com dados de topografia, NDVI, tipo de vegetação e uso do solo, e climatológicos fornecidos pela página internet de apoio ao modelo RAMS, enquanto a assimilação da temperatura da superfície do mar (tsm) é feita diariamente usando estimativas de satélites fornecidas por *Remote Sensing Systems* (RSS).

No que diz respeito às parametrizações físicas, é usado o esquema de Chen & Cotton (1987; 1983) para radiação em ondas curtas e longas, utilizado a cada 20 minutos, a turbulência é resolvida com o esquema de Mellor & Yamada (1982), a microfísica de nuvens é parametrizada conforme o esquema proposto por Walko *et al.* (1995), e a interação com superfície é resolvida pelo submodelo LEAF2, descrito por Walko *et al.* (2000). Em relação à convecção, foi utilizada parametrização descrita por Kain & Fritsch (1993).

3.3 Processo de correção estatística das previsões de chuva

Como será apresentado mais adiante na discussão dos resultados, as previsões obtidas com modelos numéricos de tempo podem conter erros sistemáticos, que resultam em previsões com tendência a subestimação ou superestimação dos valores de precipitação.

Estes erros sistemáticos podem ser removidos, ou pelo menos reduzidos, através de métodos estatísticos. Um dos métodos mais utilizados de correção de previsão é aquele baseado na distribuição acumulada de probabilidades. Com base numa amostra histórica de pares de valores previstos e observados de chuva, é possível corrigir o valor de previsão fornecida pelo modelo. A correção é feita de modo que a probabilidade de excedência do valor previsto corrigido, estimada

com base na distribuição acumulada dos valores observados no período histórico, seja igual à probabilidade de excedência do valor previsto, estimada com base na distribuição acumulada dos valores previstos no período histórico. Este tipo de correção resulta, no longo prazo, em previsões, cuja distribuição acumulada de probabilidades é compatível com a distribuição acumulada dos valores observados no período histórico.

Como o período disponível de dados de previsão é bastante curto, apenas dois anos, não seria razoável reservar parte do período para servir de base para a construção do procedimento de correção. Num primeiro momento, o que se fez foi corrigir as previsões de chuva do primeiro ano utilizando os dados disponíveis do segundo ano para construção das distribuições acumuladas de probabilidades, e vice-versa. O importante é que o valor de previsão a ser corrigido seja independente dos pares de valores utilizados para a construção das distribuições acumuladas de probabilidade das previsões e dos valores observados. Tendo isso em mente, o que se fez foi gerar, para cada dia do período de verificação, valores corrigidos de previsão com base nas distribuições acumuladas de probabilidades estimadas empregando todos os valores de previsão e observação disponíveis, com exceção, é claro, do par previsão-observação que estava sendo corrigido. O que se fez foi utilizar a mesma estratégia de estudos de validação cruzada, usualmente empregada para avaliar o desempenho de metodologias desenvolvidas, por exemplo, para estimar quantis de cheia em rios. Acredita-se que esta estratégia, além de preservar de forma satisfatória o tamanho da amostra para o estudo de verificação, representa de forma adequada qual será o comportamento das previsões corrigidas no futuro, que é o que se deseja obter para a realização do estudo de verificação.

4- DISCUSSÃO DOS RESULTADOS

O estudo de verificação apresentado foca numa análise agregada das previsões de chuva sobre as 39 bacias hidrográficas estudadas. Uma análise estratificada está em andamento e deverá ser publicada em breve. A análise agregada, além de fornecer uma idéia geral da qualidade da previsão de chuva no estado, devido ao maior número de pares de observação e previsão, permite estimar com mais precisão diversas métricas de verificação, de modo que seja mais fácil identificar, através de testes de hipóteses formais, se o modelo de previsão utilizado é de fato melhor do que metodologias alternativas de previsão, por exemplo, como aquelas baseadas na climatologia ou na persistência.

Esta seção foca nos resultados de três modelos de previsão: o CPTEC-RAMS sem correção estatística, o CPTEC-RAMS com correção baseada na distribuição acumulada de probabilidades, e o modelo de persistência.

A primeira avaliação da análise agregada consiste simplesmente em verificar a capacidade do sistema de previsão em prever corretamente a ocorrência de chuva nas bacias hidrográficas. O limite de precipitação empregado na análise para caracterizar a ocorrência de chuva foi de 1 mm.

A Tabela 2 apresenta os resultados de diversas métricas de verificação, incluindo intervalos de confiança de 95% para H, F e FAR. Constatou-se que o modelo CPTEC-RAMS possui uma tendência a superestimar a precipitação para os três horizontes de previsão avaliados. Isto pode ser inferido a partir da probabilidade marginal do modelo prever chuva acima de 1 mm, $p(f)$. Por exemplo, no período de verificação, o $p(f)$ é igual a 0,78, enquanto que a probabilidade marginal de se observar chuva acima de 1 mm, $p(o)$, é de apenas 0,59. Portanto, as previsões do modelo CPTEC-RAMS possuem uma razão de *bias* (B) de aproximadamente 1,33 para o horizonte de 24 horas, e pode-se observar que B aumenta com o horizonte de previsão.

Tabela 2: Resultados da análise agregada da previsão obtida com o modelo CPTEC-RAMS sem e com correção, e da previsão baseada na persistência. Análise com base em duas categorias: chuva ($p > 1\text{mm}$) e não chuva. A probabilidade marginal de se observar chuva acima de 1 mm neste período foi de 0.59.

	24 horas			48 horas			72 horas		
	Prev. s/ corr.	Prev. c/ corr.	Persist.	Prev. s/ corr.	Prev. c/ corr.	Persist.	Prev. s/ corr.	Prev. c/ corr.	Persist.
$p(f)$	0.78	0.59	0.58	0.88	0.59	0.58	0.88	0.60	0.58
B	1.33	1.00	1.00	1.49	1.00	0.99	1.46	1.00	0.98
PC	0.67	0.73	0.70	0.64	0.71	0.64	0.60	0.68	0.63
TS	0.61	0.62	0.59	0.61	0.61	0.54	0.58	0.58	0.52
H	0.89 [0.88 0.89]	0.77 [0.75 0.78]	0.74 [0.73 0.75]	0.94 [0.93 0.95]	0.76 [0.75 0.77]	0.69 [0.68 0.71]	0.90 [0.89 0.91]	0.73 [0.72 0.74]	0.68 [0.67 0.69]
F	0.63 [0.61 0.65]	0.33 [0.32 0.35]	0.36 [0.34 0.37]	0.80 [0.78 0.81]	0.35 [0.34 0.37]	0.43 [0.41 0.44]	0.84 [0.83 0.85]	0.40 [0.39 0.42]	0.44 [0.43 0.46]
FAR	0.34 [0.32 0.35]	0.23 [0.22 0.25]	0.25 [0.24 0.27]	0.37 [0.36 0.38]	0.24 [0.23 0.25]	0.30 [0.29 0.31]	0.38 [0.37 0.40]	0.27 [0.26 0.28]	0.31 [0.30 0.32]
Heidke	0.27	0.44	0.38	0.17	0.41	0.27	0.09	0.33	0.24
Pierce	0.26	0.44	0.38	0.15	0.41	0.27	0.08	0.33	0.24

O fato do modelo CPTEC-RAMS frequentemente prever chuva acima de 1 mm resulta num valor elevado de taxa de acerto (*Hit Rate* – H), 0,89 para previsões de 24 horas. Entretanto, a taxa

de alarme falso (F) e a razão de alarme falso (FAR) são bastante altas, 0,63 e 0,34, respectivamente, para horizonte de previsão de 24 horas. Estes valores aumentam de forma considerável para horizontes maiores, especialmente no caso de F, chegando a $F = 0,84$ e $FAR = 0,38$ para horizonte de previsão de 72 horas.

Pode-se observar também valores razoáveis de proporção de acertos (PC) e *Threat Score* (TS) para as previsões sem correção. Além disso, os resultados de Heidke e Pierce *scores* mostram que o modelo CPTEC-RAMS, sem qualquer tipo de correção estatística, é melhor do que um sistema aleatório de previsão que respeite a probabilidades marginais de ocorrência e previsão de precipitação acima de 1 mm. Entretanto, estes resultados mostram que o modelo CPTEC-RAMS apresenta um nível muito elevado de taxa de falso alarme (F), o que caracteriza um problema, pois prejudica bastante a confiança que a sociedade deposita num modelo de previsão como este.

Os resultados também mostram que um modelo baseado na persistência fornece previsões com características até melhores do que as fornecidas pelo modelo CPTEC-RAMS sem qualquer tipo de correção, para todos os horizontes de previsão, com exceção, obviamente, da taxa de acerto (H), que é muito elevada no caso do modelo CPTEC-RAMS devido a sua razão de *bias* elevada. Os valores de F e FAR obtidos com o modelo de persistência são estatística, e consideravelmente, menores do que aqueles obtidos com o CPTEC-RAMS, para os três horizontes de previsão considerados, e a mesma conclusão se aplica às métricas PC, Heidke e Peirce *scores*.

Por outro lado, quando a correção baseada na função de distribuição acumulada é aplicada às previsões do modelo CPTEC-RAMS, a qualidade destas previsões melhora consideravelmente. Pode-se notar que o processo de correção é capaz de remover os erros sistemáticos, fazendo com que a razão *bias* seja reduzida à unidade ($B = 1$). A proporção de acerto neste caso é maior do que a proporção de acerto do modelo baseado na persistência, mesmo na previsão de 24 horas, quando o modelo de persistência ainda é competitivo. O mesmo vale para o *Threat Score* (TS).

A previsão corrigida obtida pelo modelo CPTEC-RAMS é também superior em termos de H, F e FAR em relação ao modelo de persistência para todos os horizontes de previsão. Para os horizontes de 48 e 72 horas, os intervalos de confiança destas três métricas obtidos com a previsão corrigida não se sobrepõem aos intervalos de confiança obtidos com o modelo de persistência. Isto mostra que os resultados são estatisticamente significativos. Mesmo no caso do horizonte de 24 horas, a sobreposição é pequena, e seria necessário realizar um teste formal de hipóteses, o que não foi feito devido ao tempo necessário para a elaboração deste artigo.

Os resultados de *Heidke* e *Peirce Skill Scores* corroboram a conclusão de que as previsões corrigidas do modelo CPTEC-RAMS são melhores do que as do modelo de persistência, e melhores do que a previsão de referência relacionada com cada *score* (ver seção 2.1).

A Figura 2 apresenta os resultados de H, F e FAR, e respectivos intervalos de confiança de 95%, para os três modelos de previsão, para horizontes de 24, 48 e 72 horas.

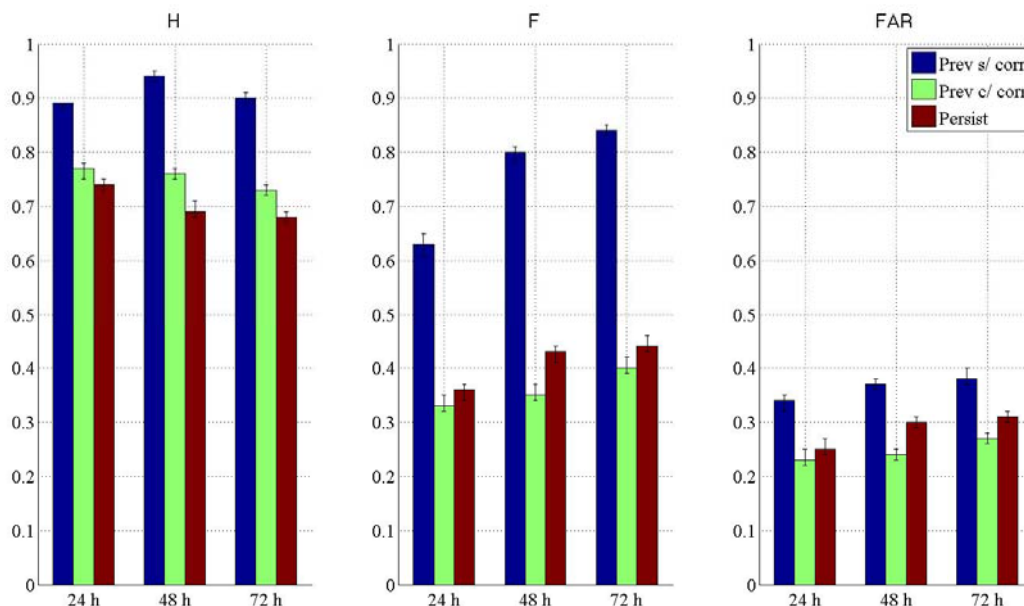


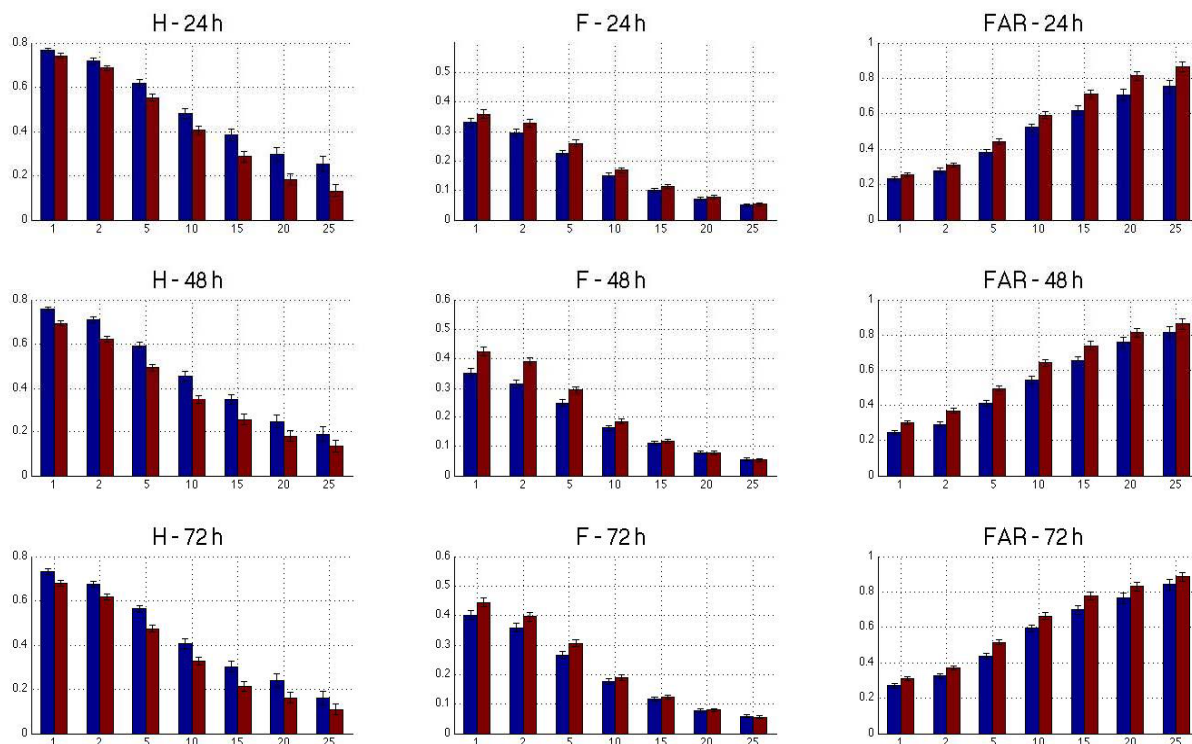
Figura 2: Resultados de Taxa de Acerto (*Hit Rate* - H), Taxa de Alarme Falso (*False Alarm Rate* - F) e Razão de Alarme Falso (*False Alarm Ratio* - FAR) da análise agregada das previsões de 24, 48 e 72 horas com o modelo CTPEC-RAMS, sem e com correção, e com a previsão baseada na persistência. Análise de verificação para duas categorias: Chuva e não-Chuva

Embora seja usual a realização de estudos de verificação utilizando como limite chuvas acima de 1mm, do ponto de vista do uso da informação para alerta de eventos extremos, ou para uso em modelos hidrológicos do tipo chuva-vazão, estes resultados possuem um valor limitado, embora sejam importantes para outros usos.

Os resultados apresentados a seguir foram baseados em diferentes valores do limite de precipitação utilizado na tabela de contingência 2 x 2. A Figura 3 apresenta os resultados de H, F e FAR obtidos tanto com a previsão corrigida do modelo CPTEC-RAMS, quanto com o modelo de persistência, para valores limites de precipitação de 1, 2, 5, 10, 15, 20 e 25 mm. Como era de se esperar, à medida que o limite aumenta, ou seja, à medida que o evento de interesse é mais intenso, a qualidade da previsão cai. Porém, do ponto de vista relativo, a previsão corrigida obtida com o modelo CPTEC-RAMS continua sendo melhor do que a obtida com o modelo de persistência.

Enquanto que para o limite de 1 mm, a taxa de acerto da previsão corrigida de 24 horas obtida com o modelo CPTEC-RAMS é igual a 0,77, para o limite de 10 mm, a taxa de acerto caiu para aproximadamente 0,50. Isto significa que em 50% dos casos onde ocorreu chuva acima de 10 mm, o modelo não foi capaz de prever corretamente. Esta perda de qualidade não é desprezível, embora

seja difícil avaliar a priori a queda associada ao valor da previsão, que como já foi observado anteriormente, depende das características da matriz de decisão de cada situação.



**Figura 3: Resultados de Taxa de Acerto (*Hit Rate - H*), Taxa de Alarme Falso (*False Alarm Rate – F*) e Razão de Alarme Falso (*False Alarm Ratio – FAR*) da análise agregada das previsões de 24, 48 e 72 horas com a previsão baseada na previsão com o modelo CPTEC-RAMS com correção (azul) e com base na persistência (vermelho).
Análise de verificação para duas categorias com diversos limites de chuva.**

A Figura 4 apresenta os resultados de *Heidke Skill Score* obtidos com as previsões corrigidas do modelo CPTEC-RAMS e aquelas obtidas com o modelo de persistência para diversos limites de precipitação. Observa-se claramente a superioridade das previsões corrigidas com o modelo CPTEC-RAMS para todos os horizontes de previsão. Ambos os modelos apresentam HSS maior do que zero, o que significa dizer que ambos possuem um desempenho melhor do que um sistema de previsão aleatório que respeite as probabilidades marginais de previsão e observação do evento em pauta. Vale notar que como os dois modelos apresentam razão de *bias* muito próxima da unidade, mesmo para limites de precipitação altos, os valores de HSS e PSS são muito próximos.

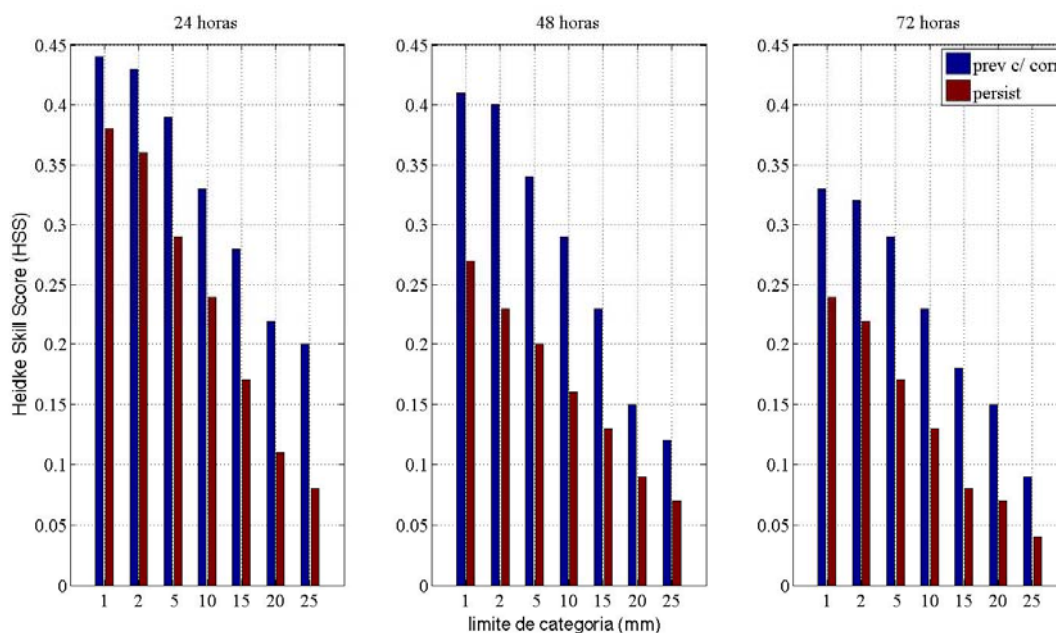


Figura 4: Resultados de *Heidke Skill Score* (HSS) da análise agregada das previsões de 24, 48 e 72 horas, para diversos limites de precipitação, obtidos com o modelo CPTEC-RAMS, com correção, e com a previsão baseada na persistência.

5- CONCLUSÕES

Este artigo apresenta os resultados de um estudo de verificação das previsões de chuva média sobre 39 bacias e sub-bacias hidrográficas do Estado do Ceará. As previsões são realizadas com o modelo regional RAMS 6.0, alimentado pelo modelo global do CPTEC, para horizontes de 24, 48 e 72 horas. As previsões utilizadas neste estudo são aquelas que foram emitidas operacionalmente pela FUNCEME durante os anos de 2008 e 2009, que foram anos bastante chuvosos.

Os resultados apresentados aqui limitaram-se à análise agregada de todas as 39 bacias hidrográficas. Um estudo de verificação com os dados estratificados está em andamento. O estudo de verificação baseou-se numa análise categórica, utilizando tabelas de contingência 2 x 2, empregando diversos limites de precipitação. Diversas métricas foram empregadas: razão de *bias* (B), proporção de acerto (PC), taxa de acerto (H), taxa de alarme falso (F), razão de alarme falso (FAR), *Heidke Skill Score* (HSS) e *Peirce Skill Score* (PSS).

O artigo apresenta uma discussão dos resultados obtidos com três modelos: CPTEC-RAMS sem correção das previsões, CPTEC-RAMS com correção estatística das previsões, e um modelo de persistência.

Os resultados mostram que as previsões brutas do modelo CPTEC-RAMS superestimam as precipitações, resultando em valores elevados de PC e H, o que aparentemente é positivo, mas resulta numa taxa de alarme falso (F) extremamente elevada, entre 65 e 80% para os horizontes de 24-72 horas, inviabilizando, de certa forma, o seu uso direto na previsão.

Quando a metodologia de correção estatística é aplicada, a razão de *bias* reduz-se praticamente à unidade. Os resultados com as previsões corrigidas indicam que o modelo CPTEC-RAMS é consistentemente melhor do que o modelo de persistência. No caso da previsão de ocorrência de chuva, com limite de 1 mm, as previsões corrigidas do modelo CPTEC-RAMS obtiveram valores bons para as métricas avaliadas. Por exemplo, a proporção de acerto (PC) para horizontes de 24, 48 e 72 horas foi igual a 0,73, 0,71 e 0,68, respectivamente. A taxa de acerto (H) para estes mesmos horizontes de previsão foi igual a 0,77, 0,76 e 0,73, enquanto que a taxa de alarme falso ficou entre 0,33 e 0,40. Além disso, tanto o *Heidke Skill Score*, quanto o *Peirce Skill Score*, das previsões corrigidas do modelo CPTEC-RAMS variaram entre 0,44 e 0,33 (para horizontes de previsão de 24 e 72 horas), valores positivos, o que indica desempenho melhor do que o sistema de referência, e maiores do que aqueles obtidos com o modelo de persistência.

À medida que o evento de interesse fica mais intenso, ou seja, à medida que o limite de chuva para a construção da tabela de contingência aumenta, a qualidade da previsão diminui, embora os resultados indiquem que as previsões corrigidas do modelo CPTEC-RAMS continuam melhores do que aquelas obtidas com o modelo de persistência.

AGRADECIMENTOS

Os autores agradecem ao CNPq pelo financiamento do projeto “Uso de Previsão de Cheias com Modelos Atmosférico-Hidrológicos na Operação Otimizada de Reservatórios”, Edital Universal 2006, e à FINEP pelo financiamento dos projetos PSICO e REPENTE.

BIBLIOGRAFIA

- ANDERSON, L. (1973). “The economics of extended-term weather forecasting”, *Monthly Weather Rev*, vol. 101 (2) pp. 115-125.
- BRADLEY, A. A., HASHINO, T., SCHWARTZ, S. S. (2003). “*Distributions-oriented verification of probability forecasts for small data samples*”, *Weather and Forecasting*, vol. 18 (5), pp. 903-917.
- BROOKS, H. E.; DOSWELL III, C. A. (1996). “*A comparison of measures-oriented and distributions-oriented approaches to forecast verification*”, *Weather and Forecasting*, vol. 11 (3), pp. 288-303.
- BUIZZA, R. (2001). “*Accuracy and potential economic value of categorical and probabilistic forecasts of discrete events*”, *Monthly Weather Rev*, vol. 129 (9) pp. 2329-2345.
- COTTON, W. R., R. A. PIELKE SR., R. L. WALKO, et al (2003). “*RAMS 2001: Current status and future directions*”. *Meteorology and Atmospheric Physics* 82, no. 1 (Janeiro 19): 5-29. doi:10.1007/s00703-001-0584-9.
- GANDIN, L. S.; MURPHY, A. H. (1992). “*Equitable skill scores for categorical forecasts*”, *Monthly Weather Rev*, vol. 120 (2) pp. 361-370.
- GERRITY Jr, J. P. (1992). “*A note on Gandin and Murphy's equitable skill score*”, *Monthly Weather Rev*, vol. 120 (11) pp. 2709-2712.
- KAIN, J. S., E J. M. FRITSCH. 1993. Convective parameterization for mesoscale models: The Kain-Fritsch scheme. *The Representation of Cumulus Convection in Numerical Models*, Meteor. Monogr. 46: 165-170.

- KERR, R. A. (2004). “Not End Yet to Forecast Advances”, *Science*, Vol. 304, pp. 948-948.
- KRZYSZTOFOWICZ, R. (1992). “*Bayesian correlation score: A utilitarian measure of forecast skill*”, *Monthly Weather Rev*, vol. 120 (1) pp. 208-220.
- KRZYSZTOFOWICZ, R.; SIGREST, A. A. (1999). “*Comparative verification of guidance and local quantitative precipitation forecasts: Calibration analyses*”, *Weather and Forecasting*, vol. 14 (3), pp. 443-454.
- MELLOR, G. & YAMADA, T. (1982). “*Development of a turbulence closure model for geophysical fluid problems*”. *Reviews of Geophysic. & Space Physics*, 20: 851–875.
- MURPHY, A. H. (1997). “*Forecast Verification*”, in *Economic Value of Weather and Climate Forecasts*. Ed. por Katz, R. W. e Murphy, A. H., Cambridge University Press, pp. 19 – 74.
- MURPHY, A. H. (1996). “*The Finley affair: A signal event in the history of forecast verification*”, *Weather and Forecasting*, vol. 11 (1), pp. 3-20.
- MURPHY, A. H. (1995). “*The coefficients of correlation and determination as measures of performance in forecast verification*”, *Weather and Forecasting*, vol. 10 (4) pp. 681-688.
- MURPHY, A. H. (1991). “*Forecast verification: Its complexity and dimensionality*”, *Monthly Weather Rev*, vol. 119 (7) pp. 1590-1601.
- MURPHY, A. H. (1988). “*Skill scores based on the mean square error and their relationships to the correlation coefficient*”, *Monthly Weather Rev*, vol. 116 (12) pp. 2417-2424.
- MURPHY, A. H. (1977). “*The value of climatological, categorical and probabilistic forecasts in the cost-loss ratio situation*”, *Monthly Weather Rev*, vol. 105 (7) pp. 803-816.
- MURPHY, A. H.; EPSTEIN, E. S. (1989). “*Skill scores and correlation coefficients in model verification*”, *Monthly Weather Rev*, vol. 117 (3) pp. 572-582.
- MURPHY, A. H.; KATZ, R. W.; WINKLER, R. L.; HSU, W. R. (1985). “*Repetitive Decision Making and the Value of Forecasts in the Cost-Loss Ratio Situation*”: A Dynamic Model, *Monthly Weather Rev*, vol. 113 (5) pp. 801-813.
- MURPHY, A. H.; WINKLER, R. L. (1987). “*A general framework for forecast verification*”, *Monthly Weather Rev*, vol. 115 (7), pp. 1330-1338.
- MURPHY, A. H.; YE, Q. (1990). “*Optimal decision making and the value of information in a time-dependent version of the cost-loss ratio situation*”, *Monthly Weather Rev*, vol. 118 (4) pp. 939-949.
- NURMI, P. (2003), *Recommendations on the verification of local weather forecasts*, Consultancy Report to ECMWF Technical Advisory Committee, ECMWF Operations Department.
- PIELKE, R. A., W. R. COTTON, R. L. WALKO, et al. 1992. A comprehensive meteorological modeling system—RAMS. *Meteorology and Atmospheric Physics* 49, no. 1 (Março 1): 69-91. doi:10.1007/BF01025401.
- CARVALHO, T. L. L.; REIS, D. S. JR.; MARIA, P. H. S.; SOMBRA, S. S.; COSTA, A. A.; SANTOS, A. C. S. (2008). “*Avaliação da Previsão de Curto Prazo de Chuva na Bacia do Alto Jaguaribe durante a Cheia de 2008*”, In: *Anais do IX Simpósio de Recursos Hídricos do Nordeste*, 25-28 de novembro, Salvador – BA.
- THOMPSON, C.; BRIER, G. W. (1955). “*The economic utility of weather forecasts*”, *Monthly Weather Rev*, vol. 83 (11) pp. 249-253.
- WILKS, D. S. (2006). *Statistical Methods in the Atmospheric Sciences*. Academic Press, 2ª Edição, 630 p.
- WILKS, D. S. (2001). “*Diagnostic Verification of the Climate Prediction Center Long-Lead Outlooks, 1995–98*”, *Journal of Climate*, vol. 13 pp. 2389-2403. (agregado)
- ZHANG, H.; CASEY, T. (2000). “*Verification of Categorical Probability Forecasts, Weather and Forecasting*”, vol. 15 pp. 80-89.
- WWRP/WGNE (2004). *Recommendations for the verification and intercomparison of QPFs from operational NWP models*. WWRP/WGNE Joint Working Group on Verification.