

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

Aplicação do Método Model Output Calibration para a Correção Estatística do Resultado da Previsão do Tempo

Carlos Amaral Hölbíg¹
Vanessa Lago Machado²

Programa de Pós-Graduação em Computação Aplicada, UPF, Passo Fundo, RS

Resumo. Este trabalho analisa a aplicação do método de calibração de modelos de simulação chamado Model Output Calibration na correção da previsão do tempo gerada pelo modelo regional de previsão numérica do tempo Eta do CPTEC/INPE. Os resultados obtidos com a correção estatística da previsão gerada pelo Eta apresentaram uma previsão do tempo com valores mais próximos aos das medições reais observadas para as variáveis de temperatura e de umidade relativa. Esses resultados mostram a possibilidade de ganhos de qualidade quando aplicados a outros modelos de simulação que utilizam como variáveis preditoras as variáveis geradas por modelos de previsão de tempo, como os modelos de culturas e os de doenças em culturas.

Palavras-chave. Correção Estatística, Previsão Numérica do Tempo, MOC, Modelo Eta.

1 Introdução

Os modelos obtidos por simulação são importantes ferramentas no dia-a-dia, atuando como instrumentos de previsão em diversas situações, possibilitando que as decisões sejam tomadas antes da ocorrência dos fatos. Porém, mesmo com toda a evolução desses modelos, o resultado das previsões, em muitos casos, ainda podem diferir consideravelmente dos dados de fato observados.

Modelos numéricos de previsão do tempo, em geral, são baseados em modelagem computacional e permitem a análise e a tomada de decisões sobre diversos fatores que estão diretamente relacionados ao tempo. Tais modelos dividem-se em “modelos globais” e “modelos regionais” ou de área “limitada”. Os modelos regionais realizam a previsão sobre uma área geográfica mais reduzida como, por exemplo, a América do Sul ou parte dela. Os modelos de previsão numérica do tempo representam seus valores em caixas de grades horizontais com níveis verticais e assumem que a atmosfera seja homogênea em cada uma dessas caixas, sendo necessário conhecer apenas os dados de um ponto por caixa. Tais modelos são conhecidos como “modelos em ponto de grade”.

Para fins desta pesquisa, o modelo de previsão numérica do tempo utilizado foi o modelo regional Eta 15km, utilizado pelo Centro de Previsão de Tempo e Estudos Climáticos

¹holbig@upf.br

²154401@upf.br

(CPTEC) do Instituto Nacional de Pesquisas Espaciais (INPE). A previsão do tempo gerada por este modelo estende-se em até 264 horas (11 dias), sendo fornecida duas vezes ao dia (às 00:00 e às 12:00 UTC – tempo universal coordenado, do inglês *Universal Time Coordinated*). Esse modelo possui, também, grades de 40km, 5km e 1km e foi desenvolvido por Mesinger et al. [9] e tornado operacional no National Centers for Environmental Prediction (NCEP) [1]. Trata-se de um modelo baseado na utilização de coordenadas verticais, o qual permanece aproximadamente horizontal em áreas montanhosas, tornando-o adequado para estudos de regiões de topografia íngreme, como, por exemplo, a cordilheira dos Andes [3]. O modelo Eta é utilizado operacionalmente pelo CPTEC/INPE desde 1997 para previsões meteorológicas sobre a América do Sul e, desde 2002, para previsões sazonais do clima [3].

Este trabalho visa analisar a aplicação do método de calibração estatística para modelos de simulação Model Output Calibration (MOC) na correção da previsão do tempo gerada pelo modelo Eta 15km do CPTEC/INPE.

2 Métodos de Correção de Modelos de Simulação

Apesar dos métodos de previsão estarem cada vez mais sofisticados, com resultados bastante precisos, muitas das variáveis previstas podem sofrer modificações quando a previsão trata de modelos que dependem de variáveis suscetíveis à mudanças, como de ações da natureza. Segundo Mao et al. [8], os modelos de previsão de temperatura são conhecidos por gerarem erros devido a representação grosseira do modelo topográfico e de possuir um modelo físico deficiente. Esses erros ocorrem pois o modelo de ponto de grade considera um ponto único por caixa, pressupondo que as variáveis climatológicas comportam-se de maneira homogênea dentro de cada caixa. Por isso, alguns métodos estatísticos denominados de “calibração” foram criados com o objetivo de corrigir ou de reduzir esses erros. A calibração dos dados é o processo de identificar possíveis problemas no modelo de previsão e de realizar ajustes nos cálculos (códigos). A seleção dos parâmetros utilizados para calibração dos dados é uma questão técnica, onde cada método estatístico trata de uma maneira própria.

Entre os principais métodos de correção dos modelos de simulação pode-se citar o Perfect Prognosis Method (PPM), também conhecido como Perfect Prog [7], o Model Output Statistics (MOS) [5] e o Model Output Calibration (MOC) [8].

3 Model Output Calibration

O MOC tem como princípio a existência de erro na previsão que pode ter sido influenciada por outras variáveis da saída do modelo de simulação. Seu esquema consiste em formular equações de regressão linear multivariada com base nos dados de séries temporais de poucas semanas das previsões e das observações, estimando e ajustando assim os erros da previsão da variável a ser corrigida. No caso dessa pesquisa estão sendo corrigidas as variáveis da temperatura do ar a 2 metros da superfície e a da umidade relativa.

A grande diferença do MOC em relação ao PPM e ao MOS é o seu curto prazo de armazenamento de previsão e de observações necessárias para realização da correção da previsão. Esse prazo varia de duas a quatro semanas enquanto o MOS, por exemplo, exige cerca de dois anos de dados observados e de previsão o que gera uma deficiência no método, pois ele não oferece a possibilidade de desenvolver grande base de dados, visto que os modelos numéricos de previsão do tempo são continuamente alterados e, para a execução do MOS, seria necessário que durante esse período de armazenamento dos dados a configuração do modelo fosse mantida sem alterações [6].

Como resultado da correção aplicada pelo MOC obtêm-se o erro da previsão em relação a uma determinada variável, conforme a equação (1):

$$\Delta T_{(ij)}(k) = T_{(ij)}(k) - T_{(ij)}^*(k) \quad (1)$$

onde $T_{(ij)}(k)$ representa uma série de previsões; (k) representa uma série de valores da variável em diferentes tempos ($k = 1, 2, 3, \dots, N$); (i, j) refere-se ao ponto da grade onde está sendo verificado os valores, ou seja, a localização; e $T_{(ij)}^*(k)$ refere-se a série de dados observados. Logo, com o valor previsto em determinado instante para determinado lugar, subtraído do valor observado, no mesmo instante e lugar, obtêm-se o erro da previsão. Assim, assume-se que $\Delta T_{(ij)}(k)$ (erro da previsão) está intrinsecamente associado com vários parâmetros e variáveis de saída do modelo.

Logo, o erro da previsão pode ser uma função de $X_{ij,l}(k)$ que representa a l -ésima quantidade de variáveis obtidas ou derivadas de uma recente saída do modelo no mesmo ponto da grade (i, j) , onde o erro previsto está correlacionado com a quantidade L de saídas do modelo. Como exemplo, Chou et al. [2] mostram que na correção estatística da previsão de temperatura há diversas variáveis associadas, tais como variáveis próximas ao solo (umidade do solo, temperatura da superfície, entre outras). Com base nos dados observados e nas previsões geradas é possível construir um modelo de regressão linear multivariada para obter o erro da previsão do modelo, conforme a equação (2):

$$\Delta T_{(ij)}(k) = \Lambda_o + \sum_{l=1}^L \Lambda_l X_{ij,l}(k) \quad (2)$$

onde $k = 1, 2, \dots, N$ e $l = 1, 2, \dots, L$. Os coeficientes Λ_o e Λ_l podem ser estimados por regressão linear multivariada. O $\Delta T_{(ij)}(k)$ é o preditando das equações, ou seja, a variável dependente, e $X_{ij,l}(k)$ são as variáveis de saída do modelo, que representam os preditores. De acordo com Mao et al. [8], a seleção de preditores no MOC ocorre a cada novo ciclo da previsão, com base em um conjunto de variáveis candidatas, ou seja, a cada nova rodada da correção os preditores são selecionados (equação (2)), diferente de outros modelos de correção estatística onde os preditores são definidos *a priori*. Consequentemente, o número de preditores selecionados (Lc), diretos ou derivados, deve ser menor que o número de variáveis candidatas (L). A calibração é aplicada após a correção $N + 1$, onde N é o tamanho da série para cálculo dos coeficientes de regressão, obtendo-se, desse modo, a previsão refinada $T_{moc}(k)$, conforme a equação (3):

$$T_{moc}(k) = T_{(ij)}(k) - \Delta T_{(ij)}(k) \quad (3)$$

onde $T_{(ij)}(k)$, representa a previsão do modelo corrente e $\Delta T_{(ij)}(k)$ representa o erro no mesmo ponto da grade. Devido ao fato do MOC necessitar de um curto período de dados para realização do ajuste da correção, é possível realizar ajustes em modelos independente de mudanças ambientais ocorridas durante o período de previsão, tornando-o uma ótima ferramenta para modelos regionais.

Mao et al. [8] mostram outras vantagens da utilização do MOC: é uma técnica simples, flexível e eficiente; baseia-se nas condições climáticas mais recentes; a seleção de preditores ocorre de forma objetiva e, de certa forma, ideal, pois após os preditandos e variáveis candidatas serem especificados, o resto do processo ocorre de forma objetiva e iterativa; o algoritmo utilizado para correção estatística do modelo de previsão é flexível e fácil de implementar para qualquer modelo, pois a equação é atualizada em cada ciclo da previsão e é menos sensível a mudanças ou reconfigurações do modelo; a informação dos resultados da seleção dos preditores pode ser utilizada em diagnósticos e na sintonização do modelo físico pois, por meio dessa seleção, é possível verificar as variáveis que estão influenciando no erro da previsão.

Ainda segundo Mao et al. [8], para avaliar se as calibrações das previsões apresentam um ganho na relação entre o valor previsto e o valor observado, são utilizadas três medidas estatísticas básicas de erro: o erro médio (BIAS) - (equação (4)), erro absoluto médio (MAE - em inglês *mean absolute error*) - (equação (5)), e a raiz quadrada do erro quadrático médio (RMSE - em inglês *root mean square error*) - (equação (6)). Por meio dessas medidas é possível determinar a qualidade da previsão e de sua calibração, além de permitir determinar pontos fracos e fortes do sistema.

$$BIAS = \frac{1}{M_{total}} \sum_{i=1}^{M_{total}} (T_i - T_i^*) \quad (4)$$

$$MAE = \frac{1}{M_{total}} \sum_{i=1}^{M_{total}} |T_i - T_i^*| \quad (5)$$

$$RMSE = \sqrt{\frac{1}{M_{total}} \sum_{i=1}^{M_{total}} (T_i - T_i^*)^2} \quad (6)$$

Nas equações (4), (5) e (6), M_{total} é a quantidade total de previsões corrigidas, T_i é o i -ésimo valor corrigido pelo MOC e T_i^* é o i -ésimo valor observado correspondente. T_i poderá indicar o valor não corrigido, conseqüentemente, sendo igual ao obtido pelo modelo de previsão. Para a análise dos resultados obtidos no BIAS e no MAE, quanto mais próximo de ZERO for o resultado melhor ele é. O resultado do RMSE é um valor não-negativo, porém, é mais sensível a grandes erros, pois, conforme [4], elevando a diferença ao quadrado, força tratar igualmente resultados negativos e positivos.

4 Análise da Aplicação do MOC na Correção de Previsões do Modelo Eta

Esta sessão analisa os resultados obtidos pela aplicação do MOC na correção da temperatura e da umidade relativa geradas pelo modelo regional Eta 15km no período de dezembro de 2015. As localidades abordadas na análise fazem parte de um conjunto de 36 localidades integrantes de um projeto entre o Programa de Pós-Graduação em Computação Aplicada (PPGCA) da Universidade de Passo Fundo (UPF) e o CPTEC/INPE destinado ao Operador Nacional do Sistema Elétrico (ONS). Esse projeto visa a melhoria da qualidade do processo de previsão de consumo de carga de energia no Brasil por meio dessas correções³. Essas localidades contemplam estações meteorológicas localizadas em aeroportos das capitais do Brasil e em algumas cidades do interior de São Paulo e da região Sul situadas em regiões onde há uma maior demanda por energia devido a agricultura e ao processo de irrigação. O sistema computacional atual foi desenvolvido para o sistema operacional Linux, utilizando *scripts shell* Linux e a linguagem Fortran. Para os cálculos estatísticos utilizou-se a linguagem estatística R.

Tabela 1: Médias diárias da temperatura ($^{\circ}C$) do ar observada a 2m da superfície, prevista pelo modelo Eta e prevista pelo modelo Eta com a aplicação da correção pelo MOC com suas correspondentes medidas estatísticas básicas de erro (BIAS, MAE e RMSE) para o mês de dezembro de 2015 para algumas localidades do Brasil

Localidade	Obs	Eta	MOC	Eta			MOC		
				Bias	Mae	Rmse	Bias	Mae	Rmse
Belém, PA	27,24	26,68	27,24	-0,56	1,23	1,68	0,00	0,93	1,35
Belo Horizonte, MG	24,55	22,15	24,66	-2,40	2,68	3,11	0,11	1,56	2,07
Brasília, DF	23,22	21,90	23,47	-1,32	2,16	2,73	0,24	1,71	2,30
Campo Grande, MS	25,21	27,47	25,04	2,26	2,67	3,52	-0,17	2,17	2,76
Chapecó, RS	23,17	21,51	22,65	-1,66	2,74	3,43	-0,52	2,30	2,88
Florianópolis, SC	23,93	23,92	23,49	0,00	1,56	2,02	-0,44	1,45	1,86
Fortaleza, CE	28,04	26,47	27,98	-1,57	1,72	1,91	-0,06	0,67	0,97
Foz do Iguaçu, PR	24,69	26,71	24,07	2,01	2,53	3,21	-0,62	2,12	2,62
Goiânia, GO	25,37	24,45	25,55	-0,93	1,99	2,49	0,17	1,76	2,36
João Pessoa, PE	26,66	26,08	26,77	-0,58	1,38	1,74	0,10	1,00	1,67
Londrina, PR	24,55	23,87	24,04	-0,68	1,99	2,52	-0,51	1,87	2,38
Maceió, AL	26,25	24,87	26,27	-1,38	1,68	2,00	0,02	0,92	1,60
Manaus, AM	28,66	27,18	28,83	-1,48	2,06	2,51	0,17	1,41	1,84
Maringá, PR	24,33	25,31	23,78	0,99	2,01	2,55	-0,54	2,07	2,59
Porto Alegre, RS	23,97	23,69	23,24	-0,29	1,84	2,30	-0,73	1,94	2,45
Ribeirão Preto, SP	25,41	24,25	25,32	-1,16	2,18	2,77	-0,09	1,85	2,51
Salvador, BA	27,73	26,21	27,65	-1,52	1,74	1,95	-0,09	0,76	1,02
Teresina, PI	31,79	29,46	31,62	-2,33	2,91	3,47	-0,17	1,28	1,92

³Os resultados dessas correções são disponibilizados diariamente pelo CPTEC/INPE em <ftp://ftp1.cptec.inpe.br/etamd1/moc/>.

Durante o período de análise dos dados foi possível identificar possíveis gargalos da previsão, onde os resultados gerados pelo modelo Eta mostraram-se distantes da observação real para algumas localidades como Belo Horizonte, Campo Grande, Foz do Iguaçu e Teresina, como pode ser observado na Tabela 1. Nessa tabela são apresentados os dados observados, as previsões geradas pelo modelo Eta e essas previsões corrigidas pelo MOC para algumas das localidades do projeto com o ONS, referente a variável da temperatura.

O refinamento obtido pelo MOC sugere uma melhora na previsão do Eta, conforme pode-se verificar, por exemplo, na Figura 1, onde as curvas azuis referem-se as medidas observadas na estação do aeroporto de Londrina, as linhas em vermelho referem-se as medidas previstas pelo modelo Eta e as linhas em verdes referem-se aos dados refinados pelo MOC. Os ciclos diurnos médios da temperatura e da umidade relativa podem indicar períodos do dia em que erros geralmente ocorrem na previsão pelo modelo Eta, onde verificou-se que, em geral, a previsão da temperatura acompanha corretamente a variação da amplitude, mas, na maioria do tempo, superestimou as medidas. Já na previsão da umidade relativa os erros maiores ocorreram entre 9:00 e 22:00 UTC quando as medidas previstas foram subestimadas.

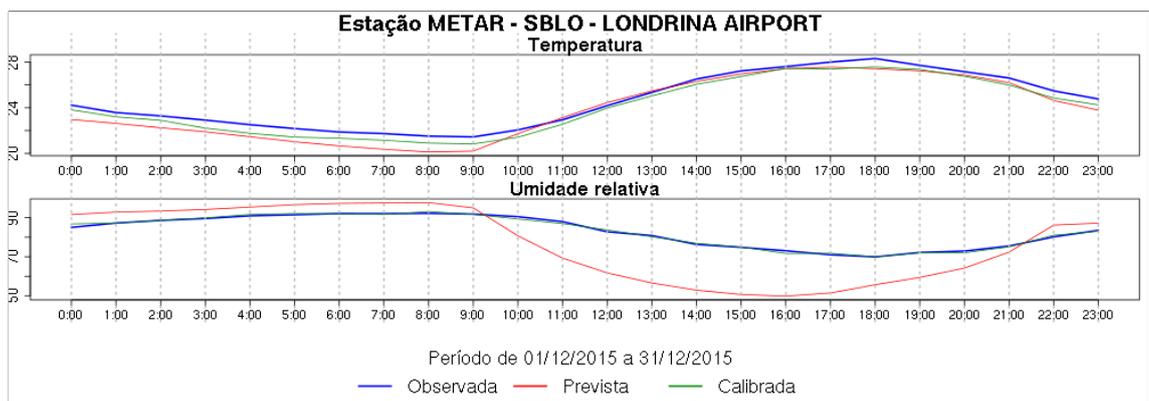


Figura 1: Ciclo diurno médio da temperatura ($^{\circ}\text{C}$) do ar a 2m e da umidade relativa (%) para a estação meteorológica localizada na cidade de Londrina (PR) no mês de dezembro de 2015, onde os dados observados estão representados pela linha azul, os dados da previsão do modelo Eta pela linha vermelha e os dados da previsão do modelo Eta corrigidos pelo MOC pela linha verde.

5 Conclusões e Trabalhos Futuros

Este artigo apresentou a análise da aplicação do MOC para o refinamento estatístico no modelo de previsão numérica do tempo Eta do CPTEC/INPE. A aplicação do refinamento apresentou um aperfeiçoamento favorável sobre a previsão original do modelo Eta, resultando em uma melhora na qualidade da previsão gerada para as variáveis de temperatura e umidade relativa. O curto período necessário para a realização da calibração e, conseqüentemente, sua fácil adaptação à mudanças no modelo a ser corrigido, é uma de suas vantagens sobre outros métodos de calibração. Após as análises realizadas

em algumas localidades do Brasil pôde-se verificar que haviam erros nos ciclos diurnos da previsão, onde em determinados momentos do dia, frequentemente, a previsão gerada era subestimada ou superestimada.

Como trabalhos futuros esta pesquisa possibilitará verificar a validade e os possíveis ganhos de se utilizar os dados de previsão do modelo Eta corrigidos pelo MOC quando eles forem utilizados nos modelos de simulação de culturas e de doenças de culturas que são utilizados em pesquisas desenvolvidas no PPGCA/UPF em parceria com a Embrapa Trigo, a Universidade da Flórida (UF) e o CPTEC/INPE. Estas pesquisas envolvem as culturas do trigo, morango e maçã e doenças como a Brusone que ataca a cultura do trigo.

Referências

- [1] T. L. Black. The new NMC mesoscale Eta model: Description and forecast examples. *Weather and Forecasting*, 9(2):265–278, 1994.
- [2] S. C. Chou, C. R. Souza, J. L. Gomes, E. F. D. Evangelista, C. Osório, and M. Cataldi. Refinamento estatístico das previsões horárias de temperatura a 2 m do modelo Eta em estações do nordeste do Brasil. *Revista Brasileira de Meteorologia*, 22(3):287–296, 2007.
- [3] S. C. Chou, A. Lyra, C. Mourão, C. Dereczynski, I. Pilotto, J. Gomes, J. Bustamante, P. Tavares, A. Silva, D. Rodrigues, D. Campos, D. Chagas, G. Sueiro, G. Siqueira, P. Nobre, and J. Marengo. Evaluation of the Eta Simulations Nested in Three Global Climate Models. *American Journal of Climate Change*, 3(05):438–454, 2014.
- [4] S. P. Ferreira, F. Varone, and R. C. M. Alvez. Uso do trmm para a correção estatística da chuva prevista pelo modelo BRAMS. *IV Simpósio Brasileiro de Ciências Geodésicas e Tecnologias da Geoinformação*, pages 01–09, 2012.
- [5] H. R. Glahn, , and D. A. Lowry. The use of model output statistics (mos) in objective weather forecasting. *J. Appl. Meteor.*, 11:1203–1211, 1972.
- [6] E. J. B. Jacks, V. J. Bower, J. P. Dagoistro, M. C. Dallavalle, and J. C. S. Erickson. New ngm-based mos guidance for maximum/minimum temperature, probability of precipitation, cloud amount, and sea surface wind. *Wea. Forecasting*, 5:128–138, 1990.
- [7] B. Klein, M. Lewis, and I Enger. Objective prediction of 5 day mean temperature during winter. *Journal of Meteorology*, 16:672–682, 1959.
- [8] Q. Mao, R. T. McNider, S. F. Muellere, and H. Juang. An optimal model output calibration algorithm suitable for objective temperature forecasting. *American Meteorological Society*, 14:190–202, 1998.
- [9] F. Mesinger, Z. I. Janjić, S. Ničković, D. Gavrilov, and D. G. Deaven. The step-mountain coordinate: Model description and performance for cases of alpine lee cyclogenesis and for a case of an appalachian redevelopment. *Monthly Weather Review*, 7(116):1493–1518, 1988.