

## CALIBRAGEM ESTATÍSTICA DAS PREVISÕES DE TEMPERATURA DO MODELO ETA

Luiz Rodrigo Lins Tozzi e Maria Gertrudes Alvarez Justi da Silva

Departamento de Meteorologia – Instituto de Geociências

Universidade Federal do Rio de Janeiro

Rio de Janeiro – RJ, Brasil

e-mail: [tozzi@ufrj.br](mailto:tozzi@ufrj.br)

[justi@acd.ufrj.br](mailto:justi@acd.ufrj.br)

José Alberto da S. Ferreira e Chou Sin Chan

Centro de Previsão de Tempo e Estudos Climáticos (CPTEC)

Instituto Nacional de Pesquisas Espaciais (INPE)

Rodovia Presidente Dutra, km 40 SP-RJ

12630-000, Cachoeira Paulista - SP, Brasil

e-mail: [alberto@cptec.inpe.br](mailto:alberto@cptec.inpe.br)

[chou@cptec.inpe.br](mailto:chou@cptec.inpe.br)

### ABSTRACT

This study make use of a statistic tool to calibrate model outputs of air temperature using a new approach called Model Output Calibration (MOC), which is based on the principle that today's forecasting error is related to last days' forecasting errors. Obtaining an estimation of forecasting error from a regression equation makes this calibration. This equation uses a matrix containing the last forecasts of the model and their respective forecasting errors. This method has improved ETA's temperature forecasts skill by 25%, in the majority of the cases analyzed.

### INTRODUÇÃO

Por melhores e eficazes que estejam se tornando, os modelos de previsão numérica não são perfeitos. Enquanto os computadores ficam cada ano mais rápidos e baratos, ainda existem problemas que permanecem mesmo quando se aumenta a resolução da grade dos modelos atmosféricos, se melhora os dados de entrada e se sofisticam as parametrizações. Isso sem contar as imprecisões associadas às equações usadas e os métodos para resolvê-las.

Esses problemas têm sido amenizados por formas alternativas de tratar os dados de entrada e saída. Alguns são métodos estatísticos que procuram uma relação entre uma variável de interesse e as demais variáveis prognosticas pelo modelo ao longo de um determinado tempo. Na literatura científica têm se usado muito duas abordagens, o MOS (Model Output Statistics) e o PPM (Perfect Prognostic Model) (Glahn & Lowry, 1972 e Matsuo, 1992).

Neste trabalho propõe-se uma nova alternativa de metodologia, uma calibragem da saída do modelo (MOC – Model Output Calibration), para suplantiar alguns problemas muito comuns no uso dos métodos acima citados (Mao et al., 1998).

A grande diferença da calibração através do MOC para os demais métodos estatísticos é a sua flexibilidade na utilização de dados. Em modelos de mesoescala, que mudam mais freqüentemente, nenhum dos dois métodos estatísticos clássicos são facilmente aplicáveis. Eles demandam vários anos de dados de previsões mantendo a mesma configuração e as mesmas características do modelo, algo que consome muito tempo para ser conseguido e quando se muda qualquer característica do modelo faz com que novos dados de previsões tenham que ser novamente conseguidos por um grande período para que sejam conseguidas equações de regressão adequadas. Já o MOC só precisa de algumas semanas de dados, o que faz dele uma alternativa ideal para o uso em modelos de mesoescala.

Neste trabalho testa-se o método MOC na melhoria da previsão da variável temperatura do ar do modelo ETA. Em cooperação com o Centro de Previsão do Tempo e Estudos Climáticos / Instituto Nacional de Pesquisas Espaciais (CPTEC/INPE), decidiu-se testar o método nas cidades de Cuiabá, Belo Horizonte e Porto Alegre, nos horizontes de previsão de 6 e 18h (aproximadamente no horário das mínimas e das máximas diárias), para ver como o MOC se comportava em diferentes regiões do país.

Este método se mostrou uma boa alternativa para implementação operacional em modelos de mesoescala, sem demandar tanto tempo e recursos com o armazenamento de anos de dados de previsões.

## METODOLOGIA

A maioria dos métodos estatísticos usados para aprimorar saídas de modelos numéricos tem como objetivo estimar o valor da variável de interesse. Isto é feito utilizando uma regressão linear múltipla, que ajusta a melhor relação entre as várias variáveis previstas pelo modelo e a variável de interesse. São utilizadas matrizes com dados observados e de previsões passadas e, após determinar quais as variáveis do modelo mais correlacionadas com a que nos interessa, elas são selecionadas para a equação de regressão.

O algoritmo MOC se diferencia em alguns pontos deste procedimento. Além do tamanho da base de dados ser muito menor (algumas semanas, ao invés de anos), o MOC não estima o valor da variável em questão, mas sim o erro de previsão da variável dado pela diferença entre o valor previsto pelo modelo ( $T_{mod}$ ) e o valor realmente observado da variável ( $T_{obs}$ ):

$$\Delta T = T_{MOD} - T_{OBS} \quad (1)$$

A idéia básica por trás desta estatística é que esse erro de previsão pode, nos últimos dias, ter sido influenciado de alguma forma por outras variáveis de saída do modelo e que este erro continuará a ter esta mesma influência nos próximos dias. Neste caso, pode-se pensar em  $\Delta T$  como uma função de algumas outras variáveis previstas pelo modelo:

$$\Delta T = F(x_1, x_2, x_3 \dots x_N) \quad (2)$$

A idéia é construir uma matriz onde uma das colunas é constituída pelos erros de previsão e as outras colunas são constituídas por um conjunto de outras variáveis previstas pelo modelo todas tomadas por um período selecionado (usaram-se aqui 21 dias). Deve-se então correlacionar esses erros com as variáveis disponibilizadas pelo modelo no mesmo horário, na mesma rodada e na mesma estação de observação. Após essa etapa, selecionam-se as variáveis mais correlacionadas e cria-se uma equação de regressão para achar a função  $F$  que determina  $\Delta T$ . A equação de regressão final é única para cada dia, horário de previsão e estação meteorológica, renovando-se todo o processo a cada nova integração do modelo. A equação de regressão tem a forma explicitada por:

$$\Delta T = F(x_1, x_2, x_3 \dots x_N) = b_0 + b_1 x_1 + b_2 x_2 + \dots \quad (3)$$

onde  $b_0$ ,  $b_1$  e  $b_2$ , ... são os coeficientes obtidos pela equação de regressão e  $x_1$ ,  $x_2$ , ... são os valores das variáveis correspondentes àqueles coeficientes.

Neste trabalho o objetivo foi calibrar as saídas do modelo a fim de minimizar os erros nas previsões de temperatura. Para isso, usou-se dados da rodada das 0h (GMT) do modelo regional ETA do CPTEC/INPE durante o todo o mês de abril de 2000, para as cidades de Cuiabá, Belo Horizonte e Porto Alegre. Este mês foi escolhido por se ter caracterizado numa mudança de regime do tempo, ocorrendo inclusive uma forte entrada de massa de ar frio sobre o continente sul-americano.

Os dados observados foram extraídos das informações de SYNOP, de forma correspondente à data e ao horário usado para na obtenção da previsão do modelo. Não houve dados de modelo nos dias 7 e 11.

As variáveis usadas da saída do modelo ETA estão listadas na Tabela 1, onde se mostra também a quantidade de vezes que cada variável foi selecionada para entrar numa equação de regressão. O critério de seleção para a equação de regressão foi ter uma correlação de mais de 0,4 (em módulo) com o erro de previsão de temperatura.

Tabela 1 – Total de vezes que as variáveis foram selecionadas para a equação de regressão

Variáveis	Belo Horizonte		Cuiabá		Porto Alegre		Total
	06	18	06	18	06	18	
Temperatura Absoluta	16	3	6	0	3	0	28
Pressão (Método Mesinger)	19	1	7	2	0	9	38
Pressão na Superfície	19	2	7	2	0	9	39
Temp. em 2m	25	26	15	9	2	0	77
Temp. do pnt. de orvalho em 2m	8	10	15	8	1	0	42
Vento Zonal 10m	25	4	0	0	13	5	47
Vento Meridional 10m	16	0	7	0	0	3	26
Precipitação Total	25	8	7	14	0	1	55
Precipitação Convectiva	25	8	7	14	0	1	55
Temperatura na Superfície	25	24	24	13	1	0	87
Temperatura no Solo	25	16	24	12	1	0	78
Ondas Curtas Receb. no Solo	25	7	21	18	1	0	72
Ondas Curtas Emit. pelo Solo	25	13	23	12	1	0	74
Ondas Longas Emit. pelo Solo	15	3	1	0	2	3	24
Albedo	19	1	7	2	0	9	38
Altura Geopotencial	25	4	1	0	15	5	50
Vento Zonal	3	0	5	0	0	3	11
Vento Meridional	5	18	0	7	20	4	54
<b>Total</b>	<b>345</b>	<b>148</b>	<b>177</b>	<b>113</b>	<b>60</b>	<b>52</b>	

## RESULTADOS E DISCUSSÃO

Para a avaliação dos resultados do MOC foram usadas técnicas estatísticas e avaliações gráficas. Nas técnicas estatísticas é usada uma medida de “skill score” do erro quadrático (SS), que determina o quanto o MOC melhorou os resultados do modelo,

$$SS = \frac{E_{ETA} - E_{MOC}}{E_{ETA}} \quad (4)$$

onde,  $E_{ETA} = \sqrt{\frac{1}{M} \sum_{i=1}^M (T_{MOD(i)} - T_{OBS(i)})^2}$ ,  $E_{MOC} = \sqrt{\frac{1}{M} \sum_{i=1}^M (T_{MOC(i)} - T_{OBS(i)})^2}$  e M é o número total de saídas de previsões verificadas.

Valores positivos do “skill score” determinam melhorias do MOC, enquanto valores negativos indicam valores piores que as previsões feitas pelo modelo. A Tabela 2 apresenta os resultados de SS em cada estação meteorológica nos dois horários analisados. Denomina-se aqui SS de índice de melhoria.

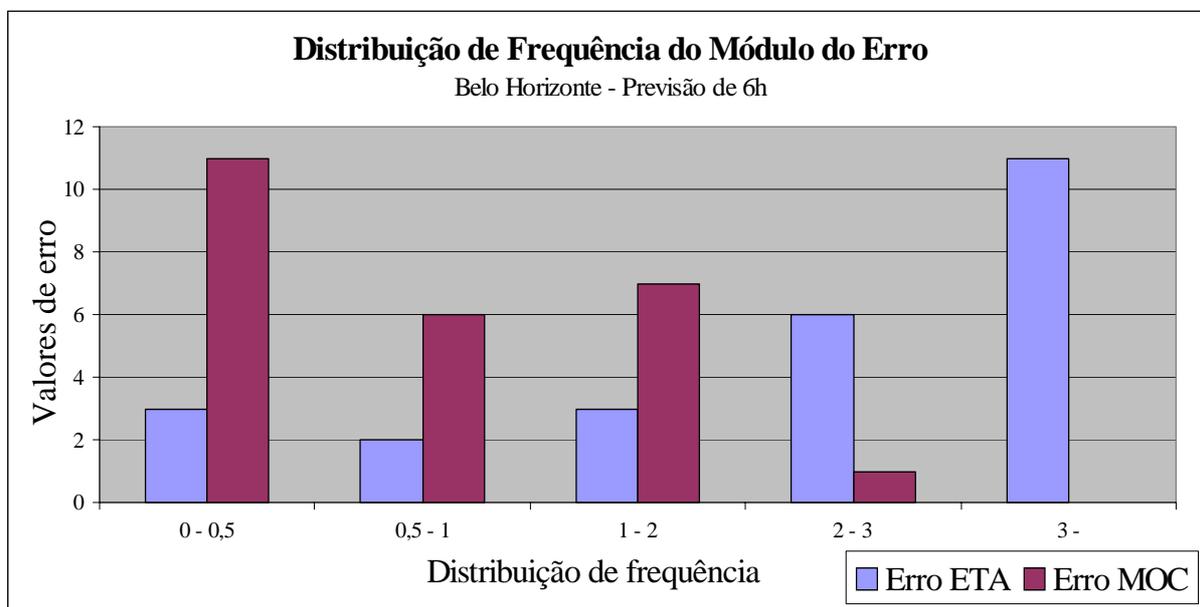
Tabela 2 – Índice de melhoria para cada local e hora de execução do modelo.

Configuração usada	Índice de Melhoria
Belo Horizonte – 06 horas	64,031%
Belo Horizonte – 18 horas	30,852%
Cuiabá – 06 horas	13,404%
Cuiabá – 18 horas	6,847%
Porto Alegre – 06 horas	47,136%
Porto Alegre – 18 horas	16,530%
<b>Média</b>	<b>29,800%</b>

Para as análises gráficas são usados dois tipos de gráficos: comparações entre os valores previstos de temperatura com e sem o MOC em relação aos dados observados e também um gráfico da distribuição de frequência do módulo do erro.

É importante ressaltar que o MOC se comporta de maneiras diferentes em função das características da estação analisada. Por isso, é interessante analisar não somente o quanto o MOC melhorou ou piorou os resultados, mas também tentar analisar o porquê dessa calibragem ter sido necessária. Para isso é útil também comparar os resultados de SS com os dois tipos de análises gráficas.

Nenhuma das três cidades tiveram valor negativo para o índice SS, o que por si só já representa um ponto



positivo para o método MOC.

Belo Horizonte foi a cidade com melhor performance do MOC. É possível ver essa melhora não somente nos resultados da Tabela 2, mas também nos gráficos relativos à esta cidade.

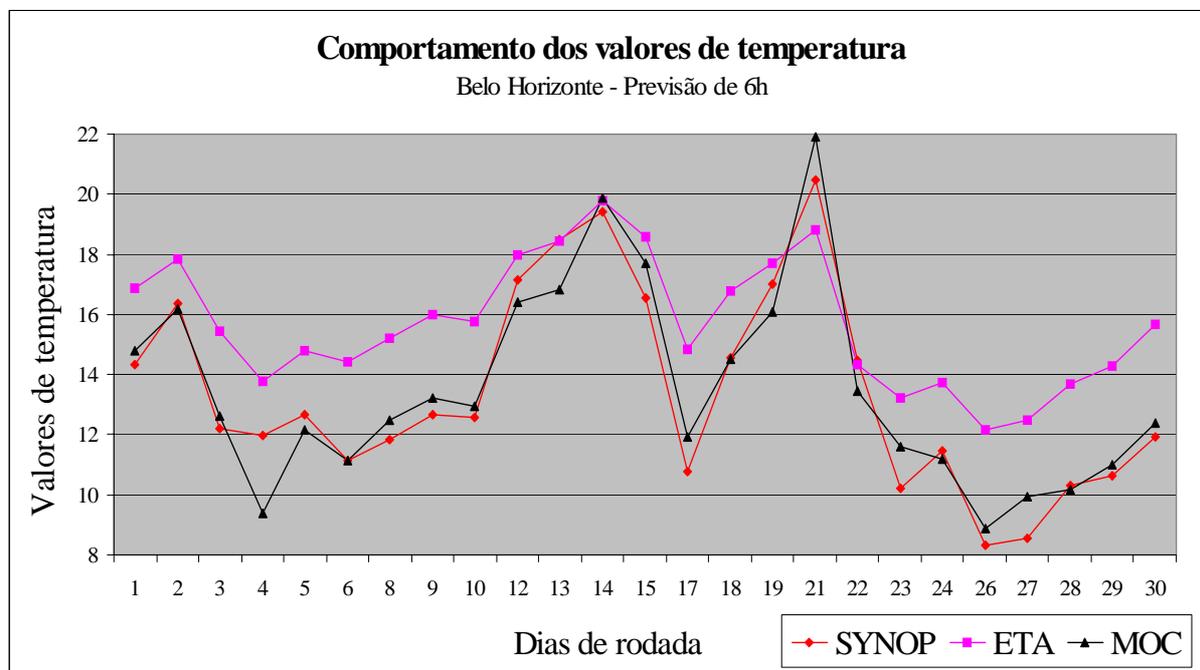
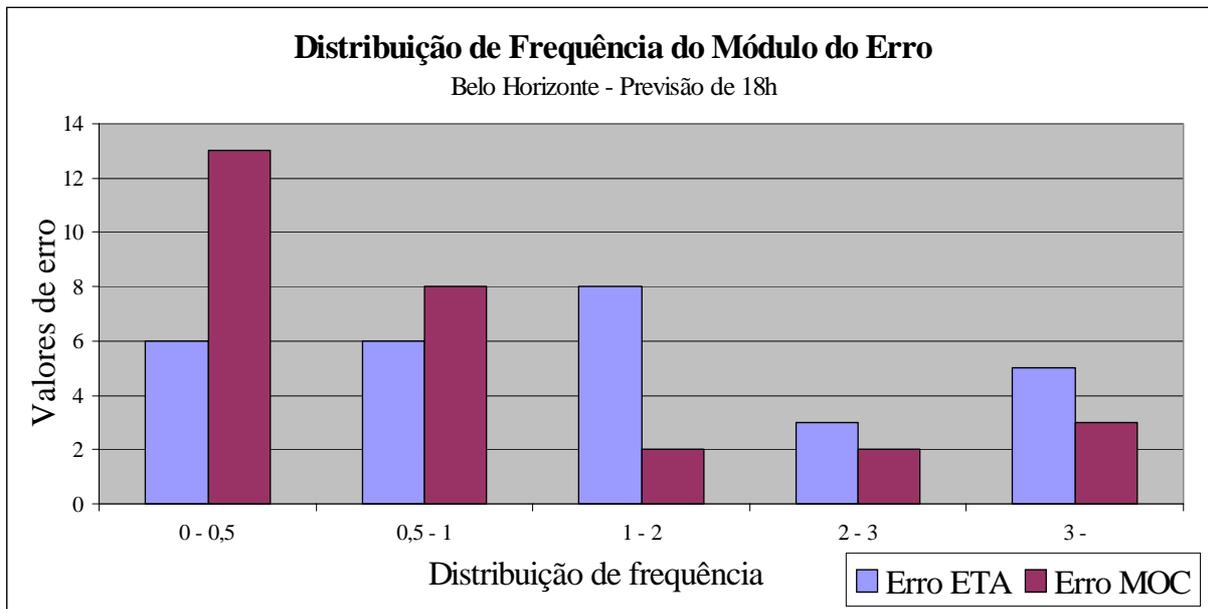


Figura 1 – Gráfico de distribuição de frequência dos módulos dos erros do modelo com e sem a calibração do MOC, na previsão de 6 horas de Belo Horizonte.

Figura 2 – Gráfico do comportamento dos valores de temperatura do modelo com e sem a calibração do MOC e os valores observados de SYNOP, na previsão de 6 horas de Belo Horizonte.

Nos gráficos de distribuição de frequência do módulo do erro de 6 e 18 horas (Figuras 1 e 3) pode-se ver uma forte concentração de erros do MOC nas faixas entre 0 e 1°C, o que indica um bom desempenho do método. Em contrapartida, os erros do ETA ficaram bem espalhados nas categorias de frequência da previsão de 18h (Figura 1) e extremamente concentrado na categoria de mais de 3°C de erro na previsão de 6h. Isso indica que o



modelo ETA, especialmente na previsão de 6h, apresentou maiores problemas e o MOC conseguiu amenizá-los.

Pode-se ver também esse problema na previsão do ETA nas Figuras 2 e 4, onde a curva do modelo nem sequer acompanha o comportamento dos valores observados do SYNOP.

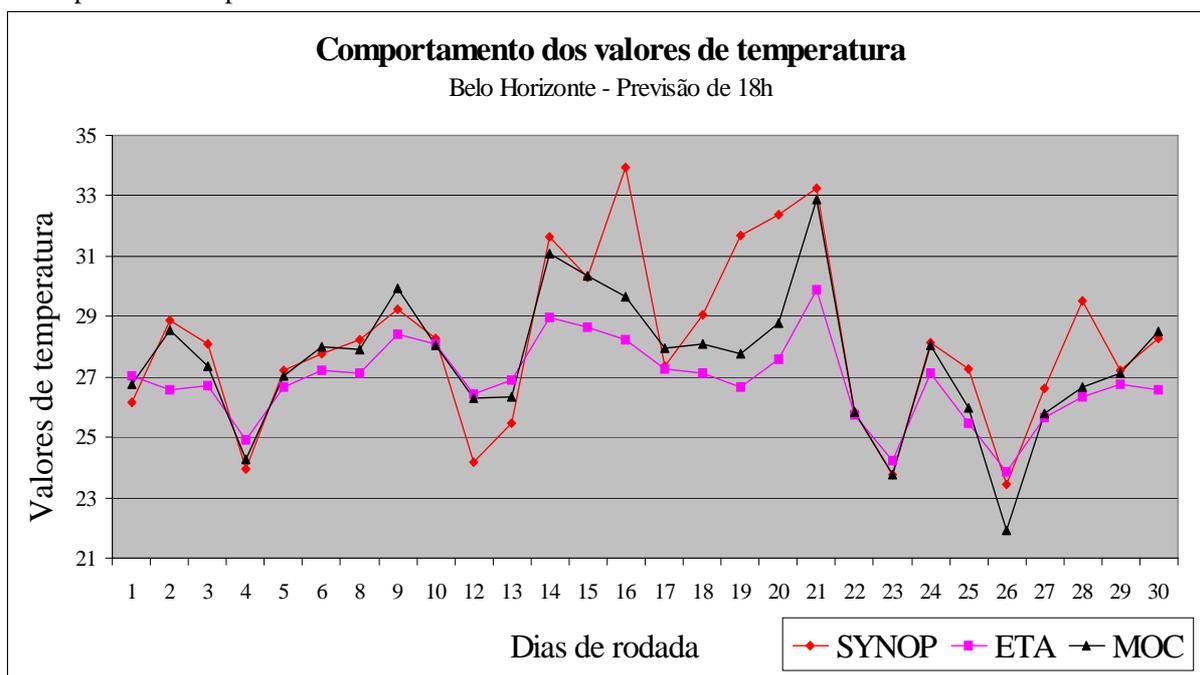
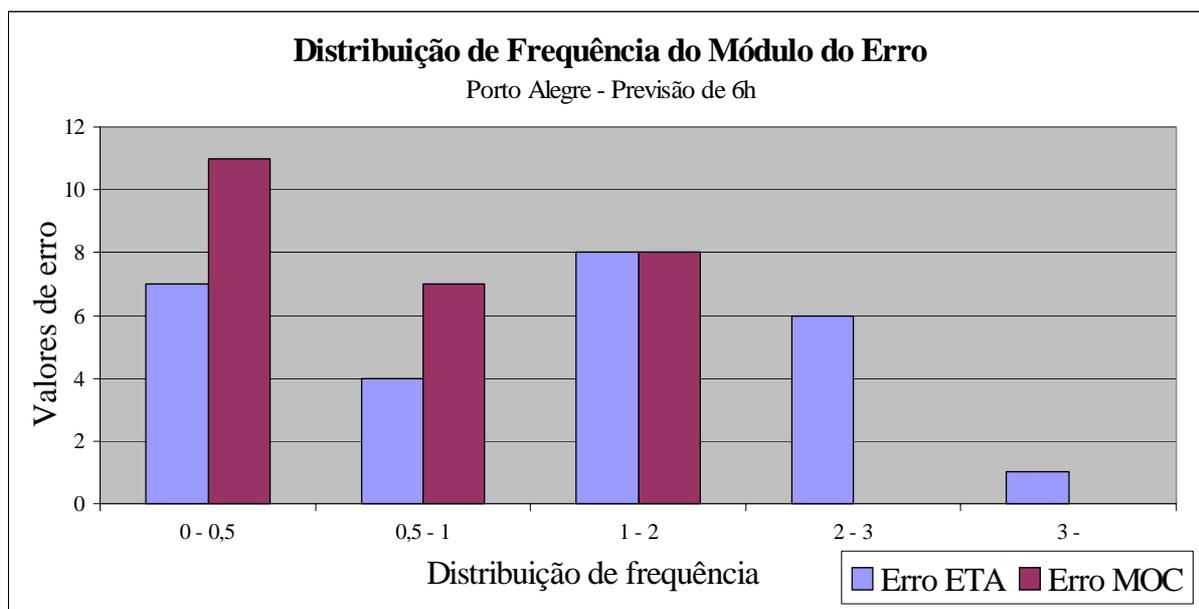


Figura 3 – Gráfico de distribuição de frequência dos módulos dos erros do modelo com e sem a calibração do MOC, na previsão de 18 horas de Belo Horizonte.

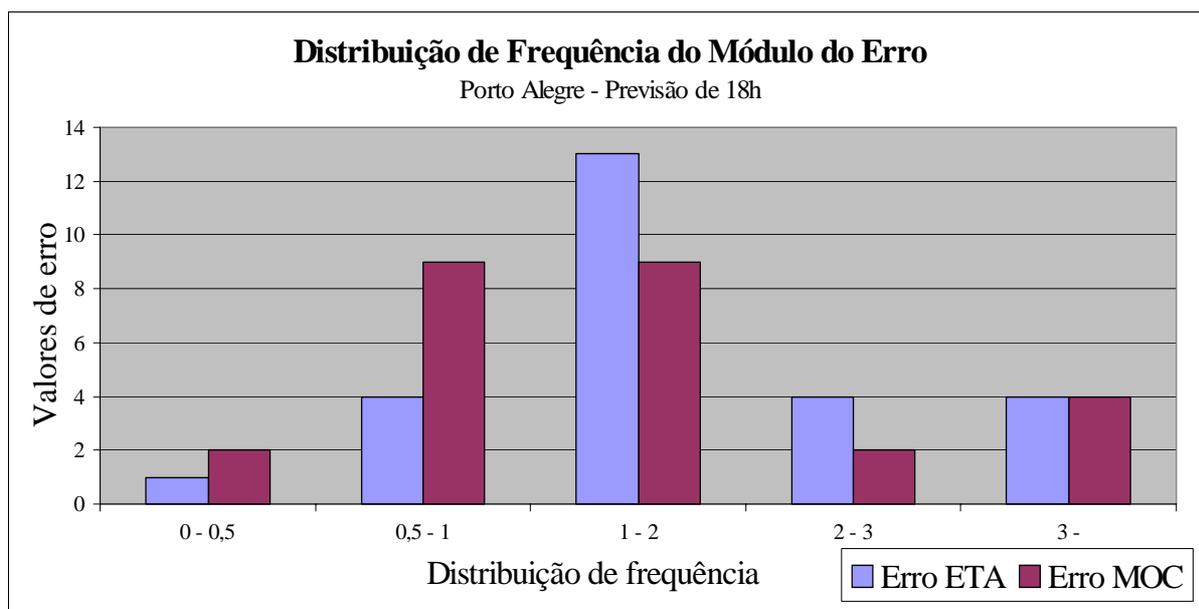
Figura 4 – Gráfico do comportamento dos valores de temperatura do modelo com e sem a calibração do MOC e os valores observados de SYNOP, na previsão de 18 horas de Belo Horizonte.

Vale lembrar que o MOC não tem como, por si só, acompanhar os valores observados. O MOC é apenas uma correção do que o modelo prevê, ou seja, se aproxima sem um compromisso físico, apenas matemático. Isso é visto por exemplo entre os dias 17 e 24 da Figura 4. Aqui, o MOC conseguiu acompanhar bem a oscilação do SYNOP, mas fica claro que não há um compromisso físico próprio. Basta ver que em certos momentos o valor fica superestimado e em outros ele é subestimado, sempre tendendo a se aproximar dos valores observados de SYNOP.

Percebe-se no caso de Belo Horizonte que muitas e diferentes variáveis foram usadas no MOC, especialmente na previsão de 6 horas. Porto Alegre foi a segunda cidade em termos de melhorias do MOC. Ao



contrário de Belo Horizonte, os gráficos de distribuição de frequência de 6 e 18hs (Figuras 5 e 6) são completamente diferentes entre si e existe uma grande diferença entre os índices de melhoria nos dois horários



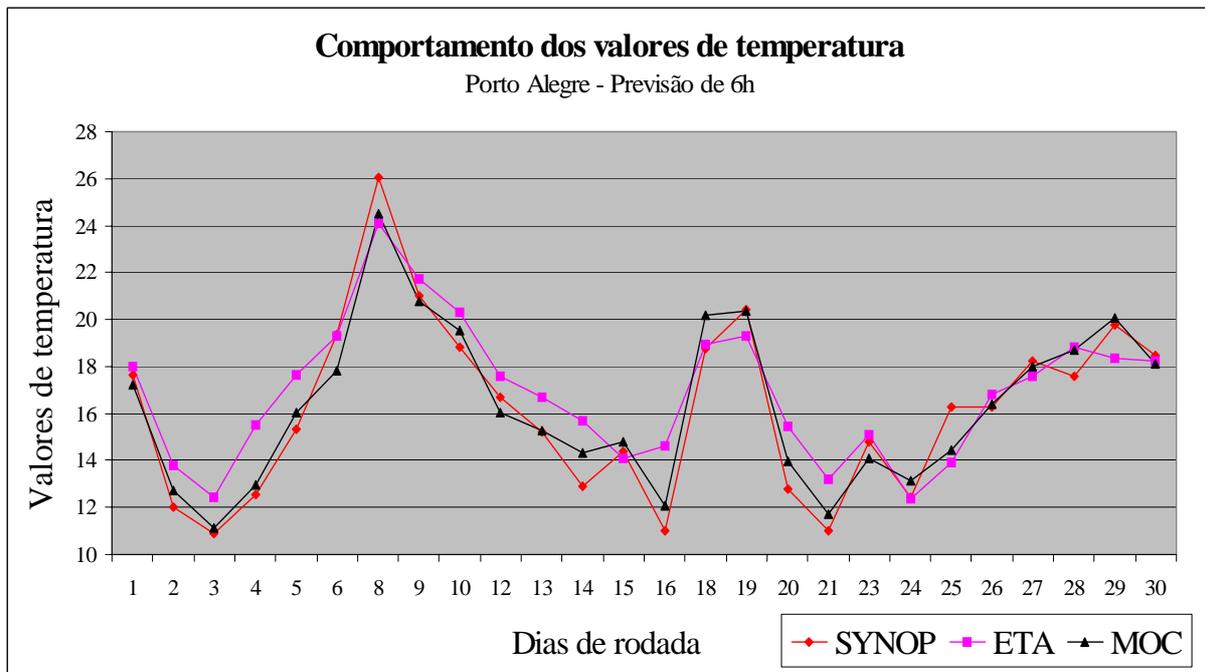
(47,13% e 16,53%, respectivamente).

Figura 5 – Gráfico de distribuição de frequência dos módulos dos erros do modelo com e sem a calibração do MOC, na previsão de 6 horas de Porto Alegre.

Figura 6 – Gráfico de distribuição de frequência dos módulos dos erros do modelo com e sem a calibração do MOC, na previsão de 18 horas de Porto Alegre.

Quando se acompanha o comportamento dos valores de temperatura nas Figuras 7 e 8, pode-se ver que no primeiro caso o MOC acompanhou bem a oscilação da curva dos valores observados, só se desviando um pouco nos últimos dias. Neste caso, o índice de melhoria chegou a 47,13%. Já na previsão das 18hs (Figura 8), nem o MOC nem o ETA foram muito bem. Ambos tiveram mais dificuldade em pegar a tendência do comportamento da temperatura observada. Existe neste caso a possibilidade de que os dados do SYNOP estejam com algum erro e isso não pode agora ser revisado.

Outro ponto importante é a análise da qualidade das variáveis selecionadas para a equação de regressão. Como se pode ver na Tabela 1, as previsões para Porto Alegre foram as que menos selecionaram variáveis. Mesmo assim, o índice de melhoria da previsão de 6 horas foi muito bom. Nota-se uma seleção de variáveis relacionadas com entradas de frente, tais como vento meridional e altura geopotencial. Já na previsão de 18 horas, estas



variáveis não foram tão significativas para a região, mostrando que outros fatores ajudam a definir a temperatura neste horário.

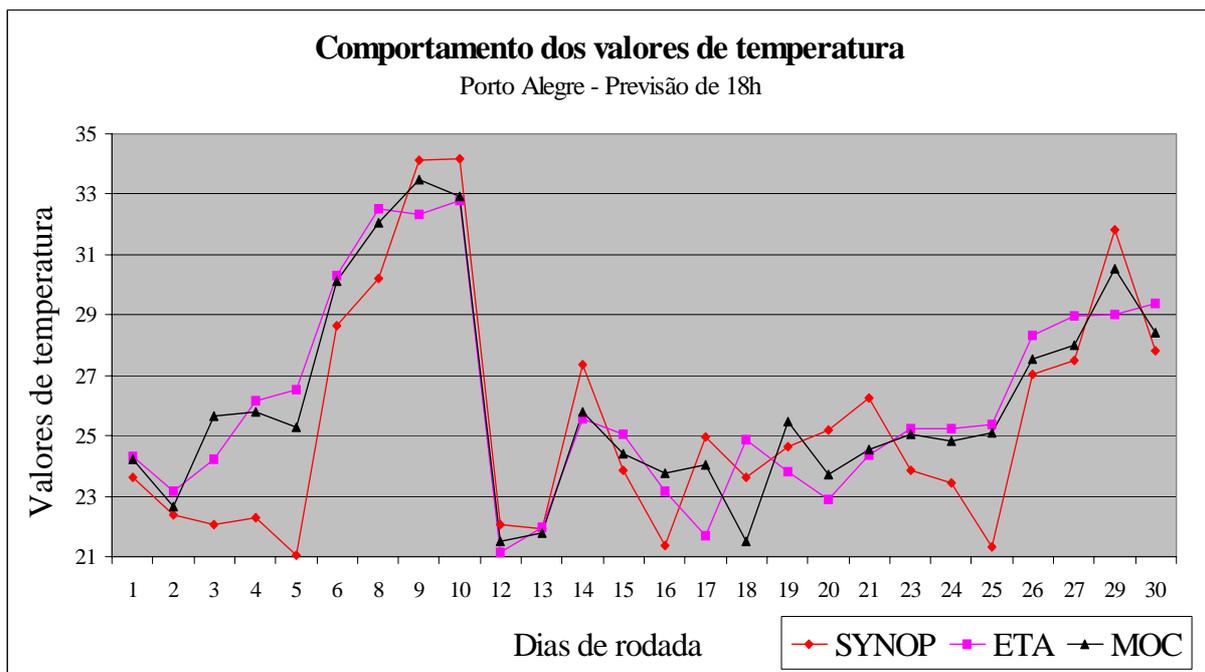


Figura 7 – Gráfico do comportamento dos valores de temperatura do modelo com e sem a calibração do MOC e os valores observados de SYNOP, na previsão de 6 horas de Porto Alegre.

Figura 8 – Gráfico do comportamento dos valores de temperatura do modelo com e sem a calibração do MOC e os valores observados de SYNOP, na previsão de 18 horas de Porto Alegre.

A pior das três cidades foi certamente Cuiabá. Não somente o índice de melhoria dos dois horários mostra isso (13,4% e 6,84%, respectivamente), mas as Figuras 9 e 10 também mostram um nivelamento muito grande na distribuição das freqüências, especialmente no horário de 18h. Em outras palavras, isso significa que a melhoria do MOC não foi tão grande quanto nas outras cidades.

Já em relação ao comportamento dos valores de temperatura, as duas figuras (Figuras 11 e 12) não mostram algo diferente do que foi dito anteriormente. O MOC não melhorou muito o desempenho do modelo, que por sua vez não conseguiu acompanhar as tendências dos valores observados.

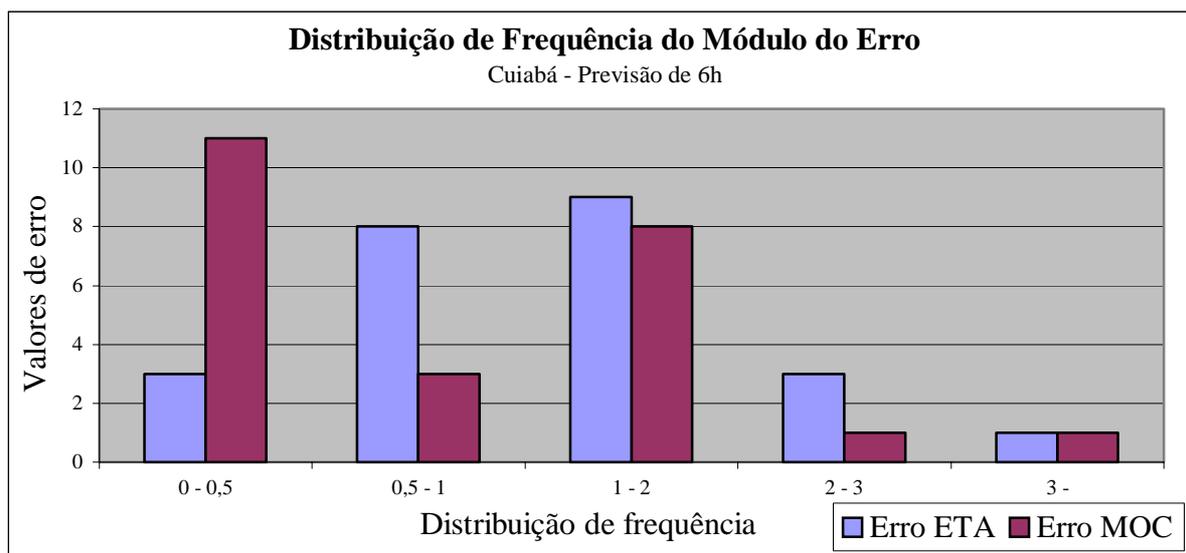
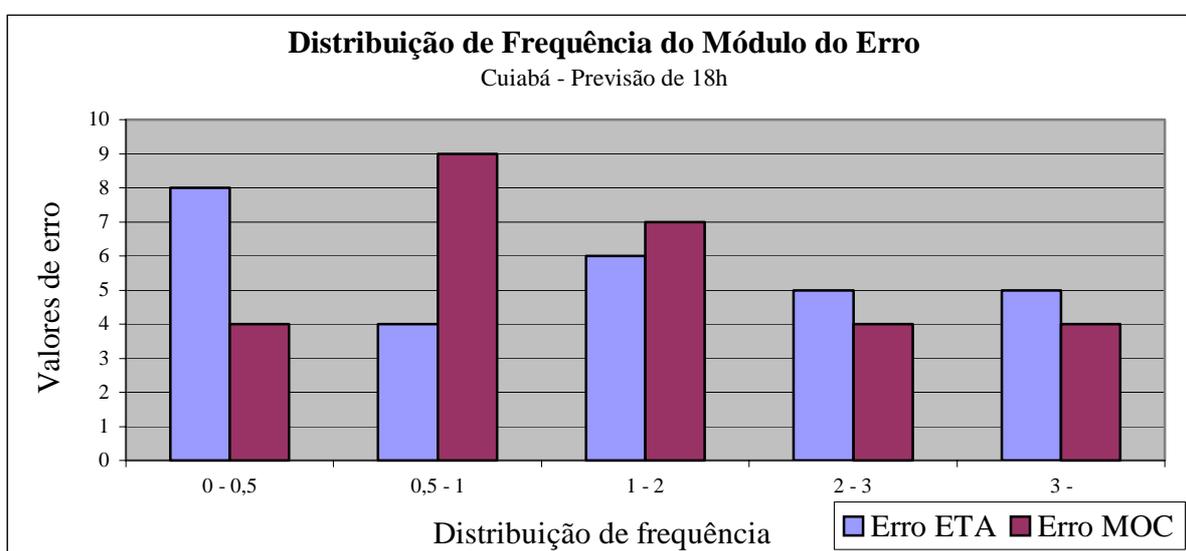


Figura 9 – Gráfico de distribuição de frequência dos módulos dos erros do modelo com e sem a calibração do



MOC, na previsão de 6 horas de Cuiabá.

Figura 10 – Gráfico de distribuição de frequência dos módulos dos erros do modelo com e sem a calibração do MOC, na previsão de 18 horas de Cuiabá.

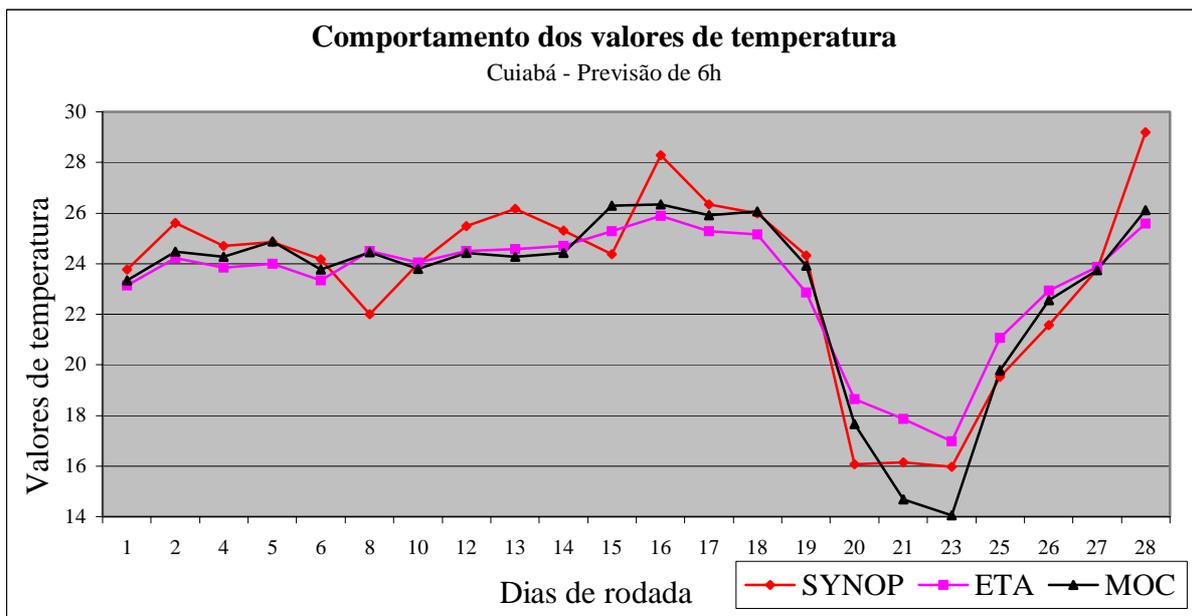
Ao analisar a Tabela 1, pode-se ver que o desempenho do MOC nem sempre tem relação tão direta com a quantidade de variáveis selecionadas. Esta cidade tem quase o dobro de variáveis selecionadas que Porto Alegre, mas tem mais da metade do desempenho da mesma.

Deve-se analisar então dois aspectos importantes: quais variáveis foram mais selecionadas e quais eram os valores da correlação destas variáveis com o erro. Isto pode nos dar uma indicação de deficiências nas parametrizações usadas no modelo ou em que direção deve-se refinar determinados dados de entrada (condições de contorno, por exemplo).

Com relação ao outro aspecto importante nesta análise, pode-se também ver as quão correlacionadas elas estavam em relação ao erro. Isso porque a equação de regressão será melhor à medida que essas variáveis selecionadas sejam mais correlacionadas com o erro.

Consultando as tabelas de dados de correlação para esta cidade (não anexos ao trabalho por não serem tão relevantes) pode-se ver que a maioria destas variáveis selecionada tinha correlações próximas ao limiar inferior de 40%.

De uma forma geral pode-se dizer que os resultados foram bem positivos. Além de melhorar o comportamento dos valores de previsão de temperatura, o MOC também conseguiu em quase todas as estações



aproximar bem o valor previsto do observado. De acordo com a Tabela 2, o índice de melhoria médio do MOC nessas 3 cidades foi de 29,8%.

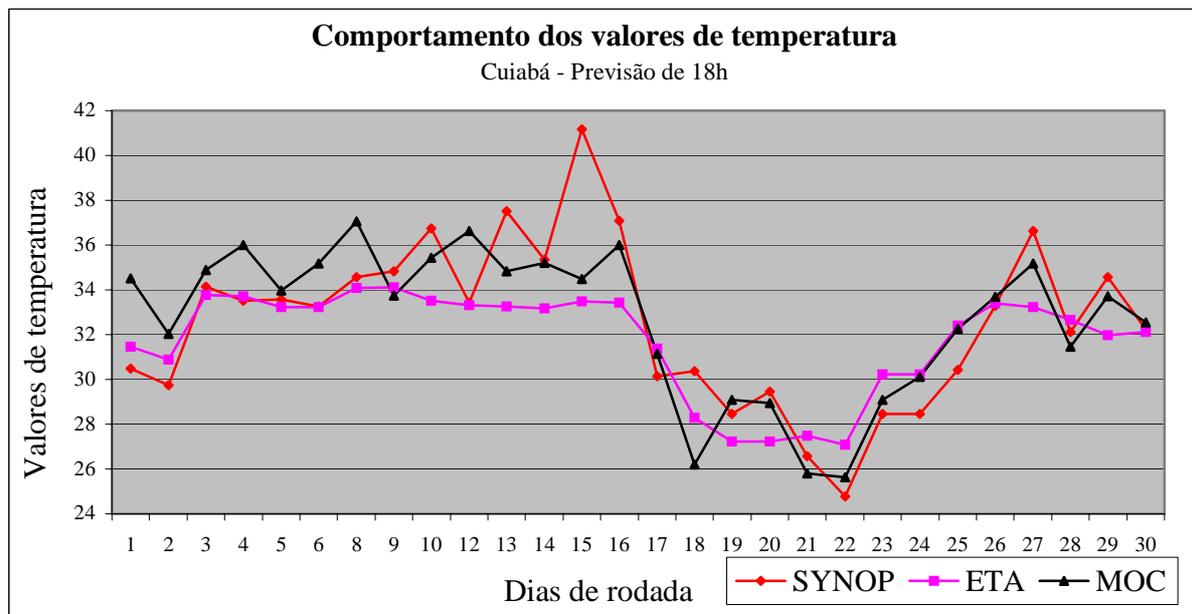


Figura 11 – Gráfico do comportamento dos valores de temperatura do modelo com e sem a calibração do MOC e os valores observados de SYNOP, na previsão de 6 horas de Cuiabá.

Figura 12 – Gráfico do comportamento dos valores de temperatura do modelo com e sem a calibração do MOC e os valores observados de SYNOP, na previsão de 18 horas de Cuiabá.

Em outras palavras, o MOC conseguiu melhorar em quase 30% as previsões de temperatura do modelo ETA do CPTEC / INPE.

## CONCLUSÕES

Este método se mostrou eficaz, mas ainda existem muitos parâmetros e aspectos a serem analisados e testados no MOC.

Devem existir variações de desempenho quando se altera a quantidade de dias para realizar a equação de regressão (atualmente fixa em 21 dias), ou quando se muda o limiar mínimo para a seleção de variáveis na equação de regressão, ou mesmo quando existe uma mudança brusca no tempo, como uma mudança de estação.

Para os resultados aqui expostos, o método foi testado apenas para o mês de abril de 2000. Já está em andamento a operacionalização do MOC começando pela experimentação e cálculo dos índices de melhoria por um período de previsões maior, para se ter uma validação adequada do método.

## **REFERÊNCIAS BIBLIOGRÁFICAS**

- Glahn, H. R. e D. A. Lowry, 1972: The use of model output statistics (MOS) in objective weather forecasting, *J. Appl. Meteor.*, **11**, 620 - 631
- Mao, Q., S. F. Mueller, H.-M. H. Juang e R. T. McNider, 1998: An Optimal Model Output Calibration Algorithm Suitable for Objective Temperature Forecasting. *Wea. Forecasting*, **14**, 190-202.
- Mass, C. F. e Y. H. Kuo, 1998: Regional real-time numerical weather prediction: current status and future potential. *Bull. Amer. Meteor. Soc.*, **79**, 253 - 263.
- Matsuo, P. T. e Dias, M. A. F., 1992: Modelos de previsão probabilística de precipitação com base em parâmetros sinóticos: Parte II - Modelagem e Avaliação, *Anais do VII Congresso Brasileiro de Meteorologia*, 665 - 669
- Vislocky, R. L. e J. M. Fritch, 1995b: Improved model output statistics forecast through model consensus. *Bull. Amer. Meteor. Soc.*, **76**, 1157 - 1164
- Vislocky, R. L. e J. M. Fritch, 1997: Performance of an advanced MOS system in the 1996-97 National Collegiate Weather Forecasting Contest. *Bull. Amer. Meteor. Soc.*, **78**, 2851 - 2857