

# UMA EXPERIÊNCIA COM MODELO ESTATÍSTICO (MOS) PARA A PREVISÃO DA TEMPERATURA MÍNIMA DIÁRIA DO AR

S. Sugahara

Uma equação de regressão múltipla MOS (da sigla em inglês para *Model Output Statistics*), para previsão da temperatura mínima diária do ar na cidade de Bauru, estado de São Paulo, é desenvolvida. A equação de regressão múltipla, obtida usando análise de regressão *stepwise*, tem quatro preditores, três do modelo numérico global do Centro de Previsão de Tempo e Estudos Climáticos (CPTEC) e um observacional da estação meteorológica do Instituto de Pesquisas Meteorológicas (IPMet), Bauru. Os preditores são prognósticos para 24 horas do modelo global, válidos para 00:00GMT, da temperatura em 1000hPa, vento meridional em 850hPa e umidade relativa em 1000hPa, e temperatura observada às 18:00GMT. Esses quatro preditores explicam, aproximadamente, 80% da variância total do preditando, com erro quadrático médio de 1,4°C, que é aproximadamente metade do desvio padrão da temperatura mínima diária do ar observada na estação do IPMet. Uma verificação da equação MOS com uma amostra independente de 47 casos mostra que a previsão não se deteriora significativamente quando o preditor observacional for desconsiderado. A equação MOS, com ou sem esse preditor, produz previsões com erro absoluto menor do que 1,5°C em 70% dos casos examinados. Este resultado encoraja a utilização da técnica MOS para previsão operacional da temperatura mínima e seu desenvolvimento para outros elementos do tempo e outras localidades.

**Palavras-Chave:** Previsão Estatística; Estatística das Saídas do Modelo Numérico; Temperatura Mínima Diária do Ar.

**AN EXPERIENCE WITH MODEL OUTPUT STATISTICS (MOS) FOR DAILY MINIMUM AIR TEMPERATURE PREDICTION-** A MOS (*Model Output Statistics*) multiple regression equation for the prediction of daily minimum air temperature at the city of Bauru, in São Paulo State, is developed. The multiple regression equation, obtained using *stepwise regression analysis*, has four predictors, three from the CPTEC (*Centre of Weather Forecast and Climate Studies*) global model and one from observational data of the meteorological station at IPMet (*Institute of Meteorological Research*), Bauru. The predictors are the model 24 hours prognosis, valid at 00:00GMT, of 1000hPa temperature, 850hPa meridional wind and 1000hPa relative humidity, and the 18:00GMT observation of temperature. These four predictors account for approximately 80 percent of the total variance of the predictand, with a root mean square error of 1.4°C, i.e., approximately half of the standard deviation of daily minimum temperature observed at the IPMet station. A verification of the MOS equation with an independent sample of 47 cases shows that the forecast value is not significantly deteriorated when the observational predictor is not considered. The MOS equation, with or without this predictor, produces forecast with absolute errors smaller than 1.5°C in 70 percent of the cases studied. This result encourages the use of the MOS technique for operational daily minimum air temperature forecasting and the development of this technique for other weather elements and other localities.

**Key words:** Statistical Forecast; Model Output Statistics; Daily Minimum Air Temperature.

Instituto de Pesquisas Meteorológicas /Universidade Estadual Paulista, Campus de Bauru  
Av. Luiz Edmundo C. Coube s/n, Vargem Limpa  
Caixa Postal: 281, CEP: 17001-970  
Tel. (014) 2303709, Fax: (014) 2303649  
e-mail: shige@ipmet1.ipmet.unesp.br

## INTRODUÇÃO

Nas últimas décadas houve um progresso significativo no campo da previsão de tempo, graças em grande parte às facilidades proporcionadas pela tecnologia moderna, especialmente de computação eletrônica e de sistemas de comunicação de dados (Cressman, 1981). A construção de computadores com desempenho cada vez melhor possibilitou aperfeiçoamentos de modelos de simulação numérica da atmosfera, culminando durante a década de 60 com os modelos de equações primitivas. As simulações de campos meteorológicos feitas por esses modelos com precisão até então nunca vista começaram a ser exaustivamente exploradas, especialmente para as previsões de tempo propriamente dito, em particular com modelos estatísticos, que usam variáveis de entrada baseadas nessas simulações.

Os modelos estatísticos de previsão de tempo são, em geral, equações de regressão linear múltipla. De acordo com o relato de Carter et al. (1989), no final da década de 80 mais de 110.000 dessas equações estavam sendo processadas, nos EUA, pelo *National Weather Service* (NWS), proporcionando e disseminando a nível nacional as previsões de elementos meteorológicos tais como as temperaturas máxima e mínima diárias do ar e probabilidade de precipitação.

Dado o sucesso alcançado, especialmente com os métodos chamados Estatísticas das Saídas do Modelo ou MOS (sigla em inglês para *Model Output Statistics*) e Prognóstico Perfeito (PP), que usam como preditores as saídas de modelos numéricos, equações de regressão semelhantes começaram a ser desenvolvidas e usadas operacionalmente nos centros de previsão de tempo de outros países, como Austrália (Tapp et al., 1986); Canadá (Wilson & Yacowar, 1980), Japão (Tatehira & Nakayama, 1976), e Holanda (Lemcke & Kruizinga, 1988). Por exemplo, na Austrália, em 1984, para a previsão das temperaturas máxima e mínima diárias do ar, probabilidade de ocorrência e quantidade de precipitação, com as equações de MOS baseadas em seu próprio modelo de equações primitivas. Na Holanda, para a previsão de probabilidades de tempestade e de precipitação congelada, brilho solar, temperaturas máxima e mínima diárias do ar, e velocidade máxima do vento, através de equações de regressão alimentadas pelos prognósticos do modelo global do *European Centre of Medium Range Forecast* (ECMWF).

No Brasil, a história da previsão numérica operacional é muito recente. Ela teve início em janeiro de 1995, quando o Centro de Previsão de Tempo e Estudos Climáticos (CPTEC) começou a “rodar” operacionalmente o modelo global e disseminar seus produtos a nível nacional, principalmente através da rede de computadores.

O presente trabalho tem como objetivo mostrar como esses produtos podem ser úteis na previsão pontual de elementos de tempo, usando técnica estatística MOS. O trabalho se restringe à previsão de temperatura mínima diária do ar no Instituto de Pesquisas Meteorológicas, situado quase no centro do estado de São Paulo, na cidade de Bauru (22° 21' S, 49° 01' W), a 620 metros do nível do mar. É importante ressaltar que o tipo de previsão que se pretende aqui é pontual, diferente portanto das previsões de dados horários ou *grid histories* fornecidas pelo CPTEC, onde as previsões para as principais cidades brasileiras são médias para quadrículas de grade. Outra diferença das previsões pelo MOS com o *grid histories* é que no primeiro método, os efeitos da altitude da estação, do clima local, assim como as imprecisões inerentes do modelo numérico são levados em consideração durante o desenvolvimento da equação (Klein, 1982). Uma revisão sobre modelo estatístico de previsão de tempo, baseado na análise de regressão linear múltipla, é dada a seguir.

## MODELOS ESTATÍSTICOS DE PREVISÃO DE TEMPO

### Regressão Múltipla

As equações de previsão estatística de tempo são usualmente de regressão linear múltipla (ver, por exemplo, Lemcke & Kruizinga, 1988), onde o elemento de tempo a ser previsto (variável dependente ou preditando) é expressa como uma função linear das quantidades que contribuem para variação desse elemento (variáveis independentes ou preditores). A descrição bem detalhada da análise de regressão pode ser encontrada, por exemplo, em Apte (1990).

O modelo de regressão linear com  $k$  preditores tem a seguinte forma:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon, \quad (1)$$

onde  $Y$  é preditando, e  $X_1, X_2, \dots, X_k$  são os preditores. Os coeficientes de regressão  $\beta_0, \beta_1, \beta_2, \dots, \beta_k$  são determinados a partir das observações da ocorrência do preditando e dos preditores; e  $\varepsilon$  representa o erro. A Eq. (1) pode ser escrita para  $n$  observações

$$\begin{aligned}
y_1 &= \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \dots + \beta_k x_{1k} + \varepsilon_1 \\
y_2 &= \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \dots + \beta_k x_{2k} + \varepsilon_2 \\
&\dots\dots\dots \\
y_n &= \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \dots + \beta_k x_{nk} + \varepsilon_n
\end{aligned} \tag{2}$$

ou na notação mais compacta:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{3}$$

onde,

$$\mathbf{y} = [y_1 \quad y_2 \quad y_2 \quad \dots \quad y_n]^T ,$$

e

$$\mathbf{X} = \begin{bmatrix}
1 & x_{11} & x_{12} & \dots & x_{1k} \\
1 & x_{21} & x_{22} & \dots & x_{2k} \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
1 & x_{n1} & x_{n2} & \dots & x_{nk}
\end{bmatrix} ,$$

são, respectivamente, matrizes com as observações de Y, e de  $X_1, X_2, \dots, X_k$ . As matrizes  $\boldsymbol{\beta}$  e  $\boldsymbol{\varepsilon}$  são

$$\boldsymbol{\beta} = [\beta_0 \quad \beta_1 \quad \beta_2 \quad \dots \quad \beta_k]^T ,$$

$$\boldsymbol{\varepsilon} = [\varepsilon_0 \quad \varepsilon_1 \quad \varepsilon_2 \quad \dots \quad \varepsilon_n]^T .$$

A soma dos quadrados dos desvios é dada por

$$s^2 = \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} = (\mathbf{y}^T - \mathbf{b}^T \mathbf{X}^T)(\mathbf{y} - \mathbf{X}\mathbf{b}) ,$$

e os estimadores  $b_0, b_1, b_2, \dots, b_k$  dos coeficientes

$\beta_0, \beta_1, \beta_2, \dots, \beta_k$ , que minimizam  $s^2$  são dados por:

$$\mathbf{b} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \tag{4}$$

onde  $\mathbf{b} = [b_0 \quad b_1 \quad b_2 \quad \dots \quad b_k]^T$ ,  $\mathbf{X}^T$  é transposta da matriz  $\mathbf{X}$ , e o sobrescrito -1 indica a matriz inversa. Na Eq. (4) assume-se que a matriz  $\mathbf{X}^T \mathbf{X}$  é não singular.

Na análise de regressão para desenvolvimento da equação MOS, as variáveis independentes da Eq. (1) podem ser transformadas, como por exemplo numa média de pontos de grade (ver, por exemplo, Woodcock, 1984) e coeficientes de funções ortogonais empíricas (Lemcke & Kruzinga, 1988). Uma aplicação interessante da técnica MOS é encontrada em Lemcke & Kruzinga (1988), que usam dois preditores que são saídas de modelos estatísticos: modelo PP e modelo de analogia. A saída do modelo estatístico PP, como preditor da equação MOS, tem sido também

usada por Vislocky & Young (1989).

### Avaliação da Equação de Regressão

A habilidade da equação de regressão é medida pelo quadrado do coeficiente de correlação múltipla dado por

$$R^2 = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2} \tag{5}$$

onde  $\bar{y}$  é a média aritmética de n valores de Y, e  $\hat{y}_i$  i-ésimo valor de Y estimado ou previsto pela equação:

$$\hat{y}_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + \dots + b_k x_{ik} \tag{6}$$

O coeficiente  $R^2$  representa a contribuição à variância de Y dada pelas variáveis  $X_1, X_2, X_3, \dots, X_k$ , tal que quanto maior  $R^2$  maior é a habilidade do modelo para prever Y. A significância estatística de  $R^2$  é usualmente examinada testando a hipótese

$H_0: \beta_0 = \beta_1 = \beta_2 = \dots = \beta_k = 0$  com a estatística de prova

$$F = \frac{R^2}{1 - R^2} \frac{n - k - 1}{k} \tag{7}$$

Se  $H_0$  for verdade, F tem distribuição F com k e n-k-1 graus de liberdade. Esse teste de hipótese, chamado teste F, apóia-se nas seguintes suposições:

1. o erro  $\varepsilon$  tem uma distribuição normal, com média zero e variância finita e constante  $\sigma^2$ , simbolicamente  $\varepsilon = N(0, \sigma^2)$ .

2. os valores de  $\varepsilon$  estão decorrelacionados linearmente entre si, ou seja  $E(\varepsilon_i \varepsilon_j) = 0 \quad i \neq j$

3. as variáveis independentes  $X$ 's, quando estocásticas, são estatisticamente independentes de  $\varepsilon$  e sua distribuição não envolve os coeficientes  $\beta$  nem  $\sigma^2$ , e são livres de erros.

Na prática, uma ou mais suposições acima são frequentemente violadas. Nesses casos o teste F é usado apenas como um teste aproximado (para uma discussão mais completa sobre essas questões, ver por exemplo Box, 1954). A performance esperada do modelo pode também ser inferida subjetivamente examinando o valor de  $R^2$ . Mas, certamente, uma ótima forma de verificar a qualidade do modelo é através da sua aplicação em amostras independentes, e comparando as suas previsões com os valores que de fato foram observados.

## SELEÇÃO DOS PREDITORES

### Regressão Stepwise

Para previsão estatística da temperatura mínima diária do ar, existem muitas variáveis que *a priori* podem ser consideradas como preditores. Entre essas variáveis estão a altura geopotencial, temperatura, advecção de temperatura, as espessuras das camadas atmosféricas, componentes da velocidade do vento, umidade relativa. As variáveis como advecção de vorticidade, vorticidade, divergência do vento têm sido também consideradas.

Essas variáveis quando são de modelos numéricos são dadas na grade, com grande número de pontos, usualmente. A grade do modelo do CPTEC considerada no presente estudo é para região da América do Sul, compreendida entre as longitudes de 101,25°W e 11,25°W e entre as latitudes de 60°S e 15°N, com resolução de 1.875° x 1.875° e sete níveis na vertical (1000, 850, 700, 500, 300, 200, 100hPa). Existem, portanto, mais de 10000 pontos nessa grade tri-dimensional fornecendo vários parâmetros atmosféricos para o desenvolvimento da equação MOS. Dessas variáveis é que os preditores dinâmicos são selecionados.

Vários pontos devem ser considerados durante o desenvolvimento da equação de regressão. A maioria dos autores recomenda que o número de preditores não deve exceder, grosso modo, 10% do número de observações. Esse critério é considerado importante para evitar variação muito grande de preditores selecionados de um período para outro, ou de amostra para amostra, ou em outras palavras, para diminuir a incerteza quanto aos preditores. A escolha dos preditores com base no raciocínio físico tem sido também sugerida, para evitar essa variação. Outro procedimento recomendado consiste na comparação com outras equações, especialmente com aquelas já em uso operacional. Um exemplo disso pode ser encontrado em Woodcock (1984), que comparou, para os mesmos preditandos, as equações MOS que ele encontrou para a Austrália com as equações usadas operacionalmente nos EUA.

O procedimento adotado no presente estudo para o selecionamento dos preditores da temperatura mínima diária é o chamado análise de regressão *stepwise*, que tem sido largamente usado pelos desenvolvedores de modelos de previsão estatística. O procedimento pode ser resumido como segue:

a) A variável com o maior coeficiente de correlação com o preditando é retida primeiro na equação de regressão. Considere que essa variável seja  $X_i$

Então com  $X_i$  já incluída na equação, os coeficientes de correlação parcial de primeira ordem de  $Y$  com  $X_j$ , denotados por  $r^2_{YX_j \cdot X_i}$ ,  $j=1, \dots, k$  com  $j \neq i$ , são calculados. A variável com maior valor de  $r^2_{YX_j \cdot X_i}$  será então incluída na equação se sua contribuição à variância de  $Y$  for estatisticamente significativa, ou seja, se  $F > F_L$ , onde  $F_L$  é o valor mínimo de  $F$  pré-estabelecido, para a variável ser incluída na equação.

c) Considerando que a variável  $X_h$  satisfaz essa condição, agora com  $X_i$  e  $X_h$  fazendo parte da equação, calculam-se os coeficientes de correlação parcial de segunda ordem com as demais variáveis

$$r^2_{YX_j \cdot X_i X_h}, j=1, \dots, k \text{ com } j \neq i, j \neq h$$

Supondo que  $X_g$  satisfaz a condição para ser incluída na equação,

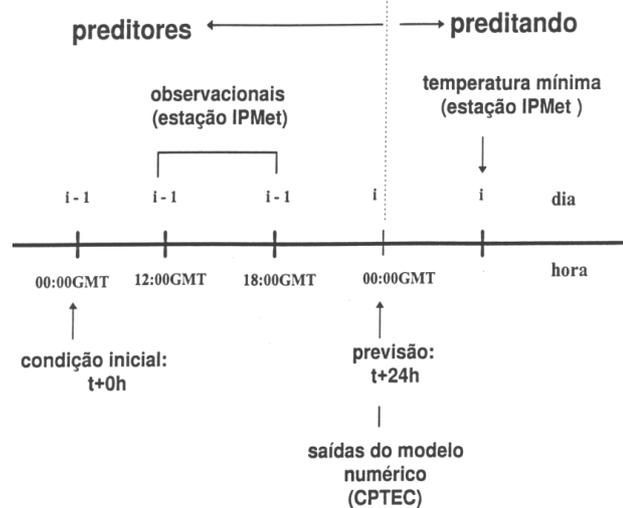
e) As variáveis  $X_i$  e  $X_h$  são então examinadas se devem ou não permanecer na equação, agora junto com  $X_g$ .

f) Supondo que  $X_i$  deva ser descartada, calculam-se então os coeficientes

$$r^2_{YX_j \cdot X_g X_h}, j=1, \dots, k \text{ com } j \neq g, j \neq h$$

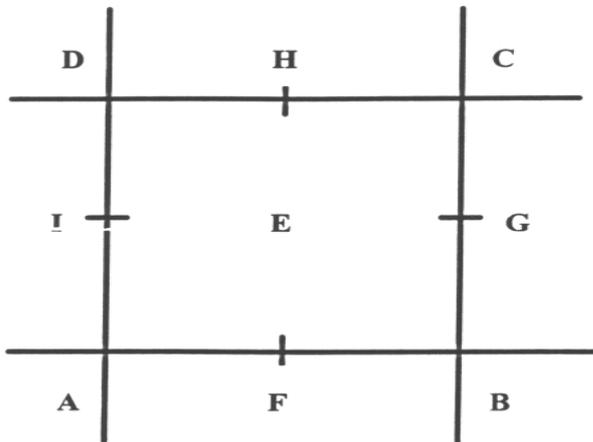
e assim por diante.

O processo de seleção é finalizado quando o



**Figura 1** - Relação temporal entre o preditando (temperatura mínima diária do ar na estação meteorológica do IPMet) e seus preditores candidatos, baseados no modelo global de previsão numérica do CPTEC e nas observações na estação do IPMet.

**Figure 1** - Temporal relationships between predictand (daily minimum air temperature at the meteorological station of IPMet) and candidates to their predictors, based on CPTEC model and on observations at the IPMet station.



**Figura 2** - Subgrade na qual os produtos do modelo do CPTEC foram interpolados para formação de preditores. O ponto E coincide com a estação do IPMet, para a qual é desenvolvida a equação de previsão da temperatura mínima diária do ar.

**Figure 2** - Subgrid to which the outputs of the CPTEC model were interpolated for the construction of predictors. The point E is the same as the station of IPMet to which the forecast equation of daily minimum air temperature is developed.

---

#### observada na estação no dia anterior

---

pressão atmosférica das 12:00 e 18:00 GMT

temperatura das 12:00 e 18:00 GMT

umidade relativa das 12:00 e 18:00 GMT

temperatura máxima

temperatura mínima

---

#### Prognósticos numéricos para 24 horas, com a condição inicial da 00:00 GMT do dia anterior

---

altura geopotencial 1000, 850, 700, 500, 300, 200 hPa

temperatura 1000, 850, 700, 500, 300, 200 hPa

umidade relativa 1000, 850, 700, 500 hPa

componente zonal da velocidade do vento 1000, 850,  
700, 500, 300 e 200hPa

componente meridional do vento 1000, 850, 700, 500,  
300 e 200hPa

espessura 1000-850hPa, 1000-500hPa

vorticidade relativa 1000, 850, 700, 500, 300, 200hPa

divergência do vento 1000, 850, 700, 500, 300, 200hPa

advecção de temperatura 1000, 850, 700, 500hPa

---

**Tabela 1** - Variáveis fornecidas ao algoritmo de regressão *stepwise* para determinar a equação MOS para previsão da temperatura mínima diária do ar.

**Table 1** - Variables used in the *stepwise regression* algorithm to determine the MOS equation for the daily minimum air temperature forecast.

critério  $F > F_L$  não justifica mais a inclusão ou exclusão de qualquer variável. Não existe, porém, nenhum critério objetivo aceito universalmente para finalização da análise de regressão *stepwise*. No presente estudo, a última variável selecionada satisfaz duas condições: (i)  $F > F_L$ , (ii) aumento não superior a 50% nos erros padrões das estimativas dos coeficientes de regressão.

## Preditores Candidatos

As previsões numéricas do CPTEC, que são preditores candidatos da temperatura mínima, são de 24 horas inicializadas à 00:00GMT. Outro conjunto de preditores candidatos são de dados observados da estação meteorológica do IPMet, das 12:00GMT e 18:00GMT, de um dia antes da ocorrência do preditando. A relação temporal entre esses preditores e o preditando é mostrada na Fig. 1.

A partir dos prognósticos numéricos, três conjuntos de preditores são preparados, seguindo procedimento semelhante ao de Woodcock (1984) e Tapp et al. (1986), que leva em consideração somente quatro pontos da grade mais próximos da estação. Os dados desses quatro pontos são interpolados numa subgrade de 9 pontos, usando função polinomial bicúbica (ver, por exemplo, Murray & Simmonds, 1991). A Fig. 2 mostra a subgrade, onde o ponto E coincide com a estação, e os pontos A, B, C e D são da grade original. O primeiro conjunto de preditor corresponde ao ponto E, o segundo, às médias para cinco pontos (D, E, F, H, B) e o terceiro, às médias para todos os 9 pontos. Como esses três conjuntos de preditores estão altamente correlacionados entre si, para a mesma variável, com coeficiente de correlação  $r$  maior do que 0,8, o conjunto das médias de 5 pontos foi arbitrariamente escolhido para a regressão *stepwise*. Os parâmetros meteorológicos considerados estão apresentados na Tab. 1.

## RESULTADOS E DISCUSSÕES

### Equações de Regressão

Todos os preditores listados na Tab. 1, para um total de 120 dias, correspondentes a maio de 1997, e de abril a setembro de 1998, foram fornecidos para o algoritmo de análise de regressão *stepwise*, com  $F_L = 4,0$ . Os parâmetros foram selecionados nesta ordem: temperatura Kelvin em 1000 hPa ( $T_{1000}$ ),

R <sup>2</sup> = 0,78	F(4,115) = 1=3,32	p = 0,00000	ε = 1,39°C
preditores e coeficientes b <sub>0</sub>	coeficientes de regressão b	e(b)	p
b <sub>0</sub>	-279,149	27,8787	0,000000
T <sub>1000</sub>	0,975	0,09358	0,000000
v <sub>850</sub>	-0,143	0,04196	0,000875
UR <sub>1000</sub>	0,042	0,01361	0,002291
t <sub>18h</sub>	0,126	0,04272	0,003815

**Tabela 2** -Resultado da análise de Regressão Múltipla com temperatura mínima diária do ar como preditand e temperatura do ar em 1000 hPa ( $T_{1000}$ ), vento meridional em 850hPa ( $v_{850}$ ), umidade relativa em 1000 hPa ( $UR_{1000}$ ), e temperatura do ar observada às 18:00 GMT ( $t_{18h}$ ), na estação meteorológica do IPMet, como preditores. Os valores de p indicam níveis de significância dos valores estimados de b, e e(b), erros padrão dos valores estimados dos coeficientes b.

**Table 2** - Result of multiple regression analysis with daily minimum air temperature as predictand and air temperature at 1000 hPa ( $T_{1000}$ ), meridional wind at 850 hPa ( $v_{850}$ ), relative humidity at 1000 hPa ( $UR_{1000}$ ) and observed temperature at 18:00 GMT at the meteorological station of IPMet ( $t_{18h}$ ) as predictors. The p-values indicate statistical significance levels, and e(b) the standard errors of the coefficients.

vento meridional em 850 hPa ( $v_{850}$ ), umidade relativa do ar em 1000 hPa ( $UR_{1000}$ ), dentre os prognósticos numéricos. O algoritmo selecionou como quarto preditor a temperatura do ar em Celsius observada às 18:00GMT ( $t_{18h}$ ). A pressão atmosférica observada às 12:00GMT foi a quinta variável, satisfazendo o critério  $F > F_L$ , mas descartada da solução final, porque notou-se que a sua inclusão triplica o erro padrão da estimativa do coeficiente  $b_0$  da equação de regressão. Os coeficientes de regressão e outras estatísticas mais relevantes dessa análise de regressão são mostrados na Tab. 2. Nessa tabela, os níveis de significância p associados aos coeficientes  $b_0, b_1, b_2, \dots, b_k$  foram calculados sob a hipótese  $H_0: \beta_i = 0, i=0, 2, 3, \dots, k$ , usando a estatística  $t = b_i / s_{b_i}$ , que tem distribuição t de Student com n-k graus de liberdade;  $s_{b_i}$  é estimativa do desvio padrão de  $b_i$  e é calculada como raiz quadrada positiva de  $c_{ii} s^2$ ,  $c_{ii}$  é i-ésimo elemento da diagonal da matriz  $(\mathbf{X}^T \mathbf{X})^{-1}$  e  $s^2$  é estimativa da variância de  $\epsilon$ .

Usando as notações acima e os coeficientes da Tab. 2, a temperatura mínima ( $T_{\min}$ ) de amanhã será dada por:

$$T_{\min} = 279.149 + 0.975T_{1000} - 0.134v_{850} + 0.043UR_{1000} + 0.126t_{18} \quad (8)$$

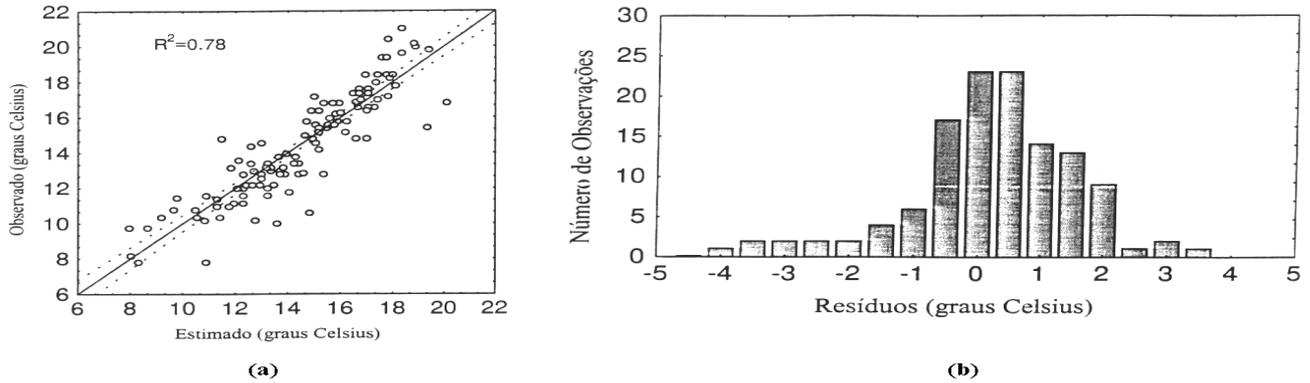
onde  $T_{1000}$ ,  $v_{850}$  e  $UR_{1000}$  são prognósticos de 24 horas válidos para 00:00GMT de amanhã e  $t_{18h}$  é temperatura observada às 18:00GMT de hoje. Durante a aplicação da Eq. (8),  $T_{1000}$  deve ser dada em Kelvin,  $v_{850}$  em m/s,  $UR_{1000}$  em percentagem, e  $t_{18h}$  em graus Celsius. Devendo ainda lembrar que os 3 primeiros preditores são médias de 5 pontos da subgrade, conforme descrito anteriormente. Considerando que a temperatura mínima na estação do IPMet ocorre tipicamente por volta das 09:00GMT (ou 06:00HL), pode-se dizer que a Eq. (8) permite previsão com antecedência de 15 horas, aproximadamente (ver Fig. 1). Usualmente os produtos do CPTEC são disponibilizados no computador do IPMet por volta das 09:00GMT (ou 06:00HL). Então a maior restrição de caráter prático sobre a Eq. (8) é imposta pelo preditor observacional das 18:00GMT.

De acordo com a Eq. (8),  $T_{1000}$  será maior (menor) quanto maior (menor) for a:

- temperatura do ar em 1000 hPa prevista para 00:00GMT
- temperatura do ar observada às 18:00 GMT
- umidade relativa em 1000 hPa prevista para 00:00 GMT

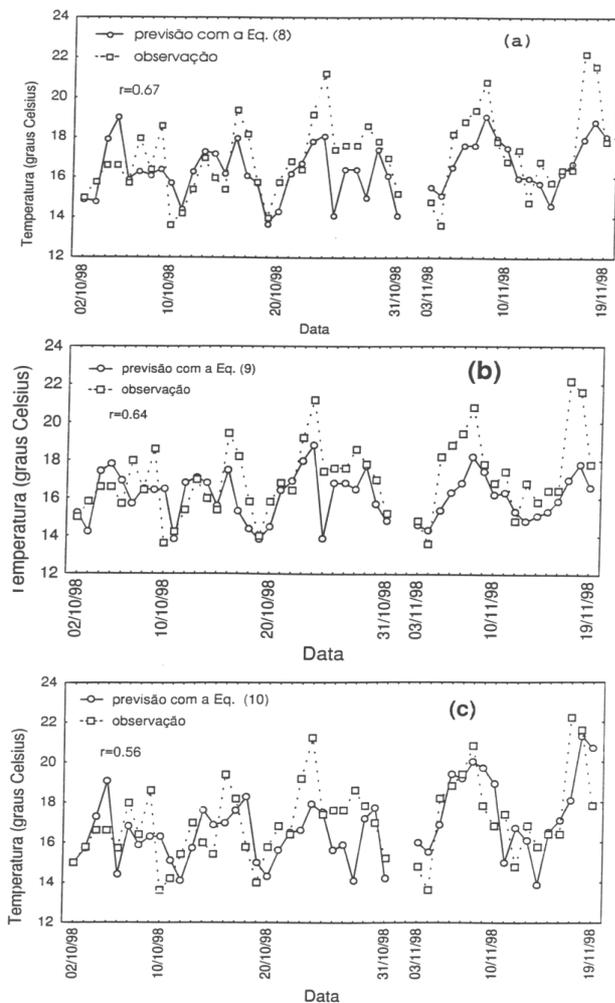
E o vento de norte contribui para aumentar e o vento de sul, para diminuir  $T_{1000}$ .

Essas quatro variáveis contribuem com 78% da variância de  $T_{1000}$ . O erro médio quadrático do modelo é 1,4 °C, que pode ser considerado baixo, comparado com o desvio padrão de  $T_{1000}$  de 2,7°C, para mesma amostra de dados. O modelo tem também um pequeno bias, de -0,04°C, e erro absoluto médio de 1,0 °C. Todos os coeficientes da equação, incluindo a constante  $b_0$ , são altamente significativos com nível de significância p menor do que 0,005, para a hipótese  $H_0: \beta_0 = \beta_1 = \beta_2 = \dots = \beta_K = 0$ . O número de preditores (4) é menor do que 5% do número de observações. Conforme mencionado anteriormente, essa relação entre o número de preditores e o número de observações é considerada adequada pela maioria dos autores. O diagrama de dispersão da temperatura mínima estimada pela equação versus temperatura mínima observada assim como a distribuição do resíduo são mostrados na Fig. 3. A concentração de pontos ao longo da reta de regressão (Fig. 3 a) e  $R^2$  de 0,78 justificam a escolha do modelo de regressão linear. O teste W de Shapiro-Wilk (1965) aplicado aos resíduos (Fig. 3 b), para o nível de significância de 0,05, indica que a hipótese da distri-



**Figura 3** - (a) Diagrama de dispersão de valores estimados pela Eq. (8) versus valores observados de temperatura mínima diária do ar, para 120 casos.  $R^2$  é quadrado do coeficiente de correlação múltipla dos quatro preditores. (b) Distribuição de resíduos (em  $^{\circ}\text{C}$ ) para Eq. (8).

**Figure 3** - Scatter diagram of estimated values from Eq. (8) versus observed values, for daily minimum air temperature, for 120 cases.  $R^2$  is the square of multiple correlation coefficient of four predictors. (b) Distribution of residues for Eq. (8), in  $^{\circ}\text{C}$ .



**Figura 4** - (a) Comparação da temperatura mínima diária do ar entre valores previstos usando Eq. (8) (contínua) e valores observados (tracejada), para 47 dias do período 2 de outubro de 1998 a 19 de novembro de 1998, em  $^{\circ}\text{C}$ . (b) Mesmo que (a), mas usando a Eq. (9). (c) Mesmo que (a), mas usando a Eq. (10).

**Figure 4** - Comparison of daily minimum air temperature between forecasted values using Eq. (8) (continuous) and observed values (dashed), for 47 days from October 2, 1998 to November 19, 1998, in  $^{\circ}\text{C}$ . (b) Same as (a), but using Eq. (9). (c) Same as (a), but using Eq. (10).

buição normal não deve ser rejeitada. Como mencionado anteriormente, a distribuição normal dos resíduos é uma das hipóteses assumidas no teste F.

Uma característica importante da Eq. (8) é que seus preditores são consistentes com a experiência sinótica. A temperatura, com a persistência comumente observada nos parâmetros meteorológicos. O componente meridional do vento, com fato de o vento de sul trazer, em média, ar mais frio e o vento de norte, ar mais quente. A umidade relativa é consistente com o fato que o ar úmido exerce maior efeito estufa do que o ar mais seco, retardando o resfriamento radiativo; e isso, por sua vez, estaria relacionado com o vento, no sentido que, na média, o vento com componente de norte é mais úmido do que o vento com o componente de sul. Mesmo a pressão atmosférica - com coeficiente de regressão negativo - que foi descartada da equação é fisicamente consistente com o fato que a pressão atmosférica mais alta (baixa) usualmente está associada com a massa de ar frio (quente), isto é, com a equação do balanço hidrostático.

Assim, apesar do método de selecionamento ter sido puramente estatístico, os preditores selecionados estão fisicamente relacionados com o preditando, e refletem os efeitos da forçante sinótica sobre a variabilidade de  $T_{\min}$ . O valor de  $R^2$  de 0,78 está próximo dos valores obtidos por outros autores. Outro aspecto interessante, particularmente dos preditores  $T_{1000}$  e  $UR_{1000}$ , é que esses são os que mais frequentemente aparecem nas equações de previsão de  $T_{\min}$  obtidas para as cidades australianas por Woodcock (1984). A importância do vento meridional, da baixa troposfera, é também revelada nas equações australianas para temperatura mínima. É importante mencionar que com amostras de maior tamanho podem ser

conseguidos coeficientes mais precisos e também incorporar mais informações acerca da relação entre o modelo e a observação. A incorporação dos efeitos da topografia e do clima local também podem ser melhorada aumentando a amostra.

A fim de verificar a importância relativa dos prognósticos numéricos do CPTEC e dos preditores observacionais, análises de regressão foram realizadas separadamente, considerando a mesma amostra de 120 casos. Com os prognósticos numéricos foi então encontrada a seguinte equação:

$$T_{\min} = -309.731 + 1.089T_{1000} - 0.156v_{850} + 0.044UR_{1000}, \quad (9)$$

onde os 3 preditores têm o mesmo significado que na Eq. (8).

E com os preditores observacionais:

$$T_{\min} = 254.6569 + 0.3238t_{18h} + 0.3980t_{\min} - 0.2676p_{12}, \quad (10)$$

onde todos os preditores são observações precedendo  $T_{\min}$ . O preditor  $t_{18h}$  é o mesmo que aparece na Eq. (8),  $t_{\min}$  é temperatura mínima, e  $p_{12h}$ , pressão das 12:00 GMT. Os valores de  $R^2$  para as Eqs. (9) e (10) são 0,76 e 0,69, respectivamente. Observa-se, portanto, melhoramento significativo na qualidade da previsão estatística quando os prognósticos numéricos são usados como preditores. Uma vantagem da Eq. (9) sobre a Eq. (8) é que, apesar da sua habilidade ligeiramente inferior, é menos limitada em termos práticos, já que pode ser usada antes da observação da temperatura das 18:00GMT. E isto pode representar um ganho de pelo menos 9 horas, considerando novamente que os prognósticos do CPTEC ficam disponibilizados até 09:00GMT.

### Experimentos com Amostra Independente

As Figs. (4a)-(4c) mostram resultados do experimento de previsão realizado com as Eqs. (8), (9) e (10), respectivamente, onde os valores previstos são comparados com os valores observados. A previsão baseada somente na técnica da persistência, em que a temperatura mínima observada hoje é repetida para amanhã, foi também examinada. As séries são mostradas para 47 dias, do período de 2 de outubro a 19 de novembro de 1998, que não foram considerados durante o desenvolvimento das equações. Nas

Figs. (4a) - (4c), os coeficientes de correlação  $r$  da série de observações com as séries de previsões são respectivamente 0,67, 0,64 e 0,56.

Considerando como previsão errada quando a sua diferença absoluta com a observação ultrapassar  $1,5^\circ\text{C}$ , que é aproximadamente a metade do desvio padrão do preditando, pode-se constatar as seguintes proporções de previsões corretas: (a) Eq. (8), 70% (b) Eq. (9), 70%, (c) Eq. (10), 56%, e finalmente, com a técnica da persistência, 60%, respectivamente, com erros absolutos médios de  $1,26^\circ\text{C}$ ,  $1,38^\circ\text{C}$ ,  $1,46^\circ\text{C}$ , e  $1,55^\circ\text{C}$ . É interessante notar que, em termos de percentagem de acerto, a Eq. (9) baseada somente em preditores do modelo numérico teve a mesma habilidade que a Eq. (8) que usa também preditor observacional. Um resultado surpreendente é o fato que o método tão simples como da persistência foi mais hábil do que a Eq. (10), que usa três preditores observacionais.

### CONCLUSÕES

Uma experiência com a técnica de previsão estatística de tempo MOS mostrou como os prognósticos numéricos, hoje facilmente conseguidos, podem ser úteis para previsão pontual de elementos de tempo. O trabalho se restringiu à previsão da temperatura mínima diária do ar na estação meteorológica do IPMet. A equação de MOS obtida com quatro preditores explica cerca de 80% da variância desse parâmetro. Dentre esses preditores, os três mais importantes são prognósticos para 24 horas do modelo do CPTEC válidos para 00:00GMT: temperatura Kelvin em 1000hPa, componente meridional do vento em 850hPa e umidade relativa em 1000hPa. O único preditor observacional é a temperatura observada na estação às 18:00 GMT. Um teste com amostra independente mostra que a equação MOS sem esse preditor observacional teve a mesma habilidade que a equação completa. Em adição, o teste evidenciou a superioridade da técnica MOS sobre outros dois métodos: o que usa somente preditores observacionais e outro baseado unicamente na persistência.

O resultado aqui obtido encoraja o desenvolvimento de equações MOS para previsão de outros elementos de tempo e para outras localidades, assim como utilização de prognósticos numéricos de prazos maiores nessas equações.

## AGRADECIMENTOS

O autor agradece às analistas de sistemas Simone Cincotto e Jaqueline Murakami do Instituto de Pesquisas Meteorológicas/Unesp pelas facilidades proporcionadas à utilização de dados, e ao CPTEC pela disponibilização de seus produtos.

## REFERÊNCIAS

- APTE, P. G. -1990**-Textbook of Econometrics. McGraw-Hill Company, 1st Ed., 570 pp.
- BOX, G. E. P. - 1954** - Some theorems on quadratic forms applied in the study of analysis of variance problems, II. Effects of inequality of variance and of correlation between errors in the two-way classification. The Annals of Mathematical Statistics, **25**: 484-498.
- CARTER, G. M., DALLAVALLE, J. P. & GLAHN, H.R.-1989** - Statistical Forecasts Based on the National Meteorological Center's Numerical Weather Prediction System. Weather and Forecasting, **4**: 401-412.
- CRESSMAN, G. P. -1981**- Man and the machine in weather forecasting of the future. Proc. Symp. on Current Problems of Weather Prediction, Viena, Austrian Meteor. Soc., 3-6.
- LEMCKE, C. & KRUIZINGA, S. -1988**- Model Output Statistics Forecasts: Three Years of Operational Experience in the Netherlands. Monthly Weather Review, **116**: 1077-1090.
- MURRAY, R. J. & SIMMONDS, I. -1991**- A numerical scheme for tracking cyclone centres from digital data. Part I: development and operation of the scheme. Aust. Met. Mag., **39**, 155-166.
- SHAPIRO, S. S. & WILK, M.B. -1965**- An analysis of variance test for normality (complete samples). Biometrika, **52**: 591-611.
- TAPP, R.G., WOODCOCK, F. & MILLS, G. -1986**- The Application of Model Output Statistics to Precipitation Prediction in Australia. Monthly Weather Review, **114**: 50-61.
- TATEHIRA, R. & NAKAYAMA, T. -1976** - Objective forecasting of heavy precipitation using numerical prediction model output. Proc. WMO Symp. on the Interpretation of Broad-Scale NWP Products for Local Forecasting Purposes, Warsaw, 122-127.
- VISLOCKY, R. L. & YOUNG, G. S. -1989**- The use of Perfect Prog Forecasts to Improve Model Output Statistics Forecasts of Precipitation Probability. Weather and Forecasting, **4**: 202-209.
- WILSON, L. J. & YACOWAR, N. -1980**- Statistical weather element forecasting in the Canadian Weather Service. Proc. WMO Symp. on Probabilistic and Statistical Methods in Weather Forecasting, Nice, 40-406.
- WOODCOCK, F. -1984**- : Australian Experimental Model Output Statistics Forecasts of Daily Maximum and Minimum Temperature. Monthly Weather Review, **112**: 2112-2121.

*Manuscript submitted May 5, 1999*  
*Revised version accepted October 20, 1999*

### AN EXPERIENCE WITH MODEL OUTPUT STATISTICS (MOS) FOR DAILY MINIMUM AIR TEMPERATURE PREDICTION

A MOS multiple regression equation for the prediction of daily minimum air temperature at the city of Bauru (22° 21'S, 49° 01'W), in São Paulo State, Brazil, is developed. The potential predictors are archived numerical outputs of CPTEC (Centre of Weather Forecast and Climate Studies) global model and observational data of the meteorological station at IPMet (Institute of Meteorological Research), Bauru. The multiple regression equation, obtained using stepwise regression analysis for 120 cases, incorporates four predictors, three from the CPTEC model and one from IPMet station data. The predictors are the model 24 hours prognosis, valid at 00:00GMT, of 1000hPa temperature, 850hPa meridional wind and 1000hPa relative humidity, and the 18:00GMT observation of temperature. These four predictors account for approximately 80 percent of the total variance of the predictand, with root mean square error (rms) of 1.4°C.

This rms value corresponds approximately half of the standard deviation of minimum temperature observed at the IPMet station. The *bias* and the absolute mean error are, respectively, -0.04°C and 1.0°C. A verification of the MOS equation with an independent sample of 47 cases shows that the forecast value is not deteriorated significantly when the observational predictor is not considered. The MOS equation, with or without this predictor, produces forecast errors smaller than 1.5°C in 70 percent of the cases studied. The comparison with other statistical models based only on observational predictors or that based only on persistence shows that the MOS technique is quite superior. The results of this study encourage the use of the MOS technique for operational daily minimum air temperature forecasting and the development of this technique for other weather elements and other localities.

## NOTES ABOUT THE AUTHORS

**Shigetoshi Sugahara**

Bacharel em Física pela Universidade Estadual de Campinas, Unicamp, 1978; Mestre em Meteorologia, 1982, pelo Instituto Nacional de Pesquisas Espaciais/MCT, São José dos Campos; Doutor em Meteorologia, 1991, Departamento de Ciências Atmosféricas - Instituto Astronômico e Geofísico, IAG-USP, Professor Assistente de Climatologia, Departamento de Ciências Atmosféricas, IAG-USP,

1982-1991; Atualmente pesquisador do Instituto de Pesquisas Meteorológicas/Unesp - Bauru, desde 1991 e Professor no curso de pós-graduação da Faculdade de Ciências, Unesp - Bauru. Áreas de interesse: previsão estatística de tempo, climatologia estatística, controle estatístico de qualidade de dados atmosféricos

## 7<sup>th</sup> International Congress of the Brazilian Geophysics Society

### 7<sup>th</sup> CISBGf Office

Instituto de Física da UFBA  
Campus Universitário de Ondina  
40210-340 Salvador, Bahia - Brasil  
Phone-Fax : +55(71) 235-0002  
E-mail: cisbgf@sbgf.org.br

The Salvador 2001 International Congress is sponsored by the Brazilian Geophysical Society (SBGf), the Society of Exploration Geophysicists (SEG), the European Association of Geoscientist and Engineer (EAGE) and the Latin American Geophysical Union (ULG). Whether your interest is in the technical conference or in the exhibition, you will find in Salvador the perfect venue to update your knowledge or do business.

Exhibition: Salvador 2001 will offer a truly outstanding marketing opportunity for Latin America. Special invitations are extend to state owned and private companies to attend and participate. Companies wishing to receive additional exhibitor information should contact the SBGf Business Office.

Technical Conference - Whether going deep into the ground or probing into outer space, geophysics is continually revealing a new wealth of understanding. Geophysicists must continuously follow the new advances in their fields of study and Salvador 2001 is the ideal place to see the recent developments, meet old friends and make new ones.



**SALVADOR 2001: COM**



7<sup>o</sup> Congresso Inte  
7th Internation