

5

MODELAGEM ESTATÍSTICA DAS ANOMALIAS DE TEMPERATURA DA SUPERFÍCIE DO MAR DO OCEANO ATLÂNTICO TROPICAL

Carlos Alberto Repelli(*) e Paulo Nobre()**

(*)Fundação Cearense de Meteorologia e Recursos Hídricos-FUNCEME

()Centro de Previsão de Tempo e Estudos Climáticos(CPTEC)**

Instituto Nacional de Pesquisas Espaciais(INPE)

1 - Introdução

A mais recente e promissora classe de métodos de previsão climática na escala sazonal utiliza modelos de circulação geral da atmosfera (MCGA) para prever as anomalias pluviométricas e de temperatura de várias regiões do Brasil com dois a três meses de antecedência. A grande vantagem deste método reside na possibilidade de se prever a distribuição espacial desses parâmetros (Barnston, 1992). Em todos os modelos, sejam estatísticos ou dinâmicos, utilizados em previsão climática, faz-se necessário o conhecimento das anomalias de TSM dos oceanos tropicais.

No caso do Nordeste, a previsibilidade das anomalias pluviométricas sazonais se deve ao caráter predominante do efeito das TSM no comportamento dos fenômenos causadores de chuva sobre a região. Assim, a confiabilidade de uma previsão das anomalias sazonais de precipitação para aquela região será tanto maior quanto melhor forem as previsões das anomalias de TSM sobre os oceanos tropicais (Ward e Folland, 1991).

Recentemente, foi desenvolvido no CPTEC, com a colaboração da FUNCEME, um modelo estatístico de previsão das anomalias de temperatura da superfície do mar (TSM) sobre a Bacia do Oceano Atlântico Tropical, baseado na técnica de Análise de Correlações Canônicas (ACC) (Repelli, 1996).

2 - Dados e Metodologia

A metodologia utilizada para a construção do sistema de modelagem, baseada em uma técnica estatística multivariada é conhecida como Análise de Correlações Canônicas (**ACC**) com um pré-filtro de Funções Ortogonais Empíricas (**EOF**) sobre os dados do preditor e do preditando. Tal técnica (**ACC**) permite encontrar combinações lineares, chamadas "variáveis canônicas", a partir de dois conjuntos de dados (por exemplo, **Y**: preditor e **Z**: preditando), de forma que a correlação entre estas variáveis seja maximizada (Barnett, 1987). A grande vantagem do uso desta técnica está no fato da mesma possibilitar a busca de padrões de oscilação entre dois ou mais campos físicos. A partir daí pode-se encontrar um conjunto de equações prognósticas e se construir um modelo preditor.

O sistema desenvolvido utiliza os campos mensais com resolução de 1 grau de latitude-longitude dos seguintes conjuntos de dados: Temperatura da Superfície do Mar, Pressão ao Nível do Mar e Ventos à Superfície nas direções zonal e meridional do Comprehensive Oceanic and Atmospheric Data Set (**COADS**) disponíveis para o período de janeiro de 1945 a dezembro de 1993.

O sistema de modelagem estatística dos oceanos (**SIMOC**), cujo diagrama de fluxo é mostrado na [Fig. 1](#), é altamente flexível, permitindo que os parâmetros relacionados aos preditores e preditando possam ser livremente modificados. O campo preditor, por exemplo, pode ser constituído por um único parâmetro físico (ventos, pressão e **TSM**) ou por uma combinação destes, sobre uma ou mais áreas do globo, com qualquer resolução espacial, e ainda para qualquer mês (ou média de alguns meses) para quaisquer faixas de anos disponíveis. O campo preditando, no caso a **TSM**, também pode ser construído a partir de uma ou mais áreas do globo, com quaisquer resoluções espaciais e períodos.

Uma vez feita a seleção dos dados, é aplicada uma rotina para se extrair os pontos sobre o continente e calculam-se as climatologias, as anomalias e os desvios padrões dos campos. Os campos de anomalias de **Y** e **Z** são então normalizados pelos respectivos desvios padrões dos correspondentes pontos de grade.

Em seguida é aplicada a metodologia de **EOF** sobre os campos e posteriormente é aplicada a **ACC**; sendo que o número de modos retidos tanto no **EOF** quanto no **ACC** pode ser controlado. Nas **EOF** o número de modos retidos é tal que a variância explicada é maior ou igual a 80%, tanto no campo preditor quanto no preditando; e em alguns casos esse número é igual a 15 para o preditor e 10 para o preditando. Na **ACC** utiliza-se o mínimo dos números de modo do preditor e preditando fornecidos pela **EOF** (cerca de 12). Neste passo, o sistema salva os parâmetros de regressão para serem utilizados posteriormente.

Como o interesse é investigar a previsibilidade das anomalias térmicas sobre a Bacia do Atlântico Tropical, os testes foram realizados com o preditando sobre esta área. O módulo de "hindcast" é utilizado para se fazer as "previsões" durante um período passado. Este período pode ser idêntico ao utilizado para o treinamento (construção da equação de regressão) do modelo, ou outro período independente. O módulo "forecast" é utilizado para se fazer previsões, somente para um período diferente do utilizado para o treinamento do modelo.

A validação do **SIMOC** foi feita a partir de vários testes de desempenho, que consistem em comparações dos campos de **TSM** previstos para diversos períodos com os campos de **TSM** observados e com anomalias persistidas de **TSM**. Assim, foram feitas as análises dos coeficientes de correlação e do erro médio quadrático (**RMS**) entre os campos observado e previsto, e posteriormente essas análises foram comparadas com as mesmas estatísticas obtidas com os campos observado e previsto.

Vários experimentos foram realizados, e foram testados, individualmente, os seguintes preditores: **TSM**, pressão e ventos (zonal e meridional), para todos os "lags" (de janeiro a dezembro), a fim de se identificar o melhor parâmetro preditor e a máxima antecedência com a qual poder-se-ia conseguir os melhores resultados de previsão através desse modelo.

A área geográfica escolhida para servir como preditora é mostrada na [Figura 2a](#) e a [Figura 2b](#) mostra a área geográfica do preditando. O número total de experimentos realizados é 342 (3 parâmetros preditores x 144 lags) e a [Tabela 1](#) mostra as faixas de anos que foram utilizadas para se construir e validar o modelo.

O período chamado de "treinamento", foi utilizado para se construir o modelo. O período seguinte, chamado de "hindcast" é um período independente do primeiro, como pode ser observado, e serviu para que fossem gerados alguns anos de prognósticos e a partir dos resultados, pudesse ser calculado um fator de escala para se ajustar as magnitudes dos campos previstos. O valor encontrado para se ajustar as magnitudes das escalas é igual a 4.3. Isto é necessário, porque os campos previstos, por serem truncados após a aplicação do EOF e da ACC, apresentam sempre suas magnitudes ligeiramente menores que a dos campos observados. No período chamado de "forecast" são feitas previsões, independentemente do primeiro e do segundo estágios.

As análises estatísticas de validação foram aplicadas igualmente sobre os períodos de "hindcast" e "forecast". Assim, pode ser testado também a consistência e estabilidade do modelo em prever períodos diferentes, ou seja, se o modelo apresenta ou não alguma tendência associada ao período de previsão.

3 - Resultados dos Experimentos

A [Fig. 3a](#) mostra a média dos coeficientes de correlação médio para os Hemisférios Norte (HN) e Sul (HS), entre as anomalias de TSM persistida e a observada e a [Fig. 3b](#) mostra a média do RMSE para os dois hemisférios para os dois períodos de "hindcast". Nota-se que a persistência possui melhor correlação para o Hemisfério Norte. Observa-se ainda que, em geral, no mês subsequente ao do preditor, o coeficiente de correlação é relativamente alto (>0.6 para o HN e >0.5 para o HS) e a partir de um mês de "lag" de previsão, há um decaimento.

O RMSE observado para todos os lags é maior que 0.5 desvios padrão e aumenta consideravelmente quando o mês preditando se afasta do preditor. Os maiores erros foram observados a partir das previsões feitas em novembro para 6 a 12 meses subsequentes de "lag". Curiosamente, observa-se que os valores de RMSE são ligeiramente maiores no HN que no HS.

Os resultados das estatísticas (coeficientes de correlação e RMSE) obtidos entre os campos observados e o previsto pelo SIMOC, utilizando-se a TSM como preditor são mostrados nas [Figura 3c](#) e [Figura 3d](#). Nota-se que o modelo consegue captar a variabilidade das anomalias de TSM melhor para o HN do que para o HS, pois os coeficientes de correlação médios para o HN são maiores. Observam-se altos coeficientes de correlação (>0.6) para o HN, para as previsões feitas a partir de dezembro, para os meses de fevereiro, março e abril. Utilizando-se o mesmo parâmetro preditor em janeiro, consegue-se coeficientes de correlação maiores que 0.6 para os meses de fevereiro a junho. Em geral, os coeficientes são altos, também para as previsões feitas a partir de fevereiro, março e abril, para um "lag" de até 3 meses.

Similarmente, pode-se observar no HS as mesmas relações, porém, neste caso os coeficientes de correlação são menores que no HN (em torno de 0.3 a 0.5). Ainda assim, são ligeiramente maiores que os obtidos com o campo de anomalia de TSM persistida. Nota-se ainda, que as magnitudes do RMSE são, em média, ligeiramente maiores no HN que as do HS.

Utilizando-se a pressão como parâmetro preditor e analisando os resultados, similares aos obtidos com os campos de vento zonal e meridional (Figuras não mostradas aqui), pode-se notar que as melhores previsões são aquelas de até 4 meses feitas a partir de dezembro, janeiro e fevereiro para o HN. No entanto, o desempenho do modelo nestes casos não foi tão bom, quando comparada com os resultados obtidos utilizando-se a TSM como parâmetro preditor.

4 - Comentários Finais

Foram testados todos os "lags" possíveis, para todos os meses do ano, e utilizou-se como preditor, os campos de **TSM**, pressão e ventos zonal e meridional. As estatísticas utilizadas foram coeficientes de correlação **RMSE**. Tais estatísticas foram comparadas com os resultados obtidos com a **TSM** persistida para cada "lag". O melhor desempenho obtido pelo modelo foi conseguido utilizando-se a **TSM** como parâmetro preditor e feitas a partir de novembro a fevereiro, com coeficientes de correlação superiores a 0.6.

Esta versão do modelo já pode ser usada de maneira operacional para se prever as anomalias térmicas do Atlântico, e seus resultados podem ser utilizados como subsídio para as previsões de precipitações sazonais da Região Nordeste do Brasil e/ou como condições de contorno do **MCGA** do **CPTEC** para as previsões climáticas para o Brasil durante este período.

Referências Bibliográficas

Barnett, T. P. and Preisendorfer, R., 1987: Origins and levels of monthly and seasonal forecast skill for United States surface air temperatures determined by canonical correlation analysis. *Monthly Weather Review*, Vol. 115, 1825-1849.

Barnston, A. G. and Ropelewski, C., 1992: Prediction of ENSO episodes using Canonical Correlation Analysis. *Journal of Climate*, Vol. 5. Nr. 11. November, 1992. 1316-1345.

Repelli, C. A. , 1996: Modelagem Estatística das Anomalias de Temperatura da Superfície do Mar sobre o Atlântico Tropical. Relatório Técnico das atividades exercidas como Pesquisador Visitante no Instituto Nacional de Pesquisas Espaciais-INPE /Centro de Previsão de Tempo e Estudos Climáticos-CPTEC. CNPq - Processo nr. 460448/95-1.

Ward, M. N. and C.K. Folland; 1991 - Prediction of seasonal rainfall in the North Nordeste of Brazil using eigenvectors of sea- surface temperature. *Int. J. Climatol.*, 11, 711- 743.
