

Apresentado no XIV Congresso Brasileiro de Meteorologia, Florianópolis, 27 novembro a 01 de dezembro 2006.

ESQUEMA ESTATÍSTICO DE COMBINAÇÃO E CORREÇÃO DE PREVISÕES CLIMÁTICAS – ECCOCLIM

*Luciano P Pezzi¹, Mary Kayano,
Ana Cleide Bezerra e Simone Shizue Tomita*

RESUMO. *Este trabalho mostra, preliminarmente, o Esquema Estatístico de Combinação e Correção de Previsões Climáticas (ECCOCLIM) que implementa o método de previsões por super-conjunto de multi-modelos, baseado na técnica de Regressão Linear Múltipla (RLM). Sobre a América do Sul o ECCOCLIM mostrou resultados melhores, reduzindo os erros, comparados a média aritmética dos diferentes modelos.*

ABSTRACT. *This study shows the preliminary results of a set of routines named Climate Forecast Combining and Correcting Statistical Scheme (ECCOCLIM). This method uses multiple linear regression to combine each individual model outputs which composes the multi-model super-ensemble. ECCOCLIM has showed encouraging results over South America reducing the errors compared to the traditional arithmetic mean method.*

Palavras-Chave: Previsão Climática, Multi-modelos, Super-conjunto, ECCOCLIM

1. INTRODUÇÃO

A previsão climática sazonal é uma informação essencial e estratégica para o planejamento das atividades do agro-negócio, recursos hídricos, energia, abastecimento e várias outras atividades que afetam diretamente a sociedade tanto do ponto de vista econômico quanto social. Um dos grandes avanços nas previsões é o uso de conjuntos de integrações com diferentes modelos, ou com configurações (parametrizações físicas) diferentes. A idéia é a de se extrair o melhor de cada modelo dinâmico e estatístico, utilizando-se essas previsões combinadas e corrigidas. A temperatura da superfície do mar (TSM) do oceano Pacífico tropical é simulada, e prevista, com uma razoável destreza. Porém com a TSM do oceano Atlântico, o problema é um pouco mais complexo devido a sua baixa previsibilidade. Vários autores mostraram que o Dipolo de TSM influencia a precipitação das regiões Norte e Nordeste do Brasil (NEB), podendo inclusive reverter ou intensificar o sinal das anomalias de precipitação destas regiões em casos onde o dipolo ocorre simultaneamente ao El Niño-Oscilação Sul (ENOS), (Moura e Shukla 1981, Pezzi e Cavalcanti 2001 e Kayano e Andreoli, 2006). Porém, as previsões climáticas por conjunto, feitas com modelos dinâmicos para a região Sudeste do Brasil, apresentam uma dispersão muito grande entre elas, resultando em previsões de baixa confiabilidade. Esforços precisam ser feitos no sentido de se melhorar esses índices de previsibilidade, tanto do Atlântico como da precipitação. Uma abordagem desse problema é o estudo da física envolvida em modelos dinâmicos, como em Pezzi e Richards (2003) que estudam a energética das

¹ Centro de Previsão do Tempo e Estudos Climáticos – CPTEC Instituto Nacional de Pesquisas Espaciais – INPE, Rod. Presidente Dutra Km 40 – 12630-00 - Cachoeira Paulista, SP – email: luciano@cptec.inpe.br

correntes equatoriais e parametrizações de mistura no Pacífico e por Pezzi *et al* (2004) onde estudam a interação oceano-atmosfera das Ondas de Instabilidade Tropicais (OIT). Ambos estudos mostram a influência desses processos na simulação da TSM. Alternativamente, um caminho promissor a ser seguido, que é o objetivo deste trabalho, é o da implementação de técnicas de pós-processamento no sentido de se corrigir estatisticamente as previsões (Coelho et al 2006) atualmente disponíveis pelos vários modelos climáticos no CPTEC. Neste trabalho é apresentado o *Esquema Estatístico de Combinação e Correção de Previsões Climáticas (ECCOCLIM)*, bem como alguns resultados preliminares.

2. O ECCOCLIM

O método implementado neste projeto, ECCOCLIM, é determinístico e baseia-se na técnica estatística de Regressão Linear Múltipla (RLM). O esquema baseia-se em uma fase de treinamento onde se aplica a RLM na determinação dos pesos estatísticos para as distintas simulações climáticas e as melhores estimativas da observação. Estes pesos descrevem o viés das previsões dos modelos em cada ponto geográfico e variáveis, para cada membro dos modelos participantes. O cálculo determina o peso de cada membro a ser usado na reconstrução das previsões finais que caracterizam o super-conjunto. Essa técnica tem sido usada com sucesso por Krishnamurti *et al.* (2001), veja estas referencias para detalhes de aplicações. De um modo geral, esses trabalhos mostram que as previsões geradas por super-conjunto, usando pesos estatísticos diferenciados para cada membro e que também variam no espaço, apresentam desempenhos superiores em relação às previsões feitas tomando-se somente a média aritmética dos membros. Esta última técnica é a que tem sido empregada rotineiramente na maioria dos centros operacionais do mundo onde são feitas previsões numéricas de tempo e clima, incluindo o CPTEC. Visando a operacionalização desta pesquisa, optou-se por implementar a metodologia proposta (ECCOCLIM) em Fortran 90. Foi desenvolvido um conjunto de rotinas baseadas no conceito de código modular (software) com a técnica estatística RLM usada nas correções das previsões dinâmicas do super-conjunto de multi-modelos. A Figura 1 ilustra esquematicamente o ECCOCLIM. O esquema é composto por módulos principais, os quais agrupam sub-rotinas que desempenham tarefas similares. No modulo **Sizes**, são encontradas as declarações dos tamanhos das variáveis globais usadas ao longo de todo o código. No módulo **Inout**, encontram-se todas as sub-rotinas que fazem leitura e escrita dos dados usados no modelo, durante o processamento. No módulo **Preproc**, são feitos os pré-processamentos dos dados, como o cálculo da climatologia, anomalias, desvio padrão e normalização das séries temporais. No módulo **Mlreglp**, são feitos os cálculos da RLM. No módulo **Sensemble**, é feito o cálculo da previsão final baseado em todos os

modelos usados na fase de treinamento, porém considerando-se o peso individual de cada um deles de acordo com os coeficientes obtidos pela RLM no módulo Mlreglp.

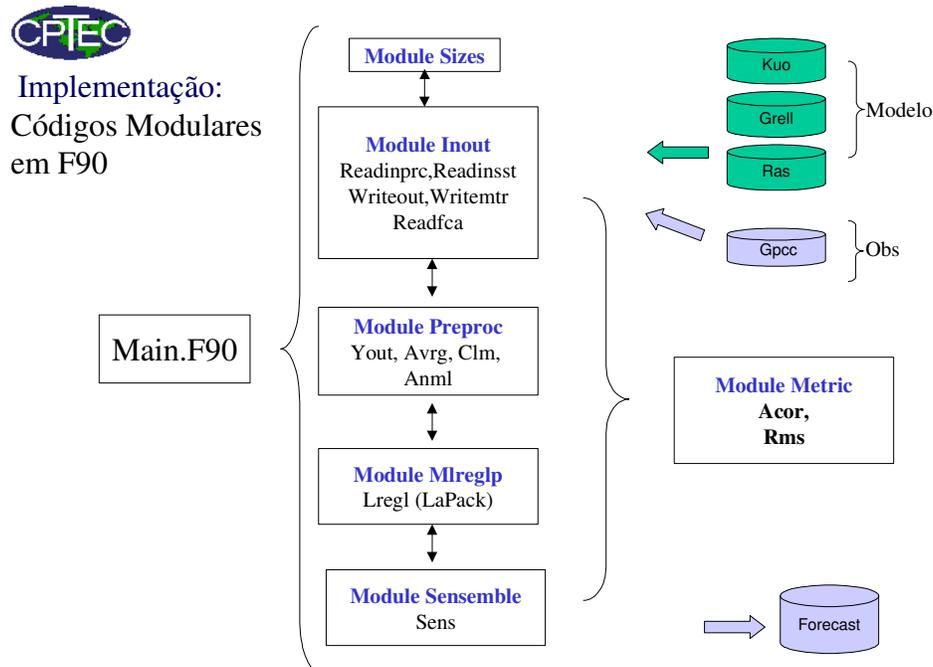


Figura 1 – Fluxograma esquemático do ECCOCLIM mostrando os principais módulos, dados de entradas provenientes das observações e dos diferentes modelos.

Finalmente no módulo **Metric**, são implementadas as rotinas para os cálculos da validação da previsão. A verificação da previsão é feita usando-se a técnica de validação cruzada, onde um ano é repetidamente retirado da série original e os coeficientes são estimados com o restante desta série. Posteriormente, a previsão é feita para este ano retirado e validado contra a observação. O ECCOCLIM, foi desenvolvido dentro do conceito de software modular. Isto significa que os módulos são independentes e usam o conceito de linguagem objeto. Desta forma é possível implementar novas técnicas de cálculo do super-conjunto ou de métricas de verificação, bastando codificar uma nova sub-rotina e inclui-la no esquema. A formulação estatística do método para a criação de um super-conjunto de multi-modelos para um dado ponto de grade é mostrada a seguir:

$$S = \bar{O} + \sum_{i=1}^N a_i (F_i - \bar{F}_i)$$

onde S = previsão do super-conjunto, \bar{O} = média temporal dos dados observados, a_i = pesos para o modelo i , N = número de modelos, \bar{F}_i = média temporal das previsões feita pelo modelo i , F_i = previsão feita pelo modelo i . Os pesos a_i são computados para cada ponto de grade e são minimizados pela seguinte função:

$$G = \sum_{t=0}^{t-\text{trein}} (S_t - O_t)^2$$

onde O_t é o estado observado, t é o tempo, $t-\text{trein}$ é o comprimento do período de treinamento. Os desempenhos das várias previsões foram avaliados calculando-se o erro quadrático médio (RMSE). As previsões geradas pelo ECCOLIM foram comparadas com um controle. A previsão de controle (MED), é a média aritmética simples dos modelos climáticos. Este é o método que tradicionalmente vem sendo usado nos centros de previsão climática, incluindo o CPTEC. No ECCOCLIM, foram usados dados de precipitação gerados por rodadas no modo climático do modelo de circulação global atmosférico (MCGA) do CPTEC. Em cada rodada, foi usado um esquema diferente de parametrização da convecção cumulus (profunda), como mostrado na Figura 2 e descritos aqui. Isto caracteriza cada grupo de integração como sendo um modelo diferente do outro. Na primeira rodada usou-se o esquema Kuo (KUO), na segunda o esquema modificado de Arakawa e Schubert (RAS) e na terceira o esquema de Grell (GRELL). Mais detalhes sobre KUO e RAS podem ser encontrados em Pezzi et al (2006). O conjunto de dados observados de precipitação usados para a correção dos modelos é o GPCC, que são provenientes do NCEP. Este conjunto de dados combina dados observados por estações de superfície com estimativas derivadas de satélite. Ao todo foram usados 20 anos de dados mensais, no período de janeiro de 1982 a dezembro 2001. Os resultados de modelo foram tomados na resolução T62L28.

3. RESULTADOS PRELIMINARES

Neste trabalho foi feita uma análise preliminar do desempenho global do ECCOCLIM, para a precipitação na América do Sul. Num segundo trabalho (Bezerra et al 2006), esta avaliação será feita em detalhes, separando-se a América do Sul em áreas de interesse e por períodos do ano. A Figura 2, sintetiza o desempenho do ECCOCLIM para América do Sul, durante o período de janeiro de 1982 a dezembro 2001, totalizando 20 anos. Nesta figura, é apresentada a média espacial observada sobre a América do Sul do erro médio quadrático (RMSE) para todo o período analisado. Estes erros são calculados em relação ao conjunto de dados de precipitação observada GPCC. Pode-se ver, de antemão, que o ECCOCLIM reduz os erros, sobre a América do Sul ao longo de todo o ano. Nos meses em que o método melhor corrige as previsões, as diferenças entre a média das previsões e as geradas pelo ECCOCLIM chegam a valores próximos a 3 mm/dia, como visto no mês de Fevereiro. A Figura 3 mostra o RMSE sazonal, calculados ponto a ponto, sobre a América do Sul. Nesta figura os erros calculados a partir dos resultados do ECCOCLIM são identificados no título por ENS e os erros calculados a partir da média aritmética dos modelos por MEAN. De um modo geral, a previsão para a América do Sul formada pelo super-conjunto de multi-modelos apresenta erros menores do que a

média principalmente sobre noroeste da América do Sul e sobre a Cordilheira dos Andes.

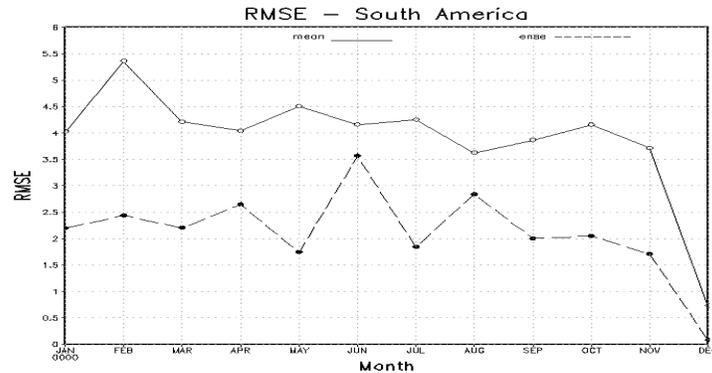


Figura 2 – A linha pontilhada mostra a media espacial observada sobre a América do Sul do erro médio quadrático (RMSE) da precipitação corrigida pelo ECCOCLIM comparada com a precipitação observada. A linha cheia é o RMSE da média aritmética dos modelos, que neste estudo é tomada como o controle. Unidades em mm/dia.

Sabe-se que os modelos dinâmicos, devido à falta de resolução e limitações em suas parametrizações físicas e na própria dinâmica têm grande deficiência em representar corretamente a precipitação sobre a região dos Andes. De uma certa forma o método está corrigindo a precipitação sobre esta área.

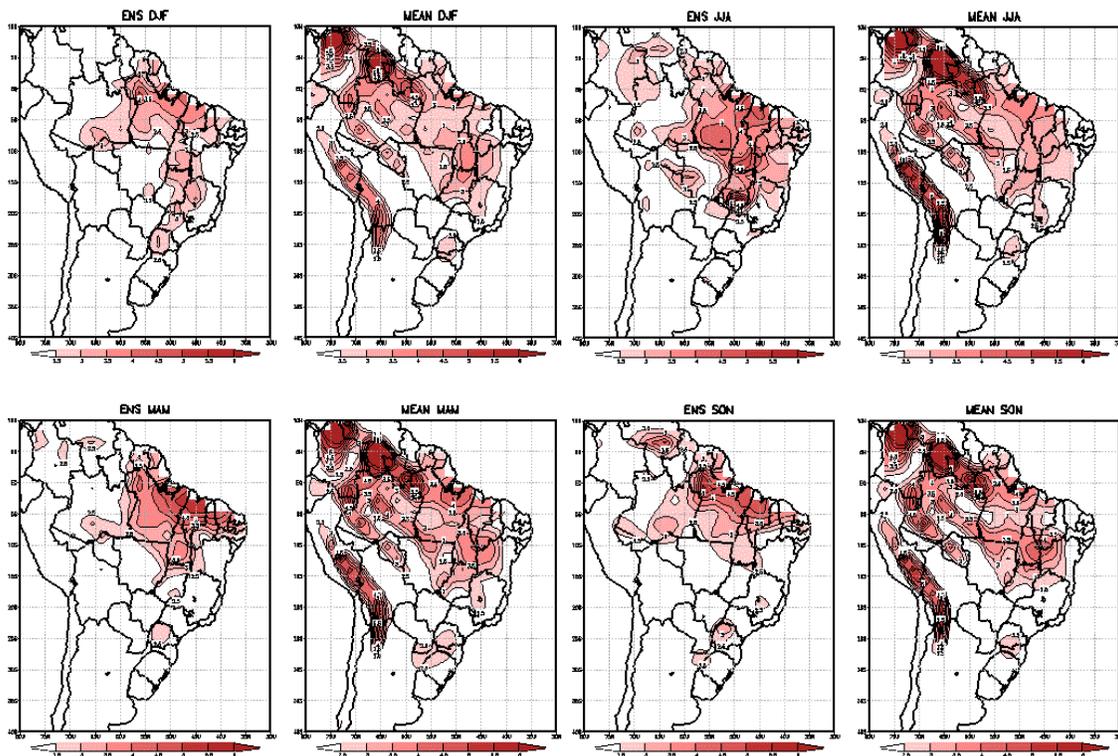


Figura 3 – Mapas sazonais de erros médios quadráticos (RMSE) sobre a América do Sul calculados a partir dos resultados do ECCOCLIM (ENS) e a partir da média aritmética dos modelos (MEAN), em relação ao conjunto de dados de precipitação observada GPCP. Unidades em mm/dia.

Os modelos tendem a superestimar também a precipitação sobre a região noroeste da América do Sul. Onde existe precipitação excessiva o método apresentado aqui tende a corrigir e apresentar melhores

resultados. Porém em algumas áreas como no caso da região norte do NEB durante sua estação chuvosa, trimestre março, abril e maio não se notou um melhor desempenho do método. Uma possível explicação é que para esta região os modelos dinâmicos têm um bom desempenho. Por outro lado, na estação chuvosa da região Sudeste do Brasil e sul do norte do NEB (pré-estação chuvosa), em dezembro, janeiro e fevereiro observa-se um pequeno ganho do método de multi-modelos comparado à média aritmética. É interessante ressaltar, que o método ainda não foi testado com outros conjuntos de dados de precipitação observada, diferentes combinações de modelos. Estes fatos poderão acarretar em mudanças dos resultados reportados aqui.

5. CONCLUSÕES

Os resultados preliminares mostram o potencial de uso do *Esquema Estatístico de Combinação e Correção de Previsões Climáticas (ECCOCLIM)*. Espera-se que o método apresentado aqui contribua para a melhora nos desempenhos das previsões sazonais climáticas de precipitação geradas pelos modelos do CPTEC. Pretende-se ainda trabalhar no refinamento da metodologia e visando melhorar a qualidade das previsões climáticas do CPTEC, atualmente disponibilizadas para tomadores de decisão e sociedade em geral. **Os códigos desenvolvidos dentro do ECCOCLIM serão colocados livremente à disposição de qualquer pesquisador ou instituição que tenham interesse no assunto.**

6. AGRADECIMENTOS

O ECCOCLIM é um trabalho financiado pelo CPCLIM (472224/2004-2) Edital Universal/CNPq (19/2004) e pelo projeto PROTIM3/FINEP (0105091200). Esta é a contribuição 012006 do CPCLIM. Agradecemos Dirceu L. Herdies, José A. Aravéquia, Ronald B. Souza e Andrea Cardoso pelas sugestões ao projeto CPCLIM.

7. REFERÊNCIAS BIBLIOGRÁFICAS

- Bezerra A. C., Pezzi, L. P., Kayano, M. e Cardoso, A (2006). Avaliação do Esquema Estatístico de Combinação e Correção de Previsões Climáticas (ECCOCLIM) sobre o Brasil. *Anais do XIV CBMet. Florianópolis, 2006*.
- Coelho, C.A.S., Stephenson D. B., Balmaseda M., Doblas-Reyes F. J. and. van Oldenborgh G. J, 2006: Towards an integrated seasonal forecasting system for South America. *J. Climate*. Vol. 19, No. 15, 3704-3721.
- Kayano, M. and Andreoli R., 2006 Relationships between rainfall anomalies over northeastern Brazil and the El Niño-Southern Oscillation *Journal of Geophysical Research* –V111, D13101, doi:10.1029/2005JD006142.
- Krishnamurti, T. N.; Surendran, S.; Shin, D. W.; Correa-Torres R J.; Vijaya Kumar, T. S. V., Williford. E., Kummerow, C, Adler, R. F., Simpson, J., Kakar, R., Olson, W. S., and Turk, R.F. (2001). Real-Time Multianalysis–Multimodel Superensemble Forecasts of Precipitation Using TRMM and SSM/I Products. *Monthly Weather Review*, v.129, p. 2861-2883.
- Moura, A. D. and Shukla, J. (1981). On the dynamics of droughts in northeast of Brazil: Observations, theory and numerical experiments with general circulation model. *Journal of the Atmospheric Sciences*, 38:2653–2675.
- Pezzi, L. P. and Cavalcanti, I. (2001). The relative importance of ENSO and tropical Atlantic sea surface temperature anomalies for seasonal precipitation over South America: A numerical study. *Climate Dynamics*, 17:205–212.
- Pezzi, L. P., Vialard, J., Richards K. J., Menkes C., and Anderson, D. (2004). Influence of ocean-atmosphere coupling on the properties of tropical instability waves. *Geophysical Research Letters*, 31, doi:10.1029/2004GL019995:L16306.
- Pezzi, L. P. and Richards, K. J. (2003). The effects of lateral mixing on the mean state and eddy activity of an equatorial ocean. *Journal of Geophysical. Research*. 108(C12):doi:10.1029/2003JC001834.