

UM MODELO ESTOCÁSTICO COMBINADO DE PREVISÃO SAZONAL PARA A PRECIPITAÇÃO NO BRASIL

PAULO SÉRGIO LÚCIO¹, FABRÍCIO DANIEL DOS SANTOS SILVA², LAURO TADEU GUIMARÃES FORTES², LUIZ ANDRÉ RODRIGUES DOS SANTOS², DANIELLE BARROS FERREIRA², MOZAR DE ARAÚJO SALVADOR², HELENA TURON BALBINO², GABRIEL FONSECA SARMANHO², LARISSA SAYURI FUTINO CASTRO DOS SANTOS², EDMUNDO WALLACE MONTEIRO LUCAS², TATIANE FELINTO BARBOSA², PEDRO LEITE DA SILVA DIAS³

¹ Departamento de Estatística, Universidade Federal do Rio Grande do Norte (UFRN), Natal, RN

² Coordenação de Desenvolvimento e Pesquisa, Instituto Nacional de Meteorologia (CDP/INMET), Brasília, DF.

³ Departamento de Ciências Atmosféricas, Instituto de Astronomia, Geofísica e Ciências Atmosféricas, Universidade de São Paulo (IAG/USP), São Paulo, SP.

pslucio@ccet.ufrn.br, {fabricao.silva, lauro.fortes, luiz.santos, danielle.ferreira, mozar.salvador, helena.balbino}@inmet.gov.br, gabriel.sarmanho@hotmail.com, {larissa.santos, edmundolucas, tatiane.barbosa}@inmet.gov.br, pldsdias@master.iag.usp.br

Recebido Novembro 2007 – Aceito Setembro 2009

RESUMO

Este artigo discute um modelo de previsão combinada para a realização de prognósticos climáticos na escala sazonal. Nele, previsões pontuais de modelos estocásticos são agregadas para obter as melhores projeções no tempo. Utilizam-se modelos estocásticos autoregressivos integrados a médias móveis, de suavização exponencial e previsões por análise de correlações canônicas. O controle de qualidade das previsões é feito através da análise dos resíduos e da avaliação do percentual de redução da variância não-explicada da modelagem combinada em relação às previsões dos modelos individuais. Exemplos da aplicação desses conceitos em modelos desenvolvidos no Instituto Nacional de Meteorologia (INMET) mostram bons resultados e ilustram que as previsões do modelo combinado, superam na maior parte dos casos a de cada modelo componente, quando comparadas aos dados observados.

Palavras-chave: Previsões sazonais, Modelos estocásticos; Correlações canônicas, Modelo agregado.

ABSTRACT A COMBINED STOCHASTIC MODEL FOR SEASONAL PREDICTION OF PRECIPITATION IN BRAZIL.

This article discusses a combined model to perform climate forecast in a seasonal scale. In it, forecasts of specific stochastic models are aggregated to obtain the best forecasts in time. Stochastic models are used in the auto regressive integrated moving average, exponential smoothing and the analysis of forecasts by canonical correlation. The quality control of the forecast is based on the residual analysis and the evaluation of the percentage of reduction of the unexplained variance of the combined model with respect to the individual ones. Examples of application of those concepts to models developed at the Brazilian National Institute of Meteorology (INMET) show good results and illustrate that the forecast of the combined model exceeds in most cases each component model, when compared to observed data.

Keywords: Seasonal forecasts, stochastic models, canonical correlations, cluster model.

1. INTRODUÇÃO

As previsões são naturalmente afetadas por incertezas, pois os sistemas dinâmicos que controlam a evolução climática têm forte componente caótica. Existem, contudo, controles de baixa frequência, associados aos lentos processos oceânicos, que conferem certa previsibilidade ao clima na escala sazonal, principalmente na região tropical (Shukla, 1981; Cavalcanti et al, 2002).

Neste trabalho, uma análise estocástica é apresentada, visando reduzir o erro de projeção de séries cronológicas no tempo. Cada situação, com seu conjunto único de problemas, requer uma previsão específica, e a solução para um caso não é necessariamente a melhor para outras situações. Não obstante, certos princípios gerais são comuns para a maioria dos problemas e devem ser incorporados em qualquer sistema de previsão.

A finalidade desta pesquisa foi explorar o comportamento de séries temporais de precipitações acumuladas trimestrais, visando a elaboração de previsões sazonais, analisando as séries e os mecanismos (processos) que a geram, propondo um modelo estocástico de previsão que combine resultados de previsões por modelos clássicos de Box-Jenkins (Box e Jenkins, 1976), Holt-Winters (Holt, 1957; Winters, 1960), e fornecidos por Análise de Correlações Canônicas (ACC) (Barnston e Smith, 1996). Busca-se um modelo combinado que, para a maioria das estações meteorológicas, apresente um melhor skill (desempenho) (Coelho *et al.*, 2003; Coelho *et al.*, 2006; Stephenson *et al.*, 2005) do que estes modelos individualmente,

por meio da análise dos resíduos e da maximização da eficiência relativa.

2. DADOS EXPERIMENTAIS

Os métodos de previsão discutidos no artigo são aplicados a séries de precipitações de 160 estações meteorológicas de superfície, espacialmente distribuídas no território brasileiro (Figura 1), que possuem séries históricas longas, do Instituto Nacional de Meteorologia (INMET) para gerar, mensalmente, previsões sazonais de precipitação.

Como ilustração e estudo de caso, foram selecionadas séries temporais de precipitação mensal das seguintes capitais brasileiras: Manaus (AM), Belém (PA), Fortaleza (CE), Recife (PE), Vitória (ES), Goiânia (GO), São Paulo (SP) e Porto Alegre (RS). Estas estações foram escolhidas por sua representatividade do clima das regiões em que se localizam, e pela boa qualidade dos dados, sendo séries mensais homogêneas e contínuas (sem falhas ou valores espúrios), iniciando em janeiro de 1961 (Tabela 1). Tais séries são acumuladas trimestralmente, para gerar previsões climáticas sazonais.

3. METODOLOGIA

Para o modelo agregado desenvolvido, são utilizados resultados de modelos estocásticos do tipo Auto Regressivo Integrado a Médias Móveis (ARIMA), e um modelo de alisamento exponencial Holt-Winters. Estes modelos analisam a variável-alvo dependente como função do tempo. No caso

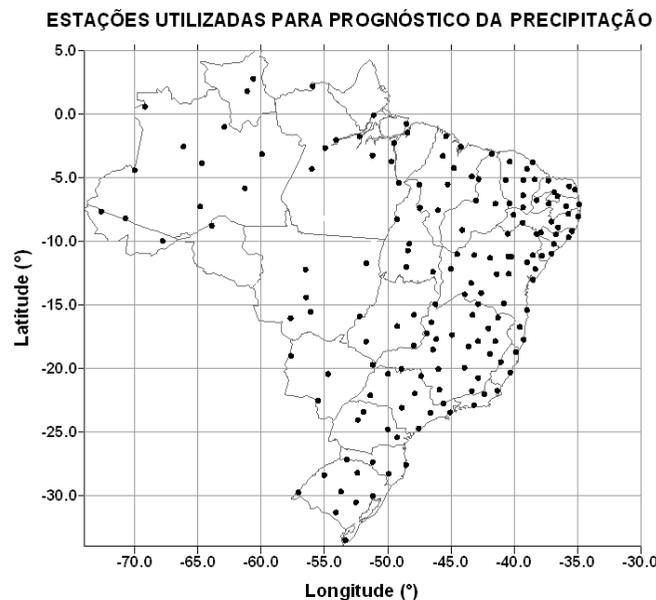


Figura 1 - Distribuição geográfica das estações meteorológicas do INMET usadas para geração de prognósticos sazonais de precipitação.

das séries temporais de precipitação, para estes dois modelos, usam-se séries acumuladas trimestrais. Estes modelos são úteis para previsibilidade de uma condição seca, normal ou chuvosa, porque a situação e, por conseguinte, a previsão para um determinado trimestre muitas vezes segue uma tendência de situações anteriores, ou seja, geralmente existe uma memória do comportamento de trimestres anteriores em relação aos trimestres seguintes, e tais condições de memória podem ser analisadas por meio de funções de autocorrelação (ACF), que medem o grau de dependência entre os valores de uma série temporal em diferentes períodos sazonais, e por meio de funções de autocorrelações parciais (PACF), que pode ser definida como a seqüência de correlações de determinados termos de uma série temporal e os elementos da série mais recentes, medindo a influência de trimestres recentes nos trimestres seguintes.

Além disso, as respostas as forçantes dinâmicas que influenciam na precipitação em várias regiões, aparecem em séries de precipitações trimestrais como períodos de maior (ou menor) acúmulo e geralmente persistem por alguns trimestres. Desta forma, são usadas séries trimestrais de precipitação nas previsões, onde são consideradas as defasagens entre trimestres, com dados mensais totalmente observados, onde os trimestres com dados observados “incompletos” são preenchidos com as normais climatológicas, reduzindo a maior variância que ocorre naturalmente mês a mês, e que é suavizada em acumulações trimestrais. Como ilustração, para se prever a chuva do trimestre junho-julho-agosto em maio, usam-se acumulados trimestrais com dados observados de toda a série até o trimestre fevereiro-março-abril. O trimestre março-abril-maio é composto pelos dados observados de março e abril e a climatologia de maio. O trimestre abril-maio-junho é composto pelos dados observados de abril com a climatologia de maio e junho e por fim, o último trimestre maio-junho-julho é representado pela sua climatologia. Tal metodologia é empregada tanto para o cálculo das previsões futuras, quanto para previsões passadas, que possibilitam a comparação a dados observados, possibilitando a obtenção das correlações para os modelos individuais ARIMA e HOLT-WINTERS.

Para a ACC, é usado o software “Climate Predictability Tool”, fornecido pelo “International Research Institute for Climate and Society” (IRI). A variável preditora utilizada para prever a precipitação acumulada trimestral, é a Anomalia de Temperatura à Superfície do Mar (ATSM) observada, obtida do “National Ocean and Atmospheric Administration” (NOAA), com resolução espacial de 1° x 1° (Reynolds et al, 2002), em uma janela que cobre boa parte das águas dos oceanos Pacífico e Atlântico (20°N-40°S, 135°W-0°E). Tais áreas são capazes de captar as influências das regiões de Niños, como a variabilidade do volume de precipitação verificada em parte das Regiões Norte, Nordeste e Sul, e dos oceanos Atlântico Tropical Norte (ATN) e Atlântico Tropical Sul (ATS), incorporando os efeitos de Dipolo, que é um importante modo de variabilidade oceano-atmosfera dominante no Atlântico Tropical (Repelli e Nobre, 2004). Este modo de variabilidade caracteriza-se pelo aparecimento de um padrão de anomalias de TSM, configurando-se especialmente com sinais opostos sobre as Bacias Norte e Sul do Atlântico Tropical (Moura e Shukla, 1981). A fase negativa do Dipolo do Atlântico Tropical, por exemplo, é definida quando o Atlântico Sul se encontra com águas mais quentes do que a média histórica, e águas mais frias no Atlântico Norte.

Para tanto, é necessário definir o número mínimo e o número máximo de modos da ACC, sendo estes limitados aos mesmos valores de máximos e mínimos das FOE (Funções ortogonais empíricas) das variáveis preditora (ATSM) e resposta (precipitação). Na prática, o CPT tem apresentado melhores resultados com máximo de 10 modos para a FOE e até de 4 modos para a ACC.

As vantagens da ACC são a fácil interpretação, a consistência espacial e a alta sensibilidade nas respostas do preditando aos preditores.

3.1 Descrição e Identificação do Modelo Estocástico

Na prática, uma série temporal pode ser analisada como uma realização parcial de um processo estocástico, cuja característica principal se fundamenta no fato das variáveis

Tabela 1 – Estações climatológicas analisadas individualmente, a título de ilustração, com respectivos códigos da OMM, coordenadas geográficas. Período de dados disponíveis: Jan/1961 – Out/2008

CIDADE/ESTADO/CÓDIGO	LAT (S)	LONG (W)	ALT (m)
MANAUS – AM – 82331	3° 08'	60° 01'	71,90
BELÉM – PA – 82191	1° 27'	48° 28'	10,00
FORTALEZA – CE – 82397	3° 46'	38° 36'	26,45
RECIFE – PE – 82900	8° 03'	34° 55'	6,90
VITÓRIA – ES – 83648	20° 19'	40° 20'	36,30
GOIANIA – GO – 83423	16° 40'	49° 15'	741,48
SÃO PAULO – SP – 83781	23° 30'	46° 37'	792,06
PORTO ALEGRE – RS – 83967	30° 01'	51° 13'	46,97

apresentarem uma estrutura de dependência. Denota-se a série temporal por $\{Y(t) \ t \in T \subset \mathbb{N}\}$, que em sua forma discretizada se transforma em: $\{Y(t) \ t \in T \subset \mathbb{N}\} = Y(1) \ Y(2), \dots, Y(n)$, onde n é o tamanho da série. Neste trabalho estudam-se séries temporais discretas, onde os dados são coletados ou convertidos em observações mensais ou trimestrais. O modelo clássico considera as séries temporais como sendo compostas de quatro elementos básicos: tendência, variações cíclicas, variações sazonais, erros aleatórios ou sistemáticos.

Os modelos autorregressivos associam a variável dependente Y_t ou $Y(t)$ com funções do tempo. Estes modelos são úteis quando os parâmetros que descrevem a série temporal a ser

prevista permanecem constantes no tempo (série estacionária).

Algumas vezes pode-se descrever uma série temporal Y_t usando um modelo de tendência. Tal modelo é definido como $Y_t = \mu + a_t$, onde Y_t é o valor da série temporal no tempo t , μ_t é o termo da tendência no período t e a_t é o erro aleatório associado ao período de tempo t .

O modelo de tendência indica que a série temporal pode ser representada pelo nível médio e pelo termo do erro sistemático. As tendências mais simples são aquelas obtidas através de um comportamento linear da série observada (neste caso, μ_t é o coeficiente angular da reta, μ , multiplicado por t). Estimativas pontuais dos parâmetros destes modelos

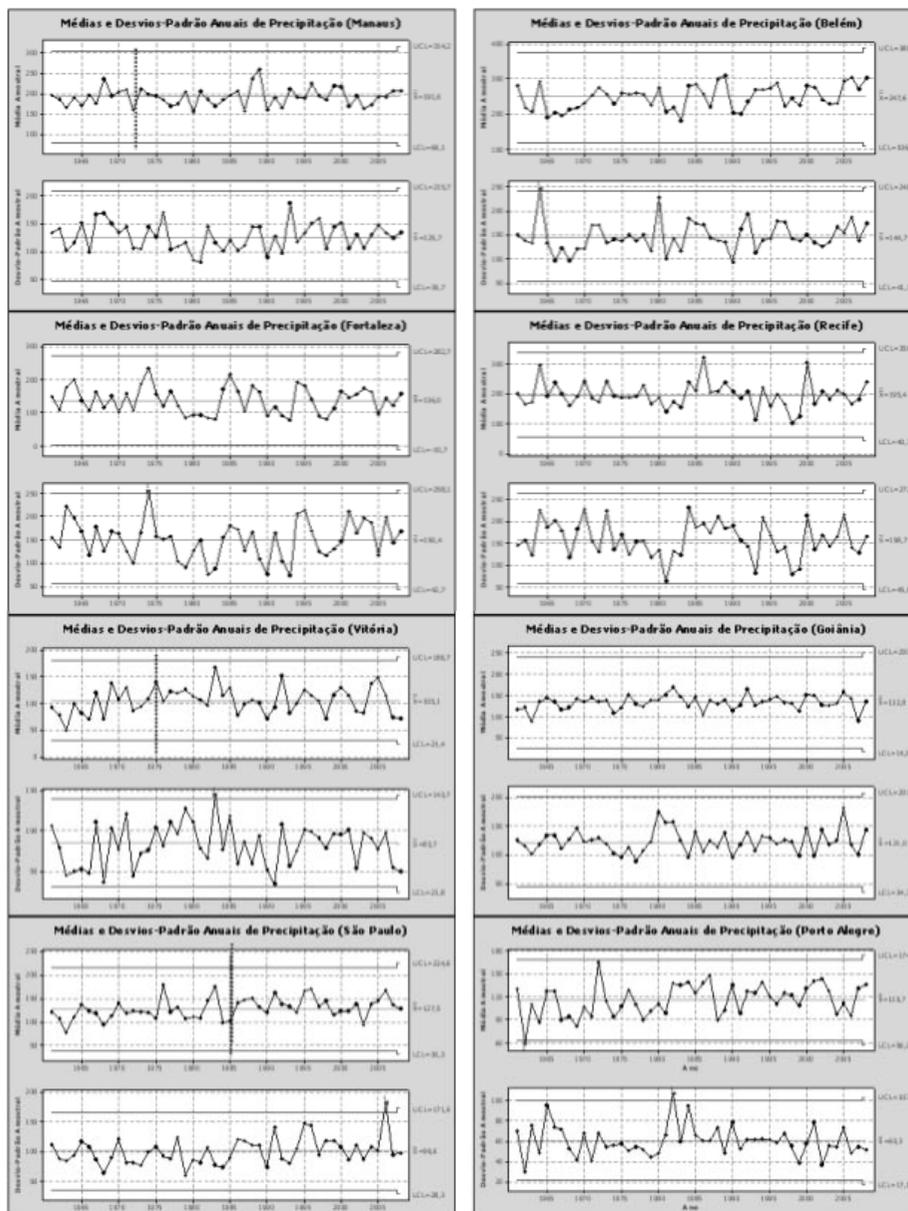


Figura 2 – Séries temporais das médias móveis (ordem 12) da precipitação mensal e respectivos desvios-padrão. Nenhuma tendência linear é detectada, apenas mudanças estruturais em algumas séries de precipitação média anual.

de tendência, pelo método dos mínimos quadrados, podem ser obtidas usando-se técnicas de regressão. Neste caso, assume-se que os termos de erro a_t satisfazem às suposições de variância constante, independência e normalidade.

As variações cíclicas são variações periódicas de amplitude superior a um ano. Para isolar variações cíclicas das outras variações (tendência e sazonal) estas devem ser removidas dos dados da série. Para remover tendência, subtrai-se de cada observação o correspondente valor de tendência. O resultado é uma série de desvios em relação à tendência, que ao ser representada graficamente revela ou não ciclos. As variações sazonais são as que ocorrem regularmente dentro do período de um ano. Há duas finalidades para se isolar a componente sazonal de uma série temporal: remover determinado padrão a fim de se estudar as variações cíclicas, e identificar os fatores sazonais de forma que eles tenham relevância na modelagem e tomadas de decisões.

Para se retirar o componente sazonal, aplica-se uma transformação a Y_t para obtenção de Y_t^* , na forma de uma média móvel centrada de 12 elementos (Figura 2). Para verificar a existência de tendência significativa, aplica-se a Y_t^* , um teste não-paramétrico, baseado no coeficiente de correlação de Spearman, \tilde{r} , que mede a relação linear entre duas variáveis - no caso a significância da tendência em relação ao tempo (Box e Jenkins, 1976). No estudo das séries de precipitação utilizadas no modelo do INMET, Y_t , a componente tendência não se mostrou significativa (ao nível de 5%) para nenhuma das estações meteorológicas analisadas, apesar das ilusões provocadas pela inspeção visual (devido aos pontos de mudança ou de intervenção – heterogeneidades) evidenciadas através das mudanças estruturais detectadas pelo Teste de Chow (Chow, 1960) (Figura 2). Desta forma, não se rejeita a hipótese nula, H_0 : não existe componente tendência! As séries livres do efeito sazonal são distinguidas por meio de um amortecimento usando uma média móvel centrada e de tamanho lag = 12.

Quanto ao teste para avaliação da sazonalidade, antes da estimação, utiliza-se o teste paramétrico com base no cálculo da estatística ES (Cordeiro, 2002) dada pela equação:

$$ES = \left(\frac{n-L}{L-1} \right) \left(\frac{\sum_{j=1}^L m_t (Y_j^* - Y^*)^2}{\sum_{j=1}^L \sum_{t=1}^{m_t} m_t (Y_{tj}^* - \bar{Y}_j^*)^2} \right),$$

onde $m_t = 4$ representa o número de estações sazonais no ano e Y_j^* é a série livre da componente tendência e $L = \text{lag} = 12$.

Os valores das estatísticas de teste são comparados aos valores da distribuição F de Snedecor (Cordeiro, 2002), rejeitando-se a hipótese nula de que não existe fator sazonal, exceto para locais como Vitória, onde a sazonalidade da precipitação, embora presente seja fraca, e Porto Alegre, onde a sazonalidade é inexistente, isto é, as autocorrelações não são

significativas (Figura 2 e Figura 3). Observa-se também, que as amplitudes da série não variam com a média, sendo isto um forte indicativo para a modelagem aditiva da componente sazonal.

As variações sazonais referem-se a movimentos similares, que uma série temporal obedece durante os mesmos meses de anos sucessivos. A Figura 3 (Índices de Sazonalidade) ilustra as séries temporais em função da importância dos fatores sazonais. O índice de sazonalidade tem por objetivo analisar o comportamento típico de uma série temporal. Para tanto, esta análise foi realizada em intervalos de tempos equidistantes, onde se calcula a média da precipitação mensal observada dividida pela média centrada. Desta forma as cargas sazonais são assim determinadas em conjunto com as contribuições mensais (Figura 3 - Contribuição da Variabilidade Mensal). Estes resultados são mostrados em gráficos do tipo Box e Whiskers (Wilks, 2006; Evans e Doswell, 2001). Através dos Box-Plot mensais (Figura 3 - Box-Plot da Precipitação Mensal) constata-se a contribuição da precipitação associada à sua variabilidade intra-sazonal, corroborada pelo cálculo da estatística ES . Uma vez retiradas às contribuições sazonais analisa-se os Box-Plot dos Resíduos (Figura 3 - Box-Plot dos Resíduos), no intuito de se obter evidências sobre os meses que mais contribuem para as incertezas das previsões, devido à alta variabilidade intrínseca do processo de precipitação.

3.2 Modelos Estocásticos Univariados

3.2.1 Modelagem Holt-Winters

Pode-se considerar o ajustamento exponencial como mais uma técnica para prever valores de séries temporais. O método de Holt-Winters é um dos mais utilizados para a previsão de curto prazo, devido a sua simplicidade, baixo custo de operação, boa precisão e capacidade de ajustamento automático e rápido a mudanças na série. Este modelo isola na série até três fatores ou coeficientes de alisamento: nível, tendência linear, fator sazonal e um elemento residual não previsível, chamado erro aleatório. Na estimação desses fatores usa-se o método de ajustamento exponencial, também chamado “suavização exponencial”. O nome “suavização” provém do fato de que a série depois de reduzida a seus componentes estruturais terá menor número de variações bruscas, mostrando um comportamento mais suave. O termo “exponencial” aparece, porque os processos de suavização envolvem as médias aritméticas ponderadas, onde os pesos decrescem exponencialmente na medida em que se avança no passado. As equações de previsão se arranjam de duas formas: aditiva ou multiplicativa, conforme a natureza da série (Holt, 1957; Winters, 1960; Box e Jenkins, 1976). Para calcular as previsões de valores futuros da série é necessário estimar o nível e a tendência da série no período atual, bem como

os valores do fator sazonal correspondente ao último período de sazonalidade. Estas estimativas são efetuadas por meio das seguintes equações:

Sazonalidade Aditiva – Nestes modelos são utilizadas as seguintes equações recursivas:

$$n_t = \alpha(Y_t - f_{t-s}) + (1 - \alpha)(n_{t-1} + b_{t-1}),$$

$$b_t = \beta(n_t - n_{t-1}) + (1 - \beta)b_{t-1}$$

$$f_t = \gamma(Y_t - n_t) + (1 - \gamma)f_{t-s};$$

Sazonalidade Multiplicativa – Nestes modelos são utilizadas as seguintes equações recursivas:

$$n_t = \alpha \frac{Y_t}{f_{t-s}} + (1 - \alpha)(n_{t-1} + b_{t-1}),$$

$$b_t = \beta(n_t - n_{t-1}) + (1 - \beta)b_{t-1}, f_t = \gamma \frac{Y_t}{n_t} + (1 - \gamma)f_{t-s},$$

onde α, β e γ são constantes de amortecimento. Ao final de cada período t , a estimativa do “passo” (tendência) e a componente

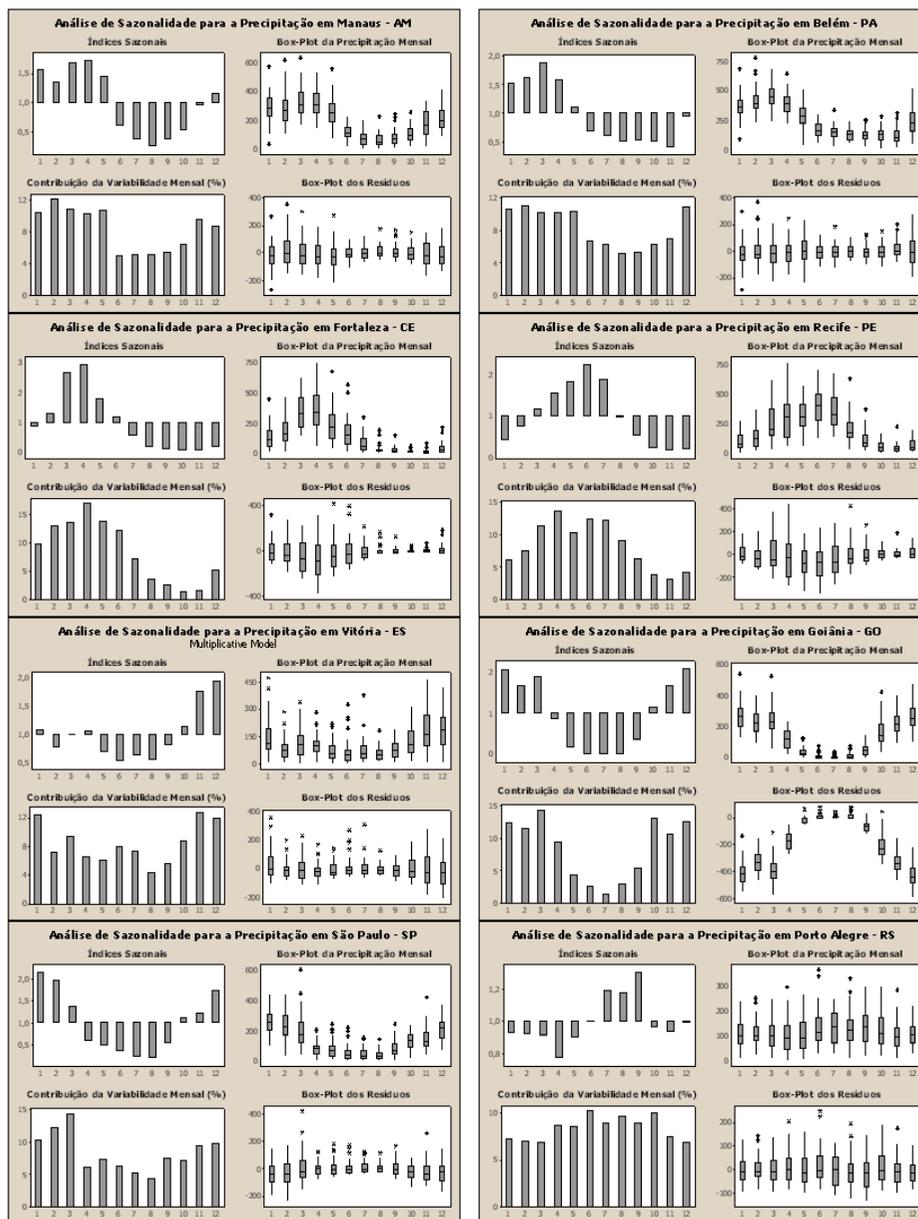


Figura 3 – Decomposição sazonal para analisar as séries temporais em termos de componentes sazonais, tendência linear e comportamento aleatório (resíduos). Emprega-se esta técnica para examinar a natureza dos componentes. Observa-se que os resíduos são maiores nas estações do inverno e da primavera.

sazonal são dadas por b_t e f_t , respectivamente. Já n_t denota a componente de nível.

Bowerman (1987) e Chatfield (1988) sugerem que o melhor critério de escolha entre os fatores multiplicativos ou aditivos é calcular o valor do Desvio Absoluto Médio (DAM), assim a opção recai no menor erro. Alternativamente pode-se utilizar como critério de decisão o Desvio Quadrático Médio (DQM).

Após as análises, tanto quantitativas quanto qualitativas, escolheu-se para as séries o amortecimento (ou suavização) exponencial sazonal aditivo de Holt-Winters. O modelo teve um bom desempenho com base nas medidas do DAM e do DQM. Na análise dos resíduos não foi constatado qualquer comportamento sistemático. Em suma, com base na estatística de Box-Pierce, aceitou-se estatisticamente, ao nível de significância de 5%, a hipótese nula para as autocorrelações residuais $r_k(\hat{a}) = 0$, para todas as defasagens múltiplas de $k = 12$.

3.2.2 Modelagem Box-Jenkins

Os modelos Box-Jenkins, denominados também de ARIMA (*Autoregressive Integrated Moving Average*), são modelos estocásticos lineares para análise de séries temporais. A base deste tipo de modelagem consiste no fato de que uma série temporal $Y(t)$ apresenta correlações seriais, e cada modelo pode ser considerado como gerado por uma seqüência de choques aleatórios independentes entre si, com determinada distribuição com média zero e variância constante, um processo com *ruído branco*. A metodologia supõe que o procedimento de filtragem do *ruído branco* pode ser decomposto em três etapas: (1) média móvel, (2) auto-regressivo estacionário e (3) diferenciação, caso não seja estacionário.

Os termos auto-regressivos correspondem a defasagens da série transformada (série estacionária induzida por diferenciação) e as médias móveis correspondem a defasagens dos erros aleatórios. O termo “integrado” refere-se ao processo de diferenciação da série original. A notação empregada para designação do modelo é normalmente ARIMA (p, d, q), onde p representa o número de coeficientes (parâmetros) auto-regressivos, d representa o número de diferenciações para que a série torne-se estacionária, e q o número de coeficientes de médias móveis. As ferramentas fundamentais na modelagem de séries temporais, baseada na metodologia de Box-Jenkins, são as funções de autocorrelação (ACF) e funções de autocorrelação parcial (PACF).

A ACF indica o grau de dependência temporal existente, medindo o quanto os pares ordenados de observações (Y_t , Y_{t+k}) estão relacionados, variando entre -1 e +1. No caso das séries de precipitação a não-estacionariedade é provocada pela componente sazonal. Esse fato é reforçado pelo correlograma da série livre da componente tendência, em que se rejeita a hipótese

nula para as autocorrelações r_k ao longo dos “lags” $k = 12$. Na Figura 5, se as linhas vermelhas são limites de significância para as autocorrelações, a hipótese nula de autocorrelação nula é rejeitada. Sendo assim, sugere-se uma transformação para a extrapolação da série, considerando-se um regime não-estacionário, quanto ao nível e sazonalidade, transformação esta, que possa induzir estacionariedade homogênea, condição necessária e suficiente para modelar uma série temporal. A PACF mede a intensidade da relação entre duas observações da série temporal, controlando (mantendo constante) o efeito das demais. Os modelos sazonais são identificados tanto na ACF, quanto na PACF, da mesma forma que os modelos auto-regressivos (AR) simples. Assim, se detecta um modelo sazonal puro, por exemplo de ordem $k = 12$, onde a ACF terá um decréscimo rápido nas defasagens múltiplas de $k = 12$ e a PACF apresentará somente um pico, na defasagem $k = 12$.

Como ilustração, pode-se dissertar sobre as características teóricas destas funções no que tange a alguns processos: 1 - AR(p), ACF infinita decrescente, exponencial e/ou senóides amortecidas, PACF finita (nula para *lags* maiores que “ p ”); 2 - MA(q), ACF finita (nula para *lags* maiores que “ q ”), PACF infinita decrescente, exponencial e/ou senóides amortecidas; 3 - ARMA(p, q), ACF infinita (exponencial e/ou senóides amortecidas para *lags* maiores que “ $p-q$ ”), PACF infinita (exponencial e/ou senóides amortecidas para *lags* maiores que “ $p-q$ ”).

Os estágios dos ciclos iterativos usados no modelo ARIMA são: 1- Uma classe geral de modelos é considerada para a análise (especificação); 2 - Identificação de um modelo, com base na análise de autocorrelações (ACF), autocorrelações parciais (PACF) e outros critérios; 3 - Estimativa dos parâmetros do modelo identificado; 4 - Verificações do modelo ajustado, através de uma análise de resíduos, para se saber se este é adequado para os fins em vista (previsão, no caso); 5 - Previsões ou projeções no tempo.

Como a maioria dos procedimentos de análise estatística de séries temporais supõe que estas sejam estacionárias, faz-se necessária a adoção de transformações, principalmente para retificar problemas, devido à variabilidade dos dados, ou seja, estabilizar a variância, melhorando significativamente as estimativas dos parâmetros do modelo. Este é um critério de decisão. Caso contrário deve-se fazer uma ou mais diferenças, até que se consiga a estacionariedade. A estacionariedade é verificada de forma primária, por inspeção visual. Desta forma, a transformação mais comum consiste em aplicar o operador de diferenças sucessivas ∇ à série original Y_t até se obter uma série estacionária. Com efeito, em geral, foi suficiente tomar até duas diferenças simples e uma diferença sazonal, estabilizando a variância e tornando-a estacionária. Após a transformação Y_t em $X(t)$, deve-se verificar o correlograma da

série transformada $X(t)$ (Figura 5a e Figura 5b). Após a análise das funções de autocorrelação das séries Y_t (Figura 4) e $X(t)$, nota-se claramente: 1 - Y_t - Existência de uma componente periódica de período ou comprimento sazonal; 2 - Y_t - As autocorrelações estimadas apresentam picos decrescentes – senóide amortecida por uma exponencial nos lags múltiplos de $k = 12$; 3 - $X(t)$ - As autocorrelações estimadas apresentam pico apenas no “lag” $k = 12$, alguns dos demais podem ser vistos como acúmulos residuais, que provavelmente serão reduzidos a partir do processo de modelagem.

Nos casos descritos neste trabalho, detectou-se que a diferença deve ser sazonal e de ordem 12 para se atingir a estacionariedade: na ACF da série das diferenças há um pico na defasagem um (1) e outro na defasagem doze (12);

na PACF há um pico na defasagem de ordem doze (12) com “decaimento” exponencial em seus múltiplos. Desta forma, há evidências de que o modelo de previsão a ser ajustado seja um $ARIMA(1,0,0)(0,1,0)_{12}$. No caso dos estudos realizados, os modelos funcionaram de forma satisfatória, com relação à condição de estacionariedade, e através da estatística de *Portemanteau* – análise de adequação (Ljung e Box, 1978), com base nos resíduos não se rejeita a hipótese nula relativa às autocorrelações; isto é, considera-se $r_k(\hat{a}) = 0$ para todas as defasagens múltiplas de $k = 12$.

Torna-se importante salientar que a inspeção gráfica da série é uma ferramenta poderosa para se fazer inferência e para uma melhor prospecção do modelo. Nestes termos, imputou-se ao modelo inicial um teste de sobre-fixação, cujo objetivo é verificar

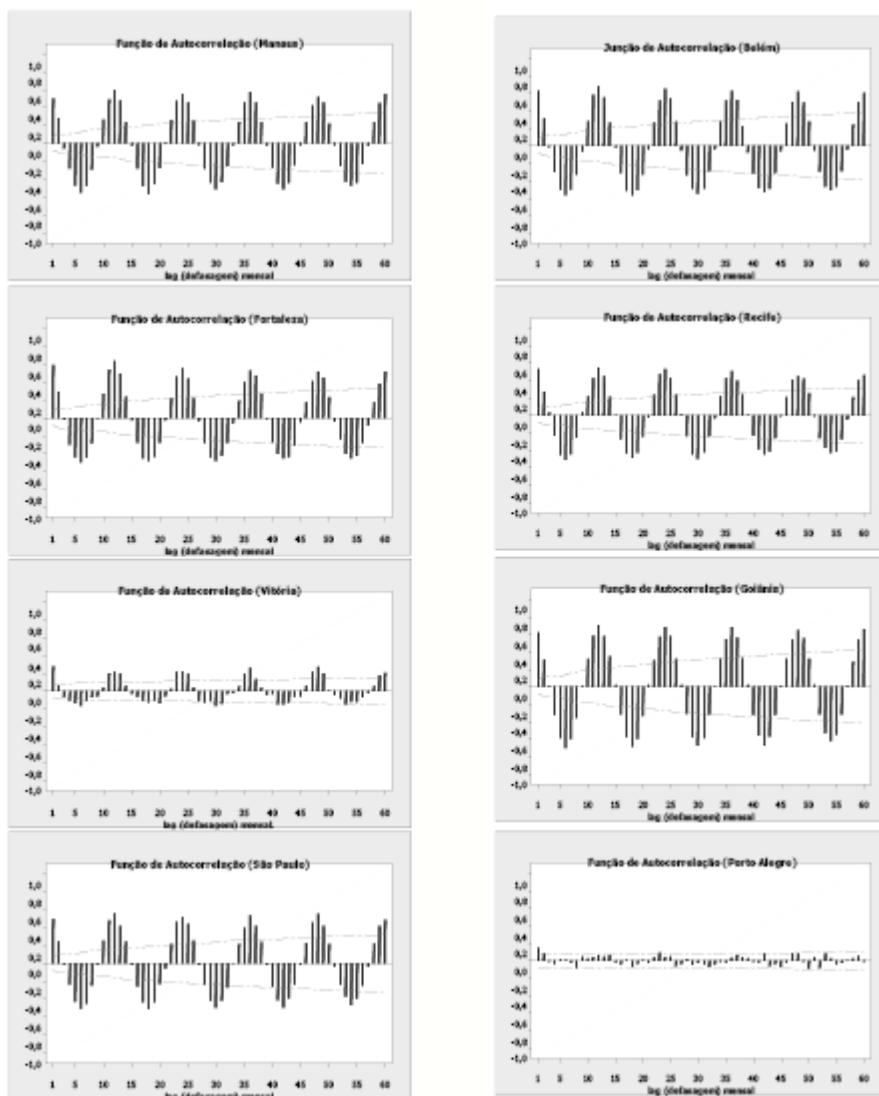


Figura 4 – A autocorrelação é a correlação entre observações da série temporal com defasagem de k unidades de tempo. O gráfico de autocorrelação é a função de autocorrelação (ACF). A ACF é um guia para a escolha dos parâmetros para o modelo ARIMA.

se o número de parâmetros é correto. Optou-se pela inclusão do parâmetro referente à média móvel: ARIMA(1,0,0)(0,1,1)12, cuja estrutura da equação é parcimoniosa no modelo Box-Jenkins. A Figura 6 exibe as previsões para o ano 2008 e seu respectivo intervalo de confiança a 95%. Com efeito, estes modelos além de atenderem às condições de estacionariedade e invertibilidade apresentam todas as qualidades estatísticas do modelo anterior ARIMA(1,0,0)(0,1,0)12, porém, com melhor desempenho, com base na análise dos resíduos, bem como, na variância residual que é minimizada.

3.3 Modelos Estocásticos Multivariados

A ACC (Johnson e Wichern, 2007) pode ser vista como

uma extensão da regressão múltipla. Na ACC existem duas ou mais variáveis dependentes. O princípio básico consiste em desenvolver uma combinação linear em cada um dos conjuntos de variáveis, de tal forma que a correlação entre os dois conjuntos seja maximizada. Na metodologia de correlação canônica não existe a distinção entre variável independente e dependente, existem somente dois conjuntos de variáveis em que se busca a máxima correlação entre ambos. A equação básica pode ser expressa por:

$$\beta_1 PREC_1 + \dots + \beta_k PREC_k = \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \dots + \alpha_n X_n.$$

Salvador *et al.* (2006) corroboram, através da aplicação do método de ACC utilizando o CPT, a existência de uma forte

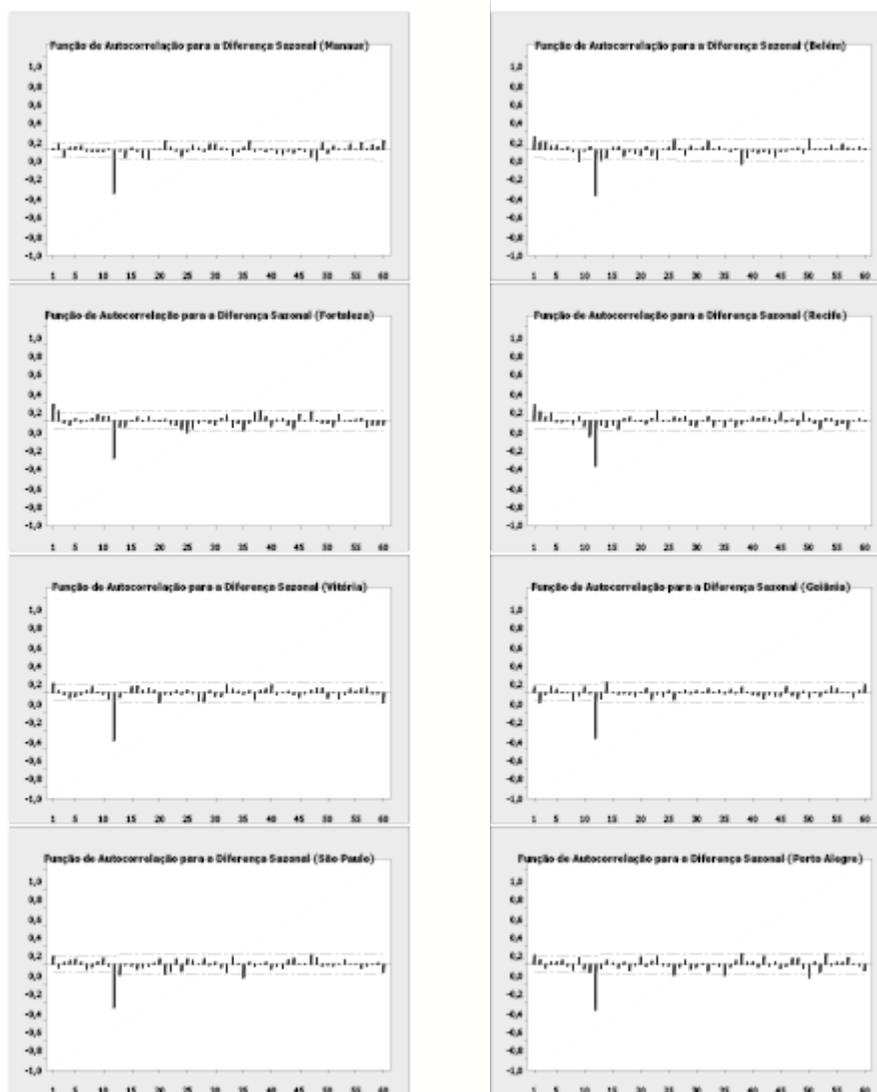


Figura 5a – O gráfico da função de autocorrelação (ACF) das diferenças sazonais (ordem 12). As diferenças sazonais são utilizadas para simplificar a estrutura de correlação na identificação de padrões como guia para a escolha dos parâmetros para o modelo ARIMA.

correlação direta entre a temperatura da superfície do mar (variável explicativa) no Oceano Atlântico Sul e a precipitação pluviométrica (variável dependente) na costa leste do Nordeste do Brasil (Oliveira, 2005 e Pezzi et al., 1998). Por exemplo, a análise de correlação canônica do CPT apresentou um único modo das Funções Ortogonais Empíricas (FOE), com o significativo valor da correlação de aproximadamente 0,73 entre a TSM e a precipitação total durante a quadra chuvosa do leste do NEB (Nordeste Brasileiro). Isso indica que o CPT consegue captar de forma eficiente a correlação entre a variabilidade da precipitação e a TSM do Atlântico.

Ferreira *et al.* (2006) aplicaram a ACC do CPT para a verificação de correlações entre a TSM dos Oceanos Pacífico e Atlântico (séries trimestrais de 1989 a 2005), e a produtividade

de soja no estado do Paraná. Notou-se claramente uma evolução dos valores dos coeficientes de correlação entre a TSM e a produtividade a partir do trimestre outubro-novembro-dezembro, e um decaimento a partir do trimestre fevereiro-março-abril, principalmente no que tange à TSM do Oceano Atlântico Sul. Os resultados mostraram que os trimestres dezembro-janeiro-fevereiro, janeiro-fevereiro-março e fevereiro-março-abril apresentaram coeficientes de correlação de 0,75 para o Atlântico e de 0,62 para o Pacífico na região conhecida como Niño 3.

3.4 Modelo Combinado de Previsão

Para cada trimestre em foco e para cada ponto de estação, são geradas previsões passadas (*hindcasts*), para um período de

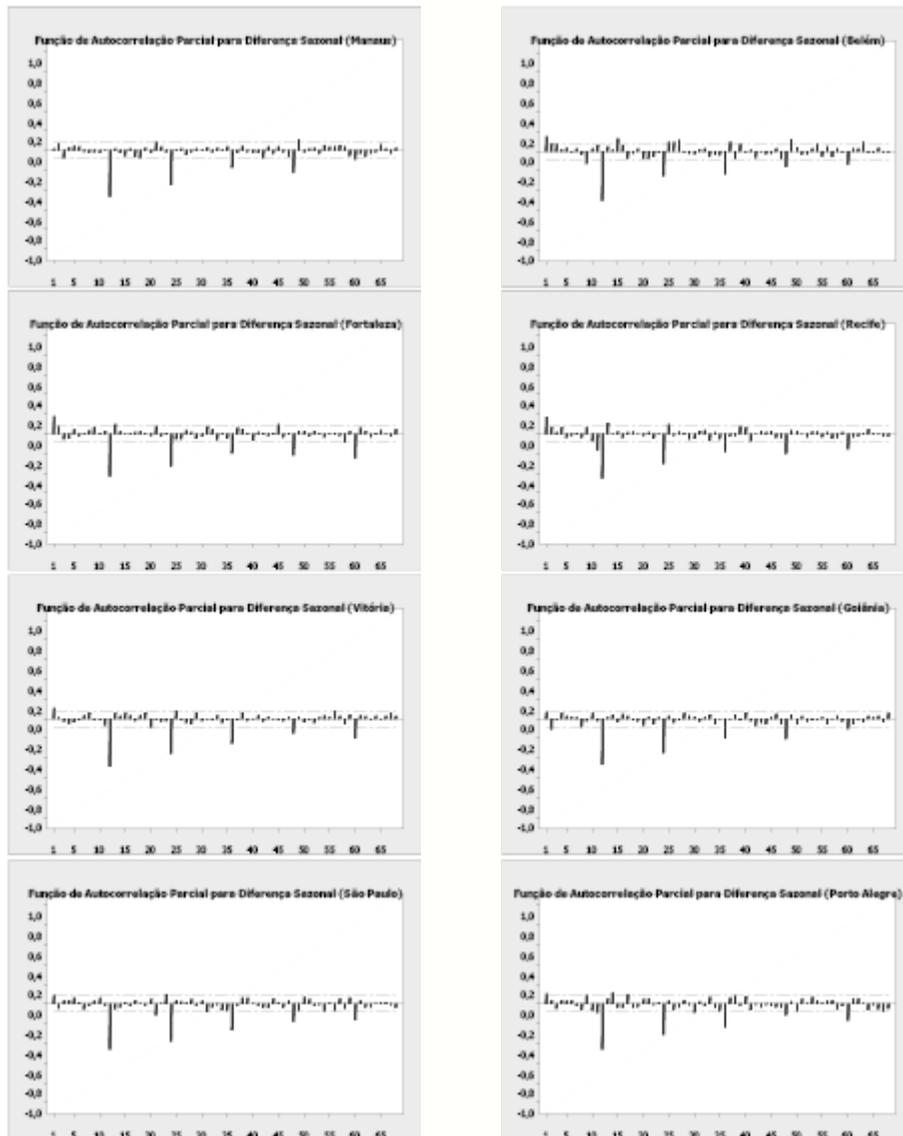


Figura 5b – O gráfico da função de autocorrelação parcial (PACF) das diferenças sazonais (ordem 12). A PACF é uma técnica de identificação de padrões que serve como guia para a escolha dos parâmetros para o modelo ARIMA.

treze anos, 1995 a 2007, comparando-as aos dados observados e possibilitando obter um valor de correlação associado (skill) a cada um dos três modelos (ARIMA, Holt-Winters e ACC).

A agregação dos resultados dos modelos é feita, ponto a ponto, por uma média ponderada, onde são atribuídos pesos proporcionais ao quadrado das correlações obtidas por cada modelo individual, sendo desconsideradas as previsões com correlação inferior a um limite de significância, a fim de realçar os resultados dos melhores modelos. Os valores finais previstos para a precipitação trimestral acumulada são classificados em categorias, de “muito seca” a “muito chuvosa”, segundo os quantis em que se enquadrem dentro da distribuição de freqüências climatológicas correspondentes. Utiliza-se um procedimento de aproximação, baseado na hipótese de

normalidade das distribuições climatológicas, gerando uma previsão probabilística dos desvios de precipitação acumulada para cada trimestre.

A formulação de modelos combinados foi inicialmente proposta por Bates e Granger (1969), através da combinação linear de previsões pontuais calculadas por meio de modelos com características distintas. Esta técnica foi aprimorada nos trabalhos de Bunn (1985) e Granger (1980), e foi amplamente revista por Clemen (1989). A metodologia proposta para a agregação dos resultados dos modelos individuais baseia-se no argumento científico de que a combinação de duas ou mais previsões pontuais, prezando pela máxima eficiência, poderão resultar em uma previsão final consensual com melhores resultados (Granger, 1980). Este argumento estende-se

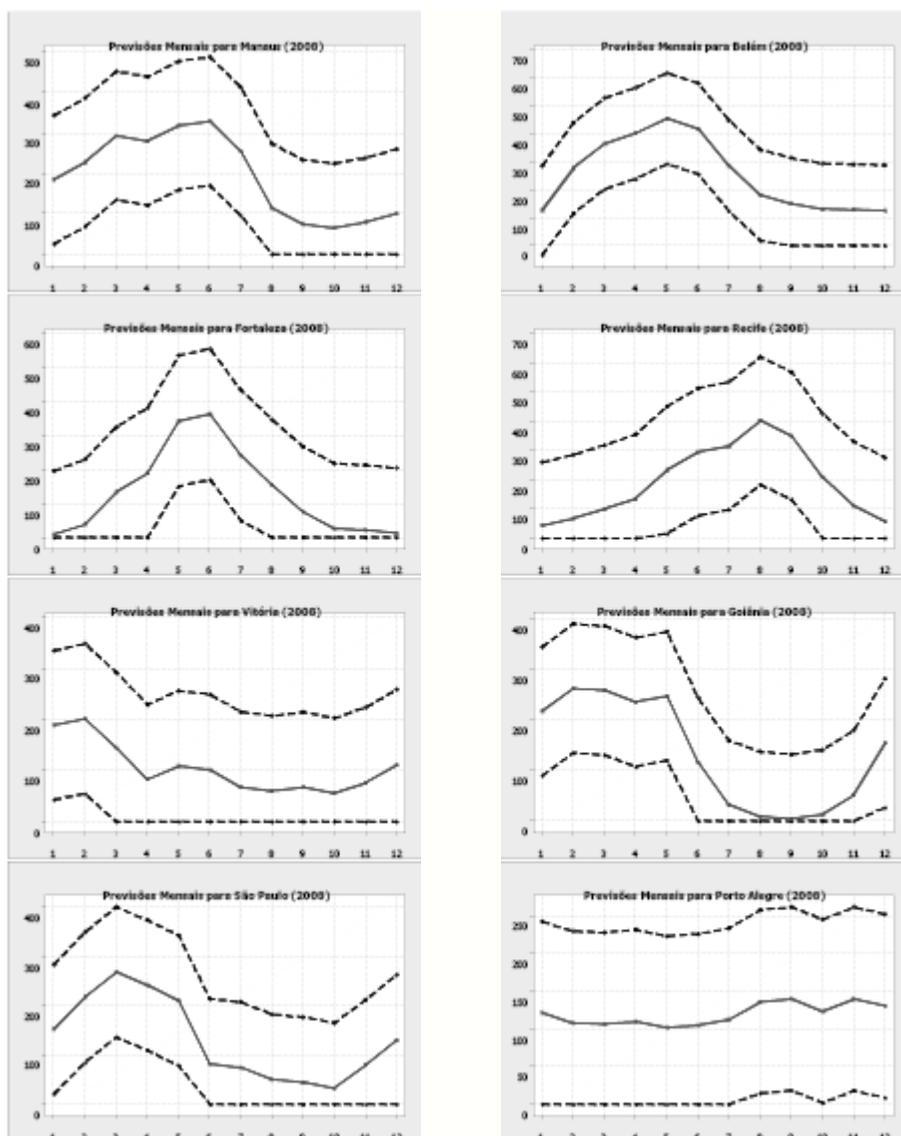


Figura 6 – Previsões ARIMA, com limites de confiança de 95%. A sazonalidade domina os perfis ou padrões de previsão.

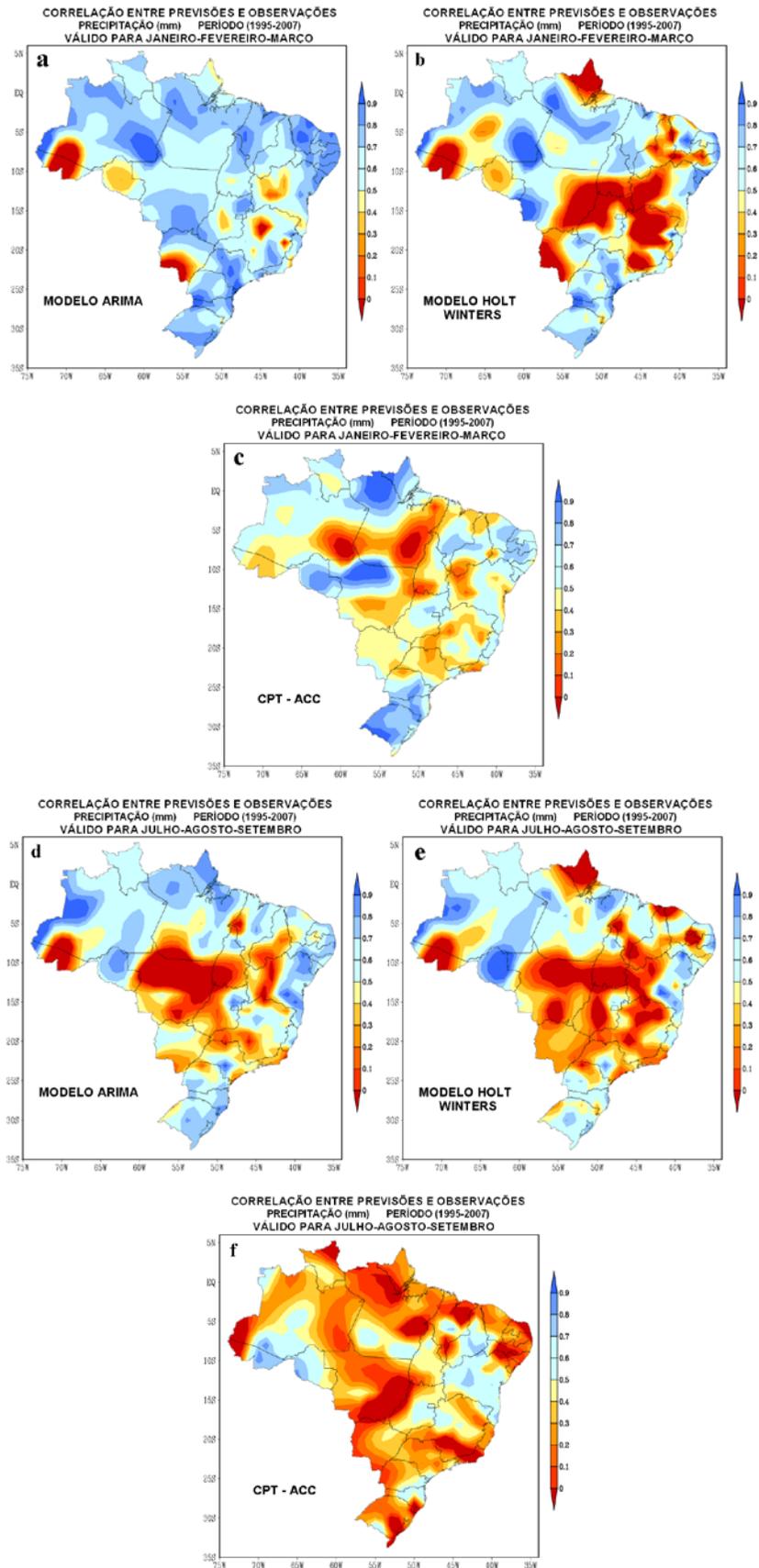


Figura 7 – Mapas de correlações dos modelos individuais para os trimestres JFM (a, b, c) e JAS (d, e, f).

facilmente para qualquer caso multivariado, ou seja, mais de dois modelos (Krishnamurti *et al.*, 2000a, Krishnamurti *et al.*, 2000b, Coelho *et al.*, 2004, Coelho *et al.*, 2005, Stephenson *et al.*, 2005).

Na agregação dos modelos, para um nível de significância de 95%, uma série de 13 elementos, como as atualmente utilizadas para os modelos ARIMA, Holt-Winters e CPT, no INMET, tem-se um valor para a correlação de corte, $r_{\min} \cong 0,48$.

4. RESULTADOS E DISCUSSÕES

Apresentam-se, a seguir, resultados obtidos pelo modelo combinado para os trimestres janeiro-fevereiro-março (JFM)

de 2008, representativo do período de verão, e julho-agosto-setembro (JAS) de 2008, representativo do período de inverno.

As Figuras 7a à 7f mostram os mapas de destreza dos modelos individuais para os trimestres JFM e JAS, respectivamente. No trimestre JFM, os modelos ARIMA e de ACC mostram boas destrezas nas regiões norte, nordeste e sul. O modelo de ACC ainda apresenta em algumas áreas das regiões centro-oeste e sul, apesar de não serem altas, correlações variando entre 0,4 e 0,6, e em outras correlações inferiores a 0,3. Do modelo de alisamento exponencial HOLT-WINTERS destaca-se a baixa destreza na região central do Brasil, com os maiores valores na região sul, partes da região norte e faixa do litoral norte da região nordeste. No trimestre JAS

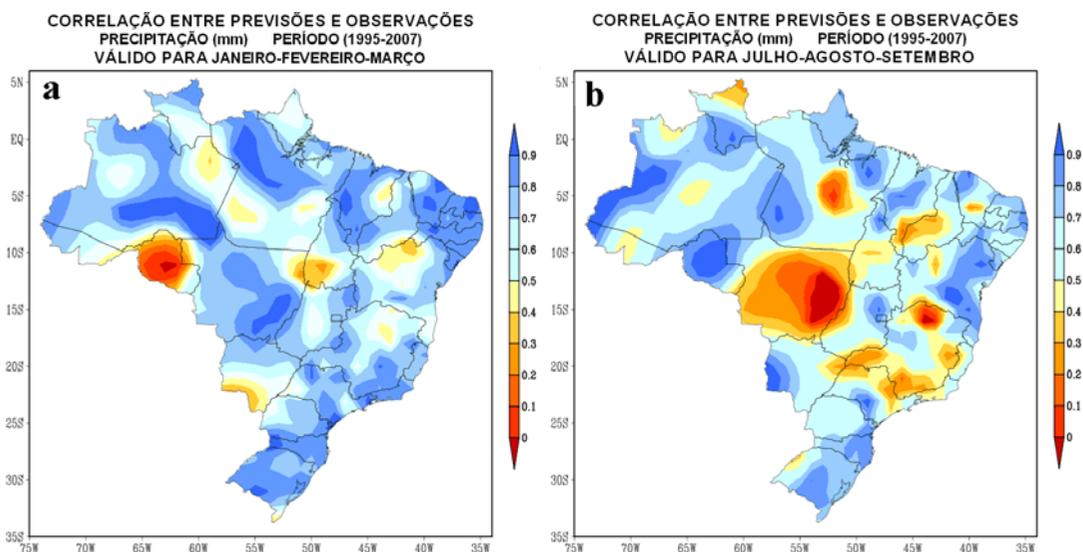


Figura 8 – Mapas de correlações do modelo agregado para precipitação dos trimestres JFM (a) e JAS (b).

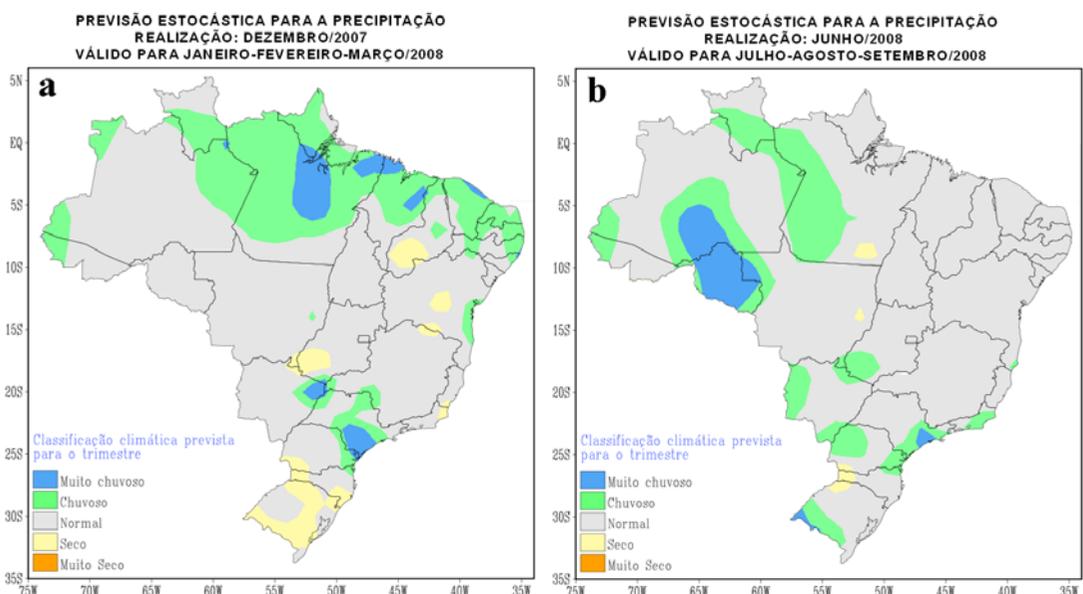


Figura 9 – Mapas das previsões classificadas em quantis de precipitação para os trimestres JFM (a) e JAS (b).

apresentam-se as menores destrezas que no período JFM, os modelos ARIMA e de HOLT-WINTERS apresentam melhores correlações na região norte, leste do nordeste e extremo sul, o modelo de ACC mostra boas correlações apenas em partes da região norte e nordeste, e em pequenas áreas do centro-oeste

e sudeste. As Figuras 8a à 8f mostram o mapa de destreza do modelo agregado das análises dos três modelos individuais, para os trimestres JFM e JAS, respectivamente. Nestes dois trimestres, a agregação conseguiu captar bem a variabilidade sazonal, ou seja, apresenta bom *skill* na maior parte das regiões

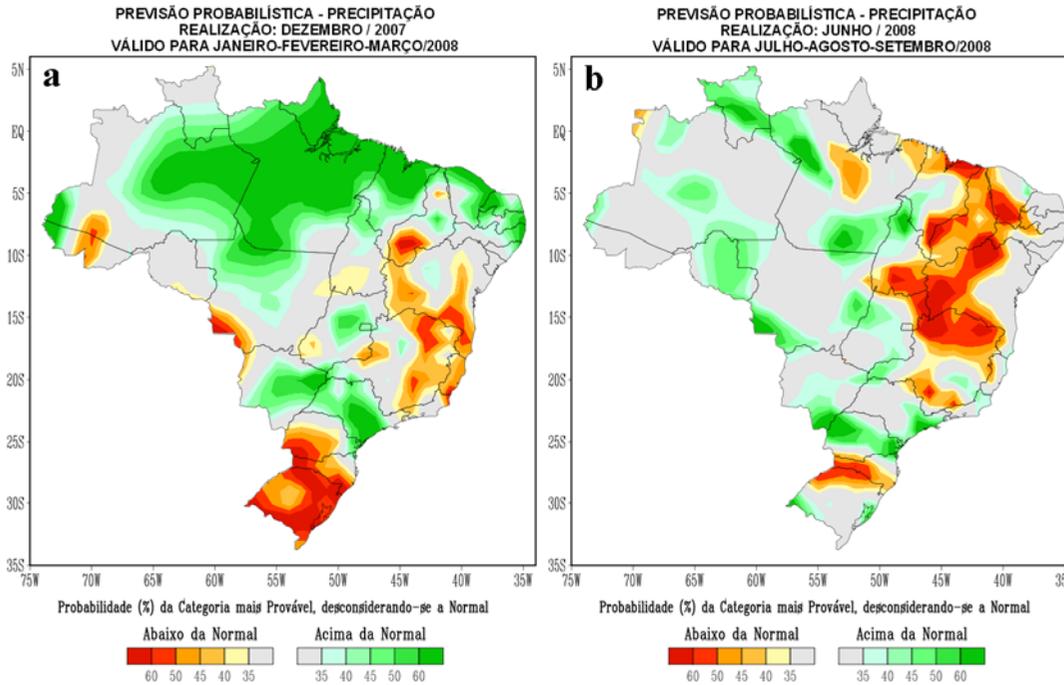


Figura 10 – Mapas de previsões probabilísticas da precipitação acumulada para os trimestres JFM (a) e JAS (b).

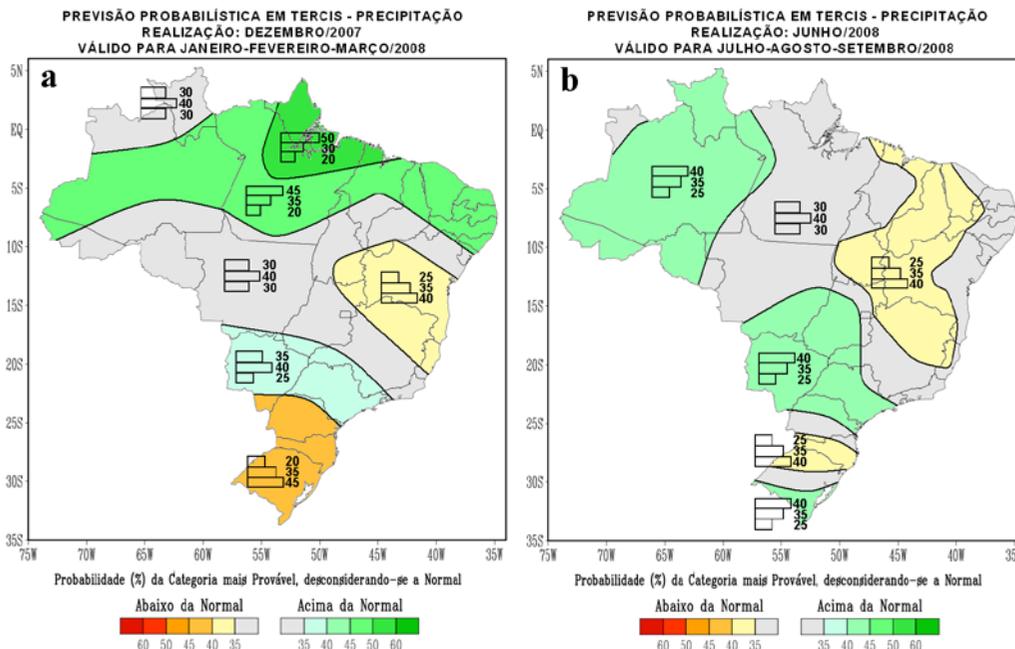


Figura 11 – Mapas de previsões probabilísticas da precipitação acumulada categorizadas em tercis para os trimestres JFM (a) e JAS (b).

brasileiras. Mapas de destreza, para outros trimestres (não mostrados), indicam que os resultados obtidos pela agregação final dos resultados, podem ser usados com confiança para gerar previsões climáticas sazonais.

Uma vez obtidos os valores previstos para a precipitação acumulada trimestral, estimativas pontuais para cada série climatológica utilizada, estes são confrontados com os respectivos valores quantílicos (Xavier *et al.*, 2002), onde chuvas entre os percentis 15 a 35 caracterizam o estado

“seco”, chuvas entre os percentis 35 e 65 caracterizam o estado “normal”, chuvas entre os percentis 65 a 85 caracterizam o estado “chuvoso” e chuvas acima do percentil 85, caracterizam o estado “muito chuvoso”. A Figura 9A mostra a previsão classificada por quantis para o trimestre JFM de 2008; A Figura 9B faz o mesmo para JAS de 2008.

Levando-se em conta a média e variância da distribuição climatológica, e as correlações entre valores observados e previstos pelo modelo, para cada série temporal de cada estação

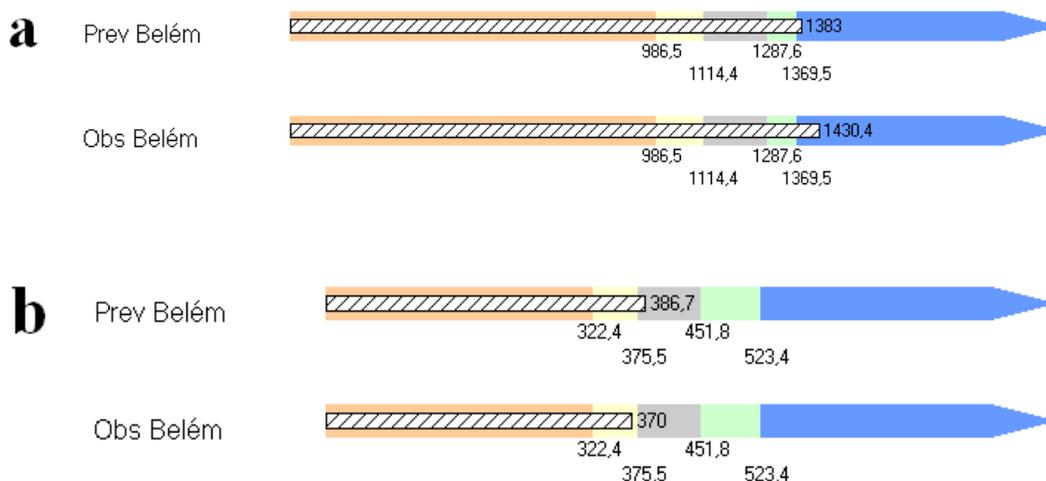


Figura 12 – Métricas mostrando os valores previstos e observados, para os trimestres JFM (a) e JAS (a), para o ano de 2008 em (mm/mês), projetados nos quantis Q1(15%), Q2(35%), Q3(65%) e Q4(85%), para Belém.

Tabela 2 – Coeficientes de correlação entre valores observados e previstos pelos modelos para o período de 1995 a 2007, para o trimestre Janeiro-Fevereiro-Março.

CIDADE/ESTADO/CÓDIGO	ARIMA	HOLT-WINTERS	Análise de Correlações Canônicas - CPT	MODELO AGREGADO
MANAUS – AM – 82331	0,74	0,64	0,54	0,51
BELÉM – PA – 82191	0,66	0,49	0,10	0,72
FORTALEZA – CE – 82397	0,87	0,84	0,63	0,94
RECIFE – PE – 82900	0,74	0,68	0,48	0,81
VITÓRIA – ES – 83648	0,73	0,68	0,62	0,74
GOIÂNIA – GO – 83423	0,41	-0,1	0,57	0,58
SÃO PAULO – SP – 83781	0,56	0,42	0,39	0,60
PORTO ALEGRE – RS – 83967	0,76	0,69	0,67	0,82

Tabela 3 – Coeficientes de correlação entre valores observados e previstos pelos modelos para o período de 1995 a 2007, para o trimestre Junho-Agosto-Setembro.

CIDADE/ESTADO/CÓDIGO	ARIMA	HOLT-WINTERS	Análise de Correlações Canônicas - CPT	MODELO AGREGADO
MANAUS – AM – 82331	0,70	0,68	0,17	0,69
BELÉM – PA – 82191	0,84	0,82	0,13	0,81
FORTALEZA – CE – 82397	0,35	0,42	0,35	0,59
RECIFE – PE – 82900	0,84	0,82	0,35	0,83
VITÓRIA – ES – 83648	0,81	0,74	0,18	0,71
GOIÂNIA – GO – 83423	0,23	-0,16	0,59	0,59
SÃO PAULO – SP – 83781	0,40	0,16	0,20	0,38
PORTO ALEGRE – RS – 83967	0,73	0,69	0,30	0,77

usada (SUGI, 2005; FORTES, 2007), podem-se calcular as probabilidades previstas, em cada ponto, para cada tercil da distribuição climatológica, e produzir mapas como os das Figuras 10A e 10B, que, por conseguinte, servem como base para a geração de uma previsão climática mais subjetiva por tercis (Figuras 11A e 11B).

As Figuras 12A e 12B mostram régua (métricas) que são confeccionadas, mostrando as quantidades de precipitação prevista e observada. São mostrados tais resultados para a cidade de Belém afim de melhor explicitar como as métricas são obtidas e auxiliam na avaliação do desempenho das previsões.

Nas Tabelas 2 e 3 apresentam-se, para os mesmos postos, a avaliação da performance das previsões, com base no coeficiente de correlação entre valores observados e previstos. Ressalta-se que estas correlações são calculadas para todas as séries temporais, para que se obtenha a melhor combinação linear para a previsão final. Embora alguns modelos para determinados casos apresentem uma melhor correlação em relação ao resultado do modelo agregado, usa-se a previsão final do modelo agregado, pelo fato da maior parte das estações utilizadas apresentarem as melhores correlações no caso do modelo agregado.

5. CONCLUSÕES

A proposta de se usar um modelo final que agregue o potencial de previsão de mais de um modelo, levando-se em

conta o *skill* relativo de cada um deles, vem sendo utilizada no âmbito do INMET para realizar previsões climáticas sazonais de precipitação (e mais recentemente também de temperatura média). A combinação de previsões mostrou-se mais robusta, para a maioria das 160 séries de pluviometrias acumuladas trimestrais utilizadas, do que as previsões de modelos isolados.

A análise das séries temporais pelos modelos estocásticos ARIMA e Holt-Winters, permite extrapolar o comportamento das séries trimestrais de precipitação segundo um processo com poucas oscilações. Tais valores, ao serem combinados com previsões provenientes da ACC, incorporam a influência dinâmica imposta pelos efeitos da TSM nas circulações atmosféricas de grande escala, que são as principais responsáveis pelos regimes de chuvas verificados no Brasil, em diferentes épocas do ano.

Em síntese, constata-se, que todas as abordagens metodológicas imputadas às séries de interesse (Modelo de Alisamento Exponencial e Modelo de Box-Jenkins) revelaram que as estimativas da precipitação são aceitáveis dentro do âmbito meteorológico. Com efeito, tanto no caso dos modelos individuais, quanto no do modelo combinado, os resultados das previsões são factíveis, permitindo extrapolar o comportamento das séries trimestrais de precipitação. As séries tendem a convergir para uma estabilidade em relação à componente sazonal. Saliente-se que essa estabilização deixaria de existir em casos extremos. Com efeito, a reincidência de eventos

associados a fortes anomalias, em qualquer ano posterior a 2008, propiciaria, sem dúvida, um desequilíbrio entre as componentes das séries. Observa-se que, no caso das séries que sofrem bruscas interferências, a componente sazonal contribui de forma importante no reconhecimento do seu padrão temporal, pois os valores das suas observações ficam muito sensíveis às variações em torno da média. Nesse caso, o efeito sazonal passa a funcionar como vetor contaminador redutor de longo prazo.

6. REFERÊNCIAS BIBLIOGRÁFICAS

- BARNSTON, A. G., SMITH, T. M. Specification and prediction of global surface temperature and precipitation from global SST using CCA. **Journal of Climate**, 9: 2660–2697, 1996.
- BATES, J. M., GRANGER, C. W. J. The combination of forecasts. **Operational Research Quarterly**, 20, 451–468, 1969.
- BOX, G. E. P., JENKINS, G. M. Time Series Analysis: Forecasting and Control. San Francisco: **Holden Day**, 1976.
- BOWERMAN, B. L. Times Series Forecasting: unified concepts and computer implementation: **Duxbury**, Boston, 1987.
- BUNN, D. W. Statistical Efficiency in the Linear Combination of Forecasting. **International Journal of Forecast**, v.1, p. 151–163, 1985.
- CAVALCANTI, I. F. A., MARENGO, J. A., SATYAMURTY, P., NOBRE, C. A., TROSNIKOV, I., BONATTI, J. P., MANZI, O., TARASOVA, T., PEZZI, L. P., D'ALMEIDA, C., SAMPAIO, G., CASTRO, C. C., SANCHES, M. B., CAMARGO, H. Global Climatological features in a simulation using the CPTEC-COLA AGCM. **Journal of Climate**, 15(21), 2965–2988, 2002.
- CHATFIELD, C. The future of time-series forecasting. **International Journal of Forecasting**, 4, 411–419, 1988.
- CHOW, G. C. “Tests of Equality Between Sets of Coefficients in Two Linear Regressions”. **Econometrica**, 28(3): 591–605, 1960.
- CLEMEN, R. T. Combining forecasts: A review and annotated bibliography. **International Journal of Forecasting**, 5, 559–583, 1989.
- COELHO, C. A. S., S. PEZZULLI, M. BALMASEDA, F. J. DOBLAS-REYES e D. B. STEPHENSON. Skill of coupled model seasonal forecasts: A Bayesian assessment of ECMWF ENSO forecasts. **ECMWF Technical Memorandum**, No. 426, 16pp, 2003.
- COELHO, C. A. S., S. PEZZULLI, M. BALMASEDA, F. J. DOBLAS-REYES e D. B. STEPHENSON. Forecast calibration and combination: A simple Bayesian approach for ENSO. **Journal of Climate**, Vol. 17, No. 7, 1504–1516, 2004.
- COELHO, C. A. S., D. B. STEPHENSON, F. J. DOBLAS-REYES e M. BALMASEDA. From multi-model ensemble predictions to well-calibrated probability forecasts: Seasonal rainfall forecasts over South America 1959–2001. **CLIVAR Exchanges No.32**. Vol. 10, No. 1, 14–20, 2005.
- COELHO, C. A. S., D. B. STEPHENSON, M. BALMASEDA, F. J. DOBLAS-REYES., G. J. van OLDENBORGH. Towards an Integrated Seasonal Forecasting System for South America. **Journal of Climate**, 19(15), 3704–3721, 2006.
- CORDEIRO, D. M. Séries temporais, análise quantitativa: teoria e aplicações. Local: **EDUPE**, p. 91–124 (Série Ciência e Tecnologia), 2002.
- EVANS, J. S.; DOSWELL, C. A. Examination of Derecho Environments using proximity soundings. **Weather and Forecasting**, 16: p.329–342, 2001.
- FERREIRA, D. B., MELLO, L. T. A., LUCIO, P. S., FORTES, L. T. G., BALBINO, H. T., SALVADOR, M. A., SANTOS, L. A. R., SILVA, F. Análise da influência da variabilidade da temperatura da superfície do mar sobre a produtividade de soja no estado do Paraná utilizando o “Climate Predictability Tool”. XIV CBMET - **Congresso Brasileiro de Meteorologia** - Florianópolis - SC, 2006.
- FORTES, L. T. Relação entre a Destreza da Previsão Climática da Precipitação Sazonal e a Probabilidade do Tercil Médio. **Nota Técnica**. CDP/INMET. Brasília, fevereiro, 2007.
- GRANGER, C. W. J. Forecasting in business and economics. In: New York, **Academic Press**, p. 131–147, 1980.
- HOLT, C. C. Forecasting seasonals and trends by exponentially weighted moving averages, **ONR Research Memorandum**, Carnegie Institute 52, 1957.
- JOHNSON, R. A., e WICHERN, D. W. Applied multivariate statistical analysis. **Prentice Hall**, 792 pp, 2007.
- KRISHNAMURTI, T. N., KISHTAWAL, C. M., SHIN, D. W. e WILLIFORD, C. E.. Improving tropical precipitation forecasts from a multi-analysis superensemble. **Journal of Climate**, 13, 4217–4227, 2000a.
- KRISHNAMURTI, T. N., C. M. KISHTAWAL, C. M., LAROW, T., BACHIOCHI, D., ZHANG, Z., WILLIFORD, C. E., GADGIL, S. e SURENDRAN, S. Multi-model superensemble forecasts for weather and seasonal climate. **Journal of Climate**, 13, 4196–4216, 2000b.
- LJUNG, G. M, e BOX, G. E. P. “On a Measure of Lack of Fit in Time Series Models,” **Biometrika** 65, pp.67–72, 1978.
- OLIVEIRA, A. C. de. Relações entre a TSM nos oceanos Atlântico e Pacífico e as condições climáticas nas regiões Sul e Sudeste do Brasil. **Tese de Doutorado em Meteorologia** - IAG-USP, 2005.
- REPELLI, C. A.; NOBRE, P. Statistical prediction of sea-surface temperature over the tropical Atlantic. **International Journal of Climatology**, v. 24, n. 1, p. 45–55, 2004.

- REYNOLDS, R.W., N.A. RAYNER, T.M. SMITH, D.C. STOKES, AND W. WANG. An Improved In Situ and Satellite SST Analysis for Climate. **Journal of Climate**, 15, 1609-1625, 2002.
- SALVADOR, M. A., BALBINO, H. T., LUCIO, P. S., MELLO, L. T. A., FERREIRA, D. B., SILVA, F., FORTES, L. T. G., REBELLO, E., SALES, N. D., PAULA, T. P., RAPOSO, K. Avaliação do aplicativo *Climate Predictability Tool* na caracterização e previsão do regime de precipitação da costa leste do Nordeste Brasileiro. XIV CBMET - **Congresso Brasileiro de Meteorologia**, Florianópolis - SC, 2006.
- STEPHENSON, D. B., C. A. S. COELHO, F. J. DOBLAS-REYES e M. BALMASEDA. Forecast Assimilation: A unified framework for the combination of multi-model weather and climate predictions. **Tellus A**, 57, 253-264, 2005.
- SHUKLA, J. Dynamical predictability of monthly means. **Journal of Atmospheric Sciences**, 38. 2547-2572, 1981.
- SUGI, M. Toward better probabilistic seasonal prediction for decision making. **Proceedings of the WMO Technical Conference on Climate as a Resource**. Beijing, China, November, 2005.
- WILKS, D. D. Statistical Methods in the Atmospheric Sciences. 2ª Edição, 627 p, 2006.
- WINTERS, P. R. Forecasting sales by exponentially weighted moving averages, **Management Science**. 6: 324-342, 1960.
- XAVIER, T. de M. B. S.; SILVA, J. de F. ; REBELLO, E. R. G. "A Técnica dos Quantis e suas Aplicações em Meteorologia, Climatologia e Hidrologia, com ênfase para as Regiões Brasileiras". **Thesaurus Editora de Brasília Ltda**, Brasília-DF, pp144, 2002.