

Modelagem Estocástica da Geração Eólica para Estudos Energéticos

Elaborado por:

ENERDADOS

Getulio B. da Silveira Filho

Para:

Agência de Cooperação Técnica Alemã - GTZ

Deutsche Gesellschaft für Technische Zusammenarbeit GmbH

Empresa de Pesquisa Energética - EPE

Outubro 2010

gtz



Programa Energia

Brasil-Alemanha



Modelagem Estocástica da Geração Eólica para Estudos Energéticos

Elaborado por: Enerdados

Autor: Getulio B. da Silveira Filho

Para: Deutsche Gesellschaft für Technische Zusammenarbeit GmbH
Empresa de Pesquisa Energética – EPE

Programa: Programa Energia, GTZ Brasil

No do Programa: 2007.2189.4-001.00

Coordenação: Torsten Schwab (GTZ),
Juarez Lopes (EPE)

Outubro 2010

Informações Legais

1. Todas as indicações, dados e resultados deste estudo foram compilados e cuidadosamente revisados pelo(s) autor(es). No entanto, erros com relação ao conteúdo não podem ser evitados. Conseqüentemente, nem a GTZ ou o(s) autor(es) podem ser responsabilizados por qualquer reivindicação, perda ou prejuízo direto ou indireto resultante do uso ou confiança depositada sobre as informações contidas neste estudo, ou direta ou indiretamente resultante dos erros, imprecisões ou omissões de informações neste estudo.
2. A duplicação ou reprodução de todo ou partes do estudo (incluindo a transferência de dados para sistemas de armazenamento de mídia) e distribuição para fins não comerciais é permitida, desde que a GTZ e a EPE sejam citadas como fonte da informação. Para outros usos comerciais, incluindo duplicação, reprodução ou distribuição de todo ou partes deste estudo, é necessário o consentimento escrito da GTZ e da EPE.
3. Em atendimento ao Termo de Confidencialidade firmado entre as partes e a cessionária das informações relativas aos parques eólicos em operação, os nomes desses parques foram substituídos por Usina X, Usina Y e Usina Z.

Conteúdo

1. Apresentação	1
2. Introdução	2
3. Análise Exploratória	5
4. Modelagem.....	12
5. Algumas Conclusões Preliminares	16
6. Estratégias de Modelagem 1: Modelos Estruturais	17
7. Estratégias de Modelagem 2: Modelos Periódicos	34
8. Considerações Finais	48

1. Apresentação

Este trabalho é o relatório referente ao projeto **Modelagem Estocástica da Geração Eólica para Estudos Energéticos**. Aqui iremos tratar de análise exploratória de algumas séries temporais referentes ao aproveitamento da energia eólica no Brasil.

Faz ainda parte deste produto um conjunto de rotinas, escritas na linguagem S, que produziu a maior parte dos gráficos e estatísticas a seguir apresentados.

Incluimos também neste relatório parte de nossos esforços iniciais de modelagem estatística das séries de fatores de capacidade compiladas pela consultoria Camargo & Schubert.

O restante deste relatório se encontra assim dividido: A seção 2 discorre rapidamente sobre a motivação do projeto; na seção 3 conduzimos a análise exploratória de algumas séries importantes para o aproveitamento da energia eólica; nas seções 4 a 7 discorreremos sobre um pequeno exercício de modelagem com as séries de fatores de capacidades e a seção 8 conclui.

2. Introdução

Nos últimos 12 anos os ventos têm sido a fonte primária de energia elétrica de maior ritmo de expansão no mundo, apresentando incremento exponencial da potência instalada. Entre 1990 e 2008, a potência instalada em aerogeradores cresceu à taxa média de 27% ao ano, alcançando 121.000 MW, dos quais mais de 54% instalados na Europa. Enquanto apresenta forte expansão na Ásia, particularmente na Índia e China, e notável crescimento nos Estados Unidos onde, só no ano de 2008 foram instalados 8.400 MW, no Brasil a expansão eólica tem sido modesta e apoiada em iniciativas governamentais como o Programa de Incentivo às Fontes Alternativas de Energia Elétrica - PROINFA (2002) e o Leilão de Energia de Reserva -LER, realizado em dezembro de 2009.

É necessário reconhecer que o impulso maior da geração eólica ocorre em países cuja geração de energia elétrica é predominantemente de base térmica, onde a atratividade econômica da fonte eólica está associada a aspectos elétricos e energéticos de curto prazo, e ambientais, vale dizer, à substituição de combustíveis fósseis e à redução das emissões de CO₂.

No Brasil, diferentemente, são os aspectos energéticos de longo prazo que definem a atratividade econômica das diversas fontes de geração da energia elétrica, pois são eles que determinam a confiabilidade do suprimento de energia elétrica em períodos de hidraulicidade desfavorável.

Pela composição do parque gerador nacional, onde a principal fonte energética - hidrelétrica - é dependente de fatores climáticos não determinísticos, a valoração econômico-energética das fontes de energia elétrica é função não apenas de seus custos intrínsecos, mas também da maior ou menor coincidência entre disponibilidades hídrica e de fontes alternativas.

No caso das fontes primárias variáveis, como a solar e suas derivadas, a hídrica, a eólica e a própria biomassa, essa associação pode ser estimada com base na modelagem estatística de séries históricas de ocorrências mensais suficientemente longas. Séries essas, obtidas a partir de medições diretas ou através de métodos de inferência estatística baseada em informações disponíveis.

Particularmente importante é a verificação dessas coincidências no período compreendido entre 1949 e 1956, definido como “período crítico” do sistema brasileiro, que serve para definir de forma determinística, o valor da “energia firme” de cada aproveitamento hidrelétrico.

A geração eólica tem desenvolvimento recente, a partir dos anos oitenta, e apenas na década de 90 viabilizou-se técnica e economicamente como alternativa segura para o suprimento da demanda de energia elétrica. Assim, as medições e registros de velocidade e direção dos ventos (a partir das quais se pode inferir a capacidade de

geração eólica local) anteriores à década de 80 em geral eram destinadas ao suporte à aviação ou a estudos meteorológicos.

No Brasil, a geração eólica é ainda mais recente e os poucos registros históricos de geração efetiva existentes abrangem período relativamente curto. Da mesma forma, os registros de velocidade e direção dos ventos são poucos e precários, de modo que a obtenção de históricos adequados para estudos energéticos de longo prazo, cobrindo o período 1949-1956, é hipótese remota.

A confessa limitação do conhecimento das características energéticas de longo prazo, evidenciada desde a época do PROINFA e, mais recentemente, nas discussões preparatórias para a formulação das regras do Leilão de Energia de Reserva - LER 2009, tem se revelado um obstáculo à valoração econômica adequada da geração eólica no Brasil.

De fato, tem-se atribuído aos parques eólicos, como benefício energético, apenas a sua geração individual média ao longo da vida útil, não sendo considerados os ganhos sinérgicos decorrentes da operação conjunta do parque eólico com o sistema predominantemente hidrelétrico. Em outras palavras, tem-se adotado no Brasil o mesmo critério de valoração econômico-energética utilizado nos países de base térmica.

Para a adequação dos procedimentos de valoração da geração eólica às características únicas do parque gerador nacional, a EPE tem realizado ações estruturantes de longo prazo e remediais de curto prazo.

Em meados de 2009, a Camargo & Schubert, conceituada empresa nacional de consultoria sobre energia eólica, com base em medições anemométricas próprias ou adquiridas em bancos de dados internacionais, inferências e correlações estatísticas, tomou a iniciativa de compor um histórico mensal de geração em quatro regiões do país, compreendendo o período de janeiro de 1976 a dezembro de 2008. Ressalta-se que o histórico reconstituído retrata situações regionais médias e não locais específicos.

A partir desse histórico a Diretoria de Estudos Econômico-Energéticos e Ambientais da EPE realizou simulações da operação do sistema energético com o modelo SUISHI-O (desenvolvido pelo Centro de Pesquisas Elétricas - CEPEL), no qual as fontes variáveis têm tratamento determinístico, e estudos de correlações temporais com vazões afluentes de diversos rios do país.

Os resultados das simulações e estudos mostram benefícios sinérgicos expressivos advindos da operação conjunta hidro-eólica, proporcionados por correlações negativas entre vazões afluentes e ocorrências de ventos nas quatro regiões analisadas.

Os ganhos de energia firme observados justificam o aprofundamento dos estudos com base em algumas séries históricas de efetiva geração eólica, ainda que relativamente curtas, cedidas, mediante termo de confidencialidade, por conceituada empresa nacional proprietária de parques eólicos em operação no estado do Ceará.

Este projeto tem como foco a modelagem estatística da geração eólica. O objetivo de maior monta é a simulação de trajetórias futuras da atividade eólica em sítios de interesse. Tais trajetórias devem ser compatíveis com o comportamento passado - representados pelas séries temporais acima citadas- e resultado da aplicação de metodologias estatísticas apropriadas. No jargão corrente no setor elétrico brasileiro, proveniente de metodologias associadas ao NEWAVE, essas trajetórias são denominadas de séries sintéticas.

Este projeto é subdividido em vários módulos. Este primeiro módulo trata da análise exploratória de séries de energia elétrica geradas a partir de 3 usinas eólicas em operação comercial localizadas no estado do Ceará. Lida também com questões de compatibilidade entre séries associadas às 3 usinas acima citadas e as séries fornecidas pela consultoria Camargo & Schubert.

À empresa cessionária das séries históricas e à Camargo & Schubert, a GTZ e a EPE agradecem a oportunidade oferecida para a evolução dos conhecimentos sobre a geração eólica no Brasil.

3. Análise Exploratória

Nesta seção conduzimos uma pequena análise exploratória dos dados em estudo. Disponemos de observações das gerações mensais relativas a 3 usinas localizadas no nordeste brasileiro. A série mais longa, referente à usina X, cobre o período Jan-1999 a Ago-2009. As duas restantes, relativas às usinas Y e Z, se estendem por períodos um pouco menores (Abr-1999 a Ago-2009 e Dez-2001 a Ago-2009, respectivamente). O gráfico a seguir exhibe a evolução conjunta dos fatores de capacidade das 3 usinas.

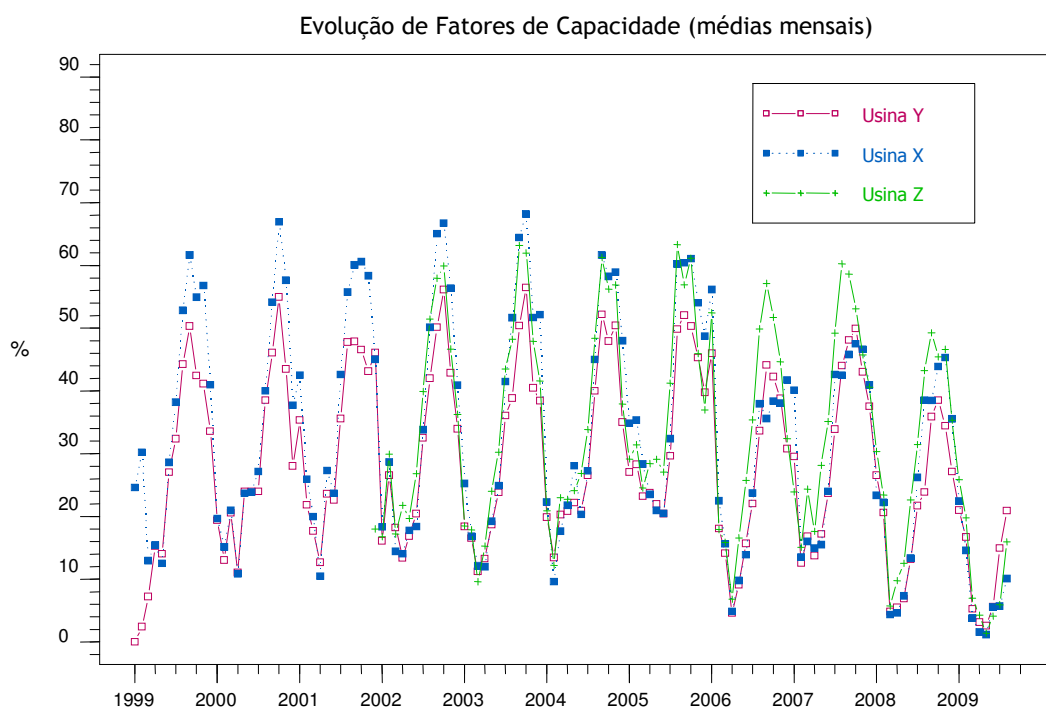


Gráfico 1: Fatores de Capacidade.

Fica clara a presença de co-movimentos importantes, basicamente relacionados ao perfil sazonal da atividade eólica. A usina Y, tipicamente, apresenta os menores fatores de capacidade. Isto é particularmente verdadeiro em meses de maior atividade eólica. As usinas X e Z possuem performances semelhantes com aparente dominância da usina Z nos anos mais recentes. Isto pode ser melhor apreciado na tabela 1 a seguir.

FATOR DE CAPACIDADE MÉDIO - ANUAL			
Ano	USINAS		
	Y	X	Z
2002	31,0%	35,7%	35,9%
2003	30,5%	37,1%	35,5%
2004	31,4%	35,8%	36,2%
2005	34,9%	40,7%	40,2%
2006	27,1%	29,0%	34,4%
2007	31,5%	33,3%	38,5%
2008	22,2%	25,8%	30,1%
Média Geral	29,8%	33,9%	35,8%

Tabela 1: Fatores de capacidade por usina; médias anuais.

Exibimos a seguir gráfico do tipo *monthplot* para a evolução da usina X (as outras plantas têm perfil *semelhante*)

Perfil sazonal dos fatores de capacidade da usina X

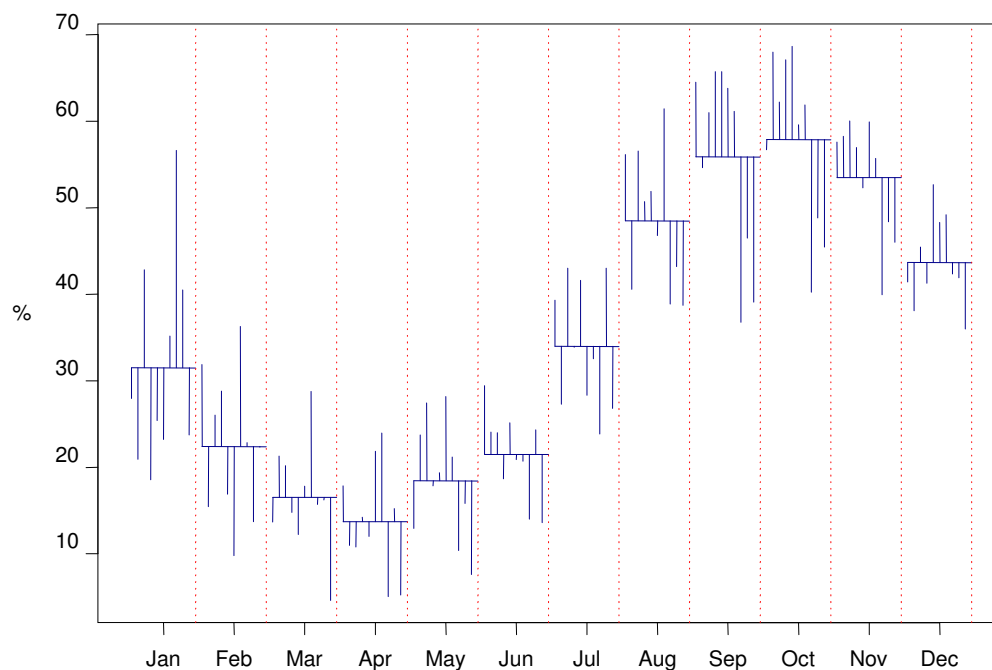


Gráfico 2: Exibição da sazonalidade via *monthplot*; usina X.

Os segmentos de reta horizontais têm alturas iguais às médias dos meses correspondentes. Concluímos, portanto, que as maiores (menores, respectivamente) atividades eólicas ocorrem tipicamente no mês de outubro (abril, respectivamente). Para cada mês, os pequenos segmentos verticais, representam a evolução, ao longo dos anos, dos fatores de capacidade referentes ao mês em questão. Aparentemente, nos últimos anos, os picos de atividade eólica (meses de setembro, outubro e novembro) vêm exibindo tendência de queda. Por economia de espaço, não exibimos

os correspondentes gráficos para as usinas Y e Z. Registramos que estas duas usinas exibem padrões similares com uma intensidade um pouco menor.

O ano de 2009 exhibe baixos fatores de capacidade. Mais ainda, para a maioria dos meses de 2009, os fatores de capacidade são os menores já observados¹. Ilustramos este comportamento na tabela 2 e no gráfico 3 a seguir. Ali vemos que os Fatores de Capacidade para os meses de Abril a Agosto de 2009 foram, uniformemente para as 3 usinas, os menores já registrados².

FATORES DE CAPACIDADE MÉDIOS: MÍNIMOS OBSERVADOS			
Ano	USINAS		
	Y	X	Z
Jan	2002	2002	2002
Fev	2007	2004	2004
Mar	2008	2009	2008
Abr	2009	2009	2009
Mai	2009	2009	2009
Jun	2009	2009	2009
Jul	2009	2009	2009
Ago	2009	2009	2009
Set	2008	2006	2008
Out	2008	2006	2008
Nov	2008	2006	2006
Dez	2008	2008	2006

Tabela 2: Anos onde o Fator de capacidade é mínimo.

¹ As observações mais recentes que dispomos datam de Ago-2009. É esperado que as performances dos fatores de capacidade para o restante do ano de 2009 tenham também sido baixas.

² Para a confecção da Tabela 5 desconsideramos o primeiro ano de operação de cada uma das 3 usinas.

Valores Mínimos Observados dos Fatores de Capacidade

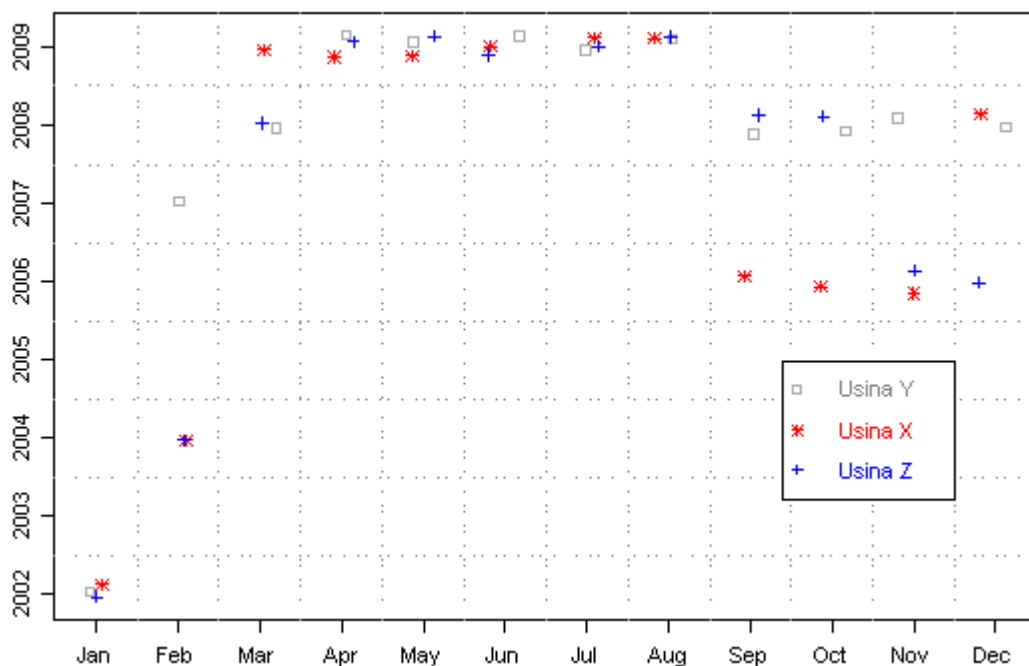


Gráfico 3: A ocorrência dos valores mínimos mensais no histórico das séries das usinas.

O fato acima pode indicar uma dificuldade relativamente à tarefa de modelagem.

Evolução de Fatores de Capacidade (médias mensais das 3 usinas)

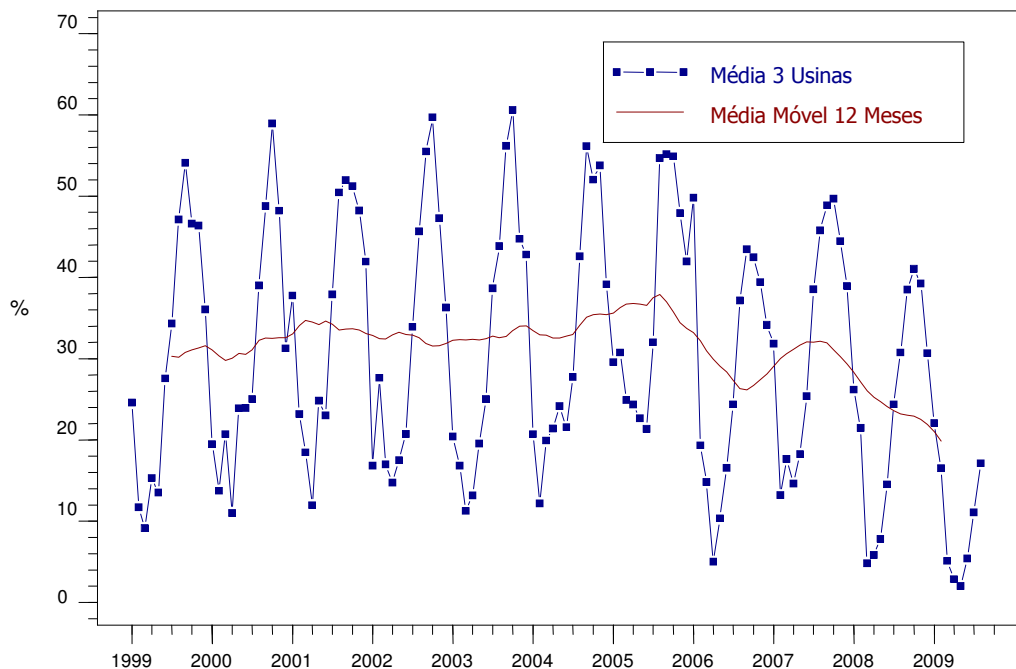


Gráfico 4: Fatores de capacidade das usinas X, Y e Z combinados.

Fonte: elaboração própria.

Como auxílio à nossa argumentação apresentamos no gráfico 4 a evolução do Fator de Capacidade Médio das 3 usinas. A aparente tendência decrescente para dados mais recentes é reflexo dos baixos níveis de ventos observados em 2009 e em 2008.

É claro que não existe justificativa para um decréscimo permanente da atividade eólica na região. Portanto, é possível que a mencionada queda faça parte de comportamento cíclico subjacente à atividade eólica da região. Nosso problema é que não dispomos de dados em quantidade suficiente para uma melhor investigação desta possibilidade.

No entanto, a Consultoria Camargo & Schubert disponibilizou para a EPE estimativas de fatores de capacidade relativos às principais (relativamente ao potencial eólico) unidades da federação. Esta base de dados possui informações desde janeiro de 1976, em base mensal, para os estados de Ceará, Bahia, Rio Grande do Norte e Rio Grande do Sul. Por facilidade de referência iremos doravante denominá-las de Fatores de Capacidade Estaduais.

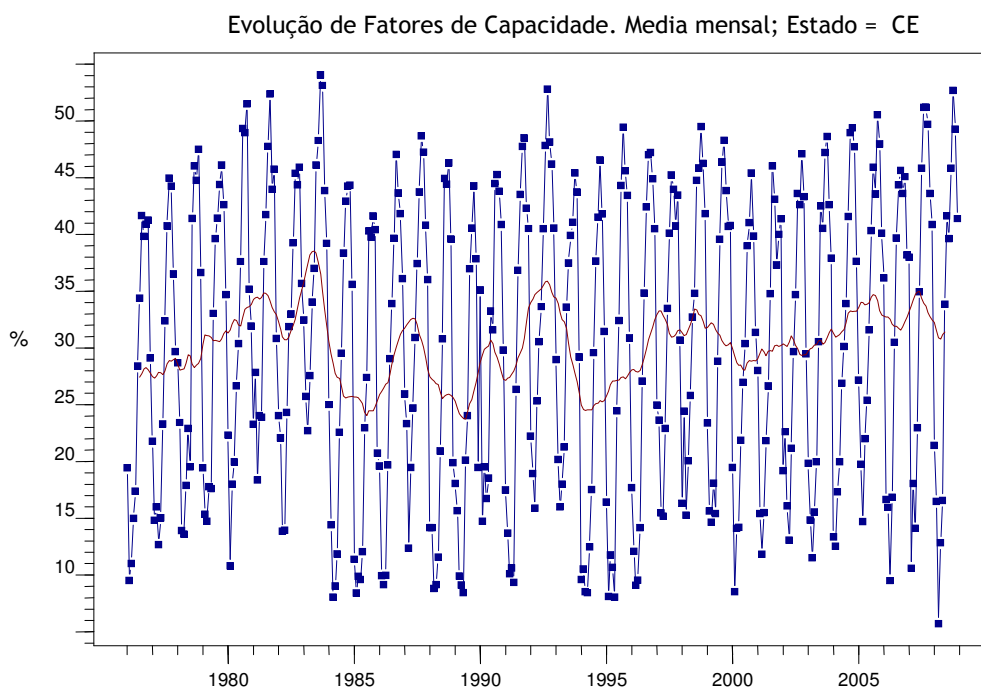


Gráfico 5: Fatores de Capacidade do Ceará. *Fonte:* Camargo & Schubert.

Nesta seção estaremos particularmente interessados no comportamento dos Fatores de Capacidade do Ceará, onde se localizam as 3 usinas de interesse. O gráfico 5 exhibe a evolução dos fatores de capacidade estaduais do Ceará.

Incluimos no gráfico 5, a exemplo do que fizemos com a média das 3 usinas (gráfico 5), a média móvel de 12 meses. Aparentemente, mesmo após a suavização induzida pela aplicação da média móvel, os fatores de capacidade exibem flutuações de

magnitude considerável. Por outro lado, a presença de comportamento cíclico com período maior que o anual deve ser investigada com técnica mais apurada.

De qualquer forma, comportamentos como o apontado no gráfico 5 ocorrem com bastante frequência na série estadual. A fim de avaliarmos a compatibilidade entre os dados das usinas e os estaduais exibimos nos gráficos 6a e 6b os BoxPlots dos fatores de capacidade segmentados pelo mês de ocorrência.

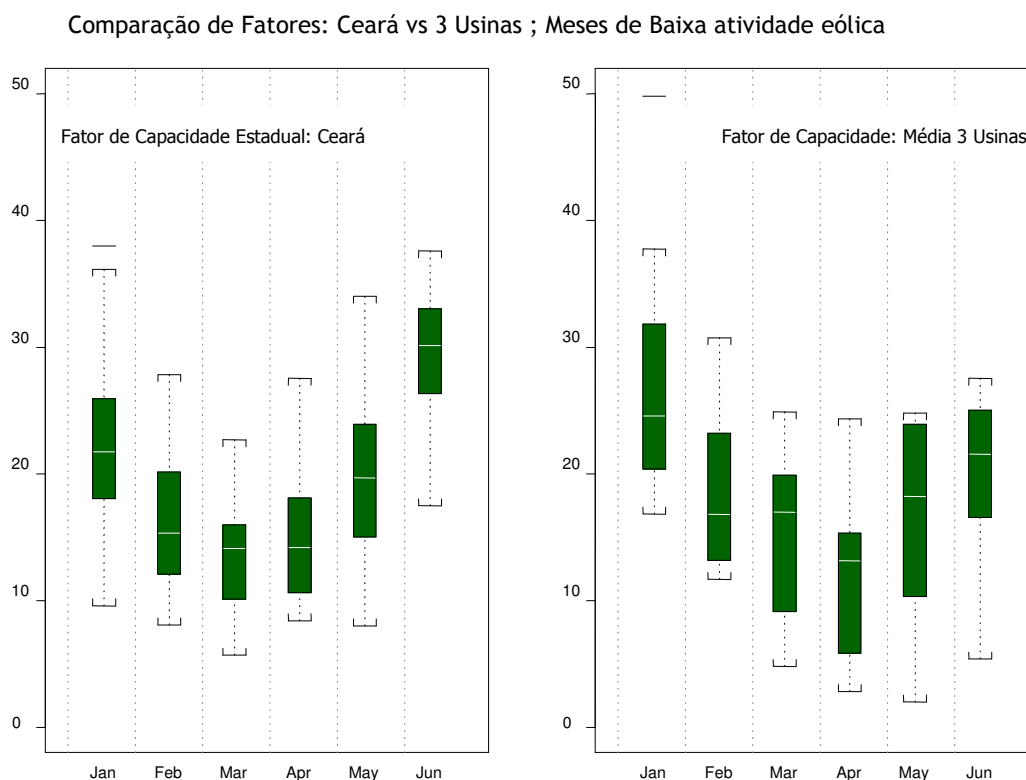


Gráfico 6a: Comparação entre as distribuições dos Fatores de Capacidade: 1º semestre.

Fica evidente que, no primeiro semestre, período de ventos de baixa intensidade, a série de Fatores estaduais é melhor comportada que a relativa às 3 usinas. Para cada um dos meses, a dispersão, medida pela altura dos caixotes, é menor nos fatores estaduais. É possível que isto se deva a questões de metodologia (maior abrangência geográfica, quantidade de postos de observação, etc.). Também, a configuração assumida pelos fatores de capacidade medianos (os segmentos horizontais brancos no interior dos caixotes) estaduais é mais suave e próxima do que esperaríamos do que o seu correspondente para as 3 usinas.

Relativamente ao segundo semestre (alta atividade eólica) valem comentários similares. Chamamos a atenção do leitor para o fato de que as maiores dispersões da série de fatores estaduais são as relativas aos meses de maio e dezembro. Uma justificativa para isto pode residir no fato daqueles serem meses de fronteira entre os períodos de alta e baixa atividade. Já com as séries das 3 usinas é digno de nota a

grande dispersão observada no mês de outubro, um dos dois meses de maior atividade.

Comparação de Fatores: Ceará vs 3 Usinas ; Meses de Alta atividade eólica

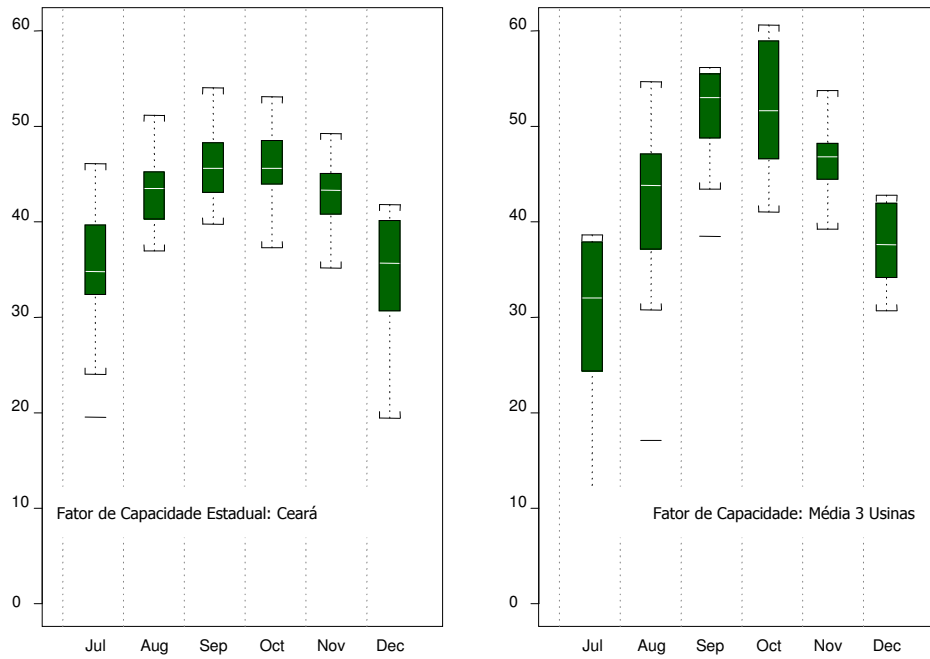


Gráfico 6b: Comparação entre as distribuições dos Fatores de Capacidade: 2º semestre.

Para uma comparação em linhas mais tradicionais ver o gráfico 7 a seguir.

Evolução de Fatores de Capacidade (médias mensais das 3 usinas)

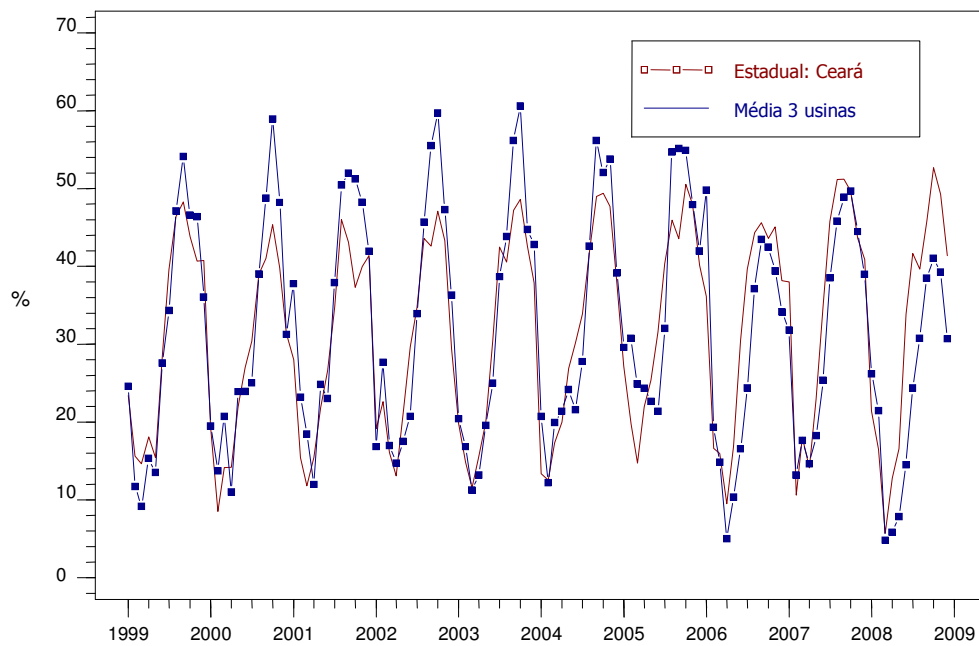


Gráfico 7: Fatores estaduais e média das 3 Usinas. O período exibido é o que dispomos de dados para ambas as séries.

4. Modelagem

Nesta seção apresentamos resultados de algumas modelagens preliminares que conduzimos com as séries em estudo. Iremos modelar a série fornecida pela Camargo & Schubert de Fatores de Capacidade para o estado do Ceará como um todo. Consideraremos também as análises e modelos para a série de Fatores Médios de Capacidade das usinas X, Y e Z (a média aritmética devidamente ponderada das 3).

Está fora do escopo deste projeto a discussão da metodologia utilizada pela Camargo & Schubert. Aqui iremos nos satisfazer em imaginar que a série por eles compilada corresponde aos fatores de capacidade de uma usina *representativa*, situada no Ceará.

Exibimos no gráfico 8 as funções de autocorrelação das séries dos Fatores do Ceará e dos Fatores médios das 3 usinas. Registramos que as correspondentes funções para cada uma das usinas individuais apresentam formas similares. Fica mais uma vez evidenciada a forte sazonalidade nas séries em questão.

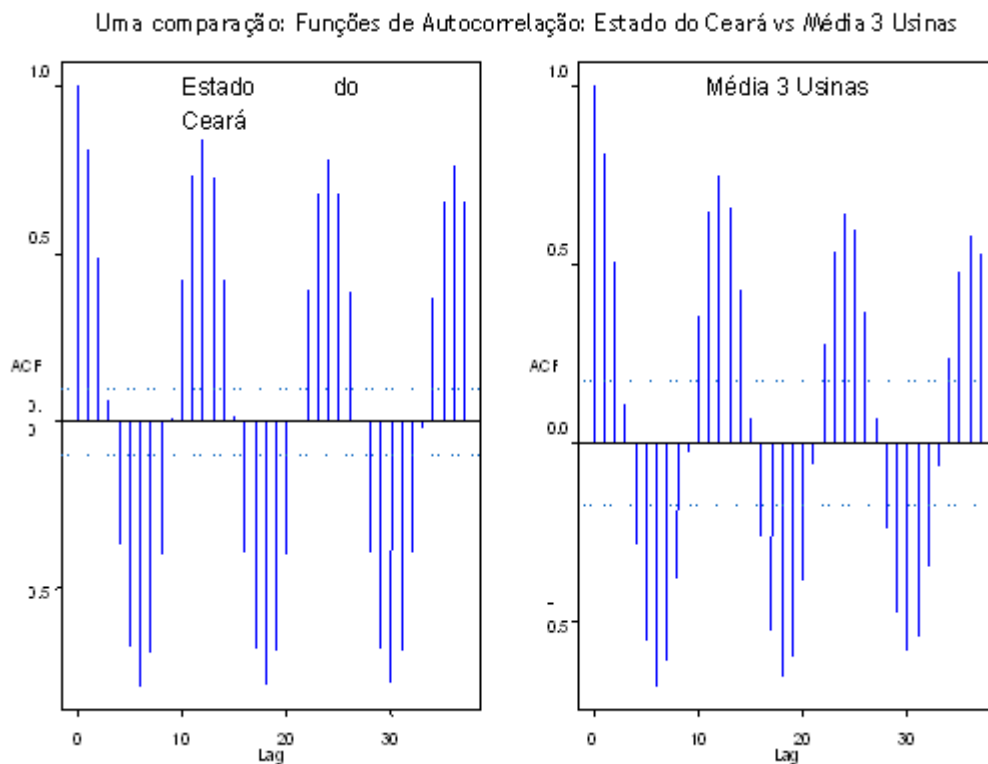


Gráfico 8: Padrões similares para a dinâmica dos fatores de capacidade do Ceará e das 3 usinas.

Uma questão importante na modelagem de séries temporais diz respeito a eventuais transformações (não-lineares) a serem aplicadas nas observações originais. Existem justificativas diversas para o uso de tais transformações. Dentre as justificativas mais importantes citamos (a) de cunho teórico [bem entendido, existe teoria subjacente ao processo sendo estudado que sugere uma determinada transformação nos dados;

e.g. logs em alguns processos econômicos], (b) ganhos na interpretabilidade de parâmetros, (c) adequabilidade a modelos estatísticos com estrutura mais simples [e.g. modelos homoscedásticos em vez de heteroscedáticos] e (d) ganhos computacionais [menos relevante nos dias de hoje].

No contexto de séries temporais uma característica importante é a estacionaridade. Isto porque, sem a estacionaridade a quantidade de modelos a serem considerados cresce muito, dificultando a tarefa de modelagem estatística. É, portanto, uma tarefa importante, a pesquisa acerca de transformações que induzam a estacionaridade.

Uma conceituação detalhada da estacionaridade foge aos objetivos deste trabalho. Iremos adotar como uma conceituação operacional para a estacionaridade a igualdade das distribuições de médias móveis ponderadas (ou seja, nem todos os pesos são iguais).

Em particular, a valer a estacionaridade, os Janeiros deveriam ter mesma distribuição que os Outubros e assim por diante³. Por clareza de exposição re-exibimos no gráfico 9 o BoxPlot dos Fatores do Ceará.

Fatores de Capacidade no Ceará: Dados Originais

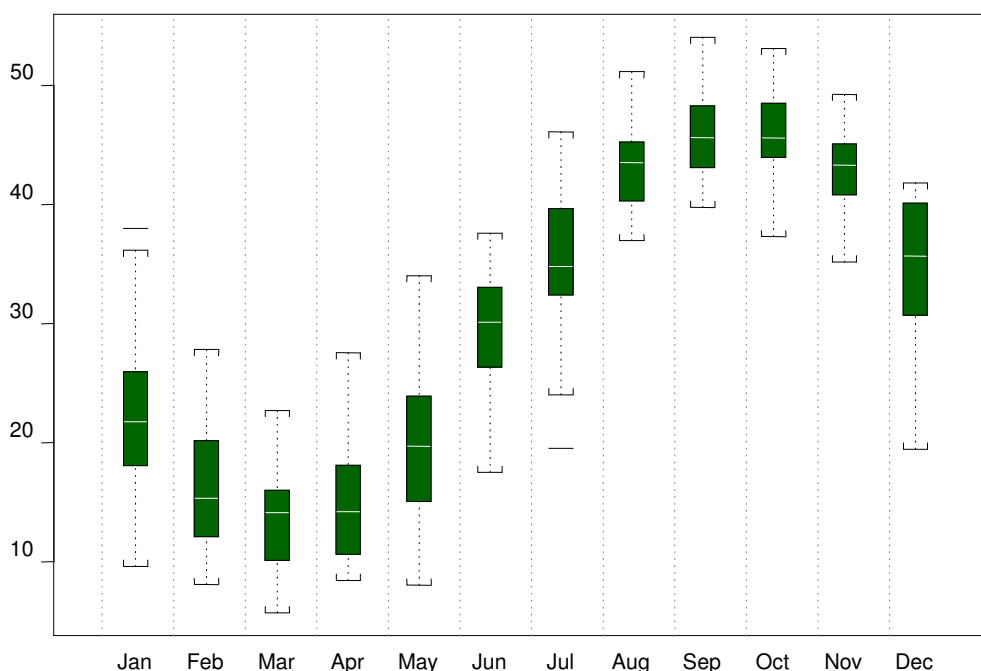


Gráfico 9: BoxPlots: Evidência de não-estacionaridade

Algumas características deste BoxPlot colocam a estacionaridade sob suspeita: (a) a maioria dos meses tem distribuição aproximadamente simétrica; exceções: Dezembro

³ Note que as observações originais forma uma média móvel de tamanho 1.

e Abril, (b) as distribuições dos meses dos cumes e vales são menos dispersas que as dos restantes meses e, a mais clara, (c) os meses evoluem ao redor de valores distintos (evidenciando a sazonalidade).

A investigação acerca da estacionaridade tem alguns protocolos clássicos. Registramos que com a série dos Fatores do Ceará conduzimos vários testes de raiz unitária (i. com e sem mudanças de regime e ii. testes de raiz unitária sazonal) e também consideramos transformações do tipo Box-Cox. Iremos alocar em apêndice no próximo módulo, algumas versões editadas destes testes. Aqui iremos nos restringir à apresentação do modelo estimado e a algumas características de interesse.

De início registramos que o uso de transformações não lineares (do tipo Box-Cox) nos dados visando simplificação da complexidade dos modelos não foi proveitoso. A diversidade sazonal das distribuições mensais persistia de forma importante após as transformações.

A pesquisa por raízes unitárias usuais segundo metodologias tradicionais se mostrou por vezes inapropriada⁴. É possível que isto tenha ocorrido devido ao caráter periódico do relacionamento entre a energia eólica no mês t , digamos, e no mês $t-1$. *Grosso modo*, quando este relacionamento probabilístico varia periodicamente com t dizemos que a série temporal segue modelo periódico.

Modelos periódicos são particularmente importantes em hidrologia. Não por acaso o modelo utilizado para o tratamento das Energias Naturais Afluentes no NEWAVE é um auto-regressivo *periódico*.

Com respeito à série de Fatores do Ceará obtivemos como modelo final uma estrutura particularmente simples. A forma final estimada é dada por

$$y_t = \varphi_t^1 y_{t-1} + \varphi_t^2 y_{t-2} + \varphi_t^3 y_{t-3} + \sigma_t \varepsilon_t$$

Onde denotamos por y_t o valor do fator de capacidade no mês t do qual subtraímos a média aritmética (todos os anos) dos valores observados para o mês t^5 . Também, os coeficientes $\varphi_t^1, \varphi_t^2, \varphi_t^3$ e σ_t variam de modo periódico (i.e. $\varphi_t^1 = \varphi_{t-12}^1$ para todo t). Finalmente, a série (não observada) $\{\varepsilon_t\}$ é um ruído branco.

Registramos que também foi este um dos melhores modelos que obtivemos para a série Média das 3 Usinas. Os valores dos parâmetros são distintos. Os testes de adequabilidade dos dados ao modelo estimado indicam uma performance pior do que para os dados do Ceará mas ainda assim bastante razoáveis.

⁴ Por exemplo testes HEGY indicaram modelo claramente super-diferenciado.

⁵ No inglês, a série y_t é a série de Fatores *periodically demeaned*.

No entanto, o fato de que um mesmo modelo foi obtido para as duas séries nos deixa em situação mais confortável relativamente ao uso da série de Camargo Schubert como *proxi* para a atividade eólica no Ceará.

Registramos no gráfico , para fins meramente ilustrativos, o ajuste do modelo para a série do Ceará.

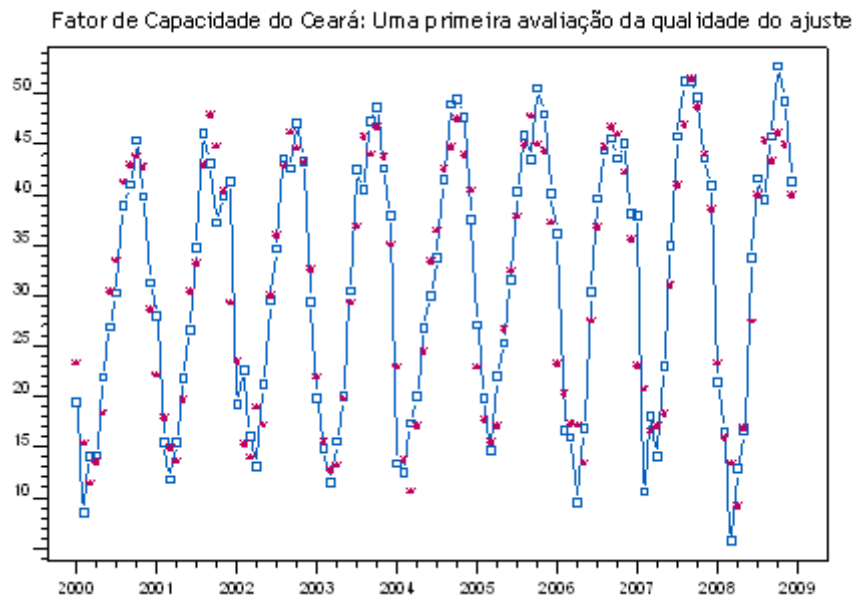


Gráfico 10: A adequabilidade da modelagem: evidências gráficas.

5. Algumas Conclusões Preliminares

Consideramos neste relatório análise exploratória de séries de fatores de capacidade de 3 usinas de geração eólica em atividade no Ceará. Estudamos também a evolução de série mais longa com cobertura tanto geográfica como temporal maiores do que as das 3 usinas.

O ano de 2009, a julgar pelos dados das 3 usinas parece ser um período de baixa atividade eólica no Ceará. Este fato não compromete a qualidade dos dados das 3 usinas. Aparentemente é um ciclo local de baixa. Tais ciclos ocorrem repetidas vezes na série de fatores de carga para o Ceará disponibilizada pela consultoria Camargo & Schubert.

Relativamente a esta última apresentamos modelagem estatística que, adiantamos, passou pela maioria dos testes de especificação clássicos. A classe de modelos que se mostrou apropriada foi a mesma que é utilizada na modelagem da Energia Natural Afluyente no NEWAVE: modelos auto-regressivos periódicos.

Esta classe também se mostrou apropriada para a média (devidamente ponderada) dos fatores de capacidade observados para as 3 usinas.

Por conseguinte temos boas indicações de que as séries de fatores de capacidade das Unidades da Federação irão se prestar satisfatoriamente como correspondente eólico da Energia Natural Afluyente modelada no NEWAVE.

6. Estratégias de Modelagem 1: Modelos Estruturais

Nesta seção conduzimos trabalho de modelagem da série de Fatores de Capacidade do estado do Ceará, fornecida pela Camargo & Schubert e também as das usinas X, Y e Z.

O enfoque aqui adotado se baseia na classe de modelos que podem ser expressos na forma de Espaço de Estados. Esta classe é bastante flexível. Comporta, por exemplo, modelos ARMA multivariados, modelos lineares com parâmetros que variam no tempo e também modelos com trocas Markovianas de regime.

O restante desta seção está assim dividido. Na subseção a seguir apresentamos a forma geral de modelos univariados em Espaço de Estados. Exibimos o caso particular de maior interesse neste projeto: os modelos estruturais de Harvey. Na subseção seguinte estendemos para o caso multivariado já tendo em vista a modelagem conjunta das 4 séries que dispomos. A última subseção apresenta a modelagem propriamente dita.

MODELOS EM ESPAÇO DE ESTADOS

Iremos denotar por $\mathbf{Y} = \{y_t, t = 1, 2, \dots, n\}$ uma série temporal univariada genérica. Dizemos que \mathbf{Y} segue modelo de Espaço de Estados quando podemos escrever⁶

$$\begin{aligned} \alpha_{t+1} &= d_t + \mathbf{T}_t \cdot \alpha_t + \mathbf{H}_t \cdot \eta_t \\ \theta_t &= c_t + \mathbf{Z}_t \cdot \alpha_t \\ y_t &= \theta_t + G_t \cdot \varepsilon_t \end{aligned}$$

$\begin{matrix} m \times 1 & = & m \times 1 & + & m \times m & \cdot & m \times 1 & + & m \times r & \cdot & r \times 1 \\ 1 \times 1 & = & 1 \times m & + & 1 \times m & \cdot & m \times 1 \\ 1 \times 1 & = & 1 \times 1 & + & 1 \times 1 & \cdot & 1 \times 1 \end{matrix}$

onde $t = 1, 2, \dots, n$ e

$$\begin{aligned} \alpha_1 &\sim \mathbf{N}(\mathbf{a}, \mathbf{P}) \\ \eta_t &\sim iid \mathbf{N}(\mathbf{0}, \mathbf{I}_r) \\ \varepsilon_t &\sim iid N(0, 1) \end{aligned}$$

Acima adotamos a convenção de que quantidades denotadas por letras gregas são estocásticas e não-observáveis. Quantidades representadas por letras latinas são fixas, possivelmente desconhecidas⁷.

⁶ Estamos adotando as notações e convenções de Zivot E. & Jiahui Wang: *Modeling Financial Time Series with S-Plus*[®]. Springer, New York.

Modelos que podem ser expressos em Espaço de Estados exibem algumas características que os tornam particularmente interessantes:

- i. Se baseiam em modelos Markovianos [não por acaso as equações que relacionam α_{t+1} e α_t são denominadas de equações de transição, jargão típico em modelos Markovianos].
- ii. Dispõem de algoritmos eficientes para os processos de estimação de quantidades desconhecidas [o filtro de Kalman, entre outros].
- iii. Dispõem de algoritmos eficientes para a previsão dos termos estocásticos tanto *ex-ante* [usualmente denominado o *problema de suavização*] como *ex-post* [*problema de previsão*, propriamente dito].
- iv. É possível relaxar-se a hipótese de normalidade dos termos de erro η_t e ε_t possibilitando, por exemplo, a modelagem dinâmica de séries temporais de dados de contagem. [ver, e.g. Durbin J & Siem J Koopman (2001) *Time Series Analysis by State Space Methods*. Oxford University Press, Oxford.]

A modelagem de séries temporais via Espaço de Estados conta com substancial literatura. Na literatura econométrica, em particular, a difusão desta classe de modelos se deve em boa medida a desenvolvimentos de A C Harvey e colaboradores.

Harvey, nos anos 70 e 80, desenvolveu a classe dos modelos estruturais de Séries Temporais. O sucesso de tais modelos possivelmente reside na simplicidade dos conceitos ali considerados. De fato, os modelos estruturais em sua forma mais simples lidam com os conceitos - já, há muito, incorporados à nossa intuição - de tendência, sazonalidade, ciclo e componente irregular.

Mais especificamente dizemos que uma série temporal Y segue um modelo básico de séries temporais estruturais quando podemos escrever:

$$y_t = \mu_t + \gamma_t + \psi_t + \xi_t$$

onde

μ_t	=	componente de tendência
γ_t	=	componente sazonal
ψ_t	=	componente cíclica
ξ_t	=	componente irregular

⁷ Quando for assim, tipicamente serão estimadas com base em DADOS: as observações da série temporal, regressores, etc.

evoluem segundo

$$\begin{aligned}\mu_{t+1} &= \mu_t + \beta_t + \eta_t, & \eta_t &\sim iid N(0, \sigma_\eta^2) \\ \beta_{t+1} &= \beta_t + \varsigma_t & \varsigma_t &\sim iid N(0, \sigma_\varsigma^2) \\ \gamma_t &= -(\gamma_{t-1} + \dots + \gamma_{t-11}) + \omega_t & \omega_t &\sim iid N(0, \sigma_\omega^2)\end{aligned}$$

e $\{\xi_t\}$ é ruído branco.

Na formulação acima supomos dados coletados mensalmente. Se fossem coletados trimestralmente, por exemplo, na equação de evolução da componente sazonal teríamos 3 em vez de 11.

A evolução da componente cíclica associada a uma frequência λ_c é dada por

$$\begin{pmatrix} \psi_{t+1} \\ \psi_{t+1}^\dagger \end{pmatrix} = \rho \begin{pmatrix} +\cos \lambda_c & \sin \lambda_c \\ -\sin \lambda_c & \cos \lambda_c \end{pmatrix} \begin{pmatrix} \psi_t \\ \psi_t^\dagger \end{pmatrix} + \begin{pmatrix} \chi_t \\ \chi_t^\dagger \end{pmatrix},$$

onde

$$\begin{pmatrix} \chi_t \\ \chi_t^\dagger \end{pmatrix} \sim iid \mathbf{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \sigma_\psi^2 (1 - \rho^2) \mathbf{I}_2 \right).$$

ou seja, a componente cíclica é modelada como a *parte real* de um vetor complexo que do instante t para o instante $t+1$ (a) tem seu módulo multiplicado por ρ , (b) tem seu argumento acrescido de λ_c e (c) é somado a vetor aleatório complexo com componentes normais independentes e com mesma variância.

A componente de tendência tem como casos particulares uma reta (ambos, ς_t e η_t , com variância nula), um passeio aleatório ($\text{Var}(\varsigma_t) = 0$ e $B_1 = 0$) e um passeio aleatório com *drift* ($\text{Var}(\varsigma_t) = 0$ e $B_1 \neq 0$).

Quando a componente sazonal é como acima dizemos que a sazonalidade foi modelada com *dummies sazonais*. A modelagem pode ser conduzida de outra forma, com o uso de funções trigonométricas. Suponhamos, para fixar idéias, que nossos dados sejam mensais. Denote por $\lambda_1, \dots, \lambda_6$ as frequências de Fourier dadas por $\lambda_j = 2\pi j / 12, j=1, \dots, 6$.

Dizemos que Y tem sua sazonalidade modelada por funções trigonométricas quando escrevemos

$$\gamma_t = \gamma_t^1 + \dots + \gamma_t^6$$

onde

$$\begin{pmatrix} \gamma_{t+1}^j \\ \tilde{\gamma}_{t+1}^j \end{pmatrix} = \begin{pmatrix} +\cos \lambda_j & \sin \lambda_j \\ -\sin \lambda_j & \cos \lambda_j \end{pmatrix} \begin{pmatrix} \gamma_t^j \\ \tilde{\gamma}_t^j \end{pmatrix} + \begin{pmatrix} \chi_t^j \\ \tilde{\chi}_t^j \end{pmatrix}$$

com os termos de erro tais que

$$\left\{ \chi_t^j, \tilde{\chi}_t^j, j = 1, \dots, 6; t = 1, \dots, n \right\} \sim iid N(0, \sigma_\gamma^2)$$

Frisamos que na especificação acima a variância dos termos de erro de cada um dos γ_t^j é a mesma. Em geral esta imposição não é problemática. Entretanto, pode ser interessante permitir uma menor rigidez à componente sazonal. Isso pode ser obtido deixando que as variâncias dos mencionados termos de erro sejam livremente estimadas. Algumas delas podem ser nulas, o que significa que a correspondente parcela da componente sazonal é constante - não estocástica.

EXTENSÕES MULTIVARIADAS

Antes de formalizarmos a extensão multivariada da modelagem em espaço de estados é conveniente justificarmos sua necessidade. Faremos isso informalmente.

Nossos dados, relativos ao estado do Ceará são de classes distintas. Por um lado temos a série, longa, da Camargo & Schubert que se refere ao estado do Ceará como um todo. Por outro temos as 3 séries de fatores de capacidade das usinas X, Y e Z. Cada uma delas é bem menor que a série para o estado todo.

É conveniente imaginarmos a série de fatores de capacidade para o estado - *série agregada*, doravante - como uma média ponderada de um conjunto de várias outras. Estão nesse conjunto as usinas X, Y e Z. Portanto, as componentes sazonais, de tendência e de ciclo de cada uma das séries das usinas devem estar relacionadas àquelas da série agregada.

Uma forma simples de investigarmos a associação entre as componentes sazonais das usinas e a da série agregada é supormos que aquelas sejam múltiplos escalares desta. Mais formalmente: denotemos por $\gamma^A(t)$, $\gamma^M(t)$, $\gamma^P(t)$ e $\gamma^T(t)$ as componentes sazonais da série agregada, de X, Y e Z, respectivamente. A associação acima apontada equivale à existência de escalares π_M , π_P e π_T tais que

$$\gamma^M(t) = \pi_M \cdot \gamma^A(t)$$

$$\gamma^P(t) = \pi_P \cdot \gamma^A(t)$$

$$\gamma^T(t) = \pi_T \cdot \gamma^A(t)$$

As relações acima refletem a existência de co-movimentos sazonais. É também de se esperar a existência de relações semelhantes para as componentes de tendência e, eventualmente, de ciclo.

Registrarmos a conveniência de ancorarmos as componentes sazonais das usinas às da série agregada. Isso se deve à (substancialmente) maior quantidade de observações da série agregada. O que implica na melhor estimação de comportamentos sistemáticos da atividade eólica no Ceará. É claro, no entanto, que as conveniências se transformariam em inconveniências caso a qualidade dos dados agregados se deteriorasse com a idade.

A generalização multivariada de modelos em espaço de estados comporta a modelagem conjunta da série agregada e das usinas X, Y e Z nos termos acima descritos. Formalmente, para a generalização multivariada basta uma mudança de dimensões:

Iremos denotar por $\mathbf{Y} = \{\mathbf{y}_t, t = 1, 2, \dots, n\}$ uma série temporal *N-variada* genérica.

Dizemos que \mathbf{Y} segue modelo de Espaço de Estados quando podemos escrever

$$\alpha_{t+1} = d_t + \mathbf{T}_t \cdot \alpha_t + \mathbf{H}_t \cdot \eta_t$$

$m \times 1 \quad m \times 1 \quad m \times m \quad m \times 1 \quad m \times r \quad r \times 1$

$$\theta_t = c_t + \mathbf{Z}_t \cdot \alpha_t$$

$N \times 1 \quad N \times m \quad N \times m \quad m \times 1$

$$\mathbf{y}_t = \theta_t + \mathbf{G}_t \cdot \varepsilon_t$$

$N \times 1 \quad N \times 1 \quad N \times N \quad N \times 1$

onde $t = 1, 2, \dots, n$ e

$$\alpha_1 \sim \mathbf{N}(\mathbf{a}, \mathbf{P})$$

$$\eta_t \sim iid \mathbf{N}(\mathbf{0}, \mathbf{I}_r)$$

$$\varepsilon_t \sim iid \mathbf{N}(\mathbf{0}, \mathbf{I}_N)$$

A linearidade das relações acima torna simples a imposição dos co-movimentos governando a evolução de nossos dados.

MODELAGEM DA SÉRIE AGREGADA POR MODELOS ESTRUTURAIS

A fim de que possamos lançar mão da estratégia de modelagem já descrita é necessário que a série agregada possa ser bem modelada segundo a metodologia de Espaço de Estados. Mais especificamente, é necessário que o modelo estrutural seja adequado à evolução da série agregada.

Conduzimos a modelagem da série agregada segundo o modelo estrutural de Harvey. A sazonalidade foi modelada via *dummies sazonais* e também funções trigonométricas. Registramos que tanto o ajuste como a capacidade preditiva nas 2 variantes (*dummies* e funções trigonométricas) foram semelhantes. Optamos por apresentar somente os resultados da modelagem com funções trigonométricas. Exibimos nos gráficos 11 e 12 o ajuste nos primeiros e nos últimos dez anos.

Fica claro que, embora as flutuações sazonais tenham sido aparentemente explicadas pelo modelo adotado, resta ainda substancial variabilidade. Tal variabilidade pode ser avaliada através dos resíduos⁸.

Análises de resíduos se prestam para a quantificação da capacidade preditiva de um modelo e também da aderência de dados ao modelo estimado. Se prestam também, e principalmente, para verificar a adequabilidade das hipóteses subjacentes ao modelo aos dados sendo modelados.

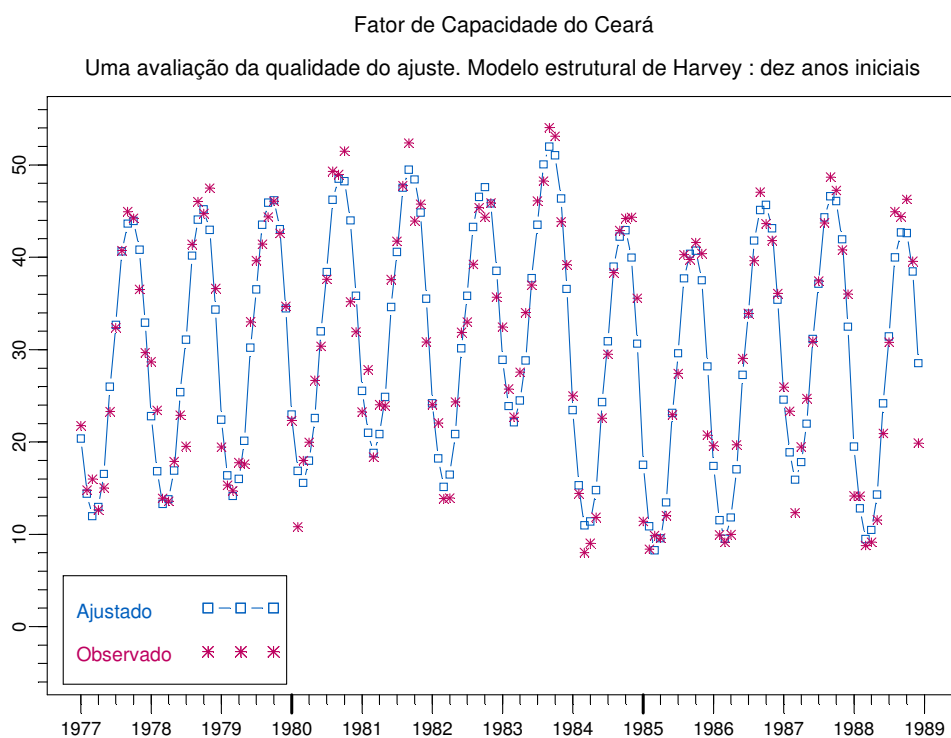


Gráfico 11: Ajuste de modelo estrutural à série da Camargo & Schubert (agregada): primeiros dez anos.

⁸ Resíduo = Valor Observado - Valor Previsto

No caso específico de modelos estruturais espera-se, como decorrência da forma assumida que a distribuição dos resíduos seja (a) aproximadamente Normal e (b) a mesma para todos os meses.

A tabela 3 abaixo fornece alguns percentis dos resíduos. Tais percentis parecem indicar que os resíduos se distribuem de maneira simétrica e, portanto compatíveis com a assumida distribuição normal dos termos de erro.

Percentis de Resíduos no Modelo Estrutural para os Dados Agregados							
Nível	5%	10%	25%	50%	75%	90%	95%
Valor	-7.14	-5.46	-2.69	+0.16	+2.93	+5.61	+7.37

Tabela 3: Resíduos do modelo estrutural para dados agregados: evidência de simetria.

Exibimos uma análise gráfica dos resíduos no gráfico 13. Lá fornecemos 2 gráficos. No primeiro, comparamos a distribuição dos resíduos com a distribuição normal através do chamado *qqplot*⁹. O padrão ali exibido evidencia uma distribuição cujos valores extremos são maiores (em módulo) do que se esperaria para a distribuição normal.

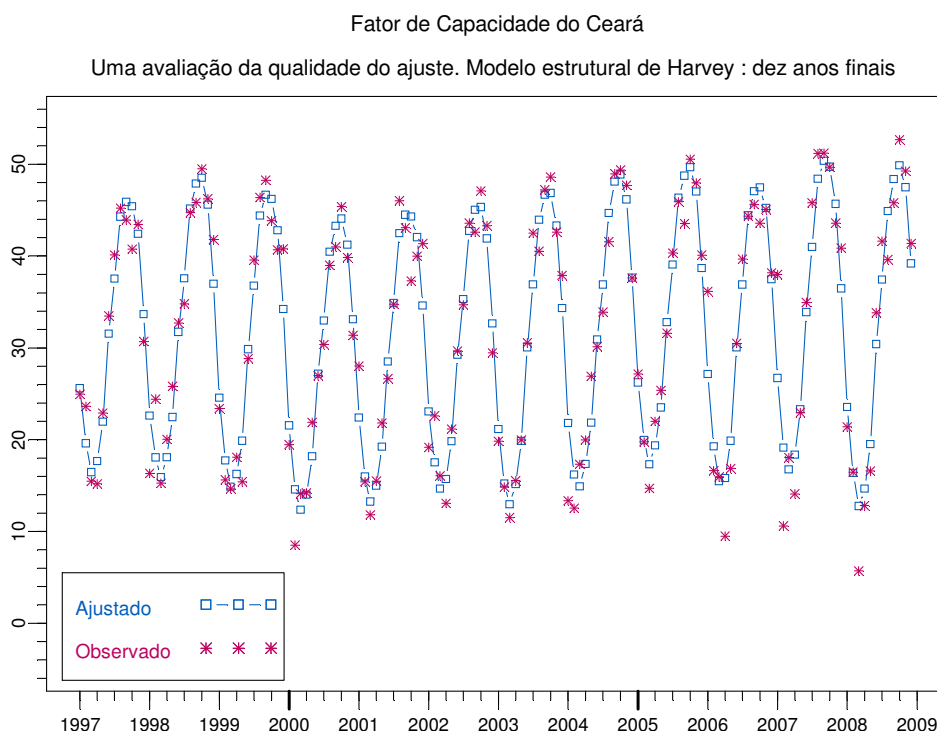


Gráfico 12: Ajuste de modelo estrutural à série da Camargo & Schubert (agregada): últimos dez anos.

⁹ Lembramos que quanto mais pontos se distanciam da reta plotada num *qqplot* menor a evidência de normalidade. Para maiores detalhes recomendamos os clássicos Tukey, J W (1977): *Exploratory Data Analysis*. Addison Wesley: New York e Cleveland, W S (1993). *Visualizing Data*. AT&T Bell Labs: Murray Hill, New Jersey.

Há várias possíveis causas para isso. Em nosso caso é importante analisarmos se a sazonalidade está, em alguma medida, associada a aquele padrão. Isso é investigado no segundo gráfico 13. Nele apresentamos os BoxPlots dos resíduos de acordo com os meses a que se referem. Ali observamos:

- i. Em alguns meses os resíduos são mais dispersos que em outros
- ii. Em alguns meses os resíduos são mais assimétricos que em outros
- iii. Nos meses de Julho, Outubro e Dezembro, a quantidade de resíduos positivos é moderadamente maior que a de negativos¹⁰.
- iv. Nos meses de Janeiro e Fevereiro a quantidade de resíduos negativos é moderadamente maior que a de positivos¹¹.
- v. A hipótese de que os resíduos mensais seguem uma mesma distribuição é de difícil sustentação.

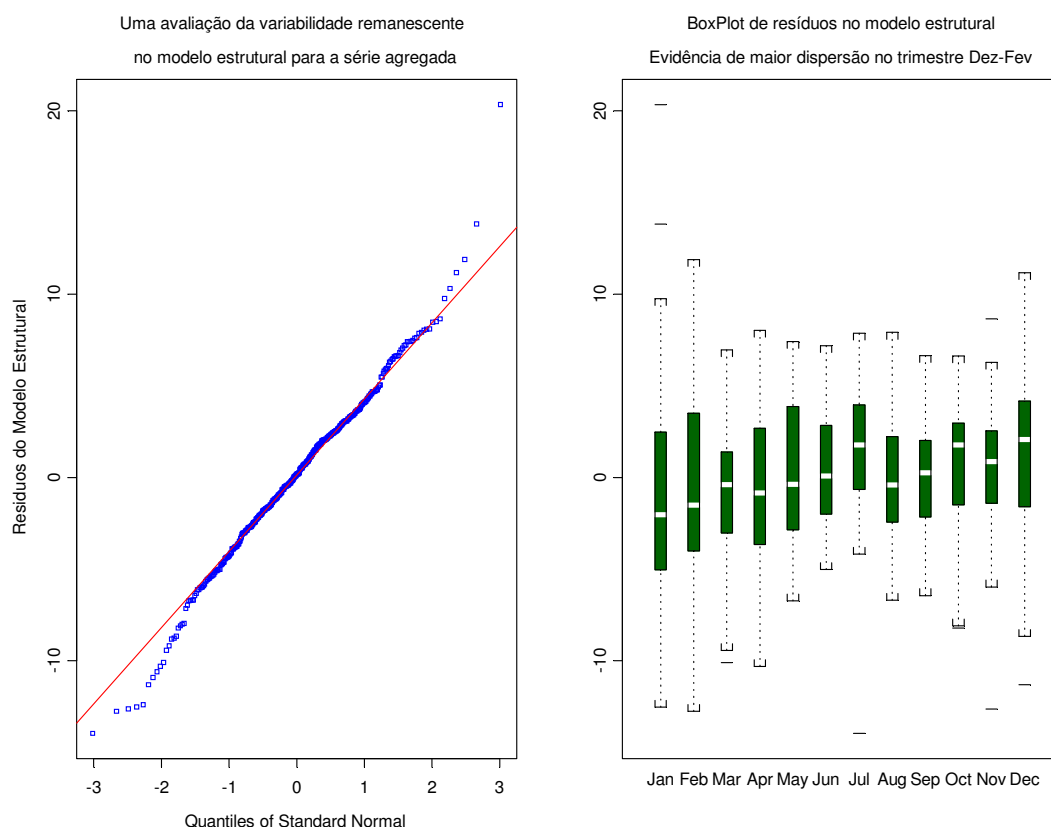


Gráfico 13: qqplots de resíduos de modelo estrutural ajustado à série agregada (Gráfico à esquerda). BoxPlots dos mesmos resíduos, segmentados pelo mês a que se referem (Gráfico à direita).

O item v, acima, implicaria a inadequabilidade, para a série agregada, dos modelos estruturais na forma anteriormente descrita. Para uma melhor avaliação, realizamos

¹⁰ Decorre daí que uma grande parcela das previsões do modelo são, naqueles meses, menores que os dados observados

¹¹ Decorre daí que uma grande parcela das previsões do modelo são, naqueles meses, maiores que os dados observados

testes formais de hipóteses: conduzimos diversos testes de normalidade¹² (Shapiro-Wilk, Shapiro-Wilk-Royston e Jarque-Bera) e também de igualdade de distribuições (Kolmogorov-Smirnov 2-amostras). A normalidade da série de resíduos -como um todo, não segmentada- foi convincentemente rejeitada (cf. Tabela 4).

Teste de Normalidade	Shapiro-Wilk	Shapiro-Wilk-Royston	Jarque-Bera
p-valor	0.004	0.003	0.000

Tabela 4: *p-valores* de testes de normalidade aplicados a resíduos não segmentados.

Quando segmentamos a série de resíduos segundo o mês subjacente, rejeitamos a normalidade dos resíduos de Julho, Outubro e Novembro (cf. Tabela 5). Também ficou claro (cf. Tabela 6) que os resíduos referentes aos meses iniciais (Jan e Fev) têm distribuições significativamente diferentes das dos resíduos dos meses de Junho e Julho e também das do último trimestre do ano.

Teste	Meses											
	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
SW	0,15	0,69	0,55	0,96	0,19	0,32	0,03	0,76	0,51	0,03	0,12	0,44
SWR	0,13	0,68	0,52	0,94	0,23	0,36	0,02	0,80	0,57	0,03	0,07	0,40
JB	0,16	0,83	0,59	0,95	0,48	0,63	0,00	0,79	0,97	0,21	0,01	0,43

Tabela 5: *p-valores* de testes de normalidade aplicados a resíduos segmentados de acordo com o mês a que se referem. Convenção para os testes utilizados: SW= Shapiro-Wilk, SWR= Shapiro-Wilk-Royston.e JB= Jarque-Bera.

¹² Temos consciência da possível perda de potência de testes quando parâmetros são estimados. Ocorre que, para os modelos estruturais, não foram ainda desenvolvidos testes mais apropriados. A referência clássica para testes com parâmetros estimados é Durbin, J. [1975] Weak convergence of the sample distribution function when parameters are estimated. *The Annals of Statistics*; Vol 1, N° 2, 279-90.

Teste de Igualdade de Distribuição dos Resíduos Mensais												
	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
Jan	1.00	0.43	0.09	0.43	0.27	0.01	0.02	0.16	0.09	0.04	0.04	0.04
Fev		1.00	0.64	0.84	0.27	0.16	0.01	0.64	0.27	0.02	0.04	0.02
Mar			1.00	0.84	0.27	0.43	0.09	0.64	0.97	0.16	0.09	0.01
Abr				1.00	0.84	0.43	0.09	0.97	0.84	0.16	0.27	0.09
Mai					1.00	0.84	0.27	0.84	0.84	0.43	0.84	0.43
Jun						1.00	0.43	0.84	0.84	0.64	0.84	0.27
Jul							1.00	0.27	0.27	0.43	0.43	0.97
Ago								1.00	0.97	0.27	0.64	0.16
Set									1.00	0.43	0.64	0.09
Out										1.00	0.64	0.27
Nov											1.00	0.27
Dez												1.00

Tabela 6: *p*-valores de testes (Kolmogorov-Smirnov 2 amostras) de igualdade de distribuição 2-a-2 dos resíduos segmentados segundo o mês a que se referem.

Concluimos, portanto, pela não adequabilidade dos modelos estruturais, na forma aqui descrita, para a série agregada do Estado do Ceará.

MODELAGEM DAS SÉRIES DAS USINAS POR MODELOS ESTRUTURAIS

Nesta subseção modelamos individualmente e conjuntamente a evolução dos Fatores de Capacidade das usinas X, Y e Z. Também aqui a metodologia utilizada é baseada em Modelos Estruturais.

Estimamos modelos estruturais individuais para as usinas X, Y e Z. Consideramos para as componentes sazonais tanto a forma de *dummies* sazonais como as funções trigonométricas. Como com a série agregada os resultados foram semelhantes. Optamos por apresentar somente resultados para a versão das funções trigonométricas.

A tabela 7 exibe os desvios padrão estimados para os termos de erro das componentes sazonais e também da componente de tendência. Exibe também os desvios padrão das componentes irregulares.

Usina	Componente		
	Irregular	Tendência	Sazonal
Z	0.046	0.039	0.001
Y	0.045	0.017	0.001
X	0.054	0.027	0.002

Tabela 7: Desvios padrão estimados das componentes dos modelos estruturais.

Os termos de erro das componentes sazonais têm desvios padrão próximos de zero. Isso indica a rigidez, quase determinística, das componentes sazonais. De resto, fica clara a semelhança das quantidades estimadas para as 3 usinas.

O ajuste pode ser conferido nos gráficos 14 a 16. Lá, a aludida rigidez das componentes sazonais pode ser observada. Em linhas gerais, os modelos estruturais individuais parecem se adequar à evolução das atividades eólicas nas 3 usinas. E, ao contrário da série agregada, o ano de 2008 não parece destoar dos anteriores.

Da mesma forma que com a série agregada, apresentamos a seguir indicações da adequabilidade -ou falta dela- dos modelos estruturais individuais aos fatores de capacidade das usinas.

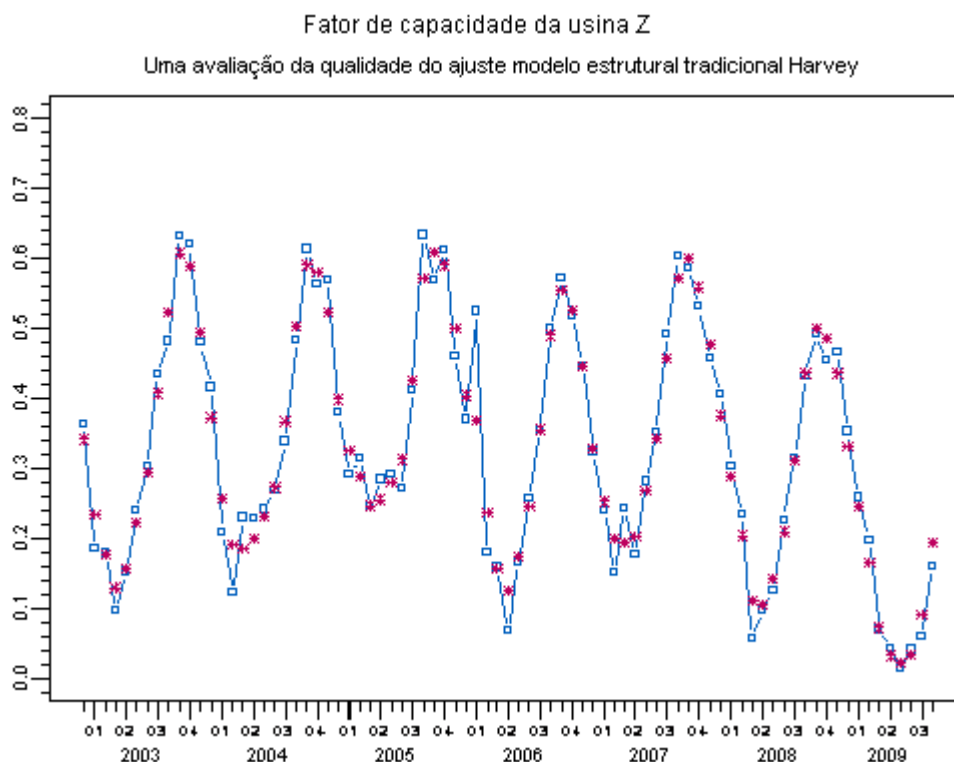


Gráfico 14: Ajuste de modelo estrutural à série de fatores de capacidade da usina Z.

Fator de capacidade da usina Y

Uma avaliação da qualidade do ajuste modelo estrutural tradicional Harvey

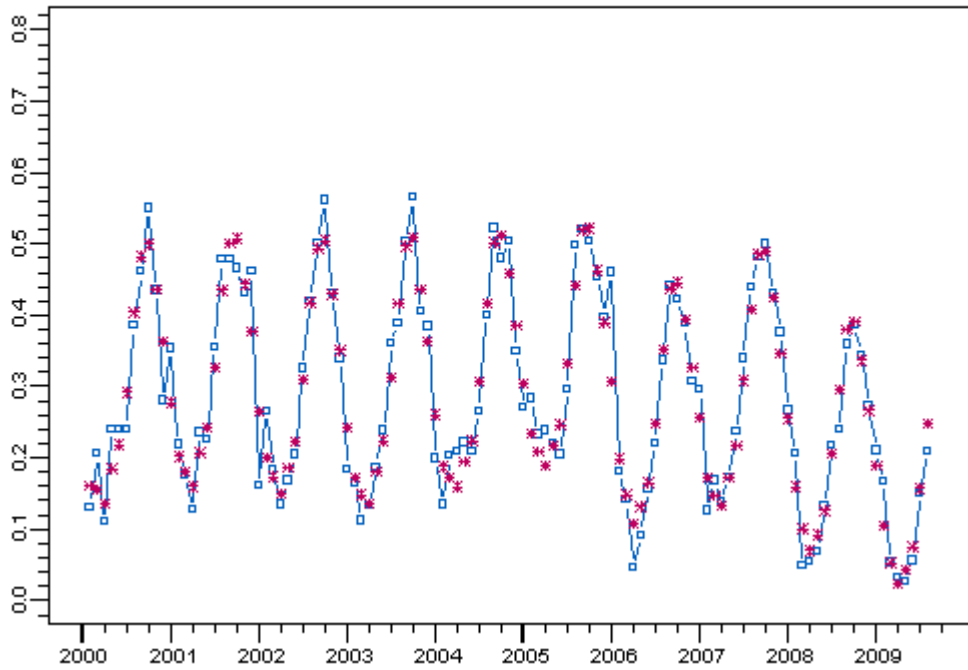


Gráfico 15: Ajuste de modelo estrutural à série de fatores de capacidade da usina Y.

Fator de capacidade da usina X

Uma avaliação da qualidade do ajuste modelo estrutural tradicional Harvey

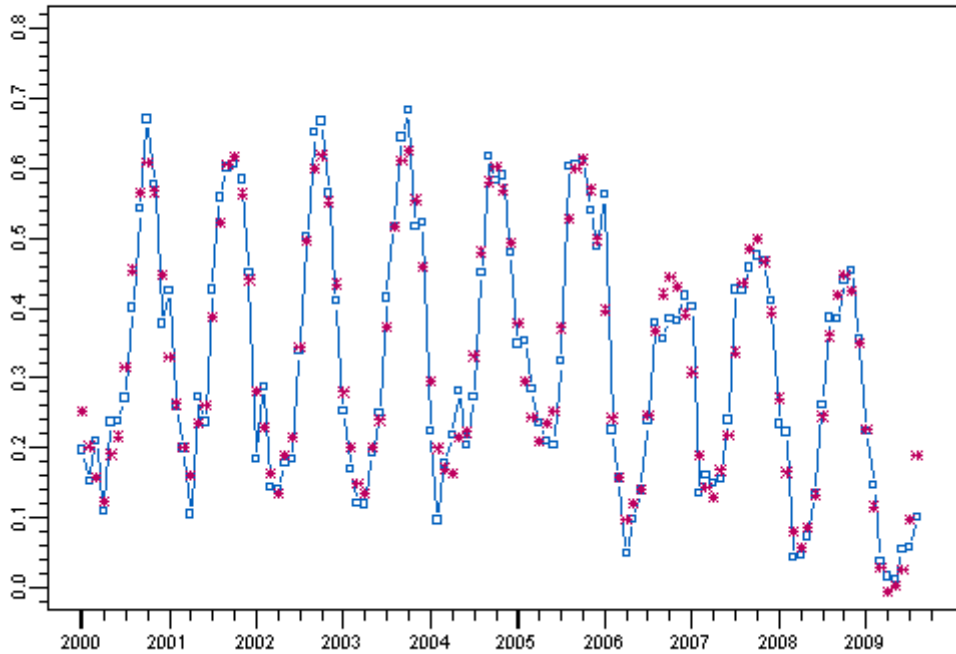


Gráfico 16: Ajuste de modelo estrutural à série de fatores de capacidade da usina X.

Da mesma forma que com a série agregada, apresentamos a seguir indicações da adequabilidade -ou falta dela- dos modelos estruturais individuais aos fatores de capacidade das usinas.

Iremos de início considerar a questão da dinâmica. Um objetivo comum a todos os modelos de séries temporais é o total aproveitamento de informações passadas para explicar o presente. Quando isso não ocorre, previsões e estimativas são, tipicamente, viesadas.

Um sintoma de que a informação passada não está sendo completamente considerada é quando os resíduos exibem dependência temporal significativa. O teste mais utilizado para a avaliação de dependências temporais -o teste de Ljung-Box- se baseia na função de autocorrelação dos resíduos¹³. Exibimos na Tabela 8 os *p-valores* para o teste de Ljung-Box aplicado aos resíduos dos modelos estruturais estimados para cada uma das séries das usinas. A hipótese nula é que não existe dependência temporal entre os termos de erro¹⁴. Fica claro que a hipótese nula não é rejeitada¹⁵.

Usina	Nº de Defasagens (k) que o teste se baseia							
	3	6	9	12	15	18	21	24
Z	0.41	0.12	0.20	0.40	0.48	0.48	0.54	0.58
Y	0.98	0.23	0.43	0.32	0.23	0.39	0.41	0.49
X	0.14	0.24	0.14	0.26	0.40	0.52	0.70	0.74

Tabela 8 : *p-valores* do teste de Ljung-Box aplicado aos resíduos dos modelos estruturais individuais.

A ausência de dependência temporal dos termos de erro não é suficiente para garantir a adequabilidade do modelo estimado. Avaliamos agora a questão da normalidade. Nos gráficos 17 e 18 podemos avaliar outro aspecto importante, já considerado na análise da série agregada.

¹³ Mais especificamente, se baseia numa soma ponderada dos quadrados das k primeiras autocorrelações estimadas

¹⁴ Aos quais os resíduos correspondem.

¹⁵ Lançamos mão do teste de Ljung-Box para seguir o protocolo da modelagem de séries temporais. No nosso contexto uma versão periódica de Ljung-Box provavelmente exibiria melhor performance (potência).

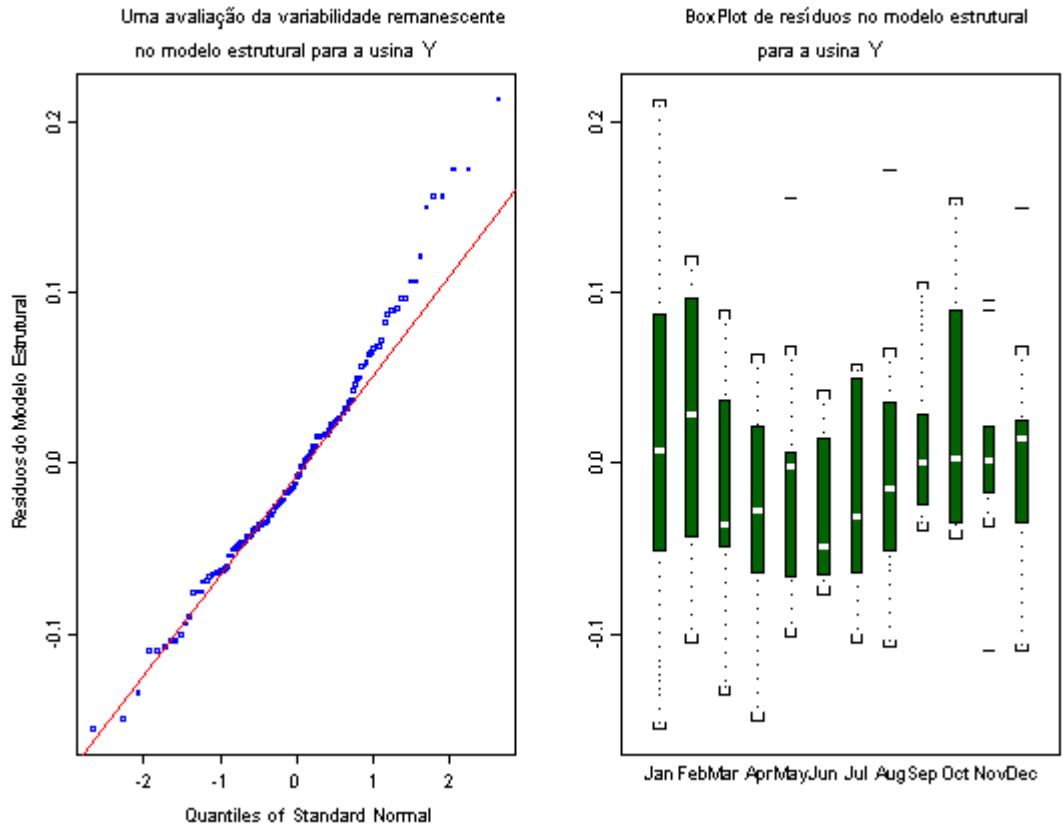


Gráfico 17: *qqplots* de resíduos de modelo estrutural ajustado à série da usina Y (Gráfico à esquerda). *BoxPlots* dos mesmos resíduos, segmentados pelo mês a que se referem (Gráfico à direita).

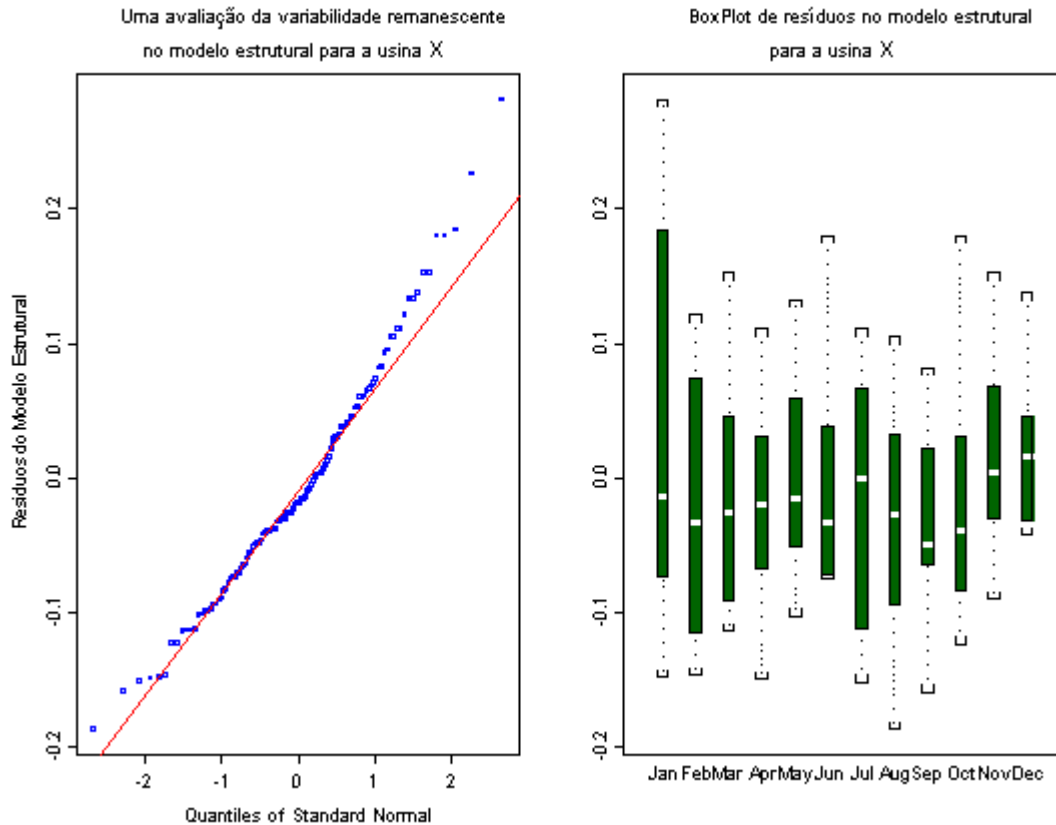


Gráfico 18: qqplots de resíduos de modelo estrutural ajustado à série da usina X (Gráfico à esquerda). BoxPlots dos mesmos resíduos, segmentados pelo mês a que se referem (Gráfico à direita).

Podemos, através das figuras 17 e 18¹⁶, colocar sob suspeita duas hipóteses assumidas nos modelos estruturais: (a) normalidade dos termos de erro e (b) os termos de erro têm mesma distribuição.

Se, de fato, (a) e (b) não forem válidas, previsões pontuais e, bem mais importante, intervalos de confiança para previsões podem ficar distorcidos. Isso decorre da incorreta consideração da variabilidade mensal. A tabela 9 exhibe os *p-valores* de testes de normalidade aplicados aos resíduos. A normalidade é convincentemente rejeitada.

Teste de Normalidade	Usinas		
	Z	Y	X
SWR	0.04	0.06	0.05
JB	0.01	0.01	0.01

Tabela 9: *p-valores* de testes de normalidade aplicados a resíduos segmentados de acordo com o mês a que se referem. Convenção para os testes utilizados: SWR= Shapiro-Wilk-Royston.e JB= Jarque-Bera.

¹⁶ Registramos que os resíduos da usina Z apresentam padrões mais irregulares que os de X e Y. Não os incluímos, pois, devido ao fato de a usina Z ter 36 observações a menos que a usina X, tais padrões irregulares devem ser analisados com maior cautela.

A não-normalidade acima concluída já era esperada, tendo em conta, por exemplo, o gráfico 18. Não iremos testar a igualdade das distribuições dos resíduos mensalmente segmentados devido à pequena quantidade de observações¹⁷. No entanto é interessante investigarmos se os resíduos (todos eles) da usina X têm a mesma distribuição que os da usina Y, por exemplo. A tabela 10 apresenta os *p-valores* para testes de igualdade de distribuição para todas as possíveis combinações 2-a-2 dos resíduos¹⁸.

Hipótese sendo testada			
	Usina Z = Usina Y	Usina Z = Usina X	Usina Y = Usina X
p-valor	0.415	0.182	0.318

Tabela 10: *p-valores* de testes de igualdade de distribuição (Kolmogorov-Smirnov 2 amostras): Resíduos de modelos estruturais univariados.

Dali obtemos indício da homogeneidade da distribuição dos fatores de capacidade das 3 usinas: não rejeitamos a igualdade de distribuição para nenhum dos pares possíveis. Ora, o teste de Kolmogorov-Smirnov compara todos os resíduos de uma usina com todos os resíduos de outra. Não leva em conta a informação do mês ao qual o resíduo se refere. Para considerar o mês, ou seja testar, por exemplo, se a distribuição dos termos de erro da usina Z em Março é igual à correspondente distribuição para a usina X, precisaríamos de amostras maiores.

No entanto podemos investigar em que medida os resíduos das usinas estão temporalmente correlacionados. Para uma análise preliminar consideramos as funções de autocorrelação dos resíduos e também as funções de correlação cruzada das séries dos 3 resíduos.

O gráfico 19 exhibe os gráficos dessas funções. Lá encontramos indícios de fortes associações instantâneas: quando o resíduo da usina X é positivo também o é (em geral), no mesmo mês, o da usina Y. Não iremos continuar nessa linha de pesquisa uma vez que ela é mais apropriada diante da estacionaridade das séries envolvidas. Não conseguimos até agora evidências de estacionaridade. Por exemplo, não conseguimos evidências de que a variância dos termos de erro dos Janeiros seja igual à dos Julhos.

¹⁷ Fizemos isso para a série agregada, onde dispúnhamos de quantidade bem maior de observações do que aqui.

¹⁸ O teste de igualdade de distribuições quando aplicado a resíduos da série agregada e da usina X tem *p-valor* $< 10^{-4}$.

Correlações entre resíduos de modelos estruturais: Auto e Cruzadas

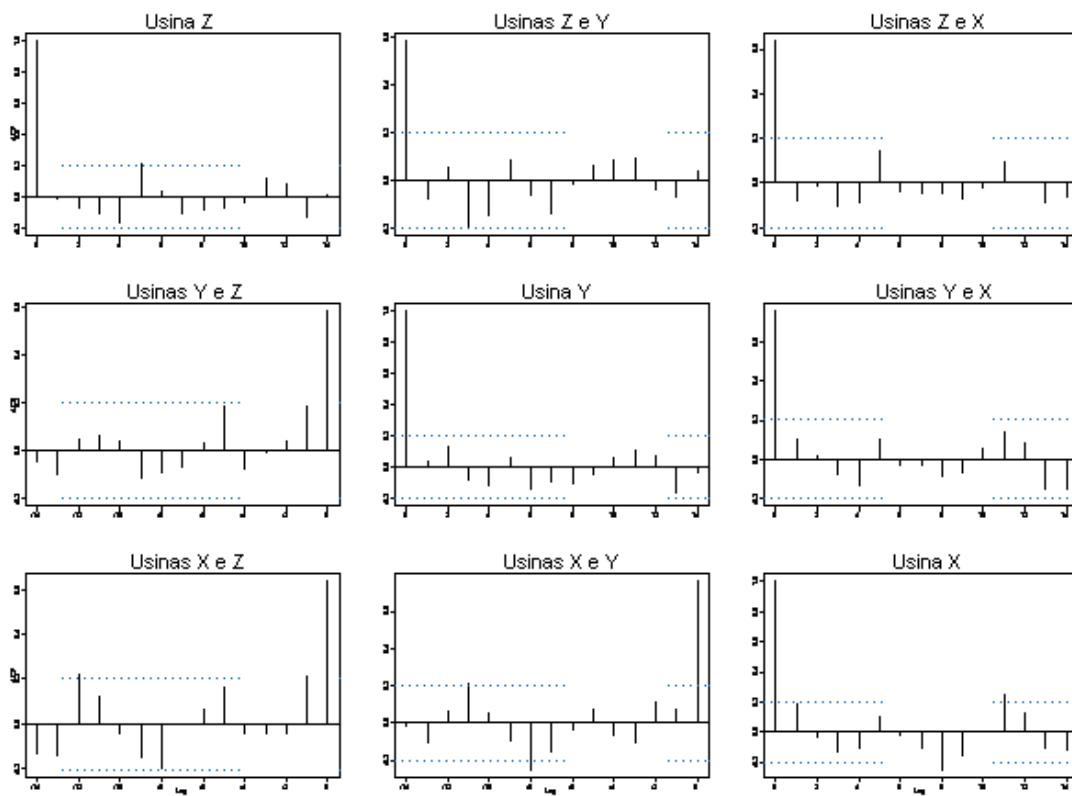


Gráfico 19: Auto-correlações e correlações cruzadas dos resíduos.

7. Estratégias de Modelagem 2: Modelos Periódicos

Nesta seção analisamos outra classe de modelos de séries temporais: os modelos periódicos. Essa classe é especialmente importante na modelagem de séries hidrológicas. Em particular os chamados modelos auto-regressivos periódicos servem de base para a modelagem das séries de Energia Natural Afluente no modelo computacional NEWAVE.

Neste trabalho iremos considerar somente os modelos auto-regressivos periódicos e uma sua generalização apropriada: os modelos auto-regressivos quantílicos periódicos¹⁹.

O restante da seção está assim dividido: na próxima subseção apresentamos os modelos PAR: auto-regressivos periódicos clássicos. Em seguida motivamos e apresentamos os modelos QAR-P, uma generalização dos modelos PAR que acomoda alguns padrões de não normalidade das séries sendo analisadas. Nas duas últimas subseções aplicamos aos dados de fatores de capacidade das usinas e também da série agregada para o estado do Ceará.

MODELOS AUTO-REGRESSIVOS PERIÓDICOS

Da mesma forma que na seção anterior começamos a descrição dos modelos periódicos para séries temporais univariadas. Mais ainda, para simplificar a exposição, iremos restringir nossa apresentação inicial aos modelos auto-regressivos periódicos de ordem 2. Iremos também convencionar que nossos dados são mensalmente coletados. Isso faz com que o período sazonal seja 12.

Diremos que uma série temporal $\{y_t\}$ segue um processo auto-regressivo periódico de ordem 2 (**Notação:** $y_t \sim \text{PAR}(2)$) quando podemos escrever

$$\tilde{y}_t = \phi_1(t) \cdot \tilde{y}_{t-1} + \phi_2(t) \cdot \tilde{y}_{t-2} + u_t$$

Onde pusemos

$$\tilde{y}_t = \frac{y_t - \nu_t}{\tau_t}$$

Com $\nu_t := E y_t$ e $\tau_t := \sqrt{\text{Var}(y_t)}$. Admitimos também que as funções

$$\mathbb{Z} \ni t \mapsto \phi_1(t) \quad \text{e} \quad \mathbb{Z} \ni t \mapsto \phi_2(t)$$

¹⁹ Registramos, entretanto, que a classe de modelos periódicos comporta modelos com estrutura bem mais geral, ARFIMAs periódicos, por exemplo.

são ambas periódicas, ou seja, para todo t e para $j= 1, 2$ temos

$$\phi_j(t) = \phi_j(t - 12)$$

Supomos, finalmente, que as u_t sejam não correlacionadas²⁰, com valor esperado nulo e que a distribuição de u_t e de u_{t-12} sejam as mesmas para todo t .

Por vezes assume-se que os termos de erro u_t sejam normalmente distribuídos. Isso não é necessário. Por exemplo, os modelos utilizados para a modelagem da Energia Natural Afluyente (ENA, doravante) no NEWAVE assumem que os termos de erro seguem distribuições *lognormais* deslocadas.

Frisamos que os modelos auto-regressivos periódicos são um exemplo de modelos periódicos. Na formulação mais geral dizemos que uma série temporal $\{ y_t \}$ segue modelo periódico quando a distribuição de (y_t, y_{t-1}, \dots) é a mesma que a de $(y_{t-12}, y_{t-13}, \dots)$, para todo t .

A formulação acima se estende de maneira natural a modelos auto-regressivos periódicos de ordem p ²¹. Aqui, por razões didáticas, não iremos formalizar essas extensões. Recomendamos ao leitor interessado o texto de Franses & Paap^{22,23}.

Iremos agora lidar com uma generalização dos modelos auto-regressivos periódicos um pouco mais sutil. Para fins de exposição, é conveniente considerarmos uma formulação simples do modelo PAR, o PAR(1), com termos de erro normais. Considere, portanto, série temporal $\{ y_t \}$ que evolui segundo

$$y_t = \phi_1(t) \cdot y_{t-1} + u_t$$

com $u_t \sim N(0, \tau^2(t))$ e tanto $\tau(t)$ como $\phi_1(t)$ sendo funções periódicas. A equação acima tem a seguinte interpretação: condicionado ao conhecimento de todo o passado a distribuição de y_t é normal com valor esperado $\phi_1(t) \cdot y_{t-1}$ e variância $\tau^2(t)$. Formalmente

$$y_t \mid y_{t-1}, y_{t-2}, \dots \sim N(\phi_1(t) \cdot y_{t-1}, \tau^2(t))$$

Especificações como acima são apropriadas em várias situações. Note, no entanto, que a valer a interpretação acima, vale também que, a distribuição “preditiva” é **simétrica** em torno de $\phi_1(t)y_{t-1}$, seu valor esperado condicionado ao conhecimento

²⁰ Isso implica (pode-se mostrar) que u_t é não correlacionado com y_{t-1}, y_{t-2}, \dots

²¹ Ou mesmo a modelos ARMA(p,q) periódicos.

²² Franses P H and R Paap (2004). *Periodic Time Series Models*. Oxford University Press, Oxford.

²³ Para análise no domínio da frequência consulte Hurd, H L and A Miamee(2007). *Periodically Correlated Random Sequences*, John Wiley & Sons Inc., New Jersey.

de todas as observações passadas. Ocorre que em várias séries climatológicas observa-se assimetrias importantes. Tais assimetrias às vezes podem ser justificadas pelas características físicas do fenômeno sendo mensurado.

Por exemplo, o fato de não existirem vazões negativas faz com que as distribuições de vazões mensais, condicionadas ou não, tenham caudas à direita mais longas do que suas caudas à esquerda. Nem sempre isso é um problema. Quando as vazões se localizam sempre bem acima de zero a limitação natural não é problemática. Entretanto quando as vazões se localizam próximas de zero a não consideração da limitação natural pode gerar previsões negativas.

O problema acima pode ser contornado de diversas maneiras. Uma delas, a implantada na modelagem da ENA em NEWAVE, postula distribuições estruturalmente assimétricas para os termos de erro U_t (e.g. *lognormais* deslocadas).

A estratégia de modelagem considerada em NEWAVE tem como vantagem o fato de que a distribuição de y_t condicionada aos valores passados pode ser escrita como a constante $\varphi_1(t)y_{t-1}$ adicionada à distribuição de u_t .

Ora a distribuição de U_t , por hipótese, independe de valores passados da série. Devido a esse fato é possível o uso de estimadores baseados nas equações de Yule-Walker ligeiramente modificadas. Essa simplicidade computacional é obtida às custas da hipótese embutida na equação acima: a distribuição condicionada de y_t só depende do passado de forma aditiva. Isso pode ser forte demais.

Uma outra possibilidade, mais sofisticada enquanto metodologia estatística, passa pelo uso de técnicas associadas a regressões quantílicas. Mais especificamente, considera a classe de modelos auto-regressivos quantílicos (QAR), desenvolvidos por Koenker.

Também aqui por simplicidade de exposição iremos nos restringir à consideração de modelo auto-regressivo quantílico bastante simples. Diremos que $\{y_t\}$ segue processo QAR(1) quando podemos escrever

$$y_t = \alpha(u_t) + \beta(u_t)y_{t-1}$$

Onde as U_t são iid $U(0,1)$ e $\alpha(\cdot)$ e $\beta(\cdot)$ são funções suaves. Algumas características dos modelos QAR

- i. AR(1) é caso particular de QAR(1) quando fazemos $\beta(u) = b = \text{constante}$ e $\alpha(u) = a + \Phi^{-1}(u)$, com $\Phi(\cdot)$ sendo a função distribuição da Normal Padrão
- ii. Uma vantagem: flexibilidade na forma da transição de y_{t-1} para y_t
- iii. Sob certas condições de regularidade

- a. $\alpha(.5) + \beta(.5)y_{t-1}$. é a mediana condicional de y_t
- b. $\alpha(.25) + \beta(.25)y_{t-1}$. é o primeiro quartil condicional y_t , etc.

Iremos considerar nesse projeto generalizações periódicas de processos QAR dadas por

$$y_t = \alpha_t(u_t) + \beta_t(u_t)y_{t-1}$$

Onde as U_t são iid $U(0,1)$ e $\alpha_t(\cdot)$ e $\beta_t(\cdot)$ são funções suaves, satisfazendo ainda $\alpha_t(\cdot) = \alpha_{t-12}(\cdot)$. e também $\beta_t(\cdot) = \beta_{t-12}(\cdot)$, para todo t . Ou seja permitimos que as funções $\alpha_t(\cdot)$ e $\beta_t(\cdot)$ que determinam medianas, quartis, decis, percentis condicionais (em suma a distribuição condicional das y_t) varie de forma periódica.

Iremos denominar essa classe de modelos de QAR-P (Auto-Regressivo Quantílico Periódico). Iremos denotar o modelo simples acima descrito por (um modelo) QAR-P de ordem 1 ou um modelo QAR-P(1). A generalização para ordens maiores que 1 é óbvia.

Devido à flexibilidade da classe acima os ganhos em generalidade podem ser bastante expressivos. Note, no entanto, que os modelos QAR-P têm como hipótese que os quantis condicionais assumem uma forma específica (linear) que certamente é, ainda, restritiva.

MODELAGEM DAS SÉRIES DE FATORES DE CAPACIDADE DAS USINAS

Antes da modelagem propriamente dita iremos investigar o caráter periódico das séries de Fatores de Capacidade das usinas. Na subseção anterior definimos os modelos auto-regressivos periódicos através de relações recursivas entre os valores da série original após padronização periódica. Mais especificamente, relações recursivas como

$$\tilde{y}_t = \phi_1(t) \cdot \tilde{y}_{t-1} + \phi_2(t) \cdot \tilde{y}_{t-2} + u_t$$

entre as

$$\tilde{y}_t = \frac{y_t - \nu_t}{\tau_t}$$

com $\nu_t := E y_t$ e $\tau_t := \sqrt{\text{Var}(y_t)}$.

É importante chamarmos a atenção para um fato implícito nos modelos PAR já definidos: a existência e invariância (periódica) de tanto τ_t quanto ν_t . É, portanto, natural iniciarmos nossa investigação da periodicidade pelas séries

padronizadas periodicamente. No gráfico 20 exibimos os fatores de capacidade da usina X padronizados periodicamente. Fica claro que boa parte da regularidade sazonal desapareceu. É difícil, ao menos visualmente, identificarmos algum comportamento sazonal.

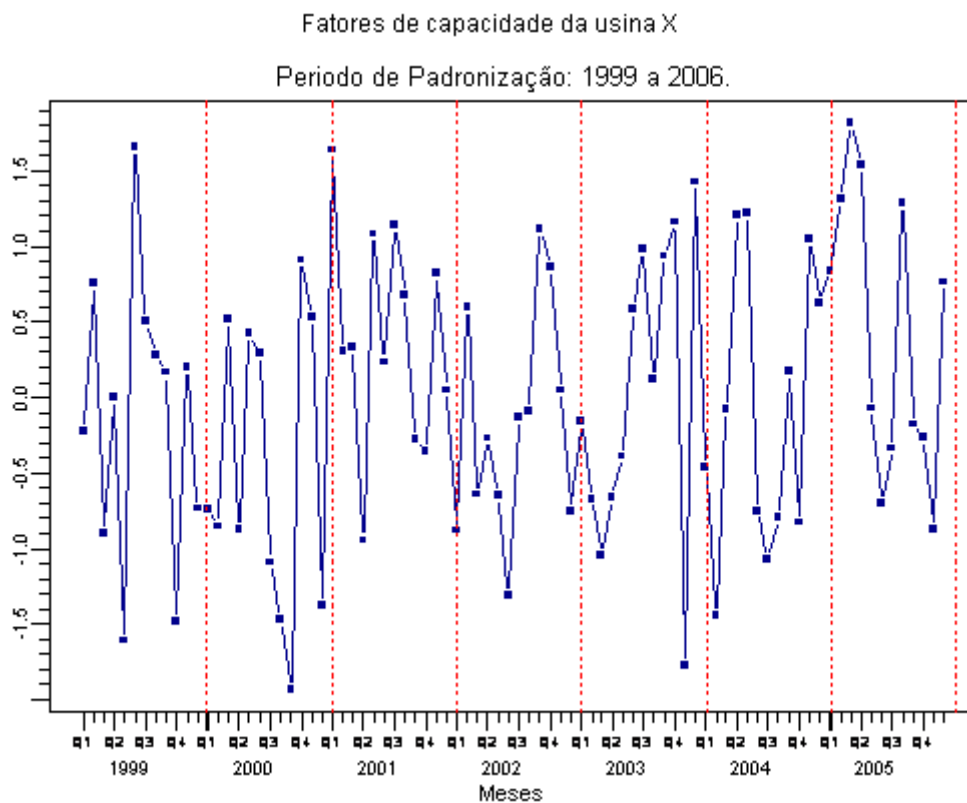


Gráfico 20: Evolução dos fatores de capacidade da usina X, padronizados periodicamente.

Grosso modo, a padronização periódica consiste, por exemplo, em subtrair de cada um dos Janeiros a média amostral dos Janeiros e em seguida dividir pelo desvio padrão dos Janeiros²⁴.

Registramos que, por ora, estão sendo considerados somente dados referentes ao período Jan-1999 a Dez-2005. Tanto para fins da montagem do gráfico como do cálculo de médias e desvios padrão. Voltamos a esse ponto mais adiante.

Na tabela 11 e no gráfico 21 investigamos a presença de autocorrelação serial na série transformada: tanto a sazonalidade como outros padrões de inércia (detectáveis por Ljung-Box) foram retirados.

²⁴ Para o cálculo do desvio padrão [dos Janeiros, por exemplo] foi necessário levar em conta a dependência serial [entre os janeiros]. Isso foi feito via ajuste de modelo auto-regressivo [aos Janeiros].

	Defasagens Consideradas			
	6	12	18	24
Valor da Estatística	2.54	8.48	12.93	18.22
p.value	0.86	0.74	0.79	0.79

Tabela 11: Testes de Ljung-Box com a usina X padronizada periodicamente: Hipótese Nula= Ausência de autocorrelação.

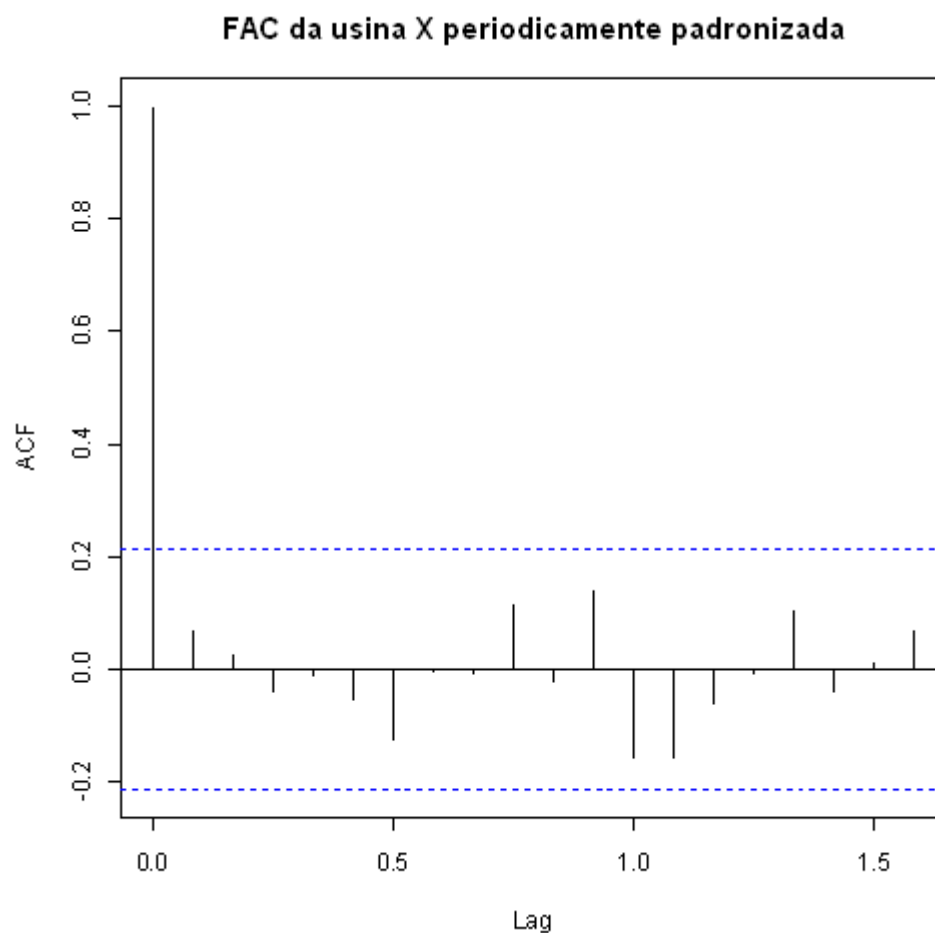


Gráfico 21: Função de autocorrelação da usina X (padronizada periodicamente). Período de cálculo: Jan-1999 a Dez-2005.

Problemas começam a ocorrer a partir de 2006. No gráfico 22 exibimos novamente a usina X padronizada periodicamente. A diferença é que agora o período utilizado para a padronização vai de Jan-2009 a Ago-2009. As retas verticais delimitam anos.

Aparentemente os anos de 2006 em diante apresentam padrão de evolução distinto daquele do período inicial, flutuando em torno de níveis mais baixos.

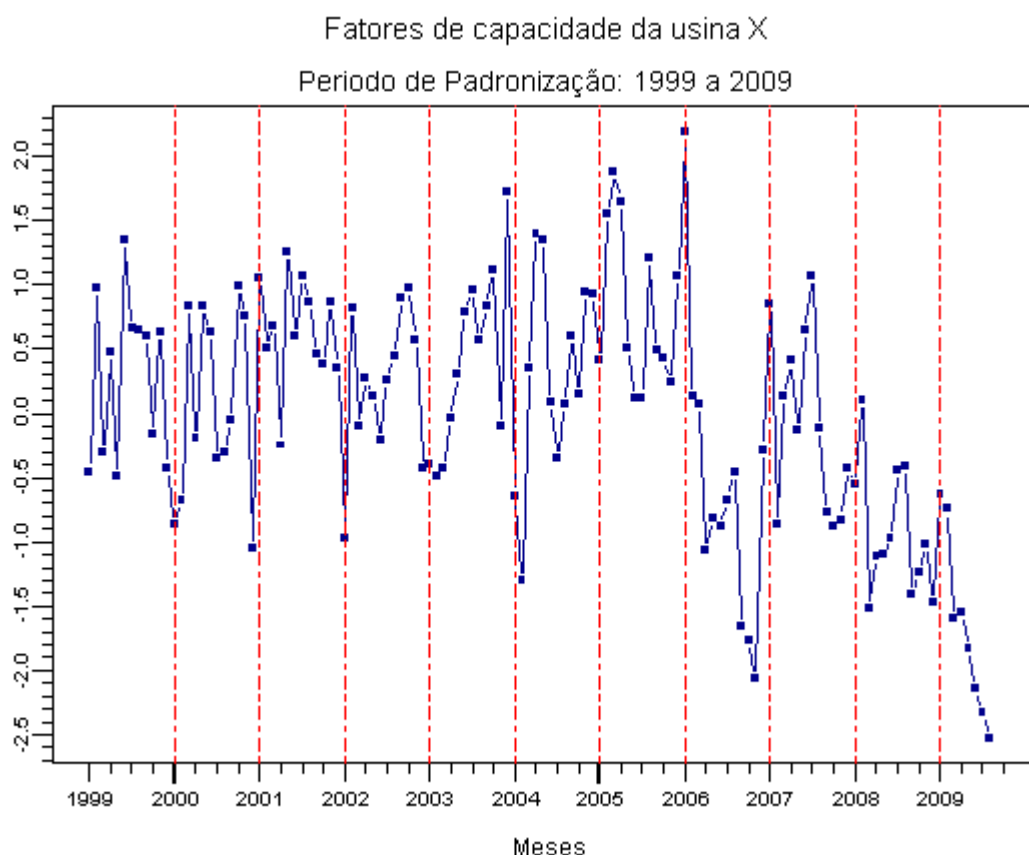


Gráfico 22: Fatores de Capacidade da usina X (padronizados periodicamente). Período de cálculo: Jan-1999 a Ago-2009.

O efeito deste novo período na dinâmica da série da usina X pode ser apreciado na função de auto-correlação (gráfico 23) estimada para o período todo. As duas funções de auto-correlação são distintas o suficiente para descartar o uso de modelos periódicos nos moldes aqui descritos (PAR).

É importante destacar que uma causa aparente para as mudanças nas funções de auto-correlação foi a mudança, para baixo, dos níveis dos fatores de capacidade. O ano de 2006, em particular, evidencia quão brusco foi esse movimento.

É interessante aqui checarmos a compatibilidade desse movimento brusco com outras fontes. Fazemos isso no gráfico 24. Lá exibimos a evolução conjunta das 3 usinas. Em linhas gerais (i) a usina Z domina a usina Y, (ii) a usina X flutua entre a usina Z e a usina Y e (iii) os fatores de capacidade mínimos anuais das 3 usinas ocorrem no mesmo mês e no mesmo patamar.

FAC da usina X padronizada periodicamente

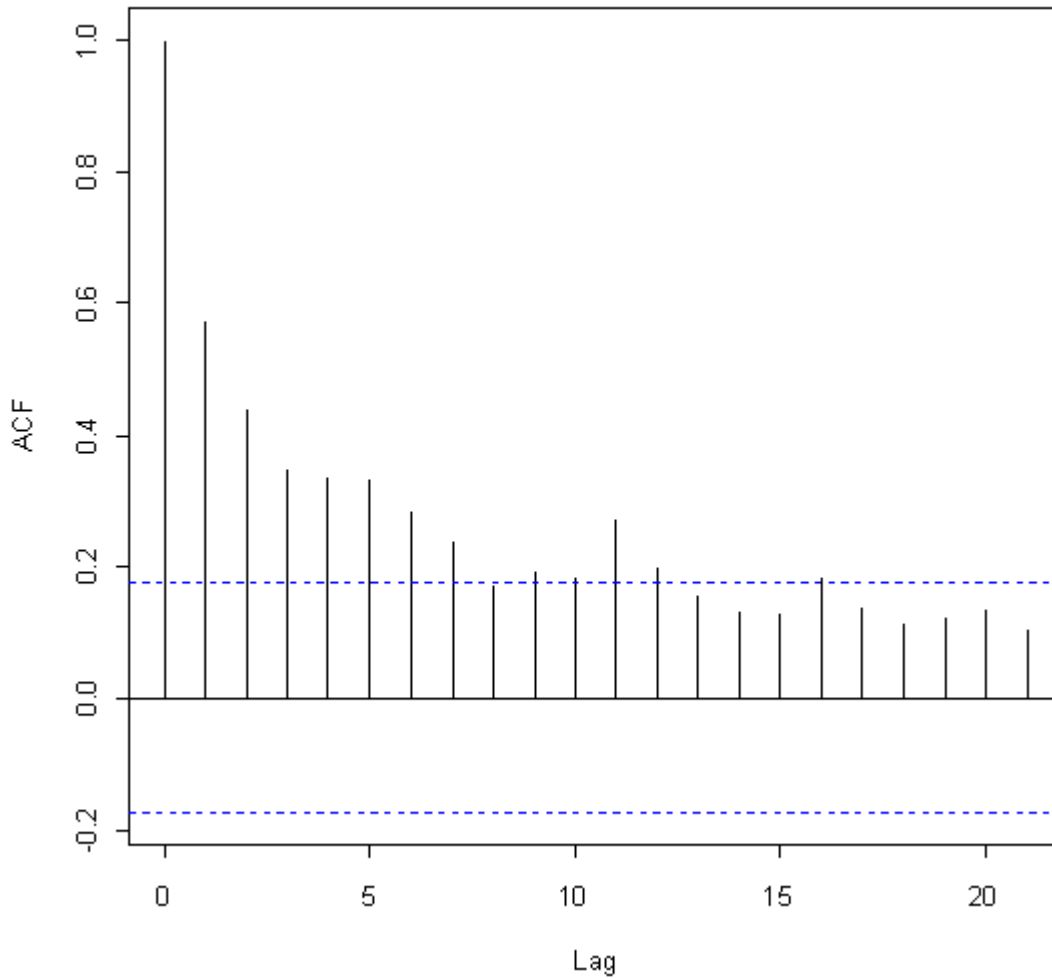


Gráfico 23: Função de autocorrelação da usina X (padronizada periodicamente). Período de cálculo: Jan-1999 a Ago-2009.

É possível, portanto, que, embora anômalo, o forte decréscimo nos fatores de capacidade da usina X em 2006 não necessariamente indique má qualidade dos dados. A conclusão acima se baseia na comparação das 3 usinas entre si.

No gráfico 25 comparamos com os dados da série agregada para o estado do Ceará [Camargo & Schubert]. Os níveis da série agregada não decrescem como na usina X. Novamente, esse fato, considerado isoladamente, não desqualifica os dados dessa usina. Deixa claro, no entanto, a magnitude do movimento ocorrido nos fatores de capacidade dessa usina.

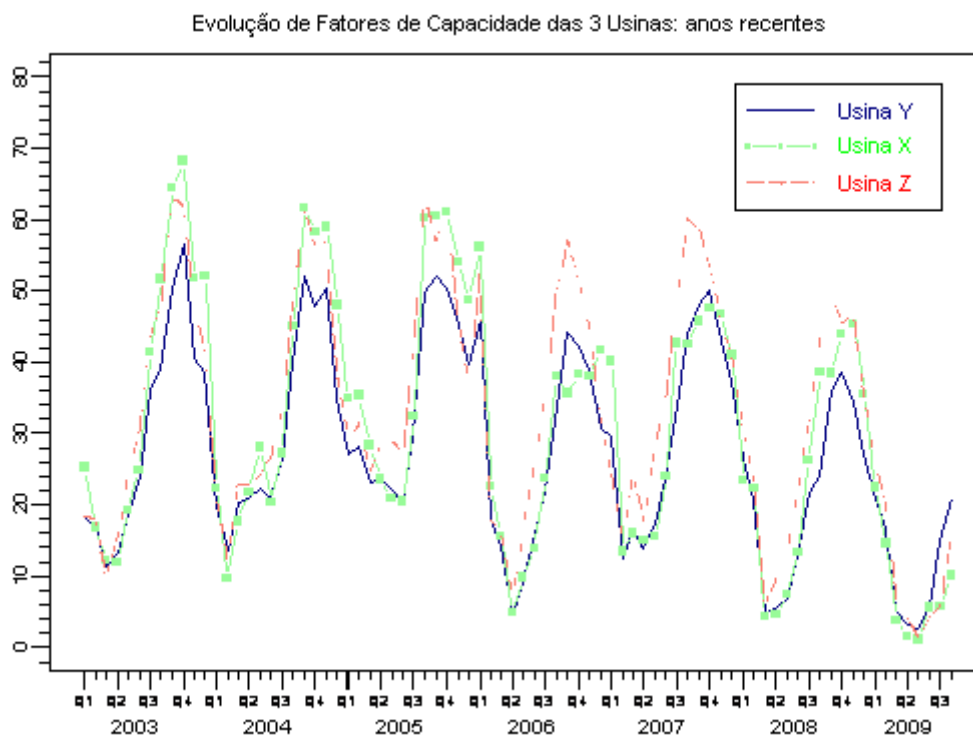


Gráfico 24: Fatores de capacidade das 3 usinas em anos recentes. Investigando a queda nos níveis da usina X em 2006.

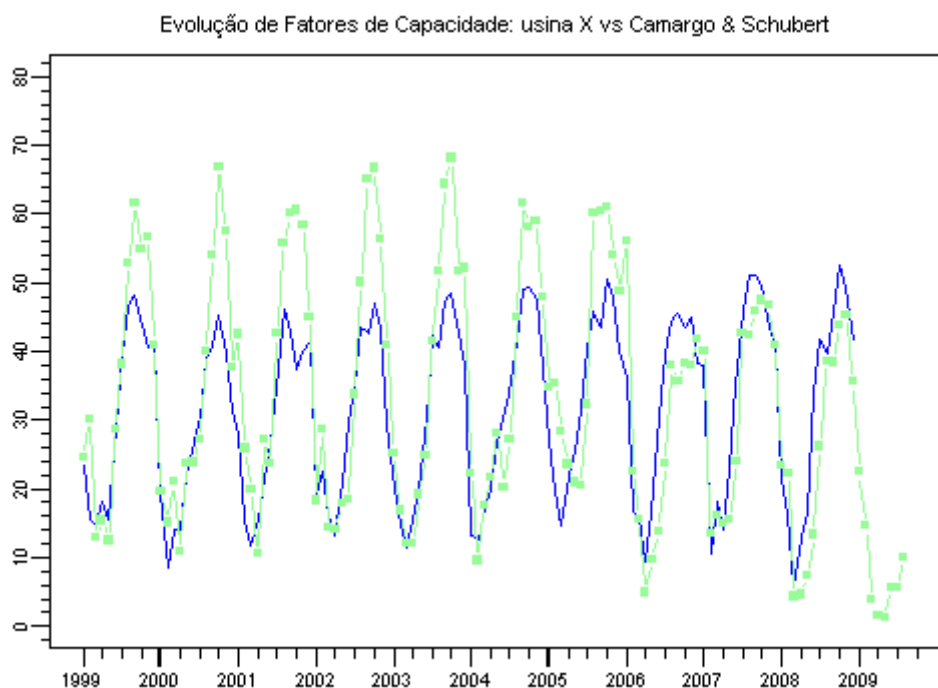


Gráfico 25: Fatores de capacidade da usina X e da série agregada.

Aparentemente, o movimento de queda geral dos fatores de capacidade observado na usina X, e, com alguma defasagem, também nas outras usinas, não ocorre na série agregada. Esse fato inviabiliza o uso da série agregada como base para a estimação

de movimentos²⁵ prevalentes nas séries das usinas, mas de difícil estimação unicamente a partir destas. É recomendável uma análise mais detalhada do processo de coleta dos dados da série agregada e das séries das usinas.

Os gráficos 26 e 27 trazem evidências adicionais acerca do comportamento de difícil modelagem exibido pelas séries das usinas. Lá exibimos gráficos equivalentes aos dos gráficos 20 e 22, para a usina Y. No gráfico 26 conduzimos a padronização periódica com dados até 2005. Observamos flutuações sem evidência aparente de anormalidade.

Fatores de Capacidade da usina Y. Período de padronização: 1999 a 2006

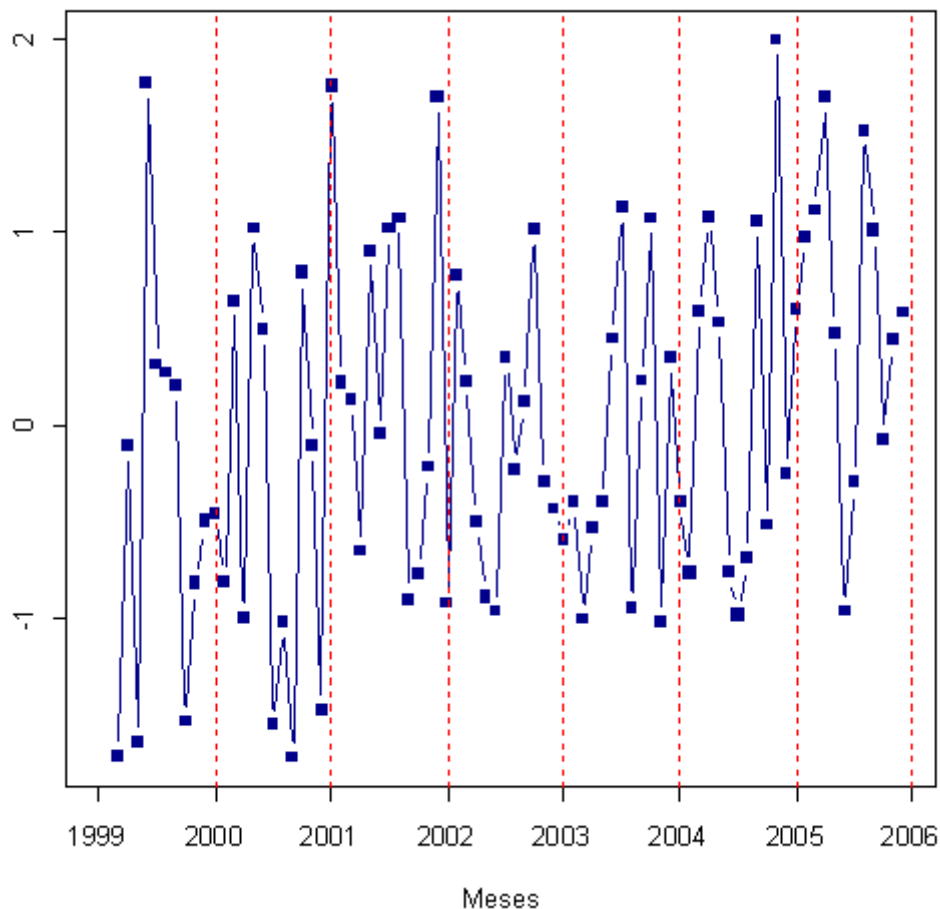


Gráfico 26: Fatores de Capacidade da usina Y (padronizados periodicamente). Período de cálculo: Jan-1999 a Dez-2005.

Por outro lado, no gráfico 27, as flutuações a partir de 2006 exibem padrão distinto do mostrado antes desse ano.

²⁵ Ciclos de baixa frequência, por exemplo.

Fatores de Capacidade da usina Y. Período de padronização: 1999 a 2009

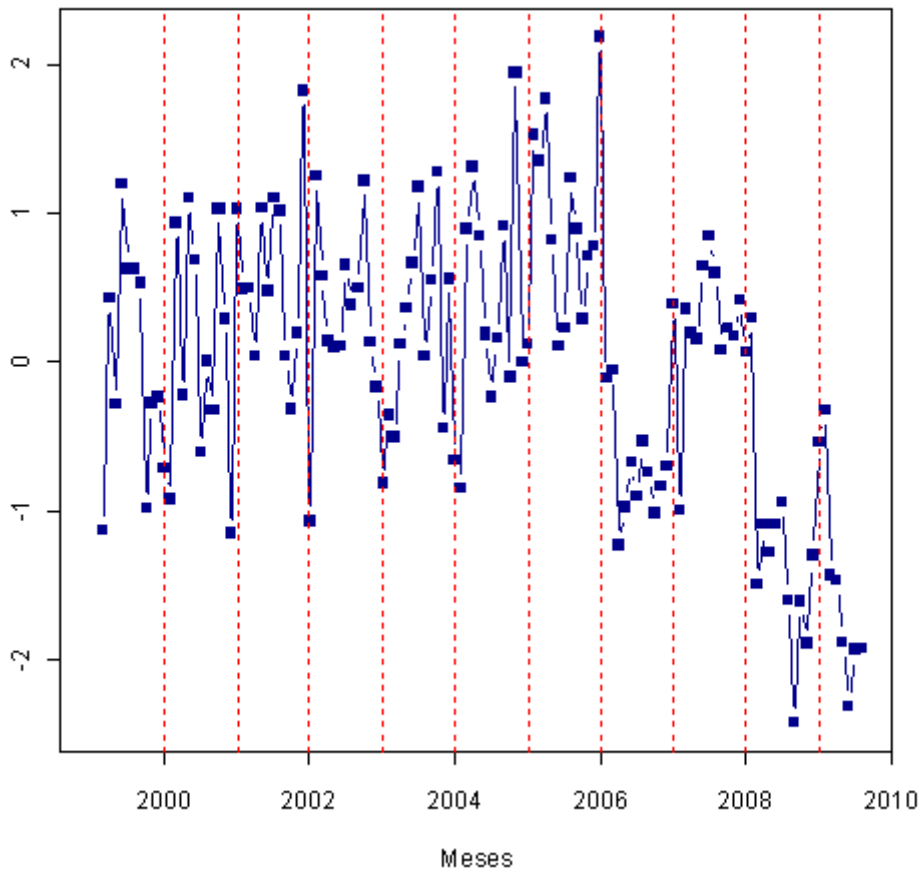


Gráfico 27: Fatores de Capacidade da usina Y (padronizados periodicamente). Período de cálculo: Jan-1999 a Ago-2009.

As considerações acima exibidas indicam que, ao menos para as usinas, os modelos auto-regressivos periódicos não são apropriados. Registramos que a principal evidência para isso é que a dinâmica pós 2005 é mais complexa do que antes desse ano.

Também, análises gráficas como a acima conduzidas apontam duas linhas de investigação: (i) análise dos processos de coleta de dados e (ii) a existência de múltiplos regimes estocásticos governando a evolução da energia eólica no estado do Ceará.

A linha (i) fica de fato como sugestão. Por outro lado é interessante considerarmos a possibilidade de múltiplos regimes. Uma análise inicial é ilustrada no gráfico 28. Ali plotamos os fatores de capacidade agregados (Camargo & Schubert) e os fatores da usina X. Ambos na forma de médias anuais efetuada após processo de padronização periódica.

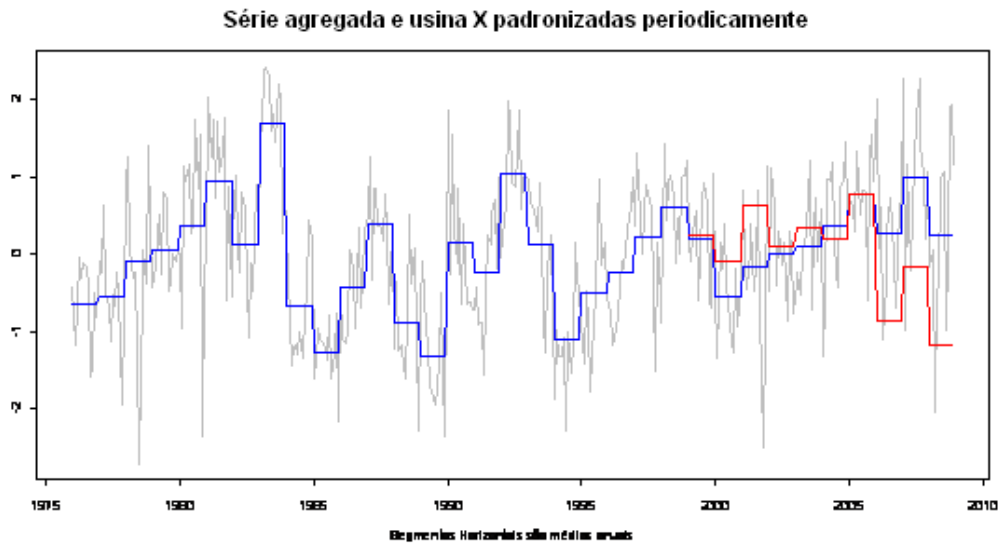


Gráfico 28: Fatores de Capacidade do Estado do Ceará e da usina X (padronizados periodicamente). Segmentos horizontais são médias anuais.

Os dados parecem de fato exibir regimes múltiplos. Num deles observamos acíves suaves ao longo de vários anos em seqüência. Num outro, declives acentuados em, no máximo, dois anos consecutivos.

Uma conseqüência importante do gráfico e análises acima é a potencial reconciliação entre dados das usinas e do estado. Uma estratégia interessante de modelagem é a consideração de modelos com trocas Markovianas de regime.

Por hora iremos seguir o caminho indicado na estratégia de modelagem II: os modelos QAR-P e PEAR. Já coletamos evidências que contra-indicam esses modelos para as séries das usinas. Iremos, portanto, nos restringir à série agregada.

Conduzimos exercício de modelagem utilizando os pacotes timsac e pear do R. As ordens auto-regressivas foram as indicadas pelo critério BIC. Após estimação simulamos a evolução de série com 1800 observações seguindo os correspondentes modelos estimados. No gráfico 29 comparamos as distribuições empíricas e estimadas segundo os dois modelos.

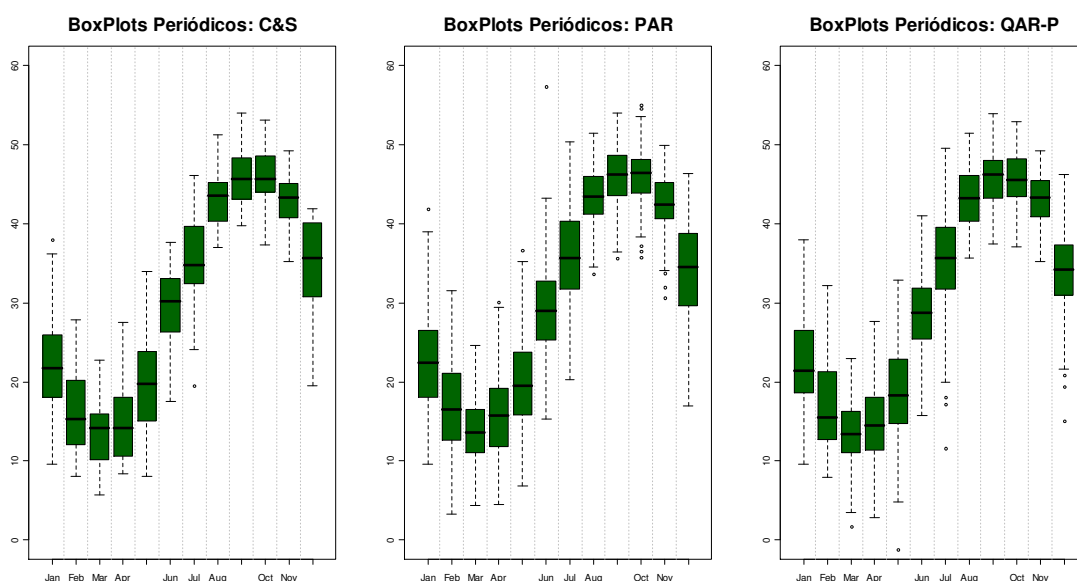


Gráfico 29: BoxPlots periódicos: Dados originais, série simulada segundo modelo PAR e segundo modelo QAR-P. A qualidade dos dois modelos é semelhante.

Fica claro que ambos parecem reproduzir de modo razoável a variabilidade dos dados da série agregada. Aparentemente o uso dos modelos QAR-P, mais sofisticados que os PAR não se justifica aqui.

Mês sendo modelado	Testes Portmanteau			
	Lags	p-valor	Lags	p-valor
Jan	4	0.96	8	0.97
Fev	4	0.62	8	0.31
Mar	3	0.44	7	0.51
Abr	3	0.59	7	0.29
Mai	4	0.54	8	0.21
Jun	4	0.17	8	0.51
Jul	4	0.41	8	0.09
Ago	4	0.22	8	0.08
Set	4	0.54	8	0.69
Out	4	0.46	8	0.57
Nov	4	0.14	8	0.06
Dez	2	0.19	6	0.31

Tabela 12: Testes *portmanteau* para a adequabilidade do modelo PAR à série agregada.

A adequabilidade do PAR pode ser checada via teste *portmanteau* (Ljung-Box adaptados ao caso periódico). A tabela 12 exhibe os p-valores correspondentes.

Uma característica desejável em qualquer modelo de séries temporais é a sua capacidade preditiva. Aqui ilustramos essa característica no gráfico 30.

Previsões Dinâmicas para série agregada: modelo QAR-P

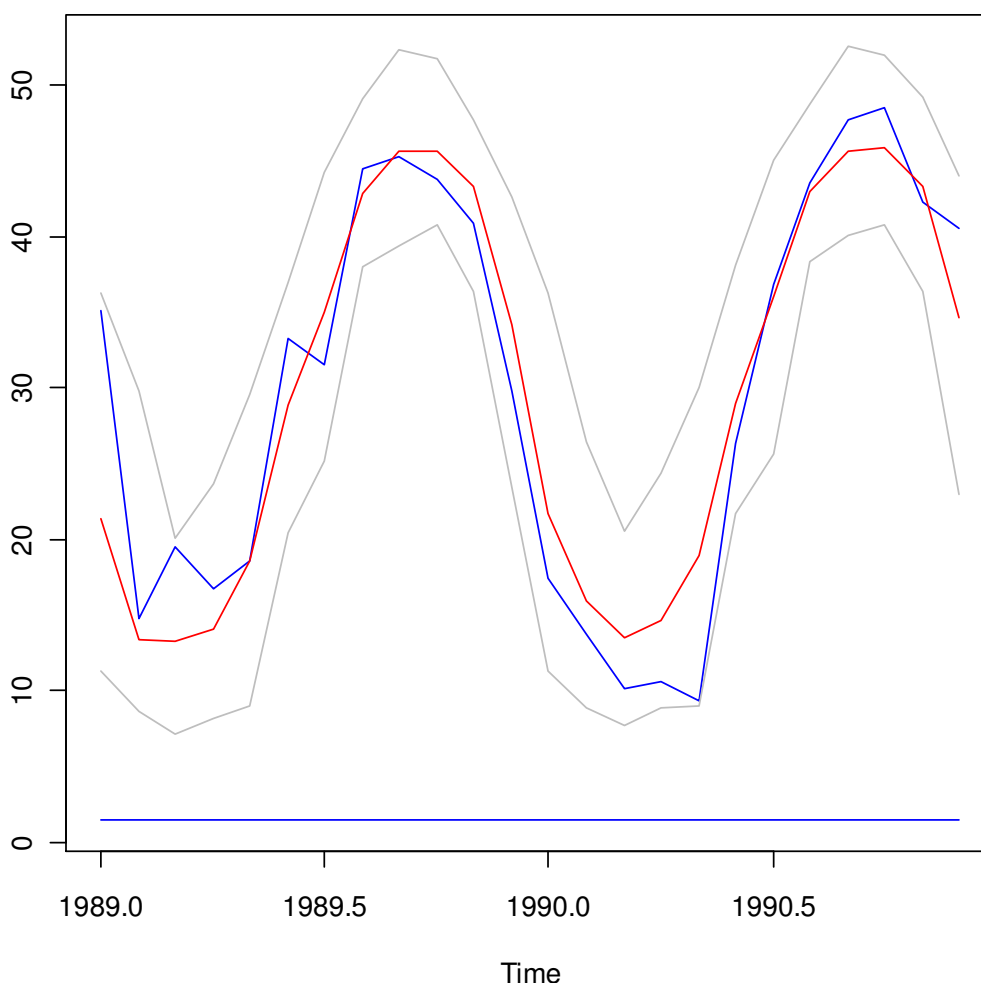


Gráfico 30: A capacidade preditiva avaliada num ano particularmente problemático: 1990. Os intervalos de confiança têm nível 90%

CONCLUSÕES

- i. Os dados de atividade eólica do Ceará apresentam dinâmicas complexas, de difícil modelagem por metodologias tradicionais.
- ii. A série agregada (Camargo & Schubert) indica a possível existência de múltiplos regimes de atividade eólica [alem das flutuações sazonais].
- iii. A modelagem isolada das séries das usinas sozinhas não deve ser conduzida. Isso porque não é factível extrairmos desses dados informações relativas aos regimes múltiplos. O uso de séries mais longas, como a da Camargo & Schubert, é fundamental.
- iv. É importante a avaliação da qualidade dos dados das séries agregadas.
- v. Os modelos PAR e QAR-P quando aplicados à série da Camargo & Schubert sem a consideração explícita de múltiplos regimes foram capazes de replicar a variabilidade da série original.

8. Considerações Finais

Nesta seção apresentamos os exercícios restantes de modelagem. Aqui buscamos seguir as indicações levantadas em seções anteriores. Mais especificamente, apresentamos aqui nossos esforços no sentido de:

- (a) considerar a existência de múltiplos regimes no processo de geração da atividade eólica no Ceará
- (b) modelar as mudanças segundo modelos de trocas Markovianas de Regime apropriados.

Adiantamos que os resultados ficaram aquém do esperado. Lembramos que a motivação para (a) e (b), acima apontados, teve por base a modelagem dos fatores de capacidade das usinas. Havíamos identificado que a dinâmica destas séries era complexa com, possivelmente, mais que um regime. Ocorre que dispomos de séries relativamente curtas para as usinas. Isto compromete quaisquer tentativas de modelagem mais sofisticada das usinas isoladamente.

Buscamos, então, lançar mão de recursos capazes de aproveitar eventuais dependências entre movimentos das usinas e da série agregada [Camargo & Schubert]. Para a série agregada dispomos de quantidade de observações bastante superior às das usinas.

Estimamos para a série agregada um modelo auto-regressivo periódico. Tanto seu ajuste como os testes usuais de adequabilidade se mostraram adequados, apesar da possibilidade da existência de múltiplos regimes.

Buscamos inicialmente estimar as trocas Markovianas de regime na presença do modelo auto-regressivo periódico. Isso gerou vários problemas de convergência dos algoritmos de estimação. Aparentemente, a história observada (numero de observações) não é suficiente para garantir estimativas com precisão razoável para os parâmetros auto-regressivos periódicos e para parâmetros subjacentes às trocas Markovianas de regimes.

Restringimos a classe de modelos. Passamos a considerar modelagens para sub-séries mensais e também para algumas agregações de interesse. O trabalho com as sub-séries foi o mais interessante no sentido de identificar consistentemente dois regimes ao longo dos 33 anos de dados.

Exibimos no gráfico 31 uma identificação típica dos dois regimes subjacentes. A observação correspondente ao tempo 1 é o ano de 1976. A última (tempo=33) se refere ao ano 2008. Frisamos que para algumas sub-séries mensais (um pouco menos da metade) não obtivemos convergência do algoritmo de estimação (função MSAR do módulo Finmetrics do pacote SPLUS).

O problema com a estrutura representada no gráfico 31 é que ela contrasta com a motivação (gráfica, confessadamente) para o aproveitamento da modelagem da série agregada para fins de uma melhor estimação da dinâmica das usinas.

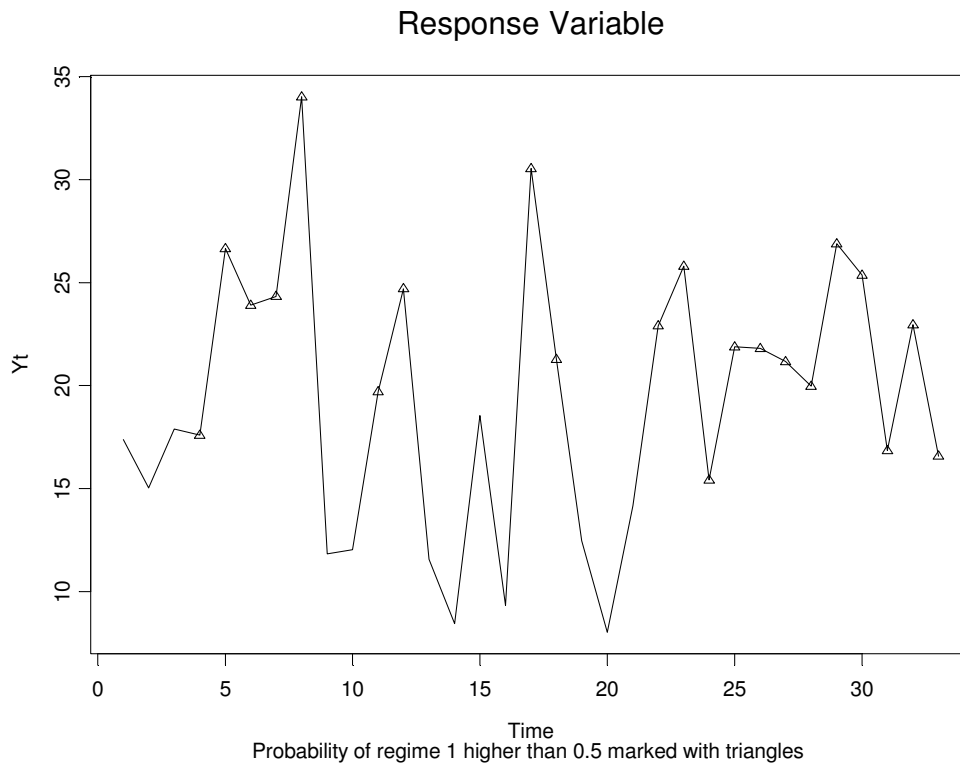


Gráfico 31: Os regimes para a série agregada: baixa mitigação de incerteza nos modelos das usinas.

Isto pode ser apreciado na figura 37: imaginávamos que nos anos mais recentes o regime prevalente fosse aquele onde a atividade eólica é mais baixa. Pois é exatamente o oposto que podemos inferir do gráfico 28.

Listamos a seguir algumas conclusões válidas pra o projeto como um todo.

CONCLUSÕES

1. Os dados de atividade eólica do Ceará apresentam dinâmicas complexas, de difícil modelagem por metodologias tradicionais.
2. A série agregada (Camargo & Schubert) é compatível com a existência de múltiplos regimes de atividade eólica [alem das flutuações sazonais]. Não é verdade, registramos, que a capacidade preditiva tenha melhorado substancialmente com a adoção de trocas Markovianas de Regimes.
3. A valer a existência de múltiplos regimes para a dinâmica das atividades eólicas nas empresas, a modelagem isolada das séries das usinas sozinhas não deve ser conduzida.
4. O uso da série da Camargo & Schubert para a explicitação / estimação dos regimes nas usinas não se mostrou razoável.
5. É importante a avaliação da qualidade dos dados das séries agregadas.
6. Os modelos PAR e QAR-P quando aplicados à série da Camargo & Schubert sem a consideração explícita de múltiplos regimes foram capazes de replicar a variabilidade da série original.