



**Indira Rodríguez Llanes**

***EMD-RQAtestes: nova abordagem para a decomposição de séries temporais em componentes estocásticas e determinísticas***

**Dissertação de Mestrado**

Dissertação apresentada como requisito parcial para Obtenção do grau de Mestre pelo Programa de Pós-Graduação em Engenharia de Produção do Departamento de Engenharia Industrial da PUC-Rio.

Orientador: Prof. Fernando Luiz Cyrino Oliveira

Rio de Janeiro  
Março de 2016



**Indira Rodríguez Llanes**

***EMD-RQA* testes: nova abordagem para a  
decomposição de séries temporais em  
componentes estocásticas e determinísticas**

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-Graduação em Engenharia de Produção do Departamento de Engenharia Industrial do Centro Técnico Científico da PUC-Rio. Aprovada pela Comissão Examinadora abaixo assinada.

**Prof. Fernando Luiz Cyrino Oliveira**

Orientador

Departamento de Engenharia Industrial PUC-Rio

**Prof. Reinaldo Castro Souza**

Departamento de Engenharia Elétrica PUC-Rio

**Prof. Plutarcho Maravilha Lourenco**

Instituto de Energia PUC-Rio

**Prof. Gheisa Roberta Telles Esteves**

Departamento de Engenharia Industrial PUC-Rio

**Prof. Marcio da Silveira Carvalho**

Coordenador Setorial do

Centro Técnico Científico – PUC-Rio

Rio de Janeiro, 3 de Março de 2016

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da autora e do orientador.

Indira Rodríguez Llanes

Graduou-se em Engenharia Industrial pelo Instituto Superior Politécnico José Antonio Echeverría (ISPJAE) de Cuba em 2010. Especialista em Gestão de Projetos Informáticos pela Universidad de las Ciencias Informáticas (UCI) em 2013, atuando principalmente na gestão da qualidade e na análise de investimentos em projetos de *software*.

Ficha Catalográfica

Rodríguez Llanes, Indira

EMD-RQAtestes : nova abordagem para a decomposição de séries temporais em componentes estocásticas e determinísticas / Indira Rodríguez Llanes ; orientador: Fernando Luiz Cyrino Oliveira. – 2016.

85 f. : il. (color.) ; 30 cm

Dissertação (mestrado)–Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Industrial, 2016.

Inclui bibliografia

1. Engenharia Industrial – Teses. 2. Série temporal. 3. Decomposição ou filtragem. 4. Componentes estocásticos. 5. Componentes determinísticos. 6. Previsão. I. Oliveira, Fernando Luiz Cyrino. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Industrial. III.

CDD: 658.5

Para meus pais, Nitza e José Manuel,  
pelo apoio e confiança infinitos.

## Agradecimentos

Agradeço a Deus por me acompanhar todos os dias, por ter guiado no transcurso da minha carreira e por ser minha fortaleza nos momentos de debilidade.

Aos meus pais José Manuel Rodríguez e Nitza de la Caridad pelo apoio incondicional em todo momento, pelos valores transmitidos, por educar-me e sobre tudo por ser um excelente exemplo a seguir.

A minha irmã por ser parte importante na minha vida e representar a unidade familiar. Por ser um exemplo de desenvolvimento profissional, por sua companhia e inesgotável ajuda.

A Yank por sua terna companhia, por apoiar-me e compartilhar meus logros, por sua paciência e amor.

Ao professor Fernando Luiz pelo apoio, confiança e por compartilhar seus conhecimentos. A meu amigo Noel por ajudar-me no desenvolvimento desta pesquisa e pelas boas ideias aportadas.

A minha amiga Lissett Daymaris por se preocupar pelo andamento desta dissertação desde a distância.

Aos amigos novos, cubanos e brasileiros, feitos no Brasil, com os quais compartilhei momentos inesquecíveis.

Aos professores do Departamento de Engenharia Industrial pelos ensinamentos adquiridos nesta etapa do mestrado.

A PUC e CAPES por permitir que eu pudesse me dedicar exclusivamente ao mestrado e por seu apoio financeiro.

Aos membros da Banca Examinadora, professores Reinaldo Castro e Plutarcho Maravilha pela sua contribuição e experiências aportadas.

A Brasil por suas boas-vindas.

Em fim, muito obrigada a todos que contribuíram de maneira direta ou indireta na conclusão desta dissertação.

## Resumo

Llanes, Indira Rodríguez; Oliveira, Fernando Luiz Cyrino. ***EMD-RQAtestes: nova abordagem para a decomposição de séries temporais em componentes estocásticas e determinísticas***. Rio de Janeiro, 2016. 85p. Dissertação de Mestrado – Departamento de Engenharia Industrial, Pontifícia Universidade Católica do Rio de Janeiro.

A análise de séries temporais tem sido de utilidade prática em cenários do mundo real, os quais apresentam uma mistura de componentes estocásticas e determinísticas. A separação dessas componentes tem grande importância no processo de previsão, pois permite modelar individualmente cada componente e conseqüentemente, obter um modelo híbrido de maior acurácia. Com base nisso, nesta dissertação se propõe uma nova abordagem para a decomposição de séries temporais em componentes estocásticas e determinísticas que usa o método de filtragem *Empirical Mode Decomposition* (EMD) para separar a série em funções chamadas *Intrinsic Mode Functions* (IMFs) e posteriormente, a cada uma delas são aplicados três testes de hipótese para conhecer a sua natureza. Estes testes: (i) número médio de pontos por linha ( $\bar{L}$ ), (ii) porcentagem de determinismo (% *DET*) e (iii) comprimento médio de linha (*ALL*) são baseados no *Recurrence Quantification Analysis* (RQA), especificamente, na análise da matriz de recorrência do *Recurrence Plot* (RP). No final, da soma dos IMFs determinísticos obtém-se a componente determinística e da soma dos IMFs estocásticos se obtém a componente estocástica. A principal vantagem da abordagem proposta é a sua capacidade para filtrar as séries independentemente das suas características de linearidade e estacionariedade. Por outro lado, a dissertação também apresenta uma modelagem híbrida que inclui a previsão individual das componentes determinísticas e estocásticas usando os modelos Polinomial e ARIMA respectivamente. Nos experimentos, a modelagem híbrida mostrou melhores resultados de previsão que os obtidos nas mesmas séries sem decompor.

## Palavras-chave

Série temporal; decomposição ou filtragem; componentes estocásticos; componentes determinísticos; previsão.

## Abstract

Llanes, Indira Rodríguez; Oliveira, Fernando Luiz Cyrino (Advisor). ***EMD-RQAtests: new approach for decomposition of time series in stochastic and deterministic components***. Rio de Janeiro, 2016. 85p. MSc. Dissertation – Departamento de Engenharia Industrial, Pontifícia Universidade Católica do Rio de Janeiro.

Time series analysis has been useful in real-world scenarios, which have a mixture of stochastic and deterministic components. The separation of the components of a time series influenced by the additive noise is very important in the forecasting process, because it allows model each component individually and consequently to get a hybrid model more accurate. Based on it, this thesis proposes a new approach to decompose the time series in stochastic and deterministic components using the Empirical Mode Decomposition (EMD) filtering method to separate the series in functions called Intrinsic Mode Functions (IMFs) and later, three hypothesis tests are applied over each IMF to know its nature. The tests are: (i) the average number of dots per line (L), (ii) the percentage of determinism (% DET), and (iii) the average line length (ALL) are based on *Recurrence Quantification Analysis* (RQA), specifically in the analysis of *Recurrence Plot* (RP) matrix. At the end, from the sum of the deterministic IMFs, the total deterministic component is obtained. Similarly, from the sum of stochastic IMFs, the total stochastic component is obtained. The main advantage of the proposed method is its ability to filter the series independently of the characteristics of seasonality and linearity. Furthermore, the dissertation also provides a hybrid modeling that includes individual prediction of deterministic and stochastic components using polynomial and ARIMA models respectively. During the experiments, the hybrid modeling showed better forecasting results than the obtained with the same series without previous decomposing.

## Keywords

Time series; decomposition or filter; stochastic components; deterministic components; forecast.

## Sumário

1	Introdução.....	13
1.1	Contexto geral.....	13
1.2	Relevância do tema, motivação e objetivos.....	13
1.3	Estrutura da dissertação .....	15
2	Revisão sistemática da literatura .....	16
2.1	Introdução .....	16
2.2	Metodologia .....	18
2.3	Análise descritiva.....	19
2.3.1	Análise quantitativa .....	19
2.3.2	Análise qualitativa .....	21
2.4	Síntese da revisão da literatura .....	29
2.4.1	Métodos de decomposição.....	30
2.4.1.1	EMD como método geral de decomposição .....	32
2.4.2	Técnicas para detectar determinismo em séries temporais.....	34
2.5	Conclusões Parciais .....	38
3	Abordagem proposta .....	39
3.1	Abordagem proposta EMD-RQA testes .....	41
3.2	Conclusões parciais.....	44
4	Avaliação da abordagem proposta. Resultados.....	46
4.1	Conjunto de dados .....	46
4.2	Modelos preditivos .....	48
4.3	Modelagem .....	50
4.3.1	Modelagem da série de treino de Consumo de Energia Elétrica do Brasil.....	51
4.3.2	Modelagem da série de treino UKDriverDeaths .....	53
4.3.3	Modelagem da série de treino ukcars .....	56
4.3.4	Modelagem da série de treino uselec .....	59
5	Conclusões e sugestões para trabalhos futuros .....	62
5.1	Conclusões .....	62
5.2	Sugestões para Trabalhos Futuros .....	64
6	Referências bibliográficas .....	65
7	Apêndices .....	70

## Lista de figuras

Figura 2-1 Exemplo de série de tempo discreta .....	16
Figura 2-2 Número de publicações por ano .....	19
Figura 2-3 Processo de filtragem.....	33
Figura 2-4 Decomposição via EMD da série de produção física de alimentos de Brasil.....	34
Figura 3-1 Abordagem proposta para decompor séries temporais em componentes estocásticas e determinísticas.....	45
Figura 4-1 Série temporal Consumo de Energia Elétrica do Brasil .....	46
Figura 4-2 Série temporal UKDriverDeaths.....	47
Figura 4-3 Série temporal ukcars .....	47
Figura 4-4 Série temporal uselec .....	48
Figura 4-5 Processo de modelagem das séries .....	49
Figura 4-6 Componente estocástica da série Consumo de Energia Elétrica do Brasil.....	51
Figura 4-7 Componente determinística da série Consumo de Energia Elétrica do Brasil.....	51
Figura 4-8 Ajuste da componente determinística da série de treino Consumo de Energia Elétrica do Brasil.....	52
Figura 4-9 Previsão híbrida da série de treino Consumo de Energia Elétrica do Brasil.....	52
Figura 4-10 Previsão simples da série de treino Consumo de Energia Elétrica do Brasil.....	53
Figura 4-11 Componente estocástica da série de treino UKDriverDeaths.....	54
Figura 4-12 Componente determinística da série de treino UKDriverDeaths.....	54
Figura 4-13 Ajuste da componente determinística da série de treino UKDriverDeaths.....	55
Figura 4-14 Previsão híbrida da série de treino UKDriverDeaths .....	55
Figura 4-15 Previsão simples da série de treino UKDriverDeaths .....	56
Figura 4-16 Componente estocástica da série de treino ukcars.....	57
Figura 4-17 Componente determinística da série de treino ukcars .....	57

Figura 4-18 Ajuste da componente determinística da série de treino ukcars .....	57
Figura 4-19 Previsão híbrida da série de treino ukcars .....	58
Figura 4-20 Previsão simples da série de treino ukcars.....	58
Figura 4-21 Componente estocástica da série de treino uselec .....	59
Figura 4-22 Componente determinística da série de treino uselec .....	59
Figura 4-23 Ajuste da componente determinística da série de treino uselec.....	60
Figura 4-24 Previsão híbrida da série de treino uselec .....	60
Figura 4-25 Previsão simples da série de treino uselec .....	60

## Lista de Tabelas

Tabela 2-1 Artigos por número de citações e recurso .....	20
Tabela 2-2 Autores com maior número de publicações .....	21
Tabela 2-3 Análise descritiva dos artigos.....	22
Tabela 2-4 Limitações das técnicas de decomposição .....	31
Tabela 4-1 Estatísticas residuais dos modelos aplicados à série de treino Consumo de Energia Elétrica do Brasil.....	53
Tabela 4-2 Estatísticas residuais dos modelos aplicados à série de treino UKDriverDeaths	56
Tabela 4-3 Estatísticas residuais dos modelos aplicados à série de treino ukcars.....	58
Tabela 4-4 Estatísticas residuais dos modelos aplicados à série de treino uselec .....	61

*La ciencia será siempre una búsqueda, jamás un descubrimiento real. Es un viaje, nunca una llegada.*

*Karl Raiumd Popper.*

# 1 Introdução

## 1.1 Contexto geral

O estudo de séries temporais tem-se convertido em uma linha de pesquisa muito difundida. A análise e modelagem têm sido estudadas para entender as relações entre as observações no tempo, sendo de aplicação prática em áreas como a biologia, a engenharia, a energia, as vendas, as comunicações, etc.

Segundo (SHUMWAY; STOFFER, 2013), a análise de uma série temporal baseada no conhecimento dos seus componentes e na interação entre eles, permite modelar o comportamento do sistema de forma eficaz. Atualmente, existem duas abordagens na análise das séries temporais. A abordagem, maiormente utilizada é a análise clássica, que se baseia na suposição de que os valores que toma a variável é consequência das componentes de tendência, sazonalidade e ruído, cuja ação conjunta resulta nos valores medidos. A outra abordagem, menos aprofundada na literatura, se baseia na análise das componentes determinísticas e estocásticas.

A análise das componentes estocásticas e determinísticas imersas em uma série de tempo permite isolar o conjunto de dados que são influenciados fortemente por processos aleatórios e que podem variar significativamente os resultados da modelagem. Segundo (ESPASA; CANCELO, 1996), a componente irregular ou estocástica são oscilações não sistemáticas que em geral afetam a série no momento em que acontecem. A introdução de uma componente estocástica em um sistema pode estar influenciada pela existência de processos puramente estocásticos, erros aleatórios na medição das variáveis, utilização de geradores aleatórios de séries, influências externas, entre outros. Dai a importância de trabalhar as séries influenciadas por ruído através da sua decomposição em componentes estocásticas e determinísticas e modelar independentemente cada componente com o objetivo de obter um modelo híbrido com maior acurácia, conforme proposto por (RIOS; DE MELLO, 2013).

## 1.2 Relevância do tema, motivação e objetivos.

A análise de séries temporais influenciadas pelo ruído aleatório torna-se de grande utilidade na análise dos sistemas reais, os quais apresentam uma mistura de componentes estocásticos e determinísticos. Estes sistemas reais estão atraindo a atenção dos pesquisadores na intenção de resolver problemas pontuais e obter melhores

resultados na modelagem devido a que a eliminação de qualquer componente, assim como a suposição ou simplificação de certos comportamentos tende a depreciar os resultados (HAN; LIU, 2009).

Diante desse contexto a busca de um método capaz de separar fielmente as componentes estocásticas e determinísticas, assume uma importância vital, servindo de base para a obtenção de previsões mais ajustadas. Neste sentido, a fim de garantir maior precisão e robustez na modelagem, algumas abordagens têm sido criadas e implementadas. Estas, geralmente se baseiam em duas etapas: a decomposição ou filtragem e a análise dos componentes obtidos na primeira etapa. Dentre das abordagens estudadas se podem citar: *Empirical Mode Decomposition – Recurrence Plot* (EMD-RP) definida em (RIOS; DE MELLO, 2013), *Empirical Mode Decomposition – Mutual Information* (EMD-MI) definida em (RIOS *et al.*, 2015), *Wavelet filtering* e algoritmos evolutivos em (MINERVA, 2010), entre outras. Não obstante, essas abordagens apresentam suas limitações, por exemplo a abordagem EMD-RP requer, logo do cálculo do determinismo para cada IMF obtido via EMD, definir um valor de corte para classificar os IMFs em estocásticos ou determinísticos. Nesse trabalho, o valor de corte foi definido visualmente, isto pode levar à introdução de erros na separação correta das componentes estocásticas e determinísticas.

Por tanto, se apresenta como objetivo deste trabalho o desenho de uma abordagem capaz de isolar eficientemente as componentes estocásticas das componentes determinísticas presentes em uma série temporal influenciada pelo ruído. Uma vez obtido, espera-se que a combinação dos modelos individuais ajustados aos componentes garanta um modelo híbrido com maior acurácia e por tanto melhores resultados nas previsões.

Com tudo, esta dissertação busca responder as seguintes questões primárias:

- i. Quais são as técnicas e/ou métodos de filtragem de séries temporais?
- ii. Quais são os métodos de análise de componentes que permitem classificá-los em estocásticos ou determinísticos?

Como questão secundária da pesquisa tem-se:

- i. Quais são os modelos mais adequados para ajustar à componente estocástica e à componente determinística independentemente?

### 1.3 Estrutura da dissertação

A dissertação está organizada em cinco capítulos. A seguir, no capítulo 2, se apresenta uma revisão sistemática da literatura sobre métodos de decomposição ou filtragem de séries temporais em componentes estocásticas e determinísticas. A revisão inclui o processo de seleção de artigos, leitura, análise descritiva, assim como uma breve exposição dos principais métodos encontrados na literatura.

Na sequência, o capítulo 3 expõe detalhadamente o método proposto.

No capítulo 4 são descritos e interpretados os experimentos realizados sobre um conjunto de séries sintéticas e séries reais para validar o método proposto. Para tais propósitos, verificou-se se a abordagem proposta gerava resultados compatíveis com conhecimentos prévios e aprofundados, existentes na literatura. Posteriormente, com o objetivo de ilustrar seu potencial na decomposição das séries em componentes estocásticas e determinísticas, ela é aplicada sobre quatro séries reais, três delas disponíveis no R (R. CORE TEAM, 2014) e a outra no site da Empresa de Pesquisa Energética do Brasil (EPE).

Por último, no capítulo 5 conclui a dissertação com as principais considerações sobre os resultados obtidos nas aplicações da abordagem proposta.

## 2 Revisão sistemática da literatura

### 2.1 Introdução

As séries de tempo se definem como uma sequência de observações, usualmente ordenadas cronologicamente no tempo (MORETTIN; TOLOI, 2006). Apresentam características particulares que permitem que possam ser classificadas segundo o intervalo entre as observações, número de variáveis coletadas, entre outros critérios.

Atendendo ao intervalo entre as observações, as séries de tempo são ditas discretas quando as observações são feitas em tempos específicos, geralmente equiespaçados. No entanto, uma série de tempo contínua é aquela cujas observações são coletadas a cada instante de tempo. Na figura 2-1 e 2-2 são representadas as séries temporais discretas e contínuas respectivamente.

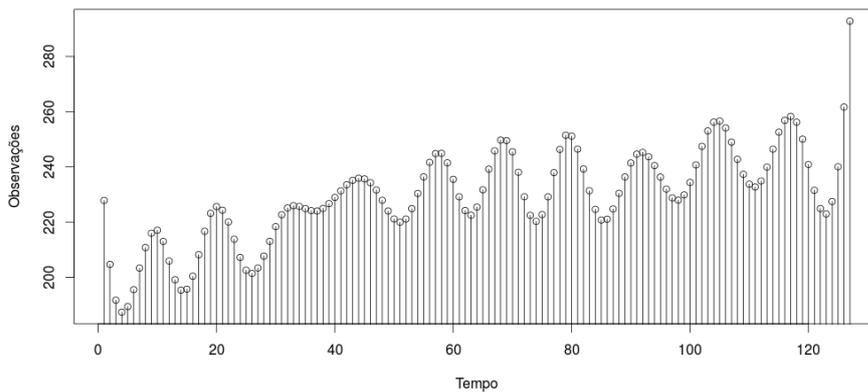


Figura 2-1 Exemplo de série de tempo discreta

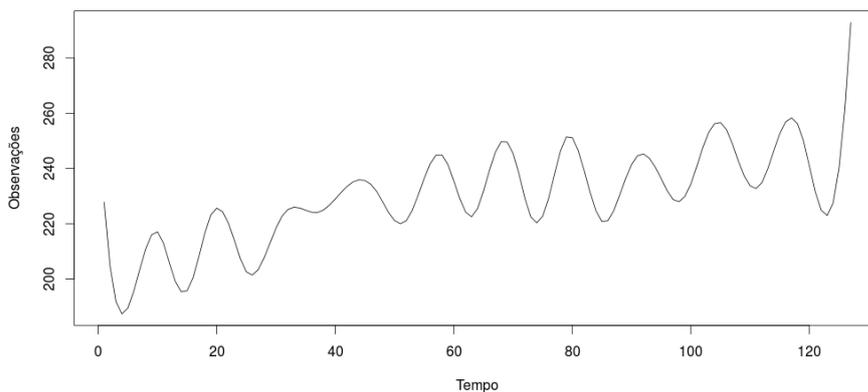


Figura 2-2 Exemplo de série de tempo contínua

Por outro lado, uma série é dita univariada se somente é controlada uma variável, pelo contrario uma serie é multivariada se são controladas mais de uma variável no mesmo instante.

Existem outras características comumente estudadas na análise de séries de tempo. Elas derivam-se da presença dos componentes embebidos no sinal, que podem ser componente linear, estacionário, tendência, ciclo, sazonal, determinístico e estocástico. Neste sentido um grande número de trabalhos tem sido realizados para separar as séries em componentes de tendência, sazonalidade e ruído. Na equação 2.1 mostra estes componentes, onde  $x(t)$  representa a série de tempo,  $T(t)$  a componente de tendência,  $S(t)$  a sazonalidade e  $\varepsilon(t)$  a componente estocástica.

$$x(t) = T(t) + S(t) + \varepsilon(t) \quad (2.1)$$

Outras investigações se focam na separação da série em componentes estáveis ou determinísticos e estocásticos ou componentes puramente aleatórios (SOARES; MEDEIROS, 2008). Uma série estocástica é um conjunto de observações e relações aleatórias que seguem funções de probabilidade e podem mudar com o tempo; geralmente envolve uma função matemática mais o termo aleatório. Por outra parte, uma série determinística é aquela que depende estritamente de observações passadas produzidas por alguma função matemática.

Existem também outros estudos baseados nas características de estacionariedade e linearidade. Diz-se que uma série de tempo é estacionaria quando suas observações estão em um estado particular de equilíbrio estatístico, no qual suas propriedades não são afetadas pela mudança na origem de tempo. Isto significa que, a distribuição de probabilidade conjunta segue sendo a mesma embora o conjunto de observações seja deslocado para frente ou para trás (REINSEL; BOX; JENKINS, 1994). Formalmente, uma série de tempo  $y_t$  é estacionaria se:

- a)  $E(y_t) = \mu_t; \forall t$
- b)  $Var(y_t) = \sigma^2; \forall t$
- c)  $Cov(y_t; y_{t+k}) = Cov(y_{t+m}; y_{t+k+m}); \forall m$

Uma serie é dita lineal quando suas observações estão compostas por combinações lineares de ocorrências passadas e ruído.

Mesmo seja muito importante a análise de todas as características das séries de tempo no processo de modelagem, este trabalho vai se focar nas características determinísticas e estocásticas presentes nas séries de tempo.

A decomposição de uma série de tempo em componentes estocásticos e determinísticos favorece o isolamento das “novidades” do esperado, sendo de relevante importância para trabalhos posteriores como a análise e previsão da série. No entanto, uma das principais dificuldades durante a decomposição de séries de tempo é delimitar

os componentes estocásticos e determinísticos (MINERVA, 2010). Com o objetivo de enfrentar esse problema algumas abordagens têm sido propostas na literatura, mas o desenvolvimento da decomposição de séries temporais para separar componentes determinísticos e estocásticos ainda é um problema em aberto (RIOS; DE MELLO, 2012). Neste contexto, resulta importante sintetizar o estado da arte, assim como identificar as abordagens de decomposição mais robustas que possam contribuir significativamente na obtenção de previsões mais exatas. Nas seções seguintes é apresentada uma revisão sistemática da literatura que prevê a busca ampla, planejada e bem estruturada dos trabalhos científicos relacionados ao objetivo principal da dissertação, assim como a identificação dos métodos de decomposição e técnicas de análise da natureza das séries, a frequência das publicações, os autores mais relevantes e outras análises.

## 2.2 Metodologia

A metodologia adotada para a realização da revisão da literatura foi baseada em (THOME *et al.*, 2012). Ela consta de cinco passos: (1) Seleção das bases de dados bibliográficas, (2) identificação das palavras chaves, (3) definição dos critérios para exclusão dos artigos, (4) leitura dos resumos e (5) revisão do texto completo.

Foram selecionadas três bases de dados bibliográficas: *IEEE Xplore*, *Web of Science* e *ScienceDirect*, levando em consideração a abrangência, o número de artigos publicados relacionados com o estudo e análise de séries temporais e a sua comum utilização pela comunidade científica. A pesquisa nessas bases de dados nos permite o acesso a artigos de revistas científicas que abordam a temática de séries temporais, além das vantagens que fornecem na localização e filtragem de artigos por palavras chaves, tipo de documento, idioma, autores, etc, podendo captar os *review* e artigos em língua inglesa. As palavras chaves foram definidas levando em consideração a sua correspondência com o tema de estudo e garantindo que os resultados não fossem limitados. De acordo com (RIOS; DE MELLO, 2012) foi preciso introduzir os termos “separação” e “filtro” como sinônimo de decomposição devido a que esses termos são usados com frequência para referir-se às técnicas que decompõem séries de tempo em componentes estocásticos e determinísticos. Finalmente as consultas foram configuradas usando os critérios de busca seguintes: “*time series*”, “*decomposition OR filter OR separation*”, “*deterministic and stochastic*”.

Com relação aos critérios de exclusão, foram eliminados do estudo aqueles artigos que: (1) não realizam separação da série em componentes estocásticos e determinísticos, (2) não se focam na modelagem da série de tempo, (3) não apresentam uma correta revisão da literatura, (4) não definem corretamente o modelo de decomposição proposto e (5) o modelo proposto é muito específico para uma série de tempo determinada. Depois dessa filtragem, foi revisada detalhadamente a bibliografia dos trabalhos resultantes, com o intuito de incorporar novos trabalhos ao estudo.

A pesquisa retornou 226 artigos, pertencendo 133 à *Web of Science*, 64 à *IEEE Xplore* e 127 à *Science Direct*. Depois de ler os resumos dos mesmos, os artigos duplicados e fora da abrangência do estudo foram eliminados, resultando 78 artigos para leitura completa. Finalmente foram aceitos 30 artigos.

### 2.3 Análise descritiva

A análise descritiva é apresentada em duas fases: análise quantitativa e análise qualitativa de cada um dos artigos aceitos.

#### 2.3.1 Análise quantitativa

Depois de julgar cada um dos 30 artigos finalmente aceitos, foi realizada uma análise com o propósito de conhecer o comportamento das publicações por ano. A figura 2.2 mostra o número de artigos publicados por ano. Embora não se observe um comportamento estável, podemos dizer que a partir do ano 2006 o número de publicações aumentou gradualmente, sendo o 2009 o ano de maior número de publicações. De acordo com a figura 2-3 percebe-se também que a partir do ano 2006 com exceção do ano 2008 o comportamento do número de publicações aumentou o qual reflete que a separação de componentes estocásticos e determinísticos tem recebido uma maior atenção dos pesquisadores.

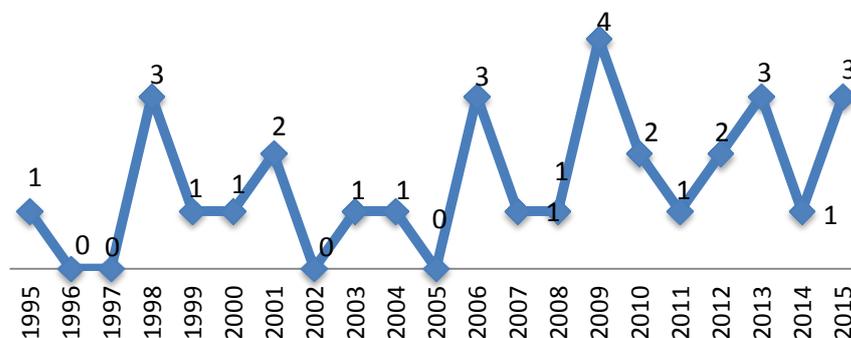


Figura 2-2 Número de publicações por ano

Os 30 artigos selecionados para a análise foram listados na tabela 2-1. De acordo com (THOME *et al.*, 2012), *Google Scholar* (GS) foi usado para quantificar o número de citações de cada artigo, dado que GS é de livre acesso na internet, é relativamente rápido e é usado em revisões científicas em diversas áreas. No total, os 30 artigos analisados foram citados 14255 vezes, deles o artigo com maior número de citações foi (HUANG *et al.*, 1998), que foi referenciado 11945 vezes, o que representa 83,79% do total de citações. Por outro lado, a revista com maior número de artigos publicados foi *Signal Processing*. Na tabela 2-2 são relacionados os autores com maior número de publicações.

Tabela 2-1 Artigos por número de citações e recurso

Referências	Número de citações	Periódico
(AMINGHAFARI; CHEZE; POGGI, 2006)	115	CSDA
(AMMAN; DAS, 2001)	32	TIE
(APARICIO; POZO; SAURA, 2008)	21	JEBO
(CHEN <i>et al.</i> , 2013)	3	JAS
(CHUNG; VIBERG; MECKLENBRÄUKER, 2010)	5	SP
(D´ALESSANDRO; YEGNANARAYANA; DARSINOS, 1995)	43	ASSP
(HUANG <i>et al.</i> , 1998)	11945	PRSLA
(HUANG <i>et al.</i> , 2001)	153	BSSA
(HUANG; CRESSIE, 2000)	44	T
(KOPSINIS; MCLAUGHLIN, 2009)	194	TSP
(LEE; QUARDA, 2012)	7	IJC
(LIMA <i>et al.</i> , 2006)	59	MBEC
(LIN; TUNG; HUANG, 2009)	28	PRE
(LISZKA; HOLMSTRÖM, 1999)	8	AASS
(LIU; LIAO, 2011)	17	ESA
(MANDIC; WU; HUANG, 2013)	45	SPM
(MINERVA, 2010)	6	NLSWA
(NOUNOU, 2006)	10	EAAI
(RIOS <i>et al.</i> , 2015)	1	RSE
(RIOS; DE MELLO, 2012)	1	JCS
(RIOS; DE MELLO, 2013)	7	SP
(RIOS; DE MELLO, 2016)	0	SP
(SAMET; MARZBANI, 2014)	0	RSER
(SMALL; TSE, 2003)	65	TCS
(SORIANO; SUYAMA; ATTUX, 2009)	9	ICASS
(TZAGKARAKIS; PAPADOPOULI; TSAKALIDES,	19	SMASWMS

2007)		
(TZAGKARAKIS; PAPADOPOULI; TSAKALIDES, 2009)	18	PE
(WU; HUANG, 2004)	1060	PRSLA
(ZBILUT; GIULIANI; WEBBER, 1998)	121	PLA
(ZBILUT; GIULIANI; WEBBER, 1998)	219	PLA

CSDA—Computational Statistic and Data Analysis, TIE—Transactions on Industrial Electronics, JEBO—Journal of Economic Behavior & Organization, JAS—Journal of Atmospheric Sciences, SP—Signal Processing, ASSP—Acoustics, Speech and Signal Processing, PRSLA—Proc. R. Soc. Lond. A, JCS—Journal of Computer Science, BSSA—Bulletin of the Seismological Society of America, T—Technometrics, TSP—Transactions on Signal Processing, IJC—International Journal of Climatology, MBEC—Med Bio Eng Comput, PRE—Physical Review E, AASS—Astronomy and Astrphysics Supplement Series, ESA—Expert Systems with Applications, SPM—Signal Processing Magazine, NLSWA—Non Linear Systems and Wavelet Analysis, EAAI—Engineering Applications of Artificial Intelligence, RSE—Remote Sensing of Environment, RSER—Renewable and Sustainable Energy Reviews, TCS—Transactions on Circuits and Systems, ICASS—Independent Component Analysis and Signal Separation, SMASWMS—Symposium on Modelling, Analysis and Simulation of Wireless and Mobile Systems, PE—Performance Evaluation, PLA—Physics Letters A,

Tabela 2-2 Autores com maior número de publicações

Autores	N. de artigos
Norden E. Huang	6
Ricardo Araújo Rios	4
Rodrigo Fernandes de Mello	4
Zhaohua Wu	3
Maria Papadopouli	2
George Tzagkarakis	2
Panagiotis Tsakalides	2
Yunxia Liu	2

### 2.3.2. Análise qualitativa

A seguir se realiza uma análise detalhada caracterizando cada artigo com base em quatro critérios: objetivo, técnicas utilizadas, resumo do método e métricas de avaliação. A seguir, na tabela 2-3 é mostrada esta análise que inclui também o título do artigo e referência.

Tabela 2-3 Análise descritiva dos artigos

Referencia	Título	Objetivo	Técnicas	Método	Métricas Avaliação
(AMINGHAFARI; CHEZE; POGGI, 2006)	<i>Multivariate denoising using wavelets and principal component analysis.</i>	Propor uma extensão multivariada do procedimento de eliminação de ruído.	Wavelet e PCA	Os autores propõem um procedimento de eliminação de ruído combinando Wavelet e PCA levando em conta a estrutura de correlação do ruído. Neste trabalho Wavelet é usado como um filtro e as saídas são analisadas usando PCA. O procedimento foi testado em gravações neurais multicanal.	SNR
(AMMAN; DAS, 2001)	<i>An efficient technique for modeling and synthesis of automotive engine sounds.</i>	Modelagem dos sons dos motores dos automóveis usando abordagens de decomposição.	SDFT	Os autores propõem um método para modelar séries temporais a partir da decomposição em componentes estocásticos e determinísticos. O método usa SDFT para decompor a série de tempo em diferentes frequências. Posteriormente, extraem o sinal estocástico através da técnica de excitação MP (multipulso) que se baseia na derivada do sinal filtrado. Logo o componente estocástico é modelado através de um modelo ARX	Análise gráfica, MSE, CR.
(APARICIO; POZO; SAURA, 2008)	<i>Detecting determinism using recurrence quantification analysis: Three test procedures</i>	Desenvolver um procedimento para detectar a existência de dependência geral em séries temporais.	RQA	Os autores propõem três testes de hipóteses a partir do indicador de porcentagem de determinismo (% DET) do RQA para classificar uma série temporal em estocástica ou determinística.	Porcentagem de sucesso dos testes propostos
(CHEN <i>et al.</i> , 2013)	<i>Detecting Signals from Data with Noise: Theory and Applications.</i>	Identificar ruído de fundo em séries temporais do mundo real.	EMD e FRT	Os autores propõem uma hipótese nula adaptativa. Essa hipótese de ruído se baseia em uma das características gerais dos processos de ruído, especificamente um processo AR1 e um ruído Gaussiano fracionário sem pressupostos lineares ou estacionários. O procedimento proposto foi testado em séries climáticas tais como: o ciclo solar e a temperatura da superfície do mar.	Análise gráfica
(CHUNG; VIBERG; MECKLENBRÄUKER, 2010)	<i>Broadband ML estimation under model order uncertainty.</i>	Propor uma solução computacionalmente eficiente para evitar	Estimadores de máxima verossimilhança	Os autores propõem um procedimento de busca máxima que se baseia no cálculo dos estimadores de máxima verossimilhança de	SNR

		o procedimento de busca completa	de banda ancha e Teste de Hipóteses	acordo com o número máximo de sinais. O procedimento seleciona os componentes relevantes através da prova de hipóteses. Também é introduzido um critério para reduzir os componentes indistinguíveis causados pela <i>overparametrization</i> .	
(D'ALESSANDRO; YEGNANARAYANA; DARSINOS, 1995)	<i>Decomposition of speech signals into deterministic and stochastic components</i>	Decompor sinais de voz naturais e sintéticos	DFT e LP	É proposto um método para decompor sinais de voz em componentes estocásticos e determinísticos. Primeiramente é feita a separação da voz nos componentes de excitação através de LP. Posteriormente, se realiza uma identificação das regiões de frequência do ruído e dos componentes determinísticos usando Cepstrum. Logo, os componentes de excitação são reconstruídos através de um algoritmo iterativo. Finalmente, os componentes estocásticos e determinísticos são obtidos mediante a combinação dos <i>frames</i> reconstruídos usando o procedimento <i>overlap-add</i> .	Análise gráfica
(HUANG <i>et al.</i> , 1998)	<i>The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis</i>	Apresentar um novo método para a análise de dados não lineares e não estacionários	EMD e HS	Propõem um novo método (EMD) para a análise de dados não lineares e não estacionários, cujas saídas (IMFs) podem ser analisadas via HS. Com HS, as frequências instantâneas dos IMFs mostram forte evidencia das estruturas embebidas.	Análise gráfica e benchmark
(HUANG <i>et al.</i> , 2001)	<i>A new spectral representation of earthquake data: Hilbert spectral analysis of station TCU129, Chi-Chi, Taiwan, 21 September 1999.</i>	Aplicar o modelo híbrido EMD-HS a dados não estacionários e não lineares.	EMD e HSA	É avaliada a vantagem do modelo híbrido EMD-HSA em dados não estacionários e não lineares (dados do terremoto do 21 de setembro de 1999 na estação TCU129 de Chin Chin, Taiwan). Os resultados obtidos foram comparados com os arrojados pelos métodos: Análise de Fourier, <i>Wavelet Transform</i> e Análise do espectro de frequência.	Benchmark, análise gráfica.
(HUANG; CRESSIE, 2000)	<i>Deterministic/stochastic wavelet decomposition for recovery of signal</i>	Prever sinais a partir de dados ruidosos	WT	É proposto um novo método bayesiano empírico de contração Wavelet para a recuperação dos sinais a partir de dados ruidosos, assumindo que	MSE

	<i>from noisy data</i>			o processo pode ser decomposto em uma tendência determinística mais um processo estocástico gaussiano de média zero mais o ruído.	
(KOPSINIS; MCLAUGHLIN, 2009)	<i>Development of EMD-based denoising methods inspired by wavelet thresholding.</i>	Propor técnicas de eliminação de ruído usando EMD.	EMD-Wavelet	São propostas três técnicas de eliminação do ruído baseado em EMD e o limiar Wavelet: “Limiar adaptado às características de EMD”, “Limiar de intervalo iterativo usando EMD”, “Limiar de intervalo iterativo limpo usando EMD”.	SNR
(LEE; QUARDA, 2012)	<i>An EMD and PCA hybrid approach for separating noise from signal, and signal in climate change detection.</i>	Propor uma abordagem de detecção de mudança climática usando EMD	EMD-PCA	Propõe-se inicialmente aplicar EMD ao conjunto de dados climáticos induzidos pelo processo de ruído branco e posteriormente, aplicar técnicas estatísticas para extrair os componentes aleatórios de alta frequência.	SNR
(LIMA <i>et al.</i> , 2006)	<i>Empirical Mode Decomposition: a novel technique for the study of tremor time series</i>	Distinguir tremor a partir de comportamentos voluntários em pessoas com problemas neurológicos.	EMD e HS	Ao aplicar o método EMD, a série é descomposta em IMFs. De acordo com os autores, o primer IMF representa o tremor involuntário e a soma dos outros IMFs representa movimentos voluntários; ou seja, o primeiro IMF representa o componente estocástico e os restantes IMFs são componentes determinísticos. O HS é utilizado para entender o comportamento da série no domínio de tempo e frequência.	MSE
(LIN; TUNG; HUANG, 2009)	<i>Data analysis using a combination of independent component analysis and empirical mode decomposition</i>	Analisar a relação entre o sinal e o ruído	ICA e EMD	É proposta uma combinação de ICA-EMD para analisar a relação entre o sinal e o ruído. A separação do ruído é feita por meio da ICA e posteriormente, é analisada a série limpa através de EMD.	Análise gráfica
(LISZKA; HOLMSTRÖM, 1999)	<i>Extraction of a deterministic component from ROSAT X-ray data using a wavelet</i>	Extrair o componente determinístico dos dados de raios X (ROSAT X-ray)	WT, PCA e filtros não lineares.	Propõem decompor a série usando WT e posteriormente, aplicam PCA para determinar o valor de corte ( <i>threshold</i> ). Tendo como referência esse valor de corte se procede a separar os componentes determinísticos e	Análise gráfica

	<i>transform and the principal component analysis.</i>			estocásticos através da comparação desse valor de corte com os coeficientes <i>wavelet</i> obtidos.	
(LIU; LIAO, 2011)	<i>Adaptive chaotic noise reduction method based on dual-lifting wavelet.</i>	Propor um método de redução de ruído para sinais caóticos que apresentam ruído não estacionário	<i>Dual-lifting Wavelet</i> , SSA e algoritmo de gradiente decente.	O método proposto usa <i>Dual-lifting Wavelet</i> . Com a finalidade de reduzir o ruído são empregados SSA e o algoritmo de gradiente decente para a análise dos coeficientes aproximados e de detalhe obtidos via <i>Dual-lifting Wavelet</i> . Os autores realizam uma análise comparativa do método proposto com o método <i>Soft-threshold Wavelet</i> e com o método <i>Basic Single Lifting Wavelet</i> .	SNR, RMSE, SNRG, NRA, DC, Função de Autocorrelação, RQA
(MANDIC; WU; HUANG, 2013)	<i>Empirical mode decomposition-based time-frequency analysis of multivariate signals: the power of adaptive data analysis.</i>	Analisar sinais multivariados através de EMD	EMD e suas extensões: EEMDe MEMD	São analisadas as limitações e vantagens do método EMD (e suas extensões) e as aplicações em cenários do mundo real.	Análise gráfica
(MINERVA, 2010)	<i>Wavelet filtering for prediction in time series analysis.</i>	Realizar uma análise preditiva de uma série de tempo financeira a partir da separação dos componentes estocásticos e determinísticos	<i>Wavelet Filtering e algoritmos evolutivos</i>	Propõem uma abordagem para decompor o sinal usando o procedimento de filtro Wavelet. Uma vez feita a filtragem foi usado um algoritmo evolutivo para determinar automaticamente o valor de corte. Posteriormente, baseado em uma regra de valor de corte ( <i>thresholding rule</i> ) é feita a separação das frequências para identificar componentes estocásticos e reconhecer varias fontes dos componentes determinísticos.	Análise gráfica
(NOUNOU, 2006)	<i>Multiscale finite impulse response modeling.</i>	Presentar um novo método multi-escala.	<i>Finite Impulse Response Model (FIR)</i>	Os autores apresentam as vantagens da construção de modelos empíricos em múltiplas escalas. Posteriormente, baseado nessas vantagens propõem um método multi-escala que melhora a precisão da estimação dos modelos FIR.	SNR, MSE
(RIOS <i>et al.</i> , 2015)	<i>Estimating determinism rates to detect patterns in geospatial datasets</i>	Detectar padrões determinísticos e estocásticos em	EMD-FT-MI	Os autores propõem uma abordagem para separar as influências estocásticas das determinísticas considerando a informação	Taxa de Determinismo (%DET) do RQA

		séries de tempo.		mutua contida no espectro de fase dos IMFs consecutivos. O método consiste em (1) aplicar EMD, (2) aplicar FT a cada IMF obtendo um conjunto de coeficientes Fourier para cada IMF, (3) calcular o espectro de fase para cada componente e (4) aplicar MI para quantificar a informação mutua no espectro de fase dos IMFs consecutivos. O algoritmo proposto é avaliado em uma série real (FAPAR). Os autores também propõem uma métrica que estima o nível de determinismo nas séries.	
(RIOS; DE MELLO, 2012)	<i>A Systematic Literature Review on Decomposition Approaches to Estimate Time Series Components</i>	Realizar uma revisão sistemática e detalhada da literatura referente às abordagens de decomposição de séries de tempo.	Systematic Literature Review (SLR)	É realizada uma análise detalhada e bem estruturada de artigos científicos que propõem abordagens de decomposição de séries de tempo em componentes estocásticos e determinísticos. A análise é feita através de uma revisão sistemática da literatura.	*Não são usadas métricas
(RIOS; DE MELLO, 2013)	<i>Improving time series modeling by decomposition and analyzing stochastic and deterministic influences</i>	Melhorar o processo de modelagem de séries de tempo mediante a separação dos componentes estocásticos e determinísticos	EMD-RP	É proposto um algoritmo para separar as influências estocásticas e determinísticas das séries de tempo discretas usando EMD-RP. O algoritmo proposto é avaliado em séries sintéticas através da modelagem de cada componente separadamente de forma que ao combiná-los formam um modelo híbrido de previsão.	MDDL
(RIOS; DE MELLO, 2016)	<i>Applying Empirical Mode Decomposition and mutual information to separate stochastic and deterministic influences embedded in signals.</i>	Aplica a abordagem proposta por (RIOS <i>et al.</i> , 2015) em séries sintéticas e reais.	EMD, FT e MI	Aplicam a abordagem proposta por (RIOS <i>et al.</i> , 2015) em duas séries sintéticas e uma série real (FAPAR). As séries sintéticas usadas foram: (1) função seno com frequência angular $2\pi$ mais uma distribuição Normal com média $\mu=0$ e desvio padrão igual a 1 e (2) sistema Lorenz com parâmetros $\sigma=10$ , $\rho=28$ e $\beta=8/3$ mais uma distribuição Uniforme no intervalo entre 0-10.	Análise gráfica
(SAMET; MARZBANI, 2014)	<i>Quantizing the deterministic</i>	Quantificar o componente	ARMA, DVV e taxa de não	É proposto um algoritmo para quantificar o componente determinístico não linear, a partir da	Análise gráfica

	<i>nonlinearity in wind speed time series.</i>	determinístico não linear em séries de tempo	linearidade.	decomposição da série nos componentes determinístico linear, determinístico não linear e componente não previsível. Primeiramente é capturado o componente determinístico linear usando modelos ARMA, posteriormente, se aplica o método DVV para obter a intensidade dos componentes determinísticos e finalmente é calculada a taxa de não linearidade.	
(SMALL; TSE, 2003)	<i>Detecting determinism in time series: The method of surrogate data.</i>	Detectar determinismo em séries de tempo usando o método de dados substitutos ( <i>Method of Surrogate data</i> )	<i>Surrogate data</i> e MDL	Os autores usam o método de dados substitutos ( <i>Method of Surrogate data</i> ) e testes de hipóteses para determinar se uma série temporal foi gerada por processos estocásticos ou determinísticos. Posteriormente, aplicam MDL para conhecer o modelo utilizado na reconstrução do dado e que permite eliminar o ruído.	Erro de previsão não linear e análise gráfica
(SORIANO; SUYAMA; ATTUX, 2009)	<i>Blind extraction of chaotic sources from white gaussian noise based on a measure of determinism.</i>	Extrair sinais caóticos de dados que apresentam fontes estocásticas.	RP e ICA	Propõem um método baseado na longitude das diagonais geradas pelo RP a partir das quais se construí uma função para separar o sinal determinístico do sinal aleatório. Os resultados obtidos pelo método foram comparados com os resultados obtidos pela ICA.	MSE
(TZAGKARAKIS; PAPADOPOULI; TSAKALIDES, 2007)	<i>Singular spectrum analysis of traffic workload in a large-scale wireless lan.</i>	Analisar a série carga de tráfego através da decomposição em componentes de baixa frequência e componentes de alta frequência.	SSA e testes de hipóteses	Os autores aplicaram SSA para decompor a série em um conjunto de cargas próprias (eigenload). Posteriormente, separam as cargas próprias principais, ou seja, os componentes determinísticos usando modelos estatísticos que mais se ajustam à série original.	Análise gráfica
(TZAGKARAKIS; PAPADOPOULI; TSAKALIDES, 2009)	<i>Trend forecasting based on Singular Spectrum Analysis of traffic workload in a large-scale wireless LAN</i>	Modelar a carga de tráfego nas redes WLAN através da decomposição em componentes de baixa frequência e componentes de alta	SSA	Aplicam SSA para decompor a série de tempo (carga de tráfego nas redes WLAN), sendo os primeiros componentes principais os responsáveis da informação principal, ou seja, os componentes determinísticos ou de baixa frequência e os componentes restantes são os que caracterizam a parte estocástica.	Erro absoluto de previsão respeito à tendência ( $E_1$ ) e Erro absoluto de previsão respeito à verdadeira carga de tráfego total ( $E_2$ )

		frequência.		Posteriormente, os autores propõem um método de previsão baseado no componente determinístico.	
(WU; HUANG, 2004)	<i>A study of the characteristics of white noise using the empirical mode decomposition method.</i>	Estudar as características do ruído branco usando EMD	EMD	Os autores primeiramente apresentam os resultados de um estudo numérico experimental de um ruído branco distribuído uniformemente usando EMD. Posteriormente, propõem um método baseado nas características estatísticas do ruído branco para pôr a prova o conteúdo da informação de um conjunto de dados com nível de ruído desconhecido.	Análise gráfica
(ZBILUT; GIULIANI; WEBBER, 1998)	<i>Recurrence quantification analysis and principal components in the detection of short complex signals.</i>	Demonstrar a utilidade de combinar RQA com PCA para avaliar probabilisticamente a presença de componentes determinísticos em séries de tempo ruidosas.	RQA e PCA	O sinal foi analisado primeiramente mediante RQA. Logo esses resultados foram introduzidos no PCA. Posteriormente, foi analisada cada uma das métricas para o componente principal sendo relevantes as métricas DET(determinismo), ENT(entropia) e DIV(divergência) na separação do sinal.	Análise gráfica
(ZBILUT; GIULIANI; WEBBER, 1998)	<i>Detecting deterministic signals in exceptionally noisy environments using cross-recurrence quantification.</i>	Detectar sinais em presença de grandes quantidades de ruído contaminante.	CRQ	Os autores propõem a técnica CRQ para extrair sinais. Esta técnica é baseada na comparação de duas séries: uma ruidosa conhecida e a série ruidosa de interesse, obtendo-se uma matriz de distância. Posteriormente, é calculada a porcentagem de recorrência que quantifica a superposição entre as duas séries.	SNR

EMD—Empirical Mode Decomposition, IMF<sub>s</sub>—Intrinsic mode function, PCA—Principal Components Analysis, SDFT--Synchronous Discrete Fourier Transform, TDF--Transformada discreta de Fourier, LP-- Linear Predictive Analysis, RQA—Recurrence Quantification Analysis, HS—Hilbert Spectral Analysis, WT—Wavelet Transform, ICA-- Independent Component Analysis, SSA—Singular Spectrum Analysis, EEMD--Ensemble Empirical Mode Decomposition, MEMD—Multivariate Empirical Mode Decomposition, FT—Fourier Transform, MI—Mutual Information, RP—Recurrence Plot, ARMA-- Autorregressive Moving Average, DVV-- Delay vector variance method, MDL-- Minimum Description Length, SNR—Signal to Noise ratio, MSE-- Mean Squared Error, CR-- Compression Ratio, RMSE-- Root Mean Square Error, SNRG—Signal to Noise ratio gain, NRA--Noise reduction accuracy, DC--Decision coefficient, MDDL--Mean Distance from the Diagonal Line, FRT--Fractional Resampling Technique, AR1—Autoregressive model, CRQ—Cross Recurrence Quantification.

## 2.4 Síntese da revisão da literatura

As abordagens de decomposição de séries temporais em componentes estocásticos e determinísticos tem sido de aplicação ampla no mundo real. Os autores (D'ALESSANDRO; YEGNANARAYANA; DARSINOS, 1995) apresentam um método de decomposição para estudar o sinal de voz no domínio do tempo. Outros trabalhos sugerem o uso da decomposição na biologia, por exemplo, para analisar a atividade do tremor no organismo humano (LIMA *et al.*, 2006).

Na meteorologia, (HUANG *et al.*, 2001) utilizam uma abordagem de decomposição baseada nos métodos EMD e HSA para analisar os dados recopilados do terremoto de Taiwan o 21 de setembro de 1999. (WU; HUANG, 2004) propõem um método baseado nas características estatísticas do ruído branco e o aplicam aos dados associados ao Southern Oscillation Index (SOI), indicador da evolução e intensidade dos eventos “El Niño” ou “La Niña” no Oceano Pacífico. (LEE; QUARDA, 2012) propõem uma abordagem de detecção de mudança climática usando EMD. (LISZKA; HOLMSTRÖM, 1999) realizam um estudo de dados dos raios X do satélite ROSAT para extrair o componente determinístico.

Na área da Ciência da Computação: (TZAGKARAKIS; PAPADOPOULI; TSAKALIDES, 2007) e (TZAGKARAKIS; PAPADOPOULI; TSAKALIDES, 2009) usam métodos de decomposição para estudar a demanda do tráfego nas redes WLAN.

Outra área na qual tem também relevante importância a aplicação da decomposição é a área financeira. (MINERVA, 2010) propõe uma abordagem de decomposição que usa o filtro de Wavelet e cadeias de Markov e a aplicam a uma série de preço de fechamento de *Benetton Italian Stocks*.

As abordagens de decomposição de séries temporais em componentes estocásticos e determinísticos geralmente usam métodos de filtragem como primeiro passo. Desses métodos os mais utilizados na literatura são: *Singular Spectrum Analysis* (SSA), *Principal Components Analysis* (PCA), Transformada de Fourier (FT) e *Empirical Mode Decomposition* (EMD). Deles, o EMD foi utilizado em 51.35% dos artigos estudados neste trabalho. Este fato está associado à natureza e propriedade adaptativa do método, que favorece sua utilização.

No entanto, estes métodos não são capazes de separar por si só as componentes estocásticas das componentes determinísticas, daí a necessidade de usar posteriormente técnicas de análise de dados mais específicas que permitem detectar determinismo em séries de tempo. Essas técnicas são: *Recurrence Quantification Analysis* (RQA), *Delay Vector Variance Method* (DVV), *Mutual Information* (MI) e *Largest Lyapunov Exponent* (LLE).

#### 2.4.1. Métodos de decomposição

De acordo com (RIOS; DE MELLO, 2012) os métodos de análise espectral são os mais comumente usados para decompor séries de tempo. Nesta classificação se encontram os seguintes métodos de decomposição: *Fourier Transform*, *Singular Spectrum Analysis*, *Principal Components Analysis*, *Wavelet Transform* e *Empirical Mode Decomposition*. A seguir cada um deles é descrito.

*Fourier Transform* (FT) é uma técnica clássica para descrever qualquer função em termos de funções trigonométricas; seno e cosseno exatamente (MALDONADO PORTILLO, 2012). O conteúdo de um sinal de frequência é normalmente considerado passando o sinal do domínio do tempo para o domínio da frequência por transformadas de Fourier definidas na equação 2.2. Os resultados da FT geralmente são apresentados como um gráfico dos valores absolutos dos coeficientes versus suas respectivas frequências. (RAMIREZ CASTRO; MONTEJO, 2011).

$$F(w) = \int_{-\infty}^{+\infty} x(t)e^{-iwt} dt \quad (2.2)$$

*Singular Spectrum Analysis* (SSA) é uma técnica robusta para decompor séries originais na soma de um pequeno número de componentes independentes e interpretáveis (GOLYANDINA; ZHIGLJAVSKY, 2013). Eles são: tendência lentamente variável, componente oscilatória e ruído sem estrutura. A decomposição da série de tempo é feita em modos estritamente ortogonais. Tem sido amplamente usada devido a sua capacidade para identificar e extrair componentes oscilatórios presentes nas séries de tempo (HASSANI, 2007).

*Principal Components Analysis* (PCA) é uma técnica não paramétrica para reduzir dimensões dos dados e extrair informação relevante deles. Intuitivamente é usada para descobrir as causas da variabilidade em um conjunto de dados e classificá-los por ordem de importância (RIOS *et al.*, 2015). PCA constrói uma transformação linear que escolhe um novo sistema de coordenadas para o

conjunto original de dados, no qual a variância maior do conjunto de dados, chamada primeira componente principal, é capturada no primeiro eixo, a segunda maior variância é o segundo eixo, e assim por diante (SMITH, 2002).

*Wavelet Transform* (WT) é uma alternativa de *Principal Components Analysis* e FT. WT é representada como uma função matemática através da equação 2.3, a qual permite a decomposição em diferentes escalas e resoluções, onde  $\varphi^*(.)$  é a função *Wavelet* básica que satisfaz certas condições gerais,  $a$  é o fator de dilatação,  $b$  é a translação da origem e  $X(t)$  é a série temporal. A decomposição faz possível analisar as séries de tempo tanto no domínio de tempo quanto no domínio da frequência. WT mantém as relações temporais e as características entre as observações das séries, permitindo sua aplicação em dados estacionários.

$$W(a, b, X, \varphi) = |a|^{-\frac{1}{2}} \int_{-\infty}^{\infty} X(t) \varphi^* \left( \frac{t-b}{a} \right) dt \quad (2.3)$$

Uma análise dos pontos fracos de cada uma das técnicas abordadas anteriormente é mostrada na tabela 2-4. Assim, como o objetivo de analisar a decomposição de séries temporais, independentemente da linearidade e estacionariedade, considera-se neste trabalho um método potente e adaptativo para a filtragem das séries chamado *Empirical Mode Decomposition* (EMD).

Tabela 2-4 Limitações das técnicas de decomposição

<b>Técnicas</b>	<b>Limitações</b>
FT	<ul style="list-style-type: none"> <li>✓ Requer condições estritas: linearidade, periodicidade e estacionariedade (HUANG <i>et al.</i>, 1998), (HUANG <i>et al.</i>, 2001)</li> <li>✓ Usa uma função básica predefinida</li> <li>✓ A acurácia depende criticamente do comprimento dos dados e da sazonalidade (MANDIC; WU; HUANG, 2013)</li> <li>✓ As transformações integrais prejudicam a noção do tempo, comprometendo a representação do sinal analítico. (MANDIC; WU; HUANG, 2013)</li> </ul>
WT	<ul style="list-style-type: none"> <li>✓ Assume que a série seja linear (HUANG <i>et al.</i>, 1998)</li> <li>✓ Usa uma função básica predefinida, ou seja, é uma técnica de natureza não adaptativa (HUANG <i>et al.</i>, 1998)</li> <li>✓ A acurácia depende criticamente do comprimento dos dados e da sazonalidade (MANDIC; WU; HUANG, 2013)</li> <li>✓ As transformações integrais prejudicam a noção do tempo, comprometendo a representação do sinal analítico. (MANDIC; WU; HUANG, 2013)</li> </ul>
SSA	<ul style="list-style-type: none"> <li>✓ Requer a seleção do comprimento de janela</li> </ul>

PCA	<ul style="list-style-type: none"> <li>✓ Assume que a série seja linear (RIOS; DE MELLO, 2012)</li> <li>✓ Não deve ser aplicada diretamente a séries temporais, devido à dependência entre as observações. (GOLYANDINA; NEKNETKIN; ZHIGLJAVSKY, 2001)</li> </ul>
-----	--

### 2.4.1.1 EMD como método geral de decomposição

EMD é um método para a análise de sinais com diferentes frequências e modulações. É usado para representar os sinais não estacionários e/ou não lineares, que permitem a análise da composição de frequência das funções que tem uma única dimensão. (MALDONADO PORTILLO, 2012). O objetivo de EMD é decompor o sinal em uma soma de funções de modo intrínsecas chamadas IMFs.

A decomposição baseia-se nas seguintes premissas: (1) o sinal tem pelo menos dois extremos, um máximo e um mínimo, (2) a escala de tempo é definida pelo lapso de tempo entre esses extremos e (3) se os dados foram totalmente desprovidos de extremos, mas contem apenas os pontos de inflexão, então se pode diferenciar uma ou mais vezes para revelar os extremos. Os resultados finais podem ser obtidos por integração dos componentes.

Com o objetivo de definir as IMFs, o método usa os *envelopes* definidos pelos máximos e mínimos locais separadamente. No início, são identificados os extremos, posteriormente, todos os máximos locais são conectados pela linha *spline cubic*, obtendo o *envelope* superior. De igual forma são conectados os mínimos locais, obtendo o *envelope* inferior. Os *envelopes* inferior e superior devem cobrir todos os dados. Logo é calculada uma media aritmética entre os envelopes definidos pelos máximos e mínimos. Essa média é designada como  $m_1$ , e a diferencia entre o sinal original e  $m_1$  é a primeira componente,  $h_1$ , ou seja:

$$h_1 = X(t) - m_1 \quad (2.4)$$

O processo de filtragem é ilustrado na figura 2-3 onde a linha verde representa a série de tempo original, a linha vermelha o *envelope* superior, a linha azul o *envelope* inferior e a negra representa a componente, ou seja, mostra a diferencia entre a série e a média conforme definido na equação 2.4. Este processo tem dois objetivos: eliminar as ondas de equitação e fazer os perfis de onda mais simétricos. Para esse fim, o processo descrito na figura 2-3 tem que ser repetido varias vezes. No segundo processo de filtragem,  $h_1$  é tratado como o sinal original da forma:

$$h_1 - m_{11} = h_{11} \quad (2.5)$$

Esse processo de filtragem é repetido  $k$  vezes até que  $h_{1k}$  cumpra as duas condições necessárias para ser considerada uma IMF, as quais são: (1) o número de extremos e o número de cruzes por zero devem ser iguais ou diferir no máximo em um e (2) em qualquer ponto, o valor médio entre o *envelope* definido pelos máximos locais e o *envelope* definido pelos mínimos locais é zero, ou seja, os máximos locais devem ser positivos e os mínimos locais devem ser negativos.

$$h_{1(k-1)} - m_{1k} = h_{1k} \quad (2.6)$$

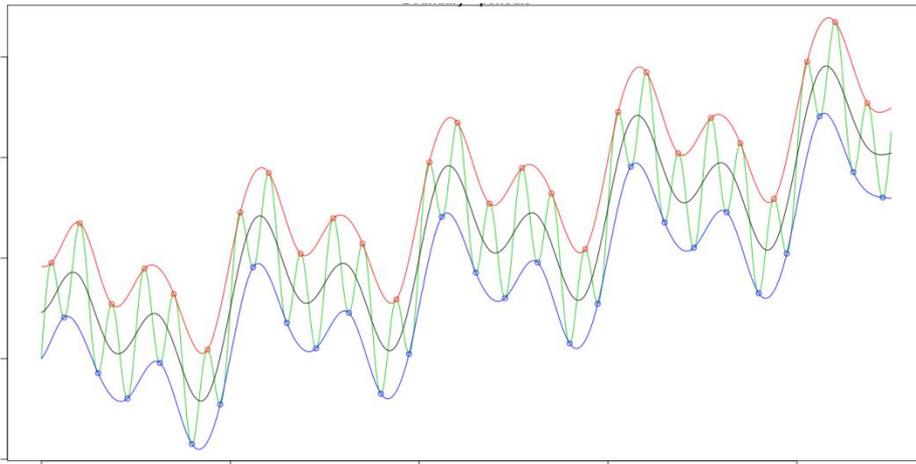


Figura 2-3 Processo de filtragem. Linha verde representa a série de tempo original, linha vermelha representa o *envelope* superior, linha azul representa o *envelope* inferior e a linha negra constitui a diferença entre o sinal original e a média.

Uma vez garantidas as condições das IMFs, então se pode dizer que  $h_{1k}$  é o primer componente IMFs da série original.

$$c_1 = h_{1k} \quad (2.7)$$

Posteriormente, o componente IMFs, é separado do resto do sinal através de:

$$X(t) - c_1 = r_1 \quad (2.8)$$

Uma vez que o resíduo,  $r_1$ , contém ainda informações de componentes de período mais longo, ele é tratado como um novo sinal e é submetido ao mesmo processo de filtragem descrito acima. Esse procedimento pode ser repetido em todos os subsequentes  $r_j$ s, e o resultado é:

$$r_1 - c_2 = r_2; \dots \dots \dots; r_{n-1} - c_n = r_n \quad (2.9)$$

O processo de filtragem pode ser interrompido por qualquer dos seguintes critérios: mesmo quando o componente,  $c_n$ , ou o resíduo,  $r_n$ , torna-se tão pequeno

que é menor que o valor predeterminado de consequência substancial, ou quando o resíduo,  $r_n$ , torna-se uma função monótona da qual não podem ser extraídos mais IMFs, ou uma função com uma extremidade só.

O resultado do processo de EMD será:  $n$  funções de modo intrínsecas IMFs  $[c_1; \dots; c_n]$  e um sinal residual:  $r_n$  como mostrado na equação 2.10. Cada IMF constitui uma série de tempo e seu comprimento é igual ao comprimento da série original. Devido à característica adaptativa desse método, a quantidade de IMFs obtidos depende só da série de tempo.

$$X(t) = \sum_{i=1}^n c_i + r_n \quad (2.10)$$

Para aplicar o método EMD, não é necessária uma referência média ou zero. EMD só precisa a localização dos extremos locais. As referências zero para cada componente é gerado pelo próprio processo de filtragem.

Para ilustrar o processo de filtragem apresentado na figura 2-4 foi usada a série temporal da “produção física de alimentos” de Brasil do período de maio de 1988 até fevereiro de 2013. Os dados tem frequência mensal e foram coletados pela Associação Brasileira das Indústrias da Alimentação (ABIA-Brasil). Os mesmos estão disponíveis na página web: <http://ipeadata.gov.br>.

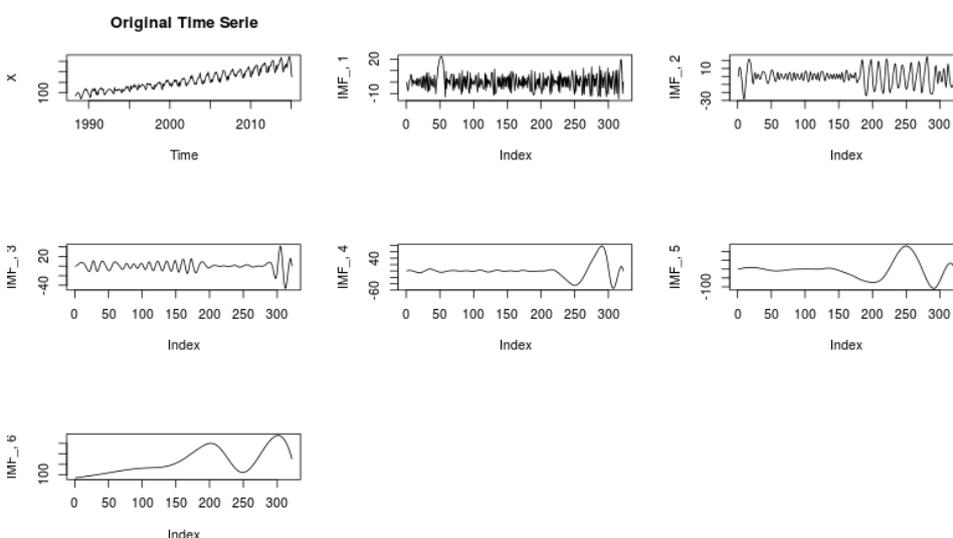


Figura 2-4 Decomposição via EMD da série de produção física de alimentos de Brasil.

#### 2.4.2 Técnicas para detectar determinismo em séries temporais

Conforme citado anteriormente, as técnicas de análise da natureza das séries de tempo mais utilizadas na literatura são: *Recurrence Quantification Analysis*

(RQA) (ZBILUT; WEBBER, 1992), *Delay Vector Variance Method* (DVV) (GAUTAMA; MANDIC; VAN HULLE, 2004), *Mutual Information* (MI) (SHANNON, 2001) e *Largest Lyapunov Exponent* (LLE) (KANTZ, 1994).

*Recurrence Plot* (RP) foi desenvolvido por (ECKMANN; KAMPHORST.; RUELLE, 1987) e constitui uma ferramenta útil na análise de séries temporais não lineares e não estacionárias (CASDAGLI, 1992). Oferece uma imagem gráfica do tempo no qual um processo permanece correlacionado com ele mesmo. A representação oferecida pelo RP está composta por pontos no plano  $i - j$  os quais refletem as relações entre as observações  $x(i)$  e  $x(j)$ . Essas relações são organizadas em uma matriz binária de duas dimensões, nomeada matriz de recorrência, a qual responde à seguinte equação, onde  $r$  é um limiar de distância preestabelecido.

$$d_{ij} = \begin{cases} 1, & \text{if } \|x_i - x_j\| < r \\ 0, & \text{if } \|x_i - x_j\| \geq r \end{cases} \quad (2.11)$$

No entanto, as estruturas criadas pelo RP não fornecem uma informação quantitativa da série (RIOS; DE MELLO, 2013), fato que motivou a criação do RQA com o objetivo de simplificar e automatizar o processo de interpretação do RP. Por tanto, RQA é um conjunto de métricas que refletem a composição do Recurrence Plot (RP). A seguir são mostradas as métricas:

- Taxa ou porcentagem de recorrência (RR)

$$\%REC = \frac{NREC}{NP} * 100 \quad (2.12)$$

$NREC$ : Número de pontos recorrentes

$NP$ : Número de pontos totais

- Taxa de determinismo (%DET)

$$DET = \frac{NPD}{NREC} * 100 \quad (2.13)$$

$NPD$ : Número de pontos nos segmentos paralelos à diagonal secundária considerando que o segmento é formado com um mínimo de, pelo menos, dois pontos adjacentes.

- Comprimento médio dos segmentos diagonais (L)

$$L = \frac{\sum_{i=l_{min}}^N lP(l)}{\sum_{i=l_{min}}^N P(l)} \quad (2.14)$$

$P(l)$ : Probabilidade de um ponto recorrente pertencer a um segmento diagonal de comprimento  $l$ .

- Comprimento do segmento diagonal mais longo ( $L_{max}$ )

$$L_{max} = \max(\{l_i\}_{i=1}^{N_l}) \quad (2.15)$$

$N_l$ : Número total de segmentos diagonais

- Divergência (DIV)

$$DIV = \frac{1}{L_{max}} \quad (2.16)$$

- Entropia Shannon ou medida da informação média contida na distribuição dos segmentos (ENT)

$$ENT = -\sum_{l=l_{min}}^N p_l \log_2 p_l \quad (2.17)$$

$p_l$ : Frequência relativa dos comprimentos dos segmentos diagonais, ou seja:

$$p_l = \frac{P(l)}{N_l} \quad (2.18)$$

- Tendência (TREND)

$$TREND = \frac{\sum_{\tau=1}^{\tilde{N}} \left(\tau - \frac{\tilde{N}}{2}\right) (RR_{\tau} - \langle RR_{\tau} \rangle)}{\sum_{\tau=1}^{\tilde{N}} \left(\tau - \frac{\tilde{N}}{2}\right)^2} \quad (2.19)$$

$\tau$ : Distância entre a diagonal secundária e qualquer segmento paralelo a ela.

- Relação entre DET e RR (RATIO)

$$RATIO = N^2 \frac{\sum_{l=l_{min}}^N lP(l)}{(\sum_{l=1}^N lP(l))^2} \quad (2.20)$$

- Laminaridade (LAM): É uma métrica análoga ao DET e se determina pela razão entre os pontos recorrentes que formam estruturas verticais e todo o conjunto de pontos de recorrência.

$$LAM = \frac{\sum_{v=v_{min}}^N vP(v)}{\sum_{v=1}^N vP(v)} \quad (2.21)$$

$v$ : Comprimento dos segmentos verticais

$P(v)$ : Probabilidade de um ponto recorrente pertencer a um segmento vertical de comprimento  $v$

- Comprimento médio das estruturas verticais (TT)

$$TT = \frac{\sum_{v=v_{min}}^N vP(v)}{\sum_{v=v_{min}}^N P(v)} \quad (2.22)$$

- Comprimento máximo dos segmentos verticais ( $L_{max}$ )

$$L_{max} = \max(\{v_l\}_{l=1}^{N_v}) \quad (2.23)$$

$N_v$ : Numero absoluto de segmentos verticais

O método *Delay Vector Variance* (DVV) analisa a prevalência das componentes determinísticas ou estocásticas em séries de tempo e examina a linearidade da série. (GAUTAMA; MANDIC; VAN HULLE, 2004). Inicialmente a série é representada como um conjunto de vetores de retardo (DV) de comprimento  $m$ . Logo depois são calculadas a média  $\mu_d$  e o desvio padrão  $\sigma_d$  para todos os pares de distâncias entre os DV. Seguidamente, se definem as distâncias  $r_d$  que são extraídas do intervalo  $[\mu_d - n_d\sigma_d; \mu_d + n_d\sigma_d]$ , onde  $n_d$  é um parâmetro de controle de distância. Para cada conjunto formado  $\Omega_k(m; r_d)$  se calcula a variância. Finalmente é calculada a variância “target”  $[\sigma^{*2}(m; r_d)]$  que constitui uma medida inversa da previsibilidade e se define como a média da variância de todos os conjuntos  $\Omega_k(m; r_d)$  dividido pela variância da série. Com a finalidade de interpretar os resultados, é feito um gráfico de distância versus variância “target”, de forma que a presença da componente determinística estará dada por variâncias “target” pequenas para distâncias pequenas, (GAUTAMA; MANDIC; VAN HULLE, 2004).

*Mutual Information* (MI) é uma métrica que quantifica a informação que uma variável aleatória contém ao redor de outra variável aleatória. Matematicamente ela é definida conforme a equação 2.24, onde  $x$  e  $y$  são as duas variáveis aleatórias discretas,  $p(x; y)$  é a probabilidade conjunta e  $p(x)$  e  $p(y)$  são as probabilidades marginais respectivamente. Quando as duas variáveis aleatórias são estritamente independentes, a MI é zero, devido a que uma variável não fornece nenhuma informação da outra, (RIOS *et al.*, 2015) e (RIOS; DE MELLO, 2016).

$$I(x; y) = \sum_{x \in X} \sum_{y \in Y} p(x; y) \log \frac{p(x; y)}{p(x)p(y)} \quad (2.24)$$

*Lyapunov Exponent* (LE) é um indicador geral da presença de caos em um sistema dinâmico. Quantifica a taxa média de divergência das trajetórias vizinhas ao longo de várias direções no espaço de fase. Dada que as condições iniciais correspondem a estados iniciais praticamente idênticos, a divergência exponencial das orbitas implica a perda de previsibilidade do sistema. Por tanto qualquer sistema que contenha pelo menos um LE positivo é definido como caótico. LE formalmente se define conforme a equação 2.25, onde  $\lambda$  é o LE.

$$\lambda = \lim_{t \rightarrow 0} \frac{1}{t} \ln \left| \frac{\Delta f^t}{\delta x_0} \right| \quad (2.25)$$

$\delta x_0$ : perturbação do ponto  $x_0$

$\Delta f^t$ : separação entre o par de trajetórias iniciais  $[x_0; x_0 + \delta x_0]$

## 2.5 Conclusões Parciais

Neste capítulo foi apresentada uma revisão sistemática da literatura sobre abordagens de decomposição de séries temporais influenciadas por ruído, em componentes estocásticas e determinísticas. Desta forma foi possível reunir e sintetizar informação e conceitos importantes relacionados com a decomposição de séries temporais, assim como identificar os métodos e técnicas, maiormente utilizados pela comunidade científica. De forma geral esta revisão constitui uma contribuição aos demais pesquisadores, pois serve de guia para pesquisas e trabalhos futuros. Essencialmente, esta revisão permitiu definir as premissas e nortear o desenvolvimento da abordagem proposta no capítulo seguinte.

### 3 Abordagem proposta

A abordagem proposta neste trabalho visa a separação das componentes estocásticas e determinísticas presentes em qualquer série de tempo, independentemente das suas características de linearidade e estacionariedade. Baseia-se nos níveis de estocasticidade ou determinismo dos IMFs. A abordagem utiliza a técnica EMD como método de filtragem da série e, posteriormente, três testes de hipótese baseados em RQA, propostos por (APARICIO; POZO; SAURA, 2008), são utilizados para caracterizar cada IMF em estocástico ou determinístico. No final, ao combinar os IMFs, obtém-se a componente estocástica e a componente determinística.

Os testes de hipótese utilizados fundamentam-se no conceito de porcentagem de determinismo e na estrutura da matriz de recorrência do RP, nomeada *Recurrence Matrix* (RM). Devido à característica simétrica da RM, os indicadores introduzidos pela RQA são obtidos da estrutura geométrica da parte triangular superior ou inferior da RM, excluindo a diagonal secundária, a qual é nomeada *Triangular Recurrence Matrix* (TRM). A análise, em termos de probabilidade, da distribuição dos segmentos que se formam dentro da TRM constitui o fundamento matemático que sustenta os testes de hipótese. O estudo dos segmentos baseia-se na variável aleatória  $L$  que aloca, a cada elemento  $y_{ij}$  da TRM, o comprimento do segmento ao qual esse elemento pertence. Assim, para cada comprimento de segmento que se forma na TRM é calculada sua probabilidade.

Com o objetivo de calcular a probabilidade de que um ponto recorrente da TRM pertença a um segmento de comprimento  $L$ , (APARICIO; POZO; SAURA, 2008) estudaram a TRM visando encontrar todos os possíveis segmentos que se podem formar para cada comprimento. A seguir é mostrado o cálculo das probabilidades nos casos mais simples para posteriormente generalizar.

- ( $L = 1$ ) Segmentos de comprimento igual a 1: implica a existência de pontos recorrentes isolados. Podem-se distinguir três tipos de segmentos de comprimento 1:
  - a) O que o elemento (1;1) formaria se for um ponto recorrente. A probabilidade de ocorrência disto é igual à quantidade de pontos recorrentes divididos pelo número de pontos totais, ou seja,  $P(a) = \frac{NREC}{NP}$

- b) Os que correspondem aos outros pontos na primeira fila e na primeira coluna da TRM, ou seja, vai se obter um segmento de comprimento 1 se os pontos da primeira fila ou da primeira coluna fossem recorrentes e os pontos contíguos não o fossem. A probabilidade de que isso ocorra é

$$P(b) = \frac{NREC}{NP} * \frac{(NP-NREC)}{NP-1}$$

- c) O que os pontos interiores da TRM formariam, ou seja, esses pontos formariam segmentos de comprimento 1 se for um ponto recorrente e os dois pontos adjacentes não o fossem. Essa probabilidade é dada por:

$$P(c) = \frac{NREC}{NP} * \frac{(NP-NREC)}{(NP-1)} * \frac{(NP-NREC-1)}{(NP-2)}$$

Logo a probabilidade de um ponto da TRM formar um segmento de comprimento igual a 1 pode-se calcular pela equação 3.1.

$$P(1) = \left[ \frac{\frac{NREC}{NP} + \left( \frac{NREC}{NP} \frac{NP-NREC}{(NP-1)} \right) 2(n-2)}{\left( \frac{NREC}{NP} \frac{NP-NREC}{(NP-1)} \frac{NP-NREC-1}{(NP-2)} \right) \frac{(n-2)(n-3)}{2}} \right] \frac{1}{NREC} \quad (3.1)$$

Onde n é o comprimento máximo que um segmento pode ter na TRM.

- ( $L = 2$ ) Segmentos de comprimento igual a 2. Fazendo uma análise similar se podem distinguir três tipos de segmentos de comprimento igual a 2 na TRM.

- a) Segmento formado pelos pontos (1;2) e (2;1) da TRM caso eles sejam recorrentes. A probabilidade de isso ocorrer é dada pela equação:  $P(a) =$

$$\frac{NREC}{NP} \frac{(NREC-1)}{(NP-1)}$$

- b) Segmentos formados pelos pontos recorrentes da primeira linha ou primeira coluna da TRM, cumprindo-se que os pontos contíguos a eles não sejam recorrentes. A probabilidade de isso ocorrer é dada pela equação:

$$P(b) = \frac{NREC}{NP} \frac{(NREC-1)}{(NP-1)} \frac{(NP-NREC)}{(NP-2)}$$

- c) Segmentos formados pelos pontos recorrentes interiores da TRM, cumprindo-se que os pontos adjacentes a eles não sejam recorrentes. A probabilidade de isso ocorrer é dada pela equação:

$$P(c) = \frac{NREC}{NP} \frac{(NREC-1)}{(NP-1)} \frac{(NP-NREC)}{NP-2} \frac{(NP-NREC-1)}{(NP-3)}$$

Logo a probabilidade de um ponto da TRM formar um segmento de comprimento igual a 2 pode-se calcular pela equação 3.2.

$$P(2) = 2 \left[ \frac{NREC}{NP} \frac{NREC-1}{NP-1} + \left( \frac{NREC}{NP} \frac{NREC-1}{NP-1} \frac{NP-NREC}{NP-2} \right) 2(n-3) \right] \frac{1}{NREC} + \left( \frac{NREC}{NP} \frac{NREC-1}{NP-1} \frac{NP-NREC}{NP-2} \frac{NP-NREC-1}{NP-3} \right) \frac{(n-3)(n-4)}{2} \frac{1}{NREC} \quad (3.2)$$

Sem perda de generalidade, de forma similar para segmentos de comprimento  $L = 3; 4; \dots; n-1$ , obtém-se a probabilidade de um ponto recorrente da TRM pertencer a um segmento de comprimento  $L$  como se mostra na equação 3.3.

$$P(L) = L \left\{ \left[ \prod_{i=0}^{L-1} \frac{NREC-i}{NP-i} \right] + \left[ \prod_{i=0}^{L-1} \frac{NREC-i}{NP-i} \frac{NP-NREC}{NP-L} 2(n-L-1) \right] \right\} \frac{1}{NREC} + \left[ \prod_{i=0}^{L-1} \frac{NREC-i}{NP-i} \prod_{i=0}^1 \frac{NP-NREC-i}{NP-L-i} \right] \frac{(n-L-1)(n-L-2)}{2} \frac{1}{NREC} \quad (3.3)$$

Essa probabilidade constitui uma ferramenta muito importante na aplicação dos três testes de hipóteses, os quais têm como objetivo comparar a distribuição dos pontos recorrentes na TRM de uma série real contra a distribuição dos pontos recorrentes de uma série puramente aleatória. Concretamente, os três testes contem as hipótese definidas como:

**$H_0$ : A série em estudo foi gerada por um processo estocástico**

**$H_1$ : A série tem sido gerada por uma função matemática.**

Especificamente, esses testes são: (1) teste baseado no valor médio da variável aleatória  $L$ , que representa o número médio de pontos por segmento, (2) teste baseado na porcentagem de determinismo e (3) teste baseado no comprimento médio do segmento (ALL).

### 3.1 Abordagem proposta *EMD-RQA* testes

Com o objetivo de entender melhor a abordagem, considere uma série de tempo  $x(t) = \{x(1); x(2); \dots \dots x(T)\}$ , onde  $T$  representa o número total de observações da série. A abordagem inicia-se com a decomposição da série  $x(t)$  usando EMD. Essa filtragem separa a série temporal em um conjunto finito de monocomponentes  $c_n(t), \forall n \in \{1, 2, \dots, N\}$  chamados IMFs, mais um resíduo  $r(t)$ . Cada IMF constitui uma série de tempo e o comprimento de  $c_n(t)$  é igual ao comprimento da série original  $x(t)$ , de forma que:  $c_n(t) = x(t) = T$ , onde  $T$  é o comprimento da série original.

O processo de filtragem da série tem a característica de retornar IMFs que são suavizados à medida que são extraídos da série, ou seja, o nível de estocasticidade tende a diminuir. Por consequência, os primeiros IMFs são “mais

estocásticos” que os últimos IMFs, até que o resíduo  $r(t)$  não constitui um resíduo em si, senão uma componente de tendência e seja considerada, automaticamente, parte da componente determinística da série.

O próximo passo é aplicar os três testes de hipótese a cada IMF. Não obstante, como mencionado no Capítulo 2, para a análise da série e, neste caso, para a análise dos IMFs através de RP e conseqüentemente RQA, precisa-se especificar os valores que serão assumidos pelas variáveis: *embedding dimension* ( $m$ ), tempo de retardo ( $\tau$ ) e limiar de distância ( $r$ ). Neste trabalho será adotado  $r$  ótimo conforme propõe (THIEL *et al.*, 2002), de forma que  $r = 5\sigma$ ; onde  $\sigma$  é o desvio padrão do ruído gaussiano. Por outro lado, (APARICIO; POZO; SAURA, 2008) demonstraram que a escolha de  $m$  e  $\tau$  não tem influencia direta no desempenho dos testes de hipótese, portanto será utilizada uma *embedding dimension* igual a 1 e um retardo igual a 10 conforme utilizado pelos próprios autores. Uma vez definidos os parâmetros iniciais necessários no RP, se prossegue para a análise da natureza dos IMFs através dos testes de hipótese. A seguir são descritos brevemente.

➤ Teste baseado em  $\bar{L}$

A estatística, Número Médio de Pontos por Segmento,  $\bar{L}$  é definida pela equação 3.4:

$$\bar{L} = \frac{1}{NREC} \sum_j L(y_j) = \sum_{L=1}^{n-1} L \frac{NREC_L}{NREC} \quad (3.4)$$

Onde  $NREC_L$  representa o número de pontos por segmento de comprimento  $L$ .

O critério de decisão se baseia no Valor Esperado de  $L$ , representado por  $E(L)$ , devido a que  $\bar{L}$  representa a média de  $L$ . Nesse sentido, se o valor de  $\bar{L}$  é significativamente maior que  $E(L)$ , a hipótese nula seria rejeitada, ou seja, a hipótese de um processo puramente aleatório seria rejeitada. Por tanto, pode-se dizer que a série e mais especificamente, o IMF, apresenta dependência geral e é classificado como determinístico.

Se o processo gerador dos dados for puramente aleatório, existirão  $NREC$  distribuídos arbitrariamente na TRM, cada um dos quais estará associado a uma realização diferente da variável aleatória  $L$ . Nessas condições os autores demonstraram que  $E(\bar{L}) = E(L)$  e  $Var(\bar{L}) = \frac{Var(L)}{NREC}$ , de modo que para um número

suficientemente alto de pontos recorrentes se cumpre que:  $\bar{L} \xrightarrow{D} N\left(E(L), \frac{Var(L)}{NREC}\right)$   
 ou  $\frac{\bar{L}-E(L)}{\sqrt{\frac{Var(L)}{NREC}}} \xrightarrow{D} N(0,1)$ .

Para a aplicação dos testes na abordagem proposta, cada IMF é considerado uma série, de forma que cada um deles pode ser classificado como estocástico ou determinístico. Nesse sentido se o valor de  $\bar{L}$  para um IMF qualquer é significativamente maior que seu  $E(L)$ , a hipótese nula seria rejeitada, o que significa que o IMF não é estocástico.

➤ Teste baseado na porcentagem de determinismo (%DET)

A estatística porcentagem de determinismo (%DET), definida na equação 2.13, é comparada com o valor esperado do %DET, definido pela equação 3.5 como a proporção de pontos na TRM que formam segmentos de comprimento maior que 1. Isto responde ao fato de que quanto mais pontos “soltos” haja na TRM, mais aleatório é o IMF. Em outras palavras, quanto menos pontos “soltos” existam na TRM, mais determinístico é o IMF.

$$E[\%DET] = \sum_{L>1} P(L) = 1 - P(1) \quad (3.5)$$

Com a finalidade de provar se a %DET da série difere significativamente do seu Valor Esperado  $E[\%DET]$ , se usa o teste  $\chi^2$ . Basicamente ele permite comprovar se a distribuição da qual é extraída uma amostra de dados é a mesma que a distribuição teórica especificada. As hipóteses do teste  $\chi^2$  e a estatística  $\chi^2$  são definidos como segue:

$H_0$ : O dado segue uma distribuição específica

$H_1$ : O dado não segue uma distribuição específica

$$\chi^2 = \frac{(NPD - NREC * E[\%DET])^2}{NREC * E[\%DET]} + \frac{\{(NREC - NPD) - NREC(1 - E[\%DET])\}^2}{NREC(1 - E[\%DET])} \quad (3.6)$$

O critério de decisão estabelece que se o processo de geração de dados for puramente aleatório, seguiria uma distribuição  $\chi^2$  com um grau de liberdade. Nesse sentido, é necessário provar se os dados da amostra respeitam a distribuição de probabilidade teórica (hipótese nula). Somente será rejeitada a hipótese nula de um processo puramente estocástico se o valor da estatística %DET é significativamente maior que seu Valor Esperado,  $E[\%DET]$ , por tanto podemos classificar a série como determinística. Pelo contrario, se %DET é menor que

$E[\%DET]$ , embora o valor de  $\chi^2$  seja alto, se aceita a hipótese nula e nesse caso a série é estocástica.

➤ Teste baseado em ALL

A estatística, Comprimento Médio de Segmento (ALL) é definida como a taxa entre o número de pontos recorrentes e o número total de segmentos na TRM. Por simplicidade, os autores propõem usar o inverso do comprimento médio do segmento, isto é  $ALL^{-1}$  mostrada na equação 3.6.

$$ALL^{-1} = \frac{1}{ALL} = \frac{1}{NREC} \sum_j \frac{1}{L(y_j)} \quad (3.6)$$

Dado que para processos puramente aleatórios se cumpre que  $E[ALL^{-1}] = E[L^{-1}]$  e  $Var[ALL^{-1}] = \frac{Var[L^{-1}]}{NREC}$ . Para um número suficientemente grande de pontos recorrentes se cumpre que:  $ALL^{-1} \xrightarrow{D} N\left(E[L^{-1}], \frac{Var[L^{-1}]}{NREC}\right)$  ou  $\frac{ALL^{-1} - E[L^{-1}]}{\sqrt{Var[L^{-1}]/NREC}} \xrightarrow{D} N(0,1)$ . Logo, o critério de decisão estabelece que se o valor da estatística  $ALL^{-1}$  é menor que o valor esperado,  $E[ALL^{-1}]$ , existe evidência para rejeitar a hipótese nula. Nesse caso o IMF seria classificado como determinístico.

Para maior certeza na classificação, assim como para um melhor desenvolvimento da abordagem, cada IMF será classificado como determinístico quando os três testes rejeitarem a hipótese nula. Caso ocorrer o contrário (os três testes aceitarem a hipótese nula) ou alguma divergência entre os resultados dos testes; então os IMFs serão classificados como estocásticos.

No final, a soma dos IMFs determinísticos mais o resíduo formarão a componente determinística da série e a soma dos IMFs estocásticos constituirá a componente estocástica. Uma explicação gráfica da abordagem proposta é mostrada na figura 3.1.

### 3.2 Conclusões parciais

Nesse capítulo foi apresentada uma nova abordagem para separar séries de tempo em componentes estocásticas e determinísticas. De forma geral, pode ser aplicado em qualquer tipo de série devido a que os métodos que a compõem não apresentam restrições. A abordagem usa EMD que permite decompor qualquer tipo de série de tempo independentemente das características de linearidade e estacionariedade e posteriormente são aplicados três testes de hipótese para

caracterizar cada “componente” (IMF), resultante da decomposição, em estocásticos ou determinísticos.

A abordagem proposta visa separar fielmente as componentes estocásticas e determinísticas da série e, a partir daí, obter melhores resultados nas previsões, devido às vantagens de modelar separadamente os componentes. O capítulo 4 avalia a abordagem proposta no domínio de previsão usando séries reais.

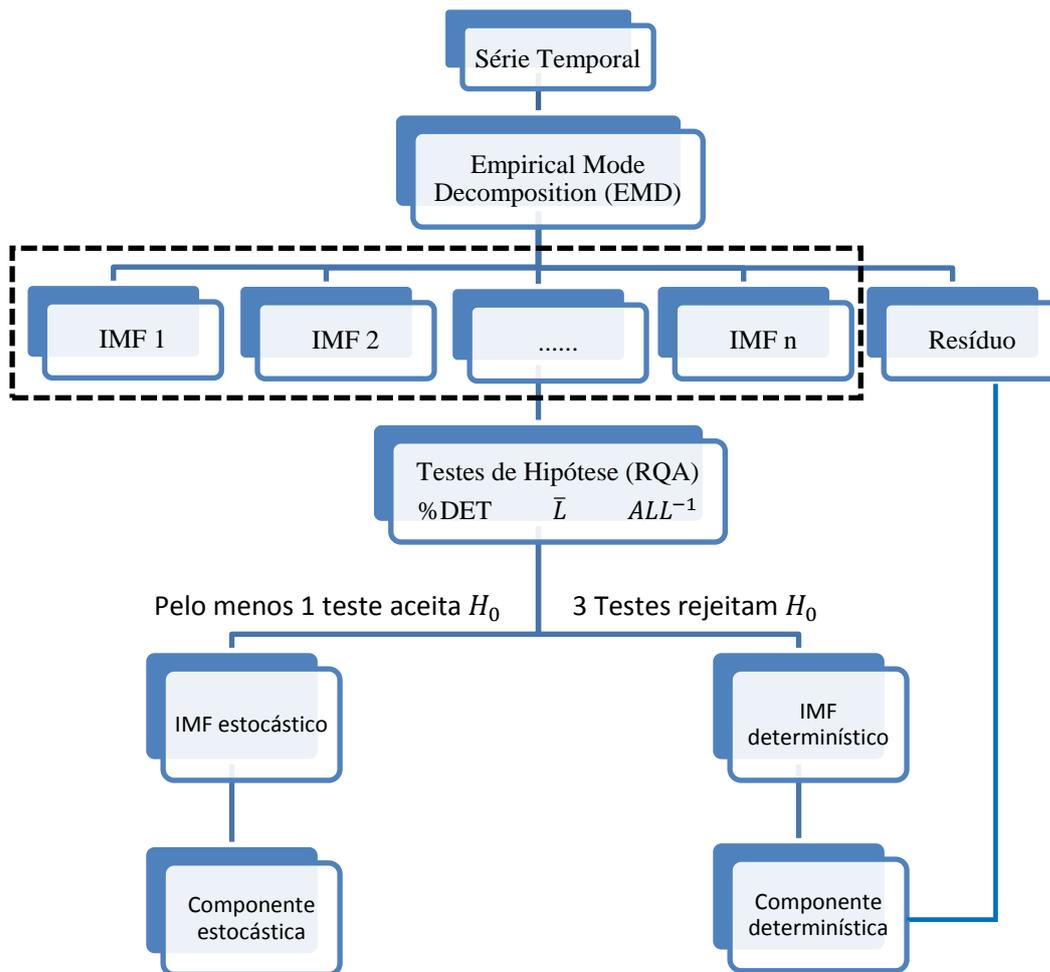


Figura 3-1 Abordagem proposta para decompor séries temporais em componentes estocásticas e determinísticas

## 4 Avaliação da abordagem proposta. Resultados

Nesse capítulo se realiza uma avaliação da abordagem proposta no domínio de previsão com o fim de mostrar as vantagens de modelar separadamente as componentes estocásticas e determinísticas extraídas das séries temporais. Inicialmente, se descrevem em detalhes as séries de tempo e os modelos preditivos usados nos experimentos. Posteriormente, se apresentam os resultados da modelagem e finalmente, se discute a eficiência da abordagem proposta.

### 4.1 Conjunto de dados

No presente capítulo são utilizados quatro conjuntos de dados reais para validar a eficiência da abordagem proposta. Estes conjuntos de dados são: (i) consumo de energia elétrica do Brasil, (ii) número de mortes de condutores de automóveis no Reino Unido, (iii) produção trimestral de ônibus no Reino Unido, e (iv) geração total de energia elétrica nos Estados Unidos. A série consumo de energia elétrica do Brasil está disponível no site da Empresa de Pesquisa Energética (EPE) e as restantes séries encontram-se nos pacotes *datasets* e *exsmooth* do software R (R. CORE TEAM, 2014).

A série temporal consumo de energia elétrica do Brasil registra mensalmente o consumo total de energia elétrica no Brasil. Os dados são medidos em megaW-hora (MWh) no período 2004.1 até 2014.1, resultando em 121 observações. Os dados estão disponíveis no site da Empresa de Pesquisa Energética (EPE) podem-se observar graficamente na figura 4-1.

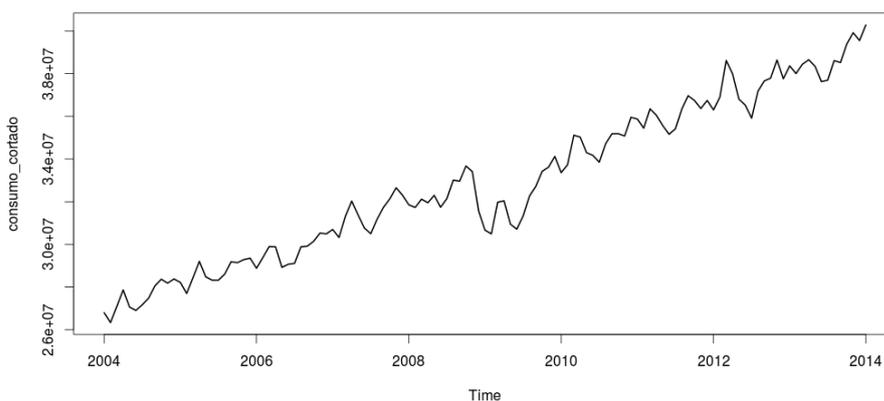


Figura 4-1 Série temporal Consumo de Energia Elétrica do Brasil

A série temporal *UKDriverDeaths* registra mensalmente as mortes de condutores de automóveis no Reino Unido. Os dados pertencem ao período de

janeiro de 1969 a dezembro de 1984, resultando em 192 observações. Na figura 4-2 pode-se observar a presença de não estacionariedade na variância da série.

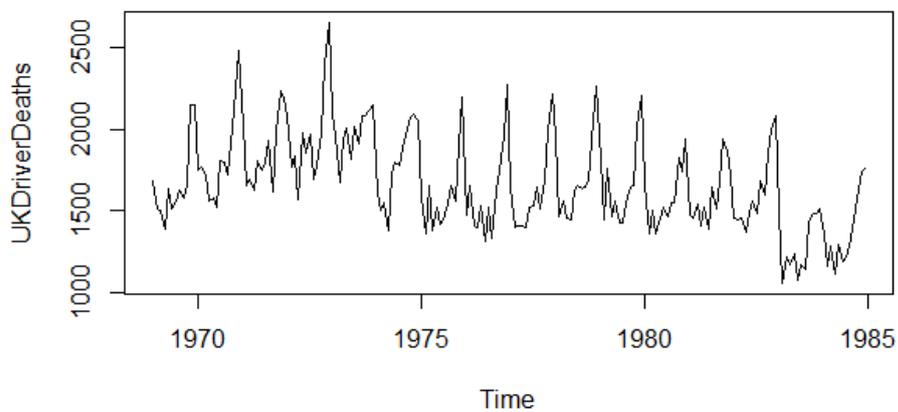


Figura 4-2 Série temporal UKDriverDeaths

Por sua vez, a série temporal *ukcars* registra a produção trimestral de ônibus no Reino Unido, medida em milhares de carros, desde o primeiro trimestre de 1977 até o primeiro trimestre de 2005, resultando em 113 observações. Os referidos dados são representados graficamente na figura 4-3, não se observando nenhum tipo de estacionariedade.

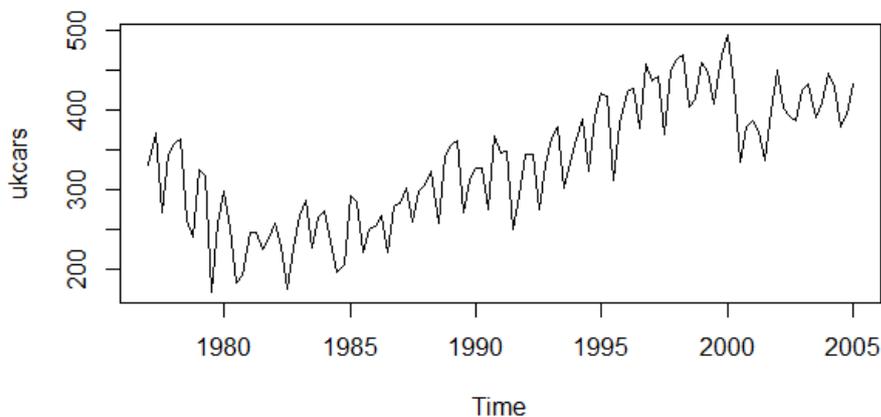


Figura 4-3 Série temporal ukcars

O quarto conjunto de dados é *uselec*. A série representa a geração total de energia elétrica nos Estados Unidos, medida desde janeiro de 1985 até outubro de 1996, sendo 142 o total de observações da série. Na figura 4-4 se mostra a gráfica da série *uselec*, onde pode-se observar que não existe estacionariedade na média, mas sim na variância.

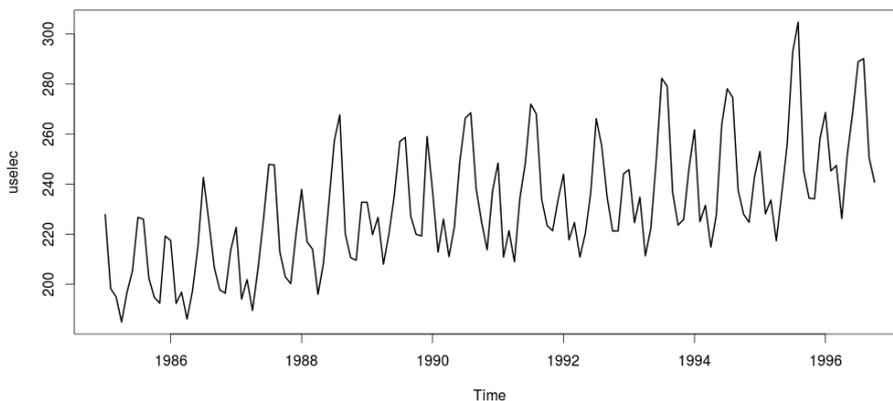


Figura 4-4 Série temporal uselec

Com o objetivo de avaliar o desempenho da modelagem que inclui a abordagem proposta no capítulo 3, as séries analisadas foram divididas em duas amostras, uma de treinamento dada pelo 90 por cento (%) das observações e uma amostra de teste formada pelo 10 por cento restante. Nesse sentido, para cada série de treino é aplicada a abordagem de decomposição proposta seguida do processo de modelagem individual dos componentes. No final, os resultados das previsões são combinados para formar o modelo híbrido e sua acurácia é avaliada e comparada com a acurácia do modelo aplicado à série sem decompor. Por outra parte, será comparado o resultado da previsão obtida pelo modelo híbrido com a amostra de teste. Nas seções seguintes será detalhado o processo de modelagem.

## 4.2 Modelos preditivos

Os modelos preditivos considerados para a avaliação da abordagem proposta foram: Polinomial (CASDAGLI, 1989) e os modelos *Box & Jenkins* (BOX; JENKINS; REINSEL, 1994). O modelo Polinomial foi desenhado para modelar dados determinísticos, sendo utilizado neste trabalho para a modelagem da componente determinística, conforme utilizado em (RIOS; DE MELLO, 2013). No entanto, os modelos *Box & Jenkins* são utilizados aqui para a modelagem da componente estocástica e para a modelagem da série de treino sem decompor. Na figura 4-5 se descreve o processo de modelagem.

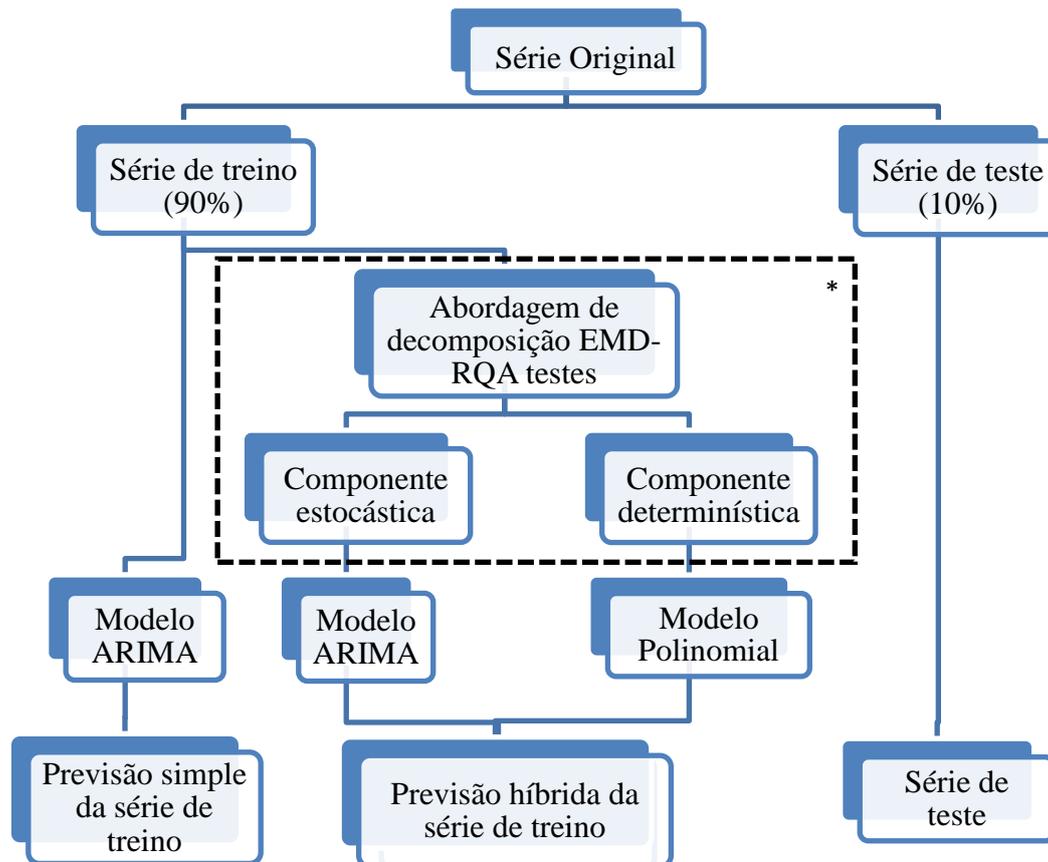


Figura 4-5 Processo de modelagem das séries (\*) Refere-se à figura 3-1

O modelo Polinomial tenta encontrar uma função de aproximação às observações da série. Essa função de aproximação é definida por:  $\pi_i f_N: \mathbb{R}^m \rightarrow \mathbb{R}, i = 1, \dots, m$ , onde  $m$  apresenta variáveis com grau no máximo  $d$ . A parametrização para esta função é realizada para minimizar a equação 4.1 (CASDAGLI, 1989).

$$\sum_{n=1}^{N-1} (\pi_i x_{n+1} - \pi f_N(x_n))^2 \quad (4.1)$$

Os modelos *Box & Jenkins*, conhecidos por ARIMA (*Autoregressive Integrated Moving Averages*) visam captar o comportamento da correlação seriada ou autocorrelação entre os valores da série temporal, e com base nesse comportamento realizar previsões futuras. A construção dos modelos Box & Jenkins é baseada em um ciclo iterativo, no qual a estrutura do modelo é escolhida baseando-se nos próprios dados, os que podem-se derivar de processos estacionários ou não estacionários e/ou sazonais.

O modelo ARIMA  $(p,d,q)$  é composto pelos processos auto-regressivo (AR) com parâmetro  $\phi_i$ , pelo processo médias móveis (MA) com parâmetro  $\theta_i$  e pelo processo de diferenciação (I) da série  $Z_t$ . Esse processo de diferenciação é usado no caso de que a série seja não estacionária com o objetivo de torná-la estacionária. Assim o modelo ARIMA  $(p,d,q)$  responde a equação 4.2, onde  $a_t$  é ruído branco.

$$\phi(\beta)\nabla Z_t = \theta(\beta)a_t \quad (4.2)$$

O modelo AR é uma particularização do modelo ARIMA  $(p,d,q)$  onde as ordens  $d$  e  $q$  são nulas. Este modelo pressupõe que seja o resultado da soma ponderada de seus  $p$  valores passados além do ruído branco  $a_t$ , como definido na equação 4.3.

$$w_t = \phi_1 w_{t-1} + \phi_2 w_{t-2} + \dots + \phi_p w_{t-p} + a_t. \quad (4.3)$$

O modelo MA é também uma particularização do modelo ARIMA  $(p,d,q)$ . Neste caso as ordens  $p$  e  $d$  são nulas. No modelo, a série  $w_t$  é resultante da combinação dos ruídos brancos do período atual com aqueles ocorridos em períodos anteriores conforme definido na equação 4.4.

$$w_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad (4.4)$$

O modelo ARMA é igualmente uma particularização do modelo ARIMA  $(p,d,q)$  onde a ordem  $d$  é nula, ou seja, não se precisa diferenciar a série pois ela é estacionária. Para construir este modelo, é preciso integrar a parte auto-regressiva com a parte média móvel conforme representado na equação 4.5.

$$\phi(\beta)Z_t = \theta(\beta)a_t \quad (4.5)$$

Por último, o modelo SARIMA é uma extensão dos modelos ARIMA e é utilizado em séries que apresentam sazonalidade. Para construir o modelo SARIMA é preciso adicionar os termos auto-regressivos sazonais  $\phi_i$  e/ou médias móveis sazonais  $\theta_i$  (Souza e Camargo, 1996) conforme representado na equação 4.6.

$$\phi(\beta)\phi(\beta^S)\nabla^d \nabla_S^D Z_t = \theta(\beta)\theta(\beta^S)a_t \quad (4.6)$$

### 4.3 Modelagem

Inicialmente, foi aplicada a abordagem de decomposição às séries de treino. Como resultado, se obtiveram duas componentes para cada série de treino: uma estocástica e outra determinística como se mostra no apêndice A, B, C e D.

Depois da decomposição, foi feita a modelagem individual das componentes extraídas e das séries sem decompor. A seguir são mostrados os resultados para cada série.

#### 4.3.1 Modelagem da série de treino de Consumo de Energia Elétrica do Brasil

Da aplicação da abordagem de decomposição *EMD-RQAtestes* na série Consumo de Energia Elétrica do Brasil, se obtiveram duas componentes, uma estocástica e outra determinística como se observa na figura 4-6 e na figura 4-7 respectivamente.

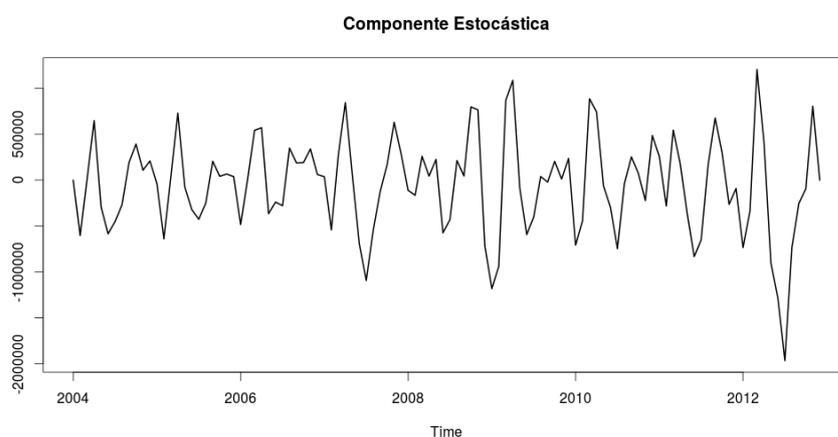


Figura 4-6 Componente estocástica da série Consumo de Energia Elétrica do Brasil

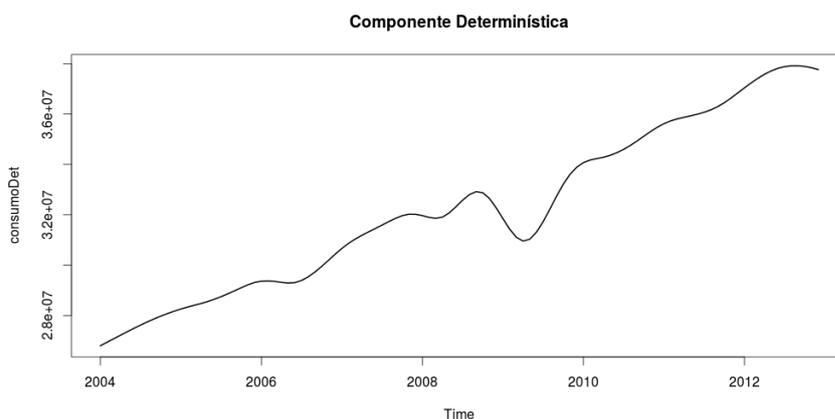


Figura 4-7 Componente determinística da série Consumo de Energia Elétrica do Brasil

Como primeiro passo do processo de modelagem, a componente determinística foi ajustada a uma função polinomial lineal definida como:  $y = 101875x + 26693700$ . O grau do polinômio foi encontrado usando a função *poly* do R. Na figura 4-8 se mostra o ajuste realizado.

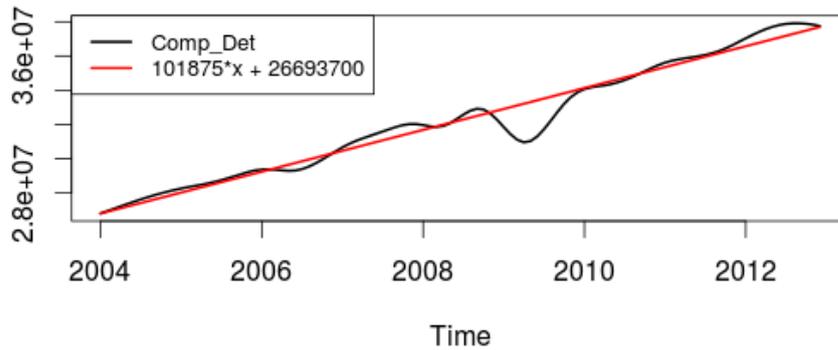


Figura 4-8 Ajuste da componente determinística da série de treino Consumo de Energia Elétrica do Brasil

Posteriormente, e sobre a função definida, realizou-se a modelagem da série 13 passos à frente e a componente estocástica foi modelada através dos modelos Box & Jenkins, usando a função `auto.arima` disponível no R (HYNDMAN; KHANDAKAR, 2007). Especificamente, a componente estocástica foi ajustada a um modelo ARIMA (0,0,1)(0,1,1)[12]. No entanto, devido a que o foco desse capítulo é a previsão híbrida, os gráficos da previsão de cada componente de forma individual foram omitidas.

Logo, os resultados da modelagem das componentes determinísticas e estocásticas foram somados para obter a previsão da série de treino Consumo de Energia Elétrica do Brasil, constituindo-se o modelo híbrido de previsão Polinomial & ARIMA. Os resultados se apresentam na figura 4-8.

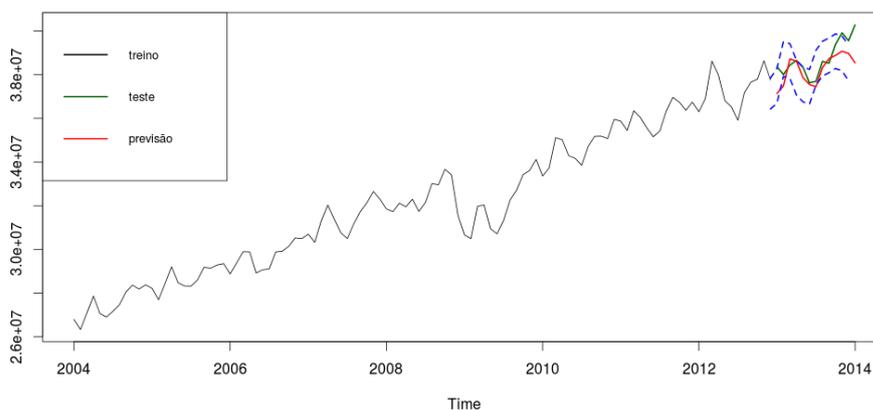


Figura 4-9 Previsão híbrida da série de treino Consumo de Energia Elétrica do Brasil

Considerando a figura, pode-se apreciar a acurácia do modelo híbrido de previsão comparando visualmente o resultado da previsão (linha vermelha) com a série de teste (linha verde).

Outra análise pode ser feita através da comparação dos resultados do

modelo híbrido e os resultados da previsão simples. Para tais fins, se aplicou o modelo ARIMA(0,1,0)(1,0,0)[12] na série de treino sem decompor, obtendo-se a previsão simples mostrada na figura 4-10.

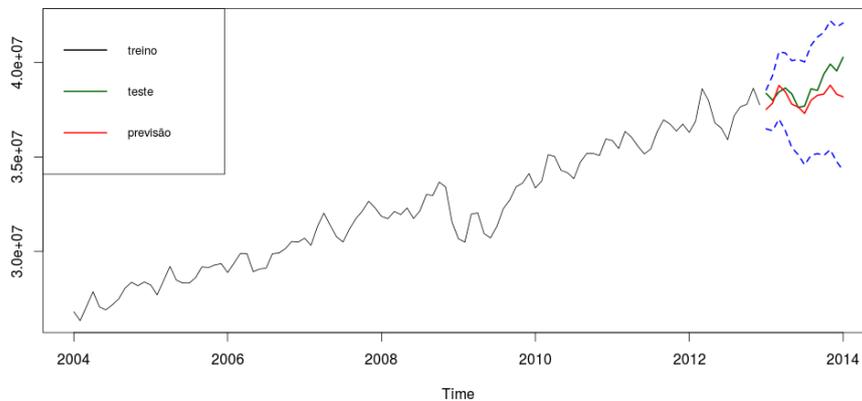


Figura 4-10 Previsão simples da série de treino Consumo de Energia Elétrica do Brasil

Outra análise comparativa da acurácia dos modelos de previsão híbrido e simples foi feita a partir das estatísticas residuais para a previsão *out of sample* RMSE, MAPE e MAE mostradas na tabela 4-1.

Tabela 4-1 Estatísticas residuais dos modelos aplicados à série de treino Consumo de Energia Elétrica do Brasil

Métricas/Modelos	Modelo Simples	Modelo Híbrido
MAPE	0.0173	0.0138
MAE	682854	540570.8
RMSE	879230.2	712621.2

Considerando as estatísticas anteriores, conclui-se que o modelo híbrido apresentou uma medida de erro menor para cada uma das métricas aplicadas com relação ao modelo simples, o que significa que o modelo híbrido conseguiu melhores resultados de previsão.

#### 4.3.2 Modelagem da série de treino UKDriverDeaths

Inicialmente, a série de treino foi decomposta em componentes estocásticas e determinísticas usando a abordagem proposta no capítulo 3. Os resultados são apresentados na figura 4-11 e na figura 4-12.

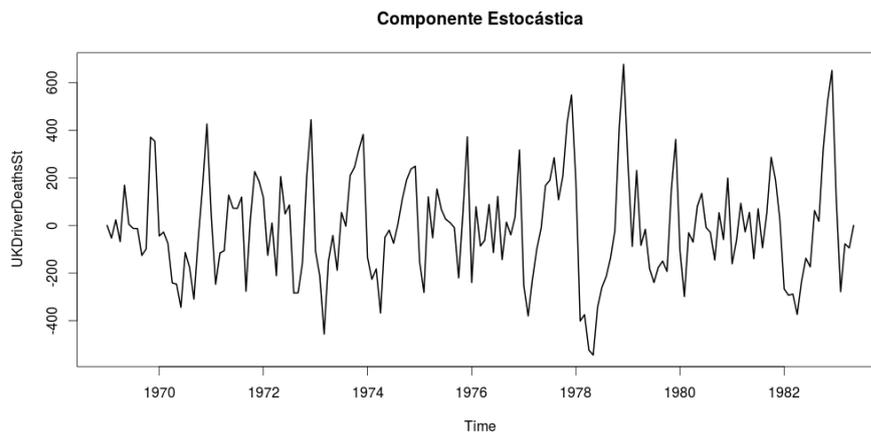


Figura 4-11 Componente estocástica da série de treino UKDriverDeaths

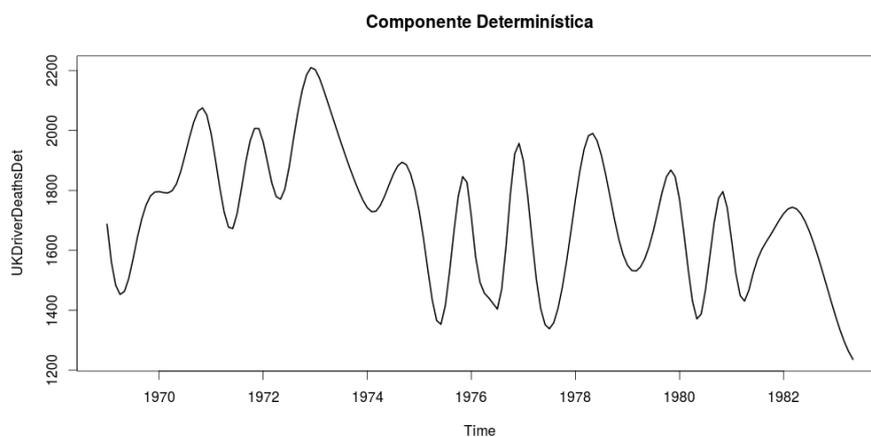


Figura 4-12 Componente determinística da série de treino UKDriverDeaths

O melhor ajuste da componente determinística da série de treino UKDriverDeaths conseguiu-se com um modelo de ordem 19. No obstante, levando em conta o critério de parcimônia e para ser consequente com os pressupostos teóricos, foi selecionado um rango para o grau do polinômio de 1 até 6. No final a função *poly* do R devolveu que o melhor ajuste nesse rango foi um modelo de grau 4. A seguir, a componente determinística é ajustada ao modelo polinomial de ordem 4 definido pela equação:  $y = 1714.6 - 1263.5x - 466.9x^2 + 591.9x^3 - 1164.4x^4$ . A figura 4-13 mostra o ajuste da componente determinística.

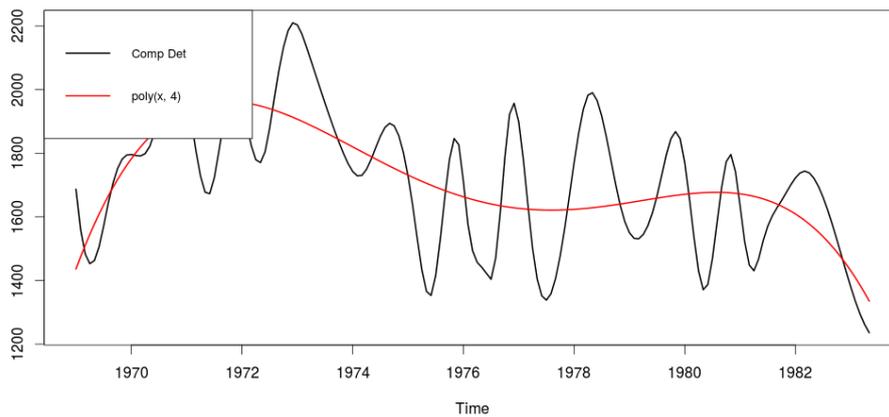


Figura 4-13 Ajuste da componente determinística da série de treino UKDriverDeaths

Com base nesse ajuste, é feita a modelagem da componente determinística. Logo, a componente estocástica foi modelada usando um ARIMA  $(4,0,2)(2,0,0)$ [12]. No final se construiu a previsão híbrida conforme se apresenta na figura 4-14.

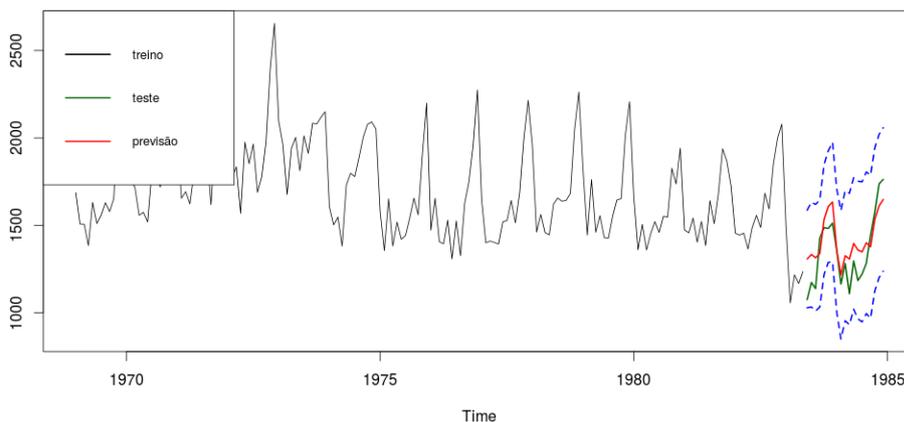


Figura 4-14 Previsão híbrida da série de treino UKDriverDeaths

Por outra parte, realizou-se a modelagem da série UKDriverDeaths sem decompor. Nesse caso foi usado um modelo ARIMA  $(1,1,1)(2,0,0)$ [12]. O resultado da previsão é mostrado na figura 4-15.

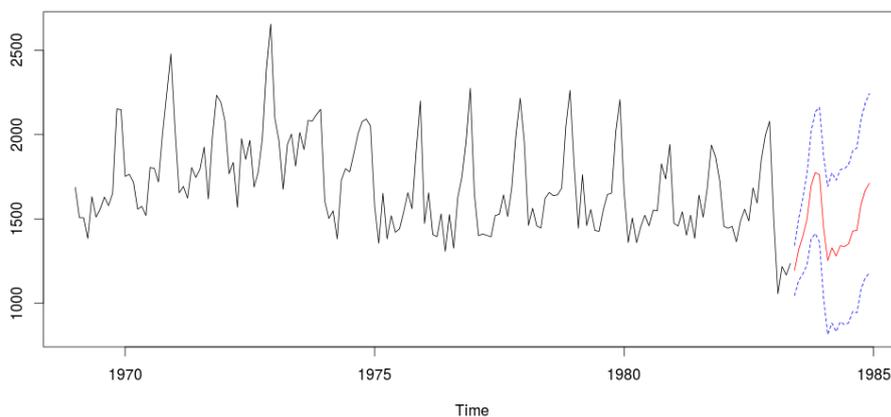


Figura 4-15 Previsão simples da série de treino UKDriverDeaths

Para a análise da acurácia dos modelos simples e híbrido sobre a série UKDriverDeaths, foi feita também uma comparação quantitativa através das estatísticas residuais para a previsão *out of sample* MAPE, MAE e RMSE. Tal e como se percebe na tabela 4-2, o melhor foi o modelo híbrido, o qual permite afirmar que tanto a abordagem de decomposição usada no modelo híbrido quanto a modelagem foram eficientes.

Tabela 4-2 Estatísticas residuais dos modelos aplicados à série de treino UKDriverDeaths

Métricas/Modelos	Modelo Simples	Modelo Híbrido
MAPE	0.1124	0.0868
MAE	143.4505	110.9902
RMSE	182.4450	125.5981

### 4.3.3 Modelagem da série de treino ukcars

Como resultado da aplicação da abordagem de decomposição *EMD-RQAtestes* na série ukcars, se obtiveram duas componentes, uma estocástica, figura 4-16, e outra determinística, figura 4-17.

Posteriormente, a componente determinística foi ajustada a uma função polinomial de ordem 4 definida pela equação:  $y = 317.879 + 581.947x + 217.684x^2 - 252.695x^3 + 6.566x^4$ . Na figura 4-18 apresenta-se o ajuste polinomial. Logo, fazendo uso do modelo é feita a previsão 8 passos à frente da componente determinística. Por outra parte, a componente estocástica foi modelada

através de um  $ARIMA(3,0,0)(2,1,1)[4]$ . Ao igual que nas séries anteriores, a previsão individual das componentes é omitida.

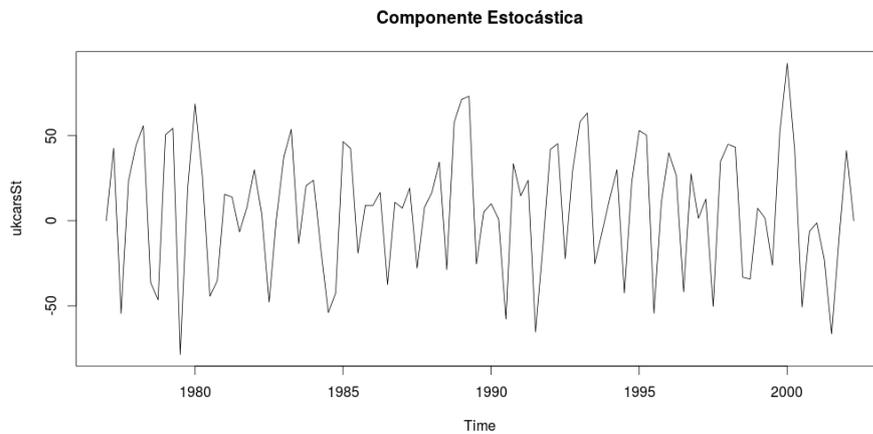


Figura 4-16 Componente estocástica da série de treino ukcars

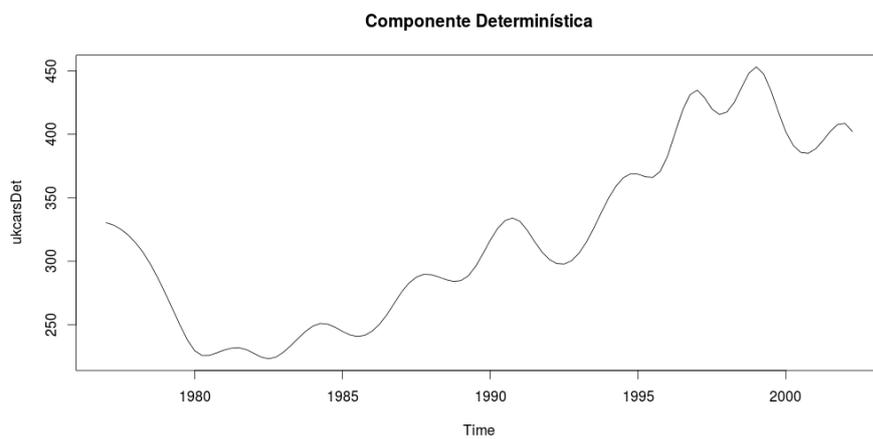


Figura 4-17 Componente determinística da série de treino ukcars

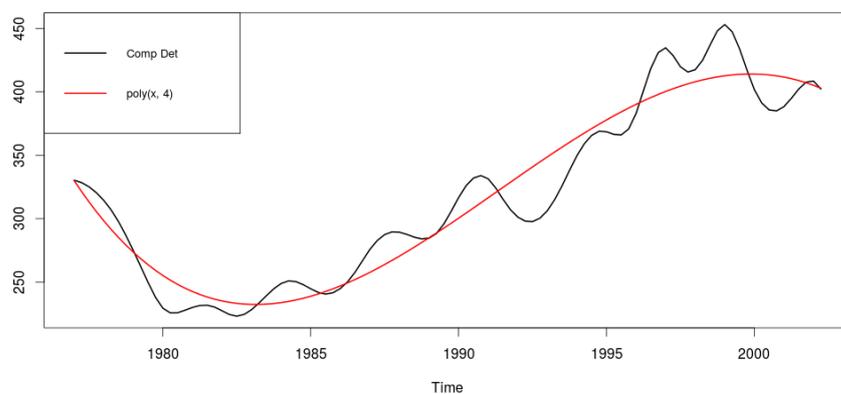


Figura 4-18 Ajuste da componente determinística da série de treino ukcars

No final, os resultados da modelagem das componentes determinísticas e

estocásticas foram somados para obter a previsão híbrida da série de treino ukcars conforme se apresenta na figura 4-19.

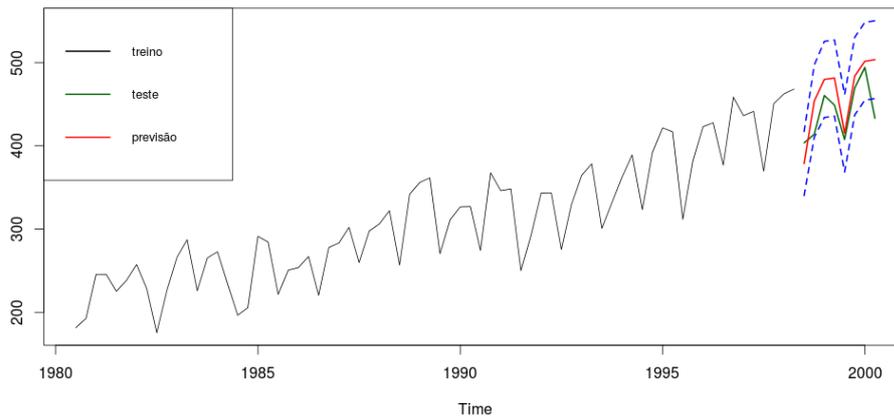


Figura 4-19 Previsão híbrida da série de treino ukcars

Com o objetivo de analisar a acurácia do modelo, foi feita uma comparação entre os diferentes tipos de previsões, ou seja, entre os resultados da previsão híbrida e da previsão simples da série de treino ukcars.

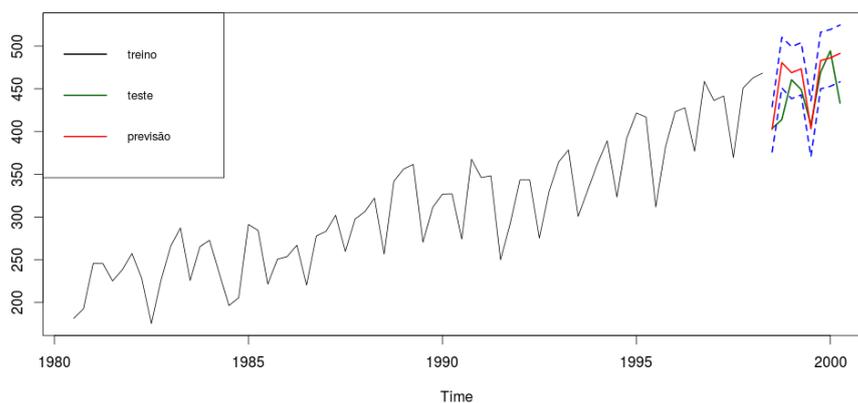


Figura 4-20 Previsão simples da série de treino ukcars

Similar às séries anteriores, se analisa a acurácia do modelo a partir das estatísticas residuais para a previsão *out of sample* MAPE, MAE e RMSE. Na tabela 4-3 apresentam-se os resultados.

Tabela 4-3 Estatísticas residuais dos modelos aplicados à série de treino ukcars

Métricas/Modelos	Modelo Simples	Modelo Híbrido
MAPE	11,8247	5,7555
MAE	48,5336	23,6230
RMSE	50,8517	27,3263

Com base nesses resultados, pode-se concluir que com o modelo híbrido a previsão foi melhor que com o modelo simples.

#### 4.3.4 Modelagem da série de treino uselec

Ao igual que na análise das séries anteriores, à série uselec foi aplicada a abordagem proposta no capítulo 3. Como resultado, se obtiveram duas componentes: uma estocástica e outra determinística como se observa na figura 4-21 e na figura 4-22

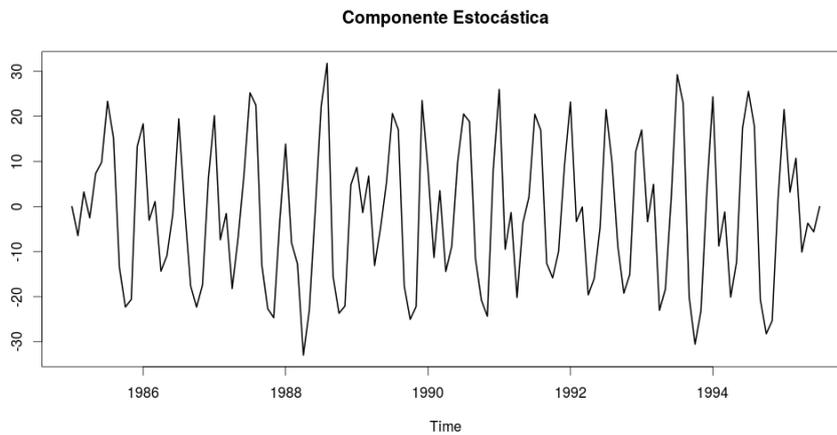


Figura 4-21 Componente estocástica da série de treino uselec

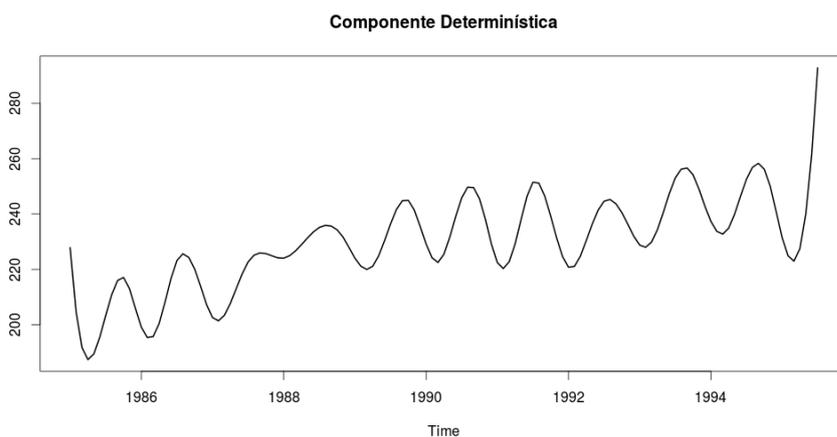


Figura 4-22 Componente determinística da série de treino uselec

Posteriormente, a componente determinística foi ajustada a um modelo polinomial de grau 4 definido pela equação:  $y = 229.75 + 142.22x - 36.11x^2 + 20.25x^3 + 17.39x^4$ . Na figura 4-23 é representado o ajuste polinomial da componente determinística.

Logo, ambas componentes foram modeladas independentemente: a componente determinística foi modelada pelo modelo Polinomial anteriormente citado e a componente estocástica pelo modelo ARIMA (1,0,0)(1,0,0)[12].

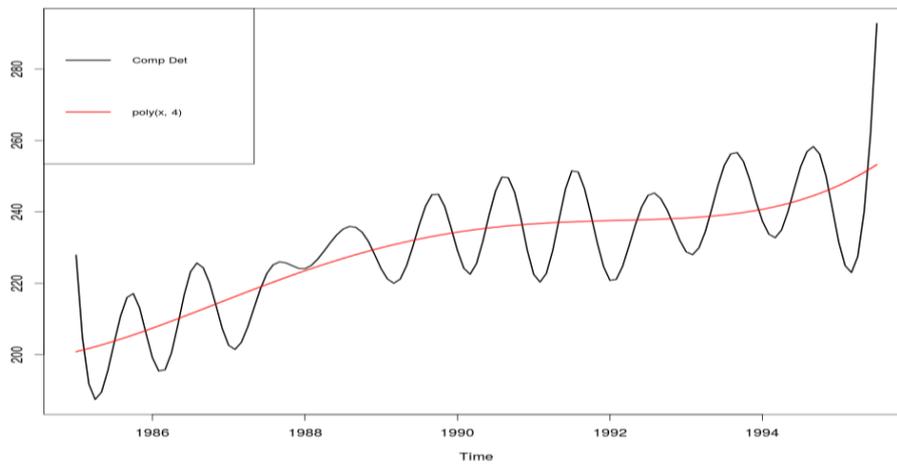


Figura 4-23 Ajuste da componente determinística da série de treino uselec

Da soma dos resultados da modelagem de ambas componentes, obtém-se a previsão híbrida da figura 4-25.

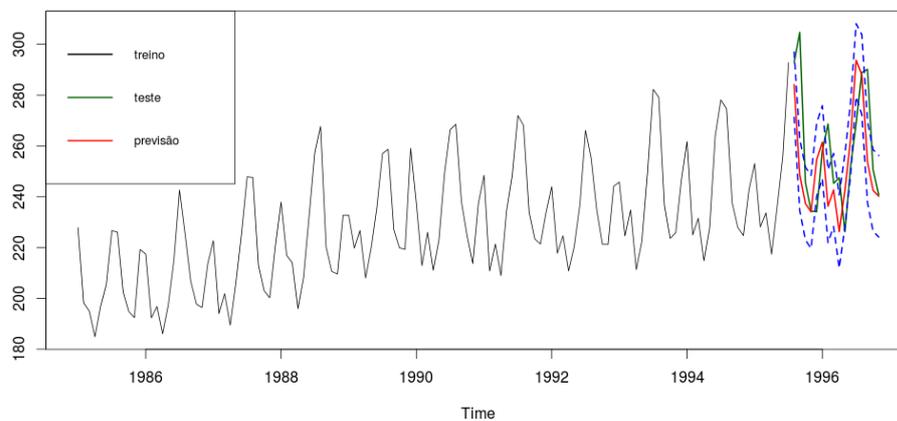


Figura 4-24 Previsão híbrida da série de treino uselec

Por outra parte, foi realizada também a modelagem da série sem decompor. Para isso foi utilizado, ao igual que nas análises das séries anteriores, a função `auto.arima`, obtendo-se como resultado a previsão mostrada na figura 4-26.

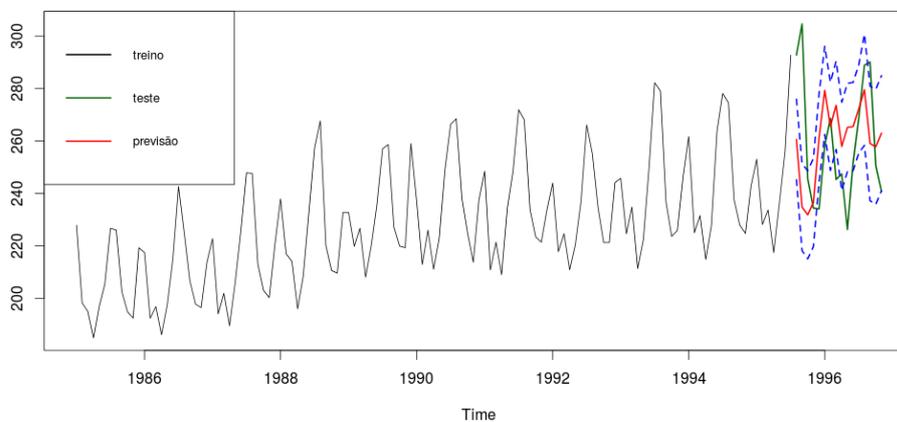


Figura 4-25 Previsão simples da série de treino uselec

Os resultados mostram a acurácia da previsão híbrida frente à previsão simples, o qual pode ser ratificado também levando em conta a análise das estatísticas residuais para a previsão *out of sample* MAPE, MAE e RMSE mostradas na tabela 4-4.

Tabela 4-4 Estatísticas residuais dos modelos aplicados à série de treino uselec

Métricas/Modelos	Modelo Simples	Modelo Híbrido
MAPE	0.0795	0.0585
MAE	20.8948	15.7154
RMSE	26.8695	21.8250

### Conclusões do capítulo

Nesse capítulo foram desenhados e analisados os experimentos com o objetivo de validar a abordagem de decomposição proposta no capítulo 3. Para isso usaram-se séries reais de: consumo de energia elétrica do Brasil, número de mortes de condutores de automóveis no Reino Unido (UKDriverDeaths), produção trimestral de ônibus no Reino Unido (ukcars) e geração total de energia elétrica nos Estados Unidos (uselec). Cada uma delas se modelou através do modelo simples e o modelo híbrido. Durante a modelagem simples, a cada série sem decompor foi ajustado um modelo ARIMA. Durante a modelagem híbrida, para cada série, a componente determinística foi ajustada ao modelo Polinomial e a estocástica foi prevista usando um modelo ARIMA, obtendo finalmente o modelo híbrido Polinomial & ARIMA. No final, os resultados foram comparados, e a análise demonstrou a importância do processo de decomposição e de modelagem individual das componentes estocásticas e determinísticas na acurácia das previsões conforme detalhado na tabela resumo 4-5. O anterior reafirma a vantagem do uso da abordagem proposta nesse trabalho.

Tabela 4-5 Tabela resumo das estatísticas residuais dos modelos de previsão *out of sample* aplicados nas séries.

Séries	Modelo Simples			Modelo Híbrido		
	MAPE	MAE	RMSE	MAPE	MAE	RMSE
Cons. Energia Elétrica do Brasil	0.0173	682854.0	879230.2	0.0138	540570.8	712621.2
UKDriverDeaths	0.1124	143.4505	182.4450	0.0868	110.9902	125.5981
ukcars	11.824	48.5336	50.8517	5.7555	23.6230	27.3263
uselec	0.0795	20.8948	26.8695	0.0585	15.7154	21.8250

## 5 Conclusões e sugestões para trabalhos futuros

### 5.1 Conclusões

Nesta dissertação foi apresentada uma nova abordagem de decomposição de séries temporais em componentes estocásticas e determinísticas. A proposta combina o método *Empirical Mode Decomposition* (EMD) e três testes de hipótese baseados no *Recurrence Quantification Analysis* (RQA). Os métodos e técnicas usados nela foram identificados através da revisão sistemática da literatura, a qual serve também de guia para futuros trabalhos na área.

A abordagem proposta baseia-se na filtragem da série e posteriormente, na análise do determinismo e/ou estocasticidade dos componentes resultantes do passo anterior. Assim, o método EMD foi utilizado para decompor o sinal em uma soma de funções de modo intrínsecas chamadas IMFs. Posteriormente, os testes baseados no número médio de pontos por segmento ( $\bar{L}$ ), na porcentagem de determinismo (%DET) e no comprimento médio de segmento (ALL), resultantes da análise da matriz de recorrência do *Recurrence Plot* (RP), são aplicados para classificar cada IMF em estocástico ou determinístico. O critério de classificação de cada IMF é robusto por quanto ele é definido levando em consideração o resultado dos três testes de hipótese. Assim, cada IMF é classificado como determinístico quando os três testes rejeitam a hipótese nula. Nos casos diferentes, os IMFs são classificados como estocásticos. No final, da combinação dos IMFs estocásticos obtém-se a componente estocástica e da soma dos IMFs determinísticos obtém-se a componente determinística.

Com relação à abordagem proposta, é importante destacar alguns pontos:

1. O uso do método de filtragem EMD não exige a introdução de nenhum parâmetro, o que constitui uma fortaleza do método. No entanto, na análise da natureza dos IMFs via testes de hipótese baseado no RP, precisa-se especificar os valores que são assumidos pelas variáveis: *embedding dimension* ( $m$ ), tempo de retardo ( $\tau$ ) e limiar de distância ( $r$ ). Para os fins desse trabalho foi adotado  $r = 5\sigma$ ; onde  $\sigma$  é o desvio padrão do ruído gaussiano,  $m = 1$  e

$\tau = 10$ , conforme sugerido por (APARICIO; POZO; SAURA, 2008).

2. Pode ser aplicada a qualquer série de tempo independentemente das características de linearidade, estocasticidade e estacionariedade.
3. O uso dos testes de hipótese fornece um padrão bem fundamentado na classificação dos IMFs e conseqüentemente na formação das componentes estocásticas e determinísticas, o que isenta a utilização da componente visual na determinação de um valor de corte (*threshold*).

Com o objetivo de demonstrar a acurácia da abordagem, foi conduzida uma avaliação da mesma no domínio de previsão visando mostrar as vantagens de modelar independentemente as componentes estocásticas e determinísticas. Para isso as séries foram modeladas de duas formas diferentes. A primeira foi feita usando a abordagem de decomposição proposta que implica modelar independentemente as componentes. Nesse caso as componentes determinísticas e estocásticas foram modeladas através dos modelos Polinomial e ARIMA respectivamente, formando o modelo híbrido Polinomial & ARIMA. A segunda foi feita na série sem decompor através do modelo ARIMA.

No processo de avaliação da abordagem foram consideradas séries reais para demonstrar a aplicabilidade da mesma em contextos reais, devido a que os cenários sintéticos muitas vezes apresentam comportamentos distantes da realidade. Nesse sentido as séries usadas nos experimentos foram consumo de energia elétrica no Brasil, número de mortes de condutores de automóveis no Reino Unido (UKDriverDeaths), produção trimestral de ônibus no Reino Unido (ukcars) e geração total de energia elétrica nos Estados Unidos (uselec).

No final, em todos os experimentos conduzidos, o modelo híbrido apresentou melhores resultados preditivos que o modelo simples. Isso demonstrou a importância da separação das componentes estocásticas e determinísticas das séries de tempo no processo de modelagem.

Embora a aplicação da abordagem proposta focou-se no processo de modelagem, considera-se que a mesma também resulta de aplicação prática no entorno de eliminação do ruído e na análise do comportamento dos sistemas.

## 5.2 Sugestões para Trabalhos Futuros

Com os resultados obtidos na realização desta dissertação, assim como o conhecimento adquirido no desenvolvimento da pesquisa, podem-se sugerir os seguintes trabalhos futuros:

- Testar a abordagem proposta em um conjunto maior de séries temporais.
- Desenvolver um algoritmo computacional que escolha automaticamente os valores ótimos das variáveis *embedding dimension* ( $m$ ), tempo de retardo ( $\tau$ ) e limiar de distância ( $r$ ).
- Com relação à modelagem: utilizar algoritmos para a obtenção automática dos modelos ajustáveis a cada componente.
- Utilizar a técnica *Bootstrap* na geração de cenários da componente estocástica.

## 6 Referências bibliográficas

AMINGHAFARI, ; CHEZE, ; POGGI, J. M. Multivariate denoising using wavelets and principal component analysis. **Computational Statistics & Data Analysis**, v. 50, n. 9, p. 2381-2398, 2006.

AMMAN, S. A.; DAS, M. An efficient technique for modeling and synthesis of automotive engine sounds. **IEEE Transactions on Industrial Electronics**, n. 48, p. 225-234, 2001. ISSN 1.

APARICIO, T.; POZO, E. F.; SAURA, D. Detecting determinism using recurrence quantification analysis: Three test procedures. **Journal of Economic Behavior & Organization**, v. 65, p. 768-787, 2008.

BASTOS, B. Q.; SOUZA, R. C.; OLIVEIRA, F. L. C. Bottom-up Long-term Forecasting of Brazilian Commercial Class Electricity Consumption: First Results. **Procedia Computer Science**, v. 55, p. 388-394, 2015.

BOX, G.; JENKINS, G. M.; REINSEL,. **Time Series Analysis: Forecasting & Control**. 3rd. ed. [S.I.]: Prentice Hall, 1994.

CASDAGLI,. Chaos and deterministic versus stochastic non-linear modelling. **Royal Statistical Society. Series B (Methodological)**, p. 303-328, 1992.

CASDAGLI, M. Nonlinear prediction of chaotic time series. **Physica D: Nonlinear Phenomena**, v. 35, n. 3, p. 335-356, 1989.

CHEN, X. et al. Detecting Signals from Data with Noise: Theory and Applications.. **Journal of the Atmospheric Sciences**, v. 70, n. 5, p. 1489-1504, May 2013.

CHUNG, P. J.; VIBERG, M.; MECKLENBRÄUKER, C. F. Broadband ML estimation under model order uncertainty. **Signal Processing**, v. 90, n. 5, p. 1350-1356, 2010.

CORDEIRO, C.; NEVES, M. M. Forecast Intervals with Boot. EXPOS. **New Advances in Statistical Modeling and Applications**. Springer International Publishing, p. 249-256, 2014.

D'ALESSANDRO, C.; YEGNANARAYANA, B.; DARSINOS, V. Decomposition of speech signals into deterministic and stochastic components. **In Acoustics, Speech, and Signal Processing, ICASSP-95., 1995 International Conference**, p. Vol 1, pp 76, 1995.

ECKMANN, J. P.; KAMPHORST, S. O.; RUELLE,. Recurrence plots of dynamical systems. **EPL (Europhysics Letters)**, v. 4, n. 9, 973, 1987.

ESPASA, A.; CANELO, J. R. Caracterización de los aspectos esenciales de un fenómeno económico mediante técnicas estadísticas de extracción de señales. **Métodos cuantitativos para el análisis de la coyuntura económica**, Madrid, n. Alianza Editorial, p. 255-324, 1996.

GAUTAMA, T.; MANDIC, D. P.; VAN HULLE, M. M. The delay vector variance method for detecting determinism and nonlinearity in time series. **Physica D**, v. 190, n. 3, p. 167-176, 2004a.

GAUTAMA, T.; MANDIC, P.; VAN HULLE, M. A novel method for determining the nature of time series. **IEEE Transactions on Biomedical Engineering**, v. 51, n. 5, p. 728-736, 2004b.

GOLYANDINA, N.; NEKNETKIN, V.; ZHIGLJAVSKY, A. **Analysis of time series structure: SSA and related techniques**. [S.l.]: Chapman & Hall / CRC, 2001.

HAN, M.; LIU, Y. Noise reduction method for chaotic signals based on dual-wavelet and spatial correlation. **Expert Systems with Applications**, v. 6, n. 36, p. 10060-10067, 2009.

HASSANI, H. Singular Spectrum Analysis: Methodology and comparison. **Journal of Data Science**, Islamic Republic of Iran, v. 5(2), p. 239-257, 2007.

HUANG, E. et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. **Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences. The Royal Society**, United States, v. Vol. 454., n. No. 1971, p. 903-995, 1998.

HUANG, H. C.; CRESSIE, N. Deterministic/stochastic wavelet decomposition for recovery of signal from noisy data. **Technometrics**, No. 3 (Aug., 2000), pp. 262-276, v. 42, n. 3, p. 262-276, agosto 2000.

HUANG, N. E. et al. A new spectral representation of earthquake data: Hilbert spectral analysis of station TCU129, Chi-Chi, Taiwan, 21 September 1999. **Bulletin of the Seismological Society of America**, 91, 5, p. 1310-1338, 2001.

HYNDMAN, R. J.; KHANDAKAR,. Automatic time series for forecasting: the forecast package for R, v. 6/07, n. Monash University, Department of Econometrics and Business Statistics., 2007.

KANTZ, H. A robust method to estimate the maximal Lyapunov exponent of a time series. **Physics letters A**, v. 185, n. 1, p. 77-87, 1994.

KLEIBER, C.; ZEILEIS,. **Applied econometrics with R**. [S.l.]: Springer Science & Business Media, 2008.

KOPSINIS, Y.; MCLAUGHLIN, S. Development of EMD-based denoising methods inspired by wavelet thresholding. **IEEE Transactions on Signal Processing**, v. 57, p. 1351-1362, April 2009. ISSN 4.

LEE, T.; QUARDA, T. B. M. J. An EMD and PCA hybrid approach for separating noise from signal, and signal in climate change detection. **International Journal of Climatology**, v. 32, n. 4, p. 624-634, 2012.

LEI, Y. et al. A review on empirical mode decomposition in fault diagnosis of rotating machinery. **Mechanical Systems and Signal Processing**, p. 108-126, 2013.

LIMA, et al. Empirical Mode Decomposition: a novel technique for the study of tremor time series. **Medical and biological engineering**, p. 569-582, 2006.

LIN, S. L.; TUNG, P. C.; HUANG, N. E. Data analysis using a combination of independent component analysis and empirical mode decomposition. **Physical Review E** **79(6)** **066705**, 2009.

LISZKA, L.; HOLMSTRÖM, J. Extraction of a deterministic component from ROSAT X-ray data using a wavelet transform and the principal component analysis. **Astronomy and Astrophysics Supplement Series**, v. 140, n. 1, p. 125-134, 1999.

LIU, Y.; LIAO, J. Adaptive chaotic noise reduction method based on dual-lifting wavelet. **Expert Systems with Applications**, v. 38, n. 3, p. 1346-1355, 2011.

MAÇAIRA, P. M.; OLIVEIRA, F. L. C.; SOUZA, R. C. **Applying SSA. Boot Procedure To Forecast Time Series**. PUC-Rio. Rio de Janeiro. 2015a.

MAÇAIRA, P. M.; SOUZA, R. C.; OLIVEIRA, F. C. Modelling and Forecasting the Residential Electricity Consumption in Brazil with Pegels Exponential Smoothing Techniques. **Procedia Computer Science**, v. 55, p. 328-335, 2015b.

MALDONADO PORTILLO, J. Instituto de Física UNAM. **Instituto de Física UNAM**, 22 junio 2012. Disponível em: <[http://www.fisica.unam.mx/noticias\\_descseriestiempo2012.php](http://www.fisica.unam.mx/noticias_descseriestiempo2012.php)>. Acesso em: 26 fevereiro 2015.

MANDIC, D. P.; WU, Z.; HUANG, N. E. Empirical mode decomposition-based time-frequency analysis of multivariate signals: the power of adaptive data analysis. **Signal Processing Magazine, IEEE**, v. 30(6), p. 74-86, 2013.

MINERVA, T. Wavelet filtering for prediction in time series analysis. **Non-Linear Systems & Wavelet Analysis**, p. 89-94, 2010. ISSN ISBN: 978-960-474-189-2.

MORETTIN, P. A.; TOLOI, C. **Análise de séries temporais**. [S.l.]: Blucher, 2006.

NOUNOU, M. N. Multiscale finite impulse response modeling. **Engineering Applications of Artificial Intelligence**, v. 19, n. 3, p. 289-304, 2006.

R. CORE TEAM. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Viena, Austria, 2014.

RAMIREZ CASTRO, I.; MONTEJO, A. Transformada de Hilbert,descomposición modal empírica y sus aplicaciones en el análisis de vibraciones libres. **Int. de Desastres Naturales, Accidentes e Infraestructura Civil**, 2011.

REINSEL, C.; BOX, G. E. P.; JENKINS, G. M. **Time Series Analysis: Forecasting and Control**. 3. ed. [S.l.]: Prentice Hall, 1994.

RIOS, R. A. et al. Estimating determinism rates to detect patterns in geospatial datasets. **Remote Sensing of Environment**, p. 11-20, 2015.

RIOS, R. A.; DE MELLO, R. F. A Systematic Literature Review on Decomposition Approaches to Estimate Time Series Components. **INFOCOMP Journal of Computer Science**, v. 11.3-4, p. 31-46, 2012.

RIOS, R. A.; DE MELLO, R. F. Improving time series modeling by decomposing and analyzing stochastic and deterministic influences. **Signal Processing**, v. 93, n. 11, p. 3001-3013, 2013.

RIOS, R. A.; DE MELLO, R. F. Applying Empirical Mode Decomposition and mutual information to separate stochastic and deterministic influences embedded in signals. **Signal Processing**, v. 118, p. 159-176, 2016.

SAMET, H.; MARZBANI, F. Quantizing the deterministic nonlinearity in wind speed time series. **Renewable and Sustainable Energy Reviews**, **39**, p. 1143-1154, 2014.

SHANNON, C. E. A mathematical theory of communication. **ACM SIGMOBILE Mobile Computing and Communications Review**, v. 5, n. 1, p. 3-55, 2001.

SHUMWAY, R. H.; STOFFER, D. S. **Time series analysis and its applications**. [S.l.]: Springer Science & Business Media., 2013.

SMALL, M.; TSE, C. K. Detecting determinism in time series: The method of surrogate data.. **IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications**, v. 50, p. 663-672, 2003. ISSN 5.

SMITH, L. I. **A tutorial of Principal Components Analysis**. [S.l.]. 2002.

SOARES, J.; MEDEIROS, M. C. Modeling and forecasting short-term electricity load: A comparison of methods with an application to Brazilian data. **International Journal of Forecasting**, v. 24, p. 630-644, 2008. ISSN 4.

SORIANO, D. C.; SUYAMA, R.; ATTUX, R. Blind extraction of chaotic sources from white gaussian noise based on a measure of determinism. **In Independent Component Analysis and Signal Separation**, n. Springer Berlin Heidelberg., p. 122-129, 2009.

THIEL, M. et al. Influence of observational noise on the recurrence quantification analysis. **Physica D: Nonlinear Phenomena**, v. 3, n. 171, p. 138-152, 2002.

THOME, A. M. T. et al. Sales and operations planning: A research synthesis. **Int. J.ProductionEconomics**, Rio de Janeiro, Brazil, v. 138, p. 1-13, 2012.

TZAGKARAKIS, G.; PAPADOPOULI, M.; TSAKALIDES, P. **Singular spectrum analysis of traffic workload in a large-scale wireless lan**. Proceedings of the 10th ACM Symposium on Modeling, analysis, and simulation of wireless and mobile systems. [S.l.]: ACM. 2007. p. 99-108.

TZAGKARAKIS, G.; PAPADOPOULI, M.; TSAKALIDES, P. Trend forecasting based on Singular Spectrum Analysis of traffic workload in a large-scale wireless LAN. **Performance Evaluation**, v. 66(3), p. 173-190, 2009.

WU, B. Z.; HUANG, N. E. A study of the characteristics of white noise using the empirical mode decomposition method.. **Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences**, v. 460, p. 1597-1611, 2004. ISSN 2046.

ZBILUT, J. P.; GIULIANI, A.; WEBBER, C. L. Detecting deterministic signals in exceptionally noisy environments using cross-recurrence quantification. **Physics Letters A**, v. 1, n. 246, p. 122-128, 1998a.

ZBILUT, J. P.; GIULIANI, A.; WEBBER, C. L. Recurrence quantification analysis and principal components in the detection of short complex signals. **Physics Letters A**, v. 3, n. 237, p. 131-135, 1998b.

ZBILUT, J. P.; WEBBER, C. L. Embeddings and delays as derived from quantification of recurrence plots. **Physics letters A**, v. 171, n. 3, p. 199-203, 1992.

ZHENG, J.; CHENG, J.; YANG, Y. Generalized empirical mode decomposition and its applications to rolling element bearing fault diagnosis. **Mechanical Systems and Signal Processing**, p. 136-153, 2013.

## 7 Apêndices

### Apêndice A: Resultados da aplicação da abordagem *EMD-RQAtestes* na série UKDriverDeaths

Inicialmente a série de treino UKDriverDeaths foi decomposta em IMFs e um resíduo através do EMD como apresentado na figura 7-1.

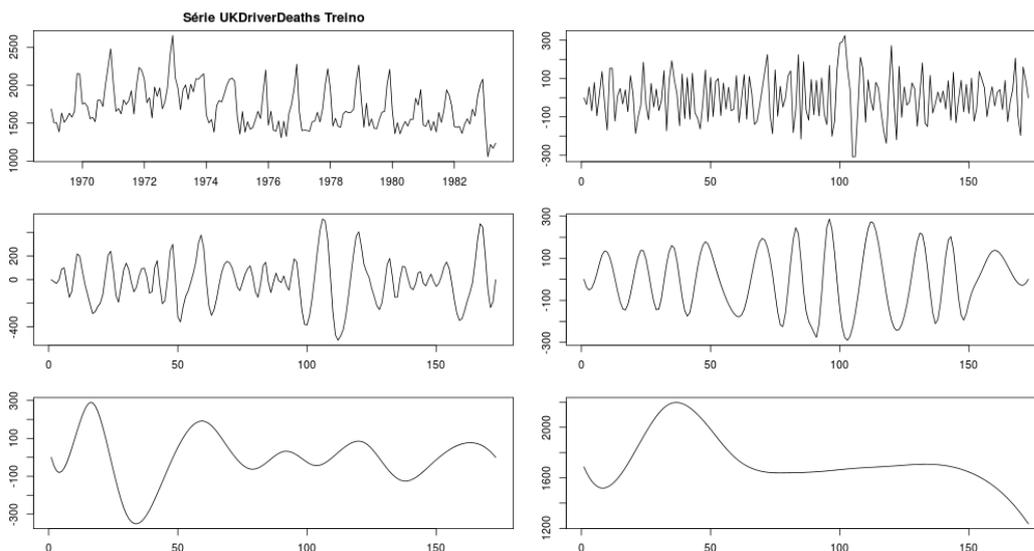


Figura 7-1 Decomposição via EMD da série UKDriverDeaths

Como resultado, se obtiveram quatro IMFs e um resíduo. Posteriormente, foi analisada a natureza de cada IMF através dos testes de hipótese com o objetivo de classificá-lo em estocástico ou determinístico. A seguir são apresentados os resultados da aplicação dos três testes de hipótese em cada IMF. O “resíduo” foi classificado automaticamente como determinístico.

#### Análise do IMF 1

Tabela 7-1 Estatísticas e seus valores esperados dos testes de hipótese para o IMF1

$\bar{L} = 1.75298$	$ALL^{-1} = 0.7492807$	$\%DET = 0.4151552$
$E(\bar{L}) = 1.381222$	$E(ALL^{-1}) = 0.7239966$	$E(\%DET) = 0.2954103$
Rejeito $H_0$	Aceito $H_0$	Rejeito $H_0$

Como estabelecido no capítulo 3, o IMF é classificado como determinístico se e só se os três testes rejeitam a hipótese nula. Nesse caso, o IMF 1 é estocástico. A seguir se apresentam os gráficos da aplicação dos testes.



Figura 7-2 Teste de hipótese do %DET do IMF1 da série de treino UKDriverDeaths

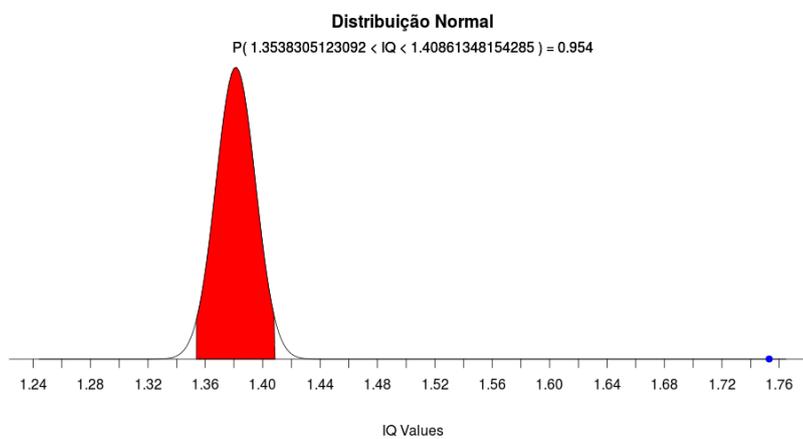


Figura 7-3 Teste de hipótese de  $L^-$  do IMF1 da série de treino UKDriverDeaths

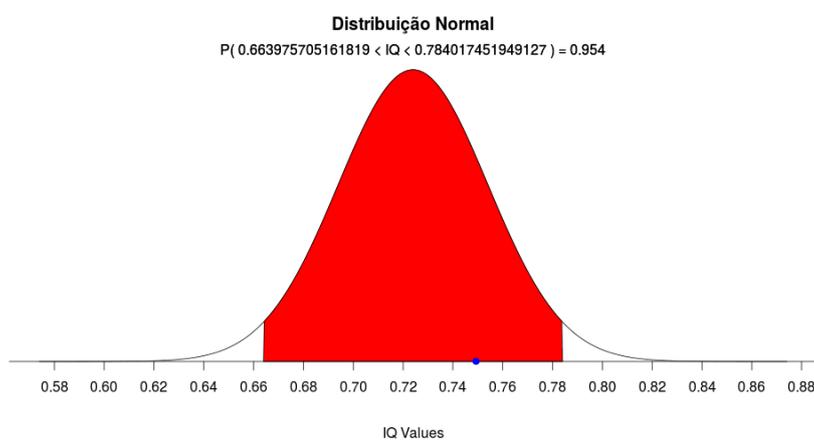


Figura 7-4 Teste de hipótese de ALL inversa do IMF1 da série de treino UKDriverDeaths

### Análise do IMF 2

Tabela 7-2 Estatísticas e seus valores esperados dos testes de hipótese para o IMF2

$\bar{L} = 2.038447$	$ALL^{-1} = 0.728932$	$\%DET = 0.4060699$
$E(\bar{L}) = 1.407722$	$E(ALL^{-1}) = 0.7103678$	$E(\%DET) = 0.3109116$
Rejeito $H_0$	Aceito $H_0$	Rejeito $H_0$

Seguindo o critério definido anteriormente, o IMF 2 é classificado como estocástico. Os gráficos dos testes se apresentam a continuação.

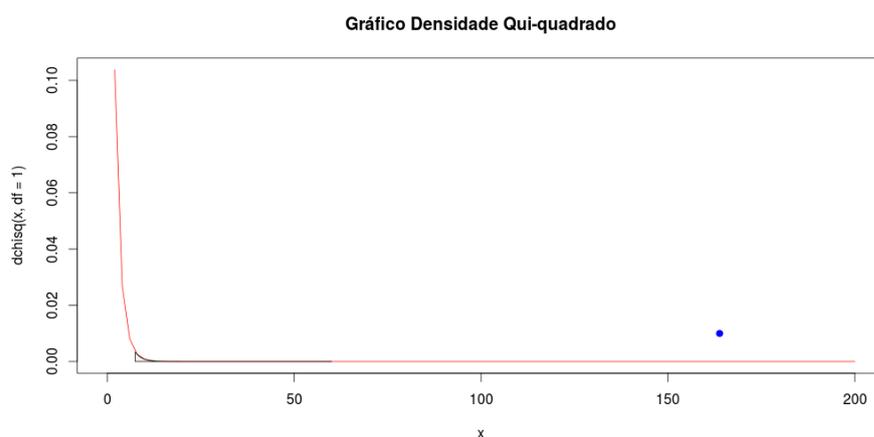


Figura 7-5 Teste de hipótese do %DET do IMF2 da série de treino UKDriverDeaths

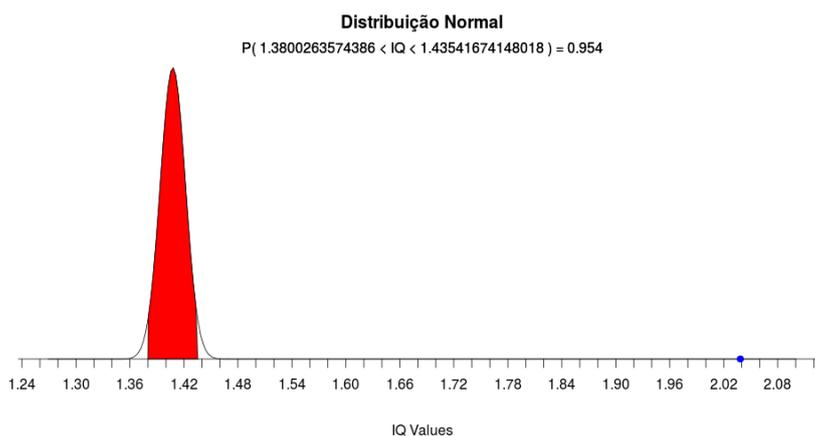


Figura 7-6 Teste de hipótese de  $\bar{L}$  do IMF2 da série de treino UKDriverDeaths

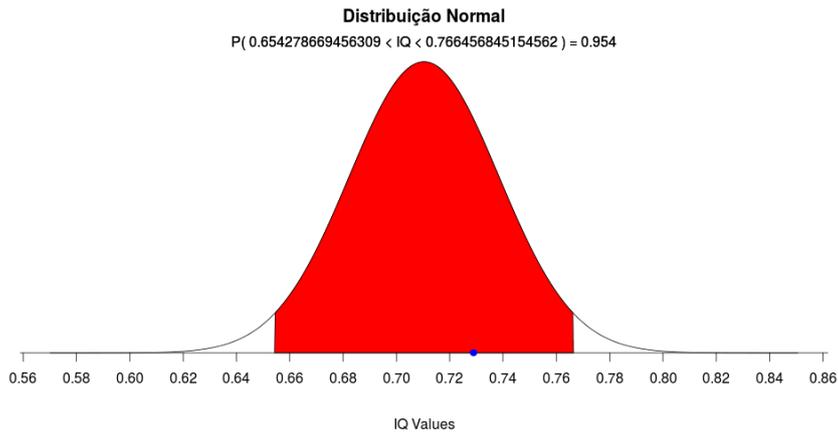


Figura 7-7 Teste de hipótese de ALL inversa do IMF2 da série de treino UKDriverDeaths

### Análise do IMF 3

Tabela 7-3 Estatísticas e seus valores esperados dos testes de hipótese para o IMF3

$\bar{L} = 2.687046$	$ALL^{-1} = 0.6528431$	$\%DET = 0.5131854$
$E(\bar{L}) = 1.36398$	$E(ALL^{-1}) = 0.7331486$	$E(\%DET) = 0.285041$
Rejeito $H_0$	Rejeito $H_0$	Rejeito $H_0$

Segundo as estatísticas, pode-se classificar o IMF3 como determinístico. A seguir são mostrados os gráficos dos testes.

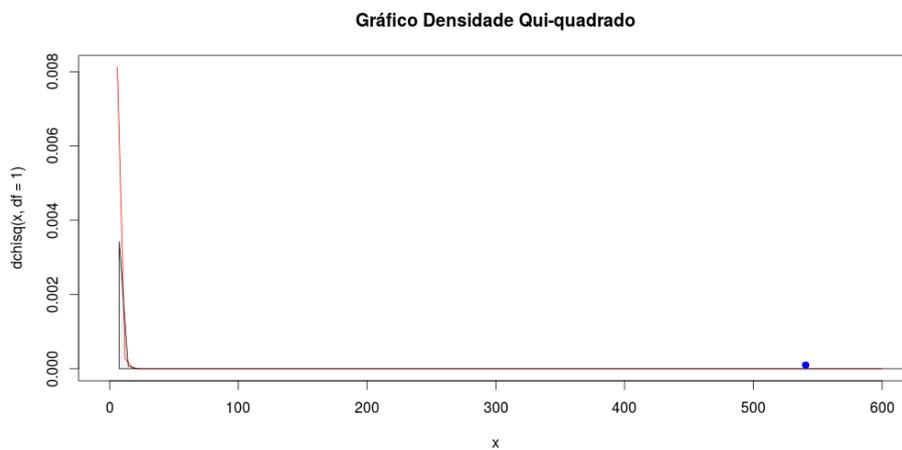


Figura 7-8 Teste de hipótese do %DET do IMF3 da série de treino UKDriverDeaths

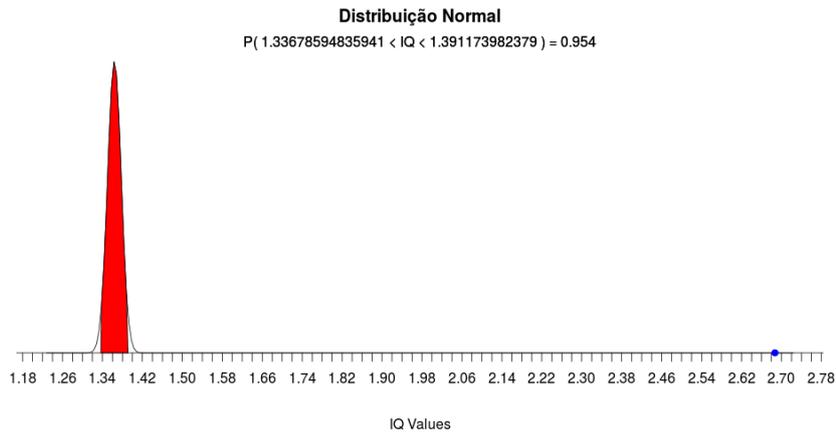


Figura 7-9 Teste de hipótese de  $\bar{L}$  do IMF3 da série de treino UKDriverDeaths

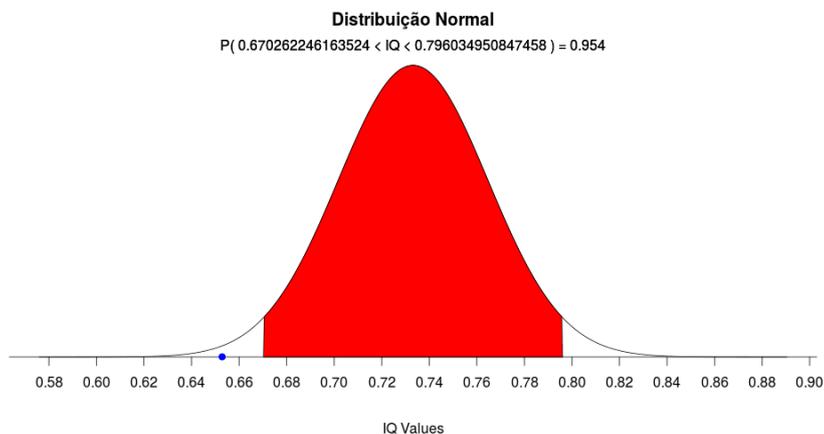


Figura 7-10 Teste de hipótese de ALL inversa do IMF3 da série de treino UKDriverDeaths

### **Análise do IMF 4**

Tabela 7-4 Estatísticas e seus valores esperados dos testes de hipótese para o IMF4

$\bar{L} = 7.656402$	$ALL^{-1} = 0.2933549$	$\%DET = 0.9378151$
$E(\bar{L}) = 1.387702$	$E(ALL^{-1}) = 0.7206156$	$E(\%DET) = 0.2992493$
Rejeito $H_0$	Rejeito $H_0$	Rejeito $H_0$

Segundo os valores das estatísticas anteriores, pode-se classificar o IMF4 como determinístico. A seguir são mostrados os gráficos dos testes.

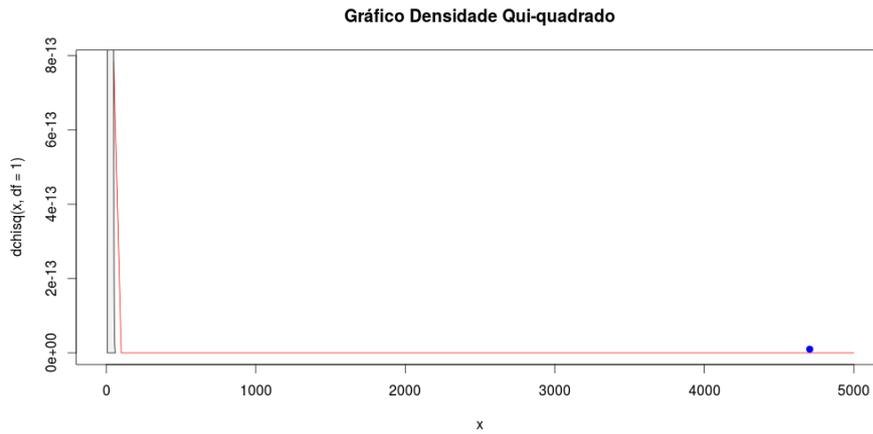


Figura 7-11 Teste de hipótese do %DET do IMF4 da série de treino UKDriverDeaths

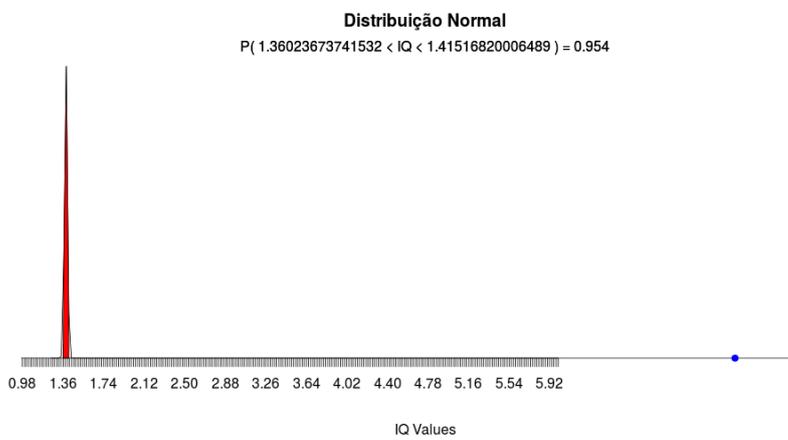


Figura 7-12 Teste de hipótese de  $L^-$  do IMF4 da série de treino UKDriverDeaths

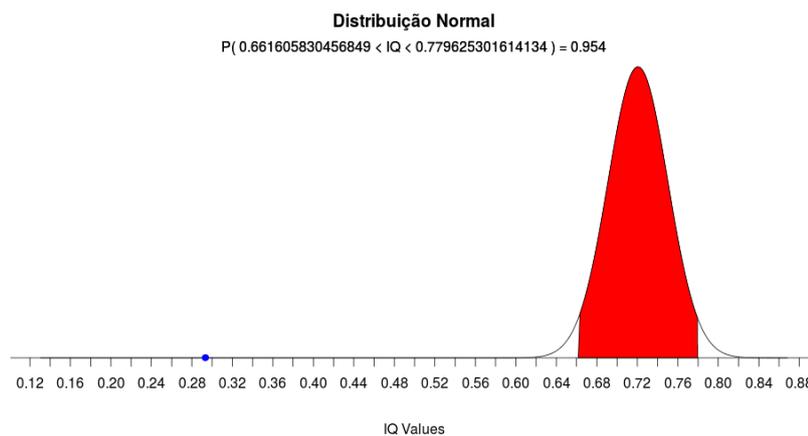


Figura 7-13 Teste de hipótese de ALL inversa do IMF4 da série de treino UKDriverDeaths

## Análise do Resíduo

Tabela 7-5 Estatísticas e seus valores esperados dos testes de hipótese para o resíduo

$\bar{L} = 44.31539$	$ALL^{-1} = 0.08118081$	$\%DET = 0.9959383$
$E(\bar{L}) = 1.85867$	$E(ALL^{-1}) = 0.5380192$	$E(\%DET) = 0.510893$
Rejeito $H_0$	Rejeito $H_0$	Rejeito $H_0$

Segundo os valores das estatísticas anteriores, pode-se classificar o resíduo como determinístico conforme sugerido por (RIOS; DE MELLO, 2013). A seguir, é mostrada a componente estocástica e a componente determinística da série UKDriverDeaths.

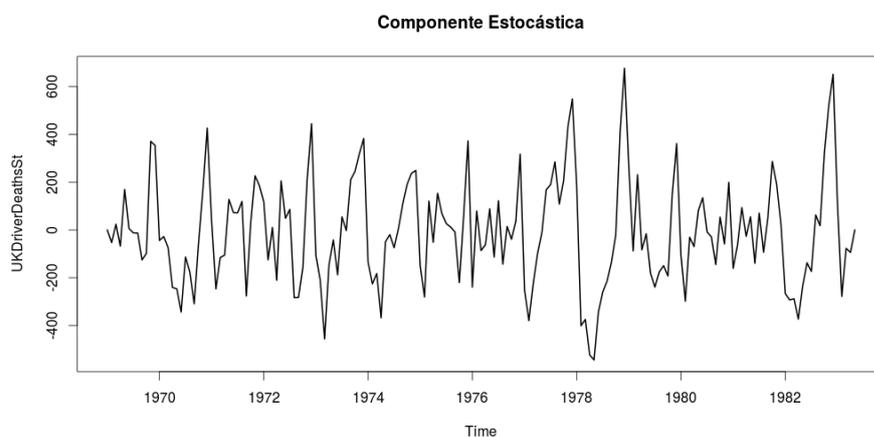


Figura 7-14 Componente estocástica da série UKDriverDeaths

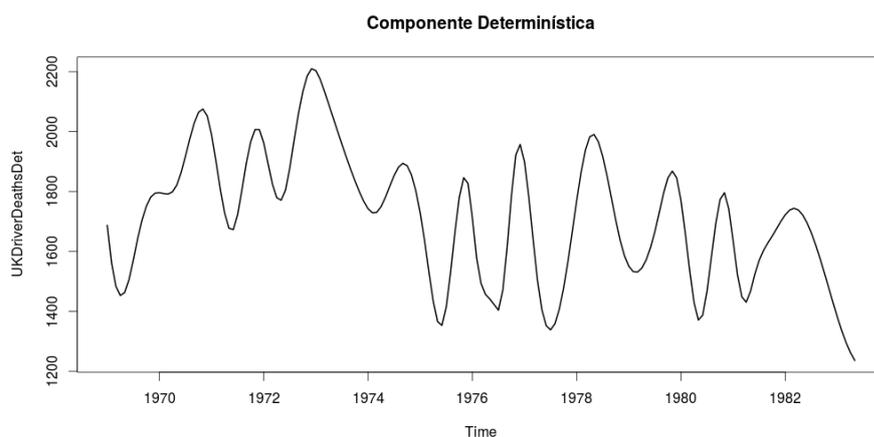


Figura 7-15 Componente determinística da série UKDriverDeaths

## Apêndice B: Resultados da aplicação da abordagem EMD-RQA testes na série consumo de energia elétrica do Brasil

Inicialmente a série de treino consumo de energia elétrica do Brasil foi decomposta em IMFs e um resíduo através do EMD como apresentado na figura 7-16.

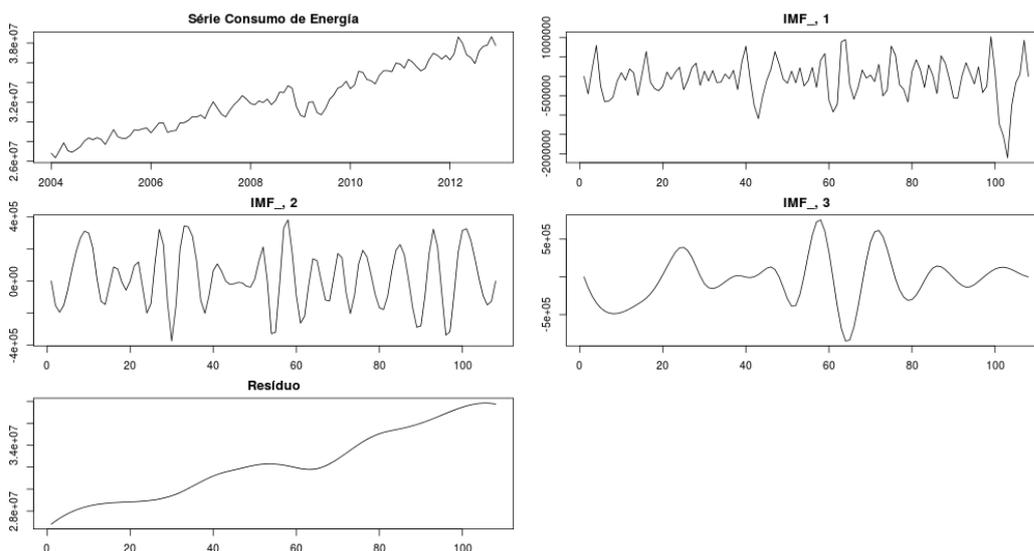


Figura 7-16 Decomposição via EMD da série Consumo de energia elétrica do Brasil

Posteriormente, foi analisada a natureza dos 3 IMFs através dos testes de hipótese com o objetivo de classificá-lo em estocástico ou determinístico. A seguir são apresentados os resultados da aplicação dos três testes de hipótese em cada IMF. Para uma simplificação serão omitidas as gráficas dos testes. O “resíduo” foi classificado automaticamente como determinístico.

### Análise do IMF 1

Tabela 7-6 Estatísticas e seus valores esperados dos testes de hipótese para o IMF 1

$\bar{L} = 1.6877$	$ALL^{-1} = 0.7713$	$\%DET = 0.3707$
$E(\bar{L}) = 1.4375$	$E(ALL^{-1}) = 0.6956$	$E(\%DET) = 0.3284$
Rejeito $H_0$	Aceito $H_0$	Aceito $H_0$

Segundo os valores das estatísticas anteriores, pode-se classificar o IMF1 como estocástico, devido a que os testes  $ALL^{-1}$  e  $\%DET$  aceitam a hipótese nula.

### **Análise do IMF 2**

Tabela 7-7 Estatísticas e seus valores esperados dos testes de hipótese para o IMF 2

$\bar{L} = 1.7293$	$ALL^{-1} = 0.7997$	$\%DET = 0.2781$
$E(\bar{L}) = 1.3518$	$E(ALL^{-1}) = 0.7397$	$E(\%DET) = 0.3080$
Rejeito $H_0$	Aceito $H_0$	Aceito $H_0$

Segundo os valores das estatísticas anteriores, pode-se classificar o IMF2 como estocástico, devido a que os testes  $ALL^{-1}$  e  $\%DET$  aceitam a hipótese nula.

### **Análise do IMF 3**

Tabela 7-8 Estatísticas e seus valores esperados dos testes de hipótese para o IMF 3

$\bar{L} = 4.3238$	$ALL^{-1} = 0.5270$	$\%DET = 0.6792$
$E(\bar{L}) = 1.3911$	$E(ALL^{-1}) = 0.7188$	$E(\%DET) = 0.3018$
Rejeito $H_0$	Rejeito $H_0$	Rejeito $H_0$

Segundo os valores das estatísticas anteriores, pode-se classificar o IMF3 como determinístico, devido a que os tres testes rejeitam a hipótese nula.

### **Análise do resíduo**

Tabela 7-9 Estatísticas e seus valores esperados dos testes de hipótese para o resíduo

$\bar{L} = 22.7580$	$ALL^{-1} = 0.0626$	$\%DET = 0.9990$
$E(\bar{L}) = 1.2603$	$E(ALL^{-1}) = 0.7934$	$E(\%DET) = 0.2179$
Rejeito $H_0$	Rejeito $H_0$	Rejeito $H_0$

Segundo os valores das estatísticas anteriores, pode-se classificar o resíduo como determinístico conforme sugerido por (RIOS; DE MELLO, 2013). A seguir, é mostrada a componente estocástica e a componente determinística da série consumo de energia elétrica do Brasil.

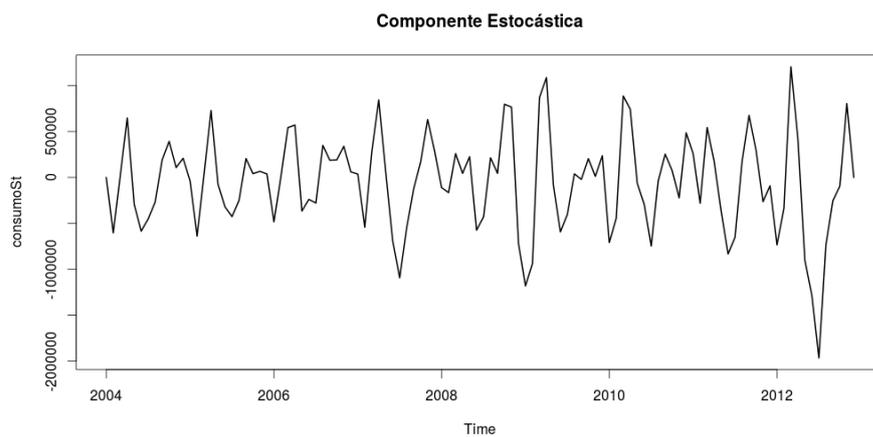


Figura 7-17 Componente estocástica da série consumo de energia elétrica do Brasil

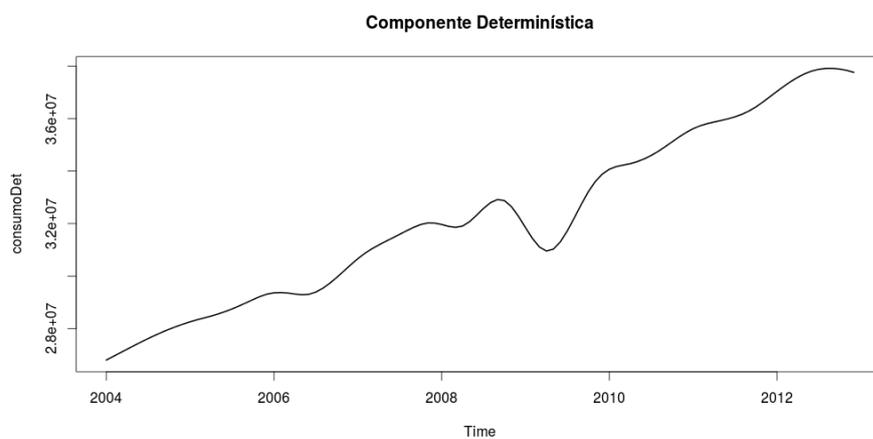


Figura 7-18 Componente determinística da série consumo de energia elétrica do Brasil

## Apêndice C: Resultados da aplicação da abordagem EMD-RQA testes na série ukcars

A série de treino ukcars foi decomposta em IMFs e um resíduo através do EMD como apresentado na figura 7-19.

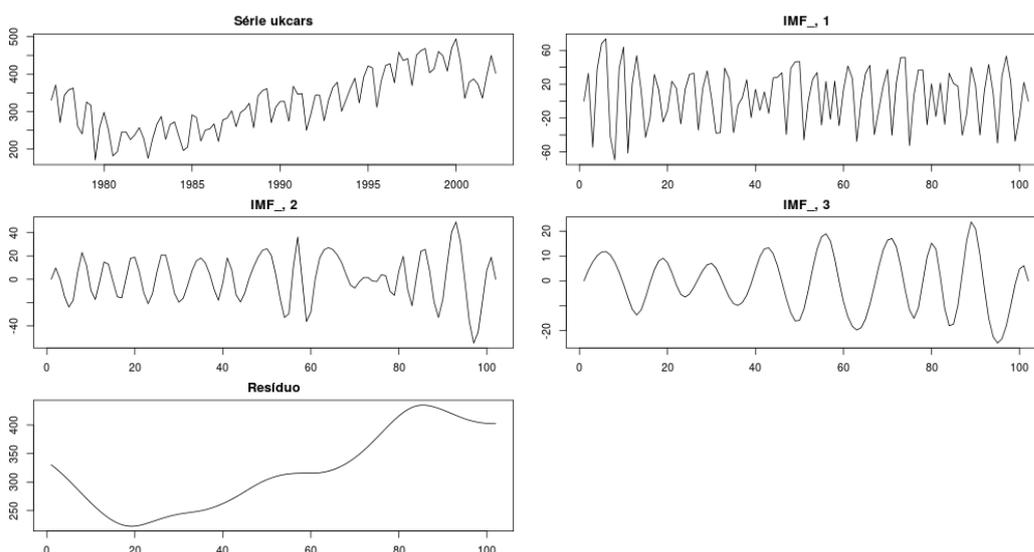


Figura 7-19 Decomposição via EMD da série ukcars

Similarmente as análises das séries anteriores, foi estudada a natureza dos 3 IMFs através dos testes de hipótese com o objetivo de classificá-lo em estocástico ou determinístico. A seguir são apresentados os resultados da aplicação dos três testes de hipótese em cada IMF. Foram também omitidas as gráficas dos testes.

### Análise do IMF 1

Tabela 7-10 Estatísticas e seus valores esperados dos testes de hipótese para o IMF1

$\bar{L} = 1.6055$	$ALL^{-1} = 0.7899$	$\%DET = 0.3606$
$E(\bar{L}) = 1.4019$	$E(ALL^{-1}) = 0.7133$	$E(\%DET) = 0.3082$
Rejeito $H_0$	Aceito $H_0$	Rejeito $H_0$

Segundo os valores das estatísticas anteriores, pode-se classificar o IMF1 como estocástico, devido a que o teste  $ALL^{-1}$  aceita a hipótese nula.

### Análise do IMF 2

Tabela 7-11 Estatísticas e seus valores esperados dos testes de hipótese para o IMF2

$\bar{L} = 1.8502$	$ALL^{-1} = 0.7700$	$\%DET = 0.3530$
$E(\bar{L}) = 1.3705$	$E(ALL^{-1}) = 0.7296$	$E(\%DET) = 0.2896$
Rejeito $H_0$	Aceito $H_0$	Aceito $H_0$

Segundo os valores das estatísticas anteriores, pode-se classificar o IMF2 como estocástico, devido a que os testes  $ALL^{-1}$  e  $\%DET$  aceitam a hipótese nula.

### **Análise do IMF 3**

Tabela 7-12 Estatísticas e seus valores esperados dos testes de hipótese para o IMF3

$\bar{L} = 2.1924$	$ALL^{-1} = 0.7164$	$\%DET = 0.4278$
$E(\bar{L}) = 1.3474$	$E(ALL^{-1}) = 0.7421$	$E(\%DET) = 0.2754$
Rejeito $H_0$	Rejeito $H_0$	Rejeito $H_0$

Segundo os valores das estatísticas anteriores, pode-se classificar o IMF3 como determinístico, devido a que os três testes rejeitam a hipótese nula.

### **Análise do resíduo**

Tabela 7-13 Estatísticas e seus valores esperados dos testes de hipótese para o resíduo

$\bar{L} = 18.0195$	$ALL^{-1} = 0.1354$	$\%DET = 0.9941$
$E(\bar{L}) = 1.3103$	$E(ALL^{-1}) = 0.7631$	$E(\%DET) = 0.2518$
Rejeito $H_0$	Rejeito $H_0$	Rejeito $H_0$

A análise do resíduo ratifica o que anteriormente foi explicado, o resíduo no constitui um resíduo em se, senão uma componente determinística. A seguir são apresentadas as gráficas da componente estocástica e da componente determinística.

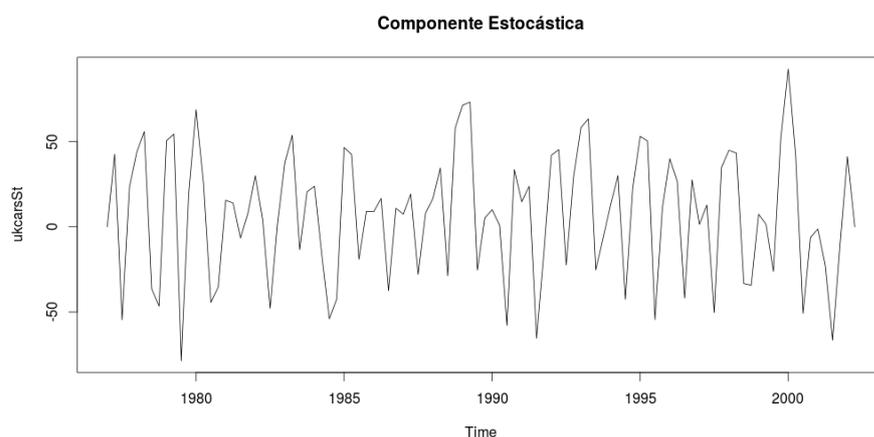


Figura 7-20 Componente estocástica da série ukcars

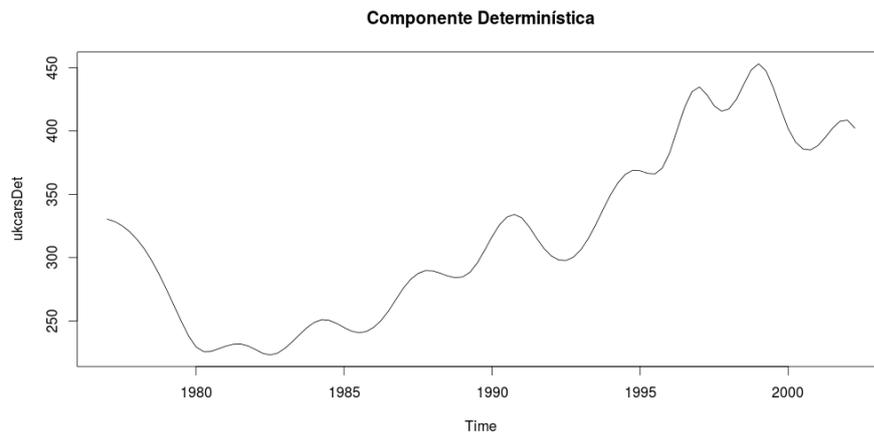


Figura 7-21 Componente determinística da série ukcars

## Apêndice D: Resultados da aplicação da abordagem EMD-RQAtestes na série uselc

A série de treino uselec foi decomposta em IMFs e um resíduo através do EMD como apresentado na figura 7-22

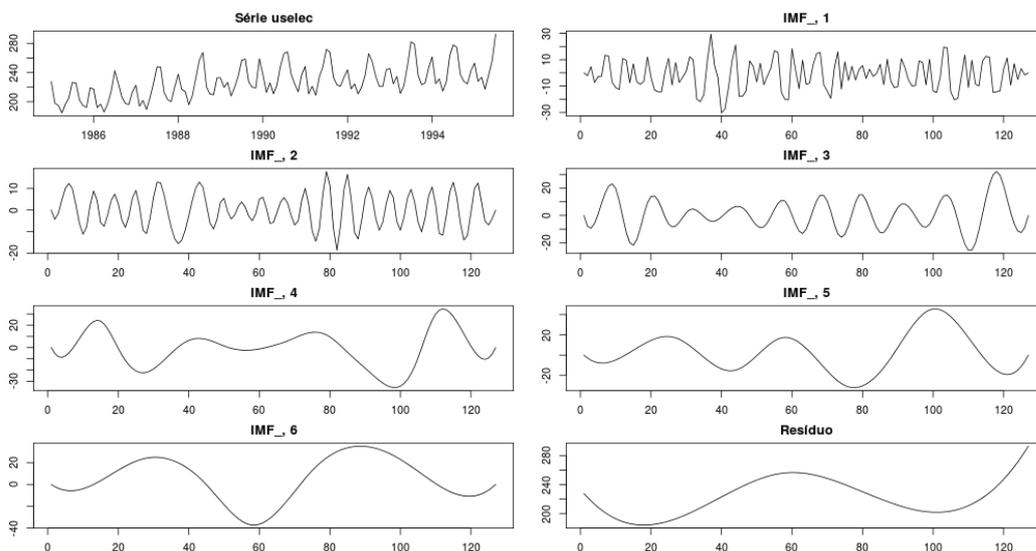


Figura 7-22 Decomposição via EMD da série uselec

Da decomposição se obtiveram 6 IMFs e um resíduo, os quais são analisados através dos testes de hipótese para determinar sua natureza. A seguir são apresentados os resultados obtidos para cada IMF. Ao igual que na série anterior as gráficas dos testes foram omitidas.

### Análise do IMF 1

Tabela 7-14 Estatísticas e seus valores esperados dos testes de hipótese para o IMF1

$\bar{L} = 1.7752$	$ALL^{-1} = 0.7347$	$\%DET = 0.4647$
$E(\bar{L}) = 1.3774$	$E(ALL^{-1}) = 0.7259$	$E(\%DET) = 0.2935$
Rejeito $H_0$	Aceito $H_0$	Rejeito $H_0$

Segundo os valores das estatísticas anteriores, pode-se classificar o IMF1 como estocástico, devido a que o teste  $ALL^{-1}$  aceita a hipótese nula.

### Análise do IMF 2

Tabela 7-15 Estatísticas e seus valores esperados dos testes de hipótese para o IMF2

$\bar{L} = 2.2819$	$ALL^{-1} = 0.7855$	$\%DET = 0.3266$
$E(\bar{L}) = 1.3610$	$E(ALL^{-1}) = 0.7347$	$E(\%DET) = 0.2835$
Rejeito $H_0$	Aceito $H_0$	Rejeito $H_0$

Segundo os valores das estatísticas anteriores, pode-se classificar o IMF2 como estocástico, devido a que o teste  $ALL^{-1}$  aceita a hipótese nula.

### **Análise do IMF 3**

Tabela 7-16 Estatísticas e seus valores esperados dos testes de hipótese para o IMF3

$\bar{L} = 2.7057$	$ALL^{-1} = 0.6897$	$\%DET = 0.4376$
$E(\bar{L}) = 1.3788$	$E(ALL^{-1}) = 0.7252$	$E(\%DET) = 0.2943$
Rejeito $H_0$	Rejeito $H_0$	Rejeito $H_0$

Segundo os valores das estatísticas anteriores, pode-se classificar o IMF3 como determinístico, devido a que os três testes rejeitam a hipótese nula.

### **Análise do IMF 4**

Tabela 7-17 Estatísticas e seus valores esperados dos testes de hipótese para o IMF4

$\bar{L} = 15.2265$	$ALL^{-1} = 0.3469$	$\%DET = 0.8451$
$E(\bar{L}) = 1.4072$	$E(ALL^{-1}) = 0.7105$	$E(\%DET) = 0.3110$
Rejeito $H_0$	Rejeito $H_0$	Rejeito $H_0$

Segundo os valores das estatísticas anteriores, pode-se classificar o IMF4 como determinístico, devido a que os três testes rejeitam a hipótese nula.

### **Análise do IMF 5**

Tabela 7-18 Estatísticas e seus valores esperados dos testes de hipótese para o IMF5

$\bar{L} = 6.0105$	$ALL^{-1} = 0.3183$	$\%DET = 0.9450$
$E(\bar{L}) = 1.3199$	$E(ALL^{-1}) = 0.7576$	$E(\%DET) = 0.2577$
Rejeito $H_0$	Rejeito $H_0$	Rejeito $H_0$

Segundo os valores das estatísticas anteriores, pode-se classificar o IMF5 como determinístico, devido a que os três testes rejeitam a hipótese nula.

### **Análise do IMF 6**

Tabela 7-19 Estatísticas e seus valores esperados dos testes de hipótese para o IMF6

$\bar{L} = 8.6966$	$ALL^{-1} = 0.2270$	$\%DET = 0.9872$
$E(\bar{L}) = 1.3051$	$E(ALL^{-1}) = 0.7662$	$E(\%DET) = 0.2481$
Rejeito $H_0$	Rejeito $H_0$	Rejeito $H_0$

Segundo os valores das estatísticas anteriores, pode-se classificar o IMF6 como determinístico, devido a que os três testes rejeitam a hipótese nula.

### Análise do resíduo

$\bar{L} = 20.7079$	$ALL^{-1} = 0.1795$	$\%DET = 0.9941$
$E(\bar{L}) = 1.3590$	$E(ALL^{-1}) = 0.7358$	$E(\%DET) = 0.2823$
Rejeito $H_0$	Rejeito $H_0$	Rejeito $H_0$

A análise do resíduo ratifica que o mesmo constitui parte da componente determinística. A seguir são apresentadas as gráficas da componente estocástica e da componente determinística.

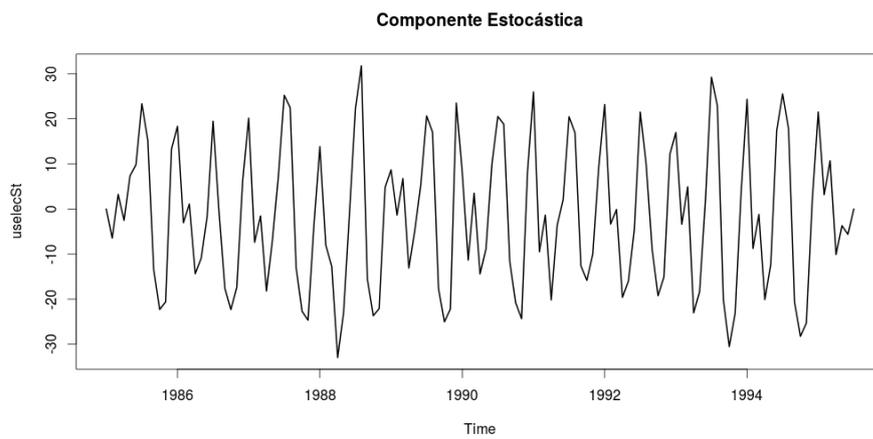


Figura 7-23 Componente estocástica da série uselec

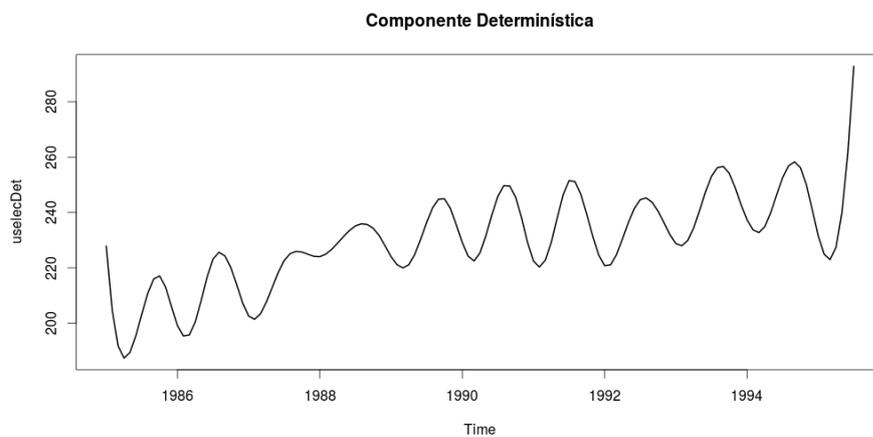


Figura 7-24 Componente determinística da série uselec