

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO  
CENTRO TECNOLÓGICO  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA AMBIENTAL

KENNEDY SCOPEL GOMES

**Modelagem INAR(p) para previsão de índices de qualidade  
do ar**

Vitória  
2009

KENNEDY SCOPEL GOMES

**Modelagem INAR(p) para previsão de índices de qualidade  
do ar**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Ambiental do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do grau de Mestre em Engenharia Ambiental, na área de concentração Poluição do ar. Orientador: Prof. Dr. Valdério Anselmo Reisen. Co-orientador: Prof. Dra Jane Meri Santos.

Vitória  
2009

Gxxx Gomes, Kennedy Scopel

Modelagem INAR(p) para previsão de índices de qualidade do ar.  
/ Kennedy Scopel Gomes – 2009. 75f.

Orientadores: Valdério A. Reisen, Jane M. Santos.

Dissertação (mestrado) - Universidade Federal do Espírito Santo,  
Centro Tecnológico.

1. Séries temporais. 2. Índices de qualidade do ar. 3.Dados de contagem  
4.Modelagem INAR. 5.Poluição do ar

KENNEDY SCOPEL GOMES

**Modelagem INAR(p) para previsão de índices de qualidade  
do ar**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Ambiental do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do grau de Mestre em Engenharia Ambiental, na área de concentração Poluição do ar

Aprovada em        de        de 2009.

**COMISSÃO EXAMINADORA**

---

Prof. Dr. Valdério Anselmo Reisen - UFES

---

Prof. Dra Jane Meri Santos - UFES

---

Prof. Dr. Neyval Costa Reis Jr. - UFES

---

Prof. Dra. Glaura Conceição Franco - UFMG

*Dedico este trabalho,  
a DEUS, pelo amor sem igual.  
A minha esposa pelo exemplo de vida.  
Aos meus pais pelo apoio incondicional.*

*“Tu és fiel Senhor  
Tu és fiel Senhor  
dia após dia com bênçãos sem fim.  
Tua mercê me sustenta e me guarda.  
Tu és fiel Senhor, fiel a mim.”*  
**T.O.Chisholm**

## Agradecimentos

Agradeço essa conquista, primeiramente a Deus, por sua graça em minha vida.

À minha esposa Ana Paula, por me mostrar como lutar e vencer. Ana Paula, te amo muito.

Aos meus Pais João e Maria, pelo exemplo de honestidade, perseverança, dignidade e pelo amor sem medida.

Aos meus Irmãos Ellen e Fernando. Amo vocês dois.

Ao meu orientador Prof. Valdério, pelas preciosas orientações, pela compreensão e força nos momentos de maiores dificuldades.

À minha digníssima Co-orientadora Jane, que me adotou nos últimos meses, sem perder a alegria de me ajudar.

Ao professor Neyval, pelas ajudas e dicas no decorrer dessa caminhada.

À professora Glaura, pela gentileza de suas preciosas contribuições.

À querida Rose Leão, por simplesmente estar sempre pronta a ajudar.

Ao meu amigo Geovane Barbosa, que graças a ele deixei de travar outras batalhas.

Aos amigos especiais: Valdir, Alexandre, Giovani Comarela, Aline, Nathaly, Edilson, Bernardo e Bartolomeu, que em momentos distintos, contribuíram para o sucesso deste trabalho.

A todos o meu muito abrigado.

## Resumo

Esta dissertação faz uso do modelo auto regressivo de valores inteiros INAR(p) para a modelagem e previsão de índices de qualidade do ar da Região da Grande Vitória (RGV). O período de análise compreendeu 01/01/07 à 19/03/07 (verão de 2007) e obteve previsões do dia 20/03/07 à 25/03/07. A Região da Grande Vitória possui oito estações de monitoramento da qualidade do ar que fornecem dados diários sobre a qualidade do ar das regiões de Laranjeiras, Carapina, Jardim Camburi, Enseada do Suá, Centro de Vitória, Ibes, Centro de Vila Velha e Cariacica. Em Laranjeiras foram investigados os poluentes  $CO$  e o  $NO_2$ , em Carapina o  $PM_{10}$ , na Enseada do Suá  $PM_{10}$ ,  $CO$  e o  $NO_2$ , no Centro de Vitória os poluentes  $PM_{10}$  e  $CO$ , no Ibes  $PM_{10}$  e  $SO_2$ , no Centro de Vila Velha os poluentes  $PM_{10}$  e  $SO_2$  e em Cariacica os poluentes  $SO_2$ ,  $NO_2$ ,  $O_3$  e  $CO$ . Para cada poluente pertencente a uma determinada estação foram gerados sete possíveis modelos. Para a escolha do modelo mais adequado foi utilizado o critério de seleção automática para modelos INAR(p) o  $AICC_{INAR}$  que seleciona a melhor ordem  $p$  para cada modelo. Todas as previsões para os índices de qualidade do ar foram classificadas como BOA conforme a Resolução CONAMA 03/90. No entanto baseados nas diretrizes da OMS 2005, a previsão do poluente  $SO_2$  da estação do Centro Vila Velha no dia 20/03/07 excedeu em termos de concentração ( $\mu g/m^3$ ) o valor de  $20 \mu g/m^3$  para uma média de 24h.

*Palavras Chave: Séries temporais de valores inteiros. Modelos INAR(p). Operador thinning. Índices de qualidade do ar. Poluição do ar.*

## Abstract

This work is based on the INteger AutoRegressive model (INAR) for air quality indices modeling and forecasting in the Grande Vitória region (RGV). The analysis period comprehended 1 january, 2007 to 19 march, 2007 (summer of 2007) and it obtained forecasts on 20 march, 2007 to 25 march, 2007. The area of Grande Vitória possesses eight air quality monitoring stations which supplies daily data about air quality in Laranjeiras, Carapina, Jardim Camburi, Enseada do Suá, downtown of Vitória, Ibes, downtown of Vila Velha, and Cariacica. In Laranjeiras pollutants as  $CO$  and  $NO_2$  were investigated, in Carapina, the  $PM_{10}$ , in Enseada do Suá,  $PM_{10}$ ,  $CO$ , and  $NO_2$ , in downtown of Vitória,  $PM_{10}$  and  $CO$ , in Ibes,  $PM_{10}$  and  $SO_2$ , in downtown of Vila Velha,  $PM_{10}$  and  $SO_2$ , and in Cariacica,  $SO_2$ ,  $NO_2$ ,  $O_3$  and  $CO$ . For each pollutant belonging to determined station, seven possible models were generated. The choice of the best model was based on the criterion of automatic selection for models INAR ( $p$ ) and the  $AICC_{INAR}$ , that selects the best order  $p$  for each model. All forecasts for air quality indices were classified as GOOD, according to the CONAMA 03/90 resolution. However, based on OMS 2005 guidelines, the forecast of the pollutant  $SO_2$  verified in downtown of Vila Velha on march 20, 2007 exceeded, in concentration terms ( $\mu g/m^3$ ), the value of  $20 \mu g/m^3$  for an average of 24hs.

*Keywords: Integer-valued time series. INAR( $p$ ) models. Thinning operation. Air Quality Index. Air pollution.*

---

## LISTA DE FIGURAS

6.1	Estações de monitoramento da qualidade do ar da RGV . . . . .	39
7.1	Box Plot $PM_{10}$ para Enseada do Suá . . . . .	45
7.2	Box Plot $PM_{10}$ para Ibes . . . . .	46
7.3	Modelagem da Série IQA- $CO$ Laranjeiras . . . . .	51
7.4	Modelagem da Série IQA- $NO_2$ Laranjeiras . . . . .	51
7.5	Modelagem da Série IQA- $PM_{10}$ Carapina . . . . .	52
7.6	Modelagem da Série IQA- $PM_{10}$ Enseada do Suá . . . . .	52
7.7	Modelagem da Série IQA- $CO$ Enseada do Suá . . . . .	53
7.8	Modelagem da Série IQA- $NO_2$ Enseada do Suá . . . . .	53
7.9	Modelagem da Série IQA- $PM_{10}$ Vitória -Centro . . . . .	54
7.10	Modelagem da Série IQA- $CO$ Vitória -Centro . . . . .	54
7.11	Modelagem da Série IQA- $PM_{10}$ VV-Ibes . . . . .	55
7.12	Modelagem da Série IQA- $SO_2$ VV-Ibes . . . . .	55
7.13	Modelagem da Série IQA- $PM_{10}$ VV-Centro . . . . .	56
7.14	Modelagem da Série IQA- $SO_2$ VV-Centro . . . . .	56
7.15	Modelagem da Série IQA- $SO_2$ Cariacica . . . . .	57
7.16	Modelagem da Série IQA- $NO_2$ Cariacica . . . . .	57
7.17	Modelagem da Série IQA- $O_3$ Cariacica . . . . .	58
7.18	Modelagem da Série IQA- $CO$ Cariacica . . . . .	58

---

# LISTA DE TABELAS

3.1	Diretrizes da Organização Mundial de Saúde 2005 . . . . .	12
3.2	Padrões nacionais de qualidade do ar (Resolução CONAMA 003 de 28/06/1990) . . . . .	13
3.3	CrITÉrios para episÓdios agudos de poluiÇo do ar . . . . .	13
3.4	Faixas de concentraÇo dos poluentes para o clculo do IQA. . . . .	14
6.1	EstaÇes de monitoramento da qualidade do ar da RGV, coordenadas UTM em metros . . . . .	39
6.2	Parmetros meteorolÓgicos e poluentes monitorados em cada estaÇo RAMQAR . . . . .	40
6.3	Método de amostragem . . . . .	41
7.1	Estatísticas descritivas dos índices de qualidade do ar dos principais poluentes das estaÇes de monitoramento da qualidade do ar da RGV, no perÍodo de 01/01/07 à 19/03/07. (78 dias) . . . . .	45
7.2	Poluentes que excederam as diretrizes da OMS 2005 . . . . .	46
7.3	SeleÇo da ordem do modelo INAR(p) baseados no critÉrio automtico $AICC_{INAR}$ , perÍodo das anlises 01/01/07 à 19/03/07. . . . .	48
7.4	Modelos INAR(p) ajustados . . . . .	49
7.5	Previso da qualidade do ar-(20/03/07) . . . . .	59
7.6	Previses da qualidade do ar-(21/03/07) à (25/03/07) . . . . .	60

---

# SUMÁRIO

Resumo . . . . .	I
Abstract . . . . .	II
Lista de Figuras . . . . .	III
Lista de Tabelas . . . . .	IV
<b>1 Introdução</b>	<b>1</b>
<b>2 Objetivo</b>	<b>5</b>
2.1 Objetivo Geral . . . . .	5
2.2 Objetivos Específicos . . . . .	5
<b>3 Qualidade do Ar</b>	<b>6</b>
3.1 Efeitos da Poluição Atmosférica . . . . .	7
3.2 Padrões de qualidade do ar . . . . .	11
3.3 Índice de qualidade do Ar . . . . .	14
<b>4 Modelo autoregressivo de valores inteiros e suas propriedades</b>	<b>16</b>
4.1 Modelo Auto-Regressivo de Valores Inteiros (INAR) . . . . .	16
4.1.1 Distribuição de Poisson . . . . .	16
4.1.2 Operador <i>thinning</i> binomial . . . . .	20
4.1.3 Modelo INAR(1) . . . . .	22
4.1.4 Estimador de mínimos quadrados condicional . . . . .	24
4.1.5 Modelo INAR(p) . . . . .	25
4.2 Critérios Automáticos de Seleção de Ordem . . . . .	28
4.2.1 Critérios automáticos . . . . .	28
4.2.2 Critério automático para modelos INAR(p) . . . . .	29
4.3 Previsão INAR . . . . .	30
4.3.1 Intervalo de previsão . . . . .	31
<b>5 Previsão da Qualidade do Ar</b>	<b>32</b>
5.1 Modelos Estatísticos para a modelagem da qualidade do ar . . . . .	33
5.2 Modelos INAR . . . . .	34

<b>6</b>	<b> Materiais e Métodos</b>	<b>38</b>
6.1	Região de estudo . . . . .	38
6.2	Rede automática de monitoramento da qualidade do ar da RGV . . . . .	38
6.3	Modelagem INAR . . . . .	41
<b>7</b>	<b> Resultados e Análises</b>	<b>42</b>
7.1	Análise descritiva dos dados . . . . .	42
7.2	Seleção da ordem dos modelos INAR(p) . . . . .	47
7.3	Modelos Ajustados . . . . .	49
7.4	Modelagem e previsão das séries dos índices de qualidade do ar . . . .	50
<b>8</b>	<b> Conclusão e Recomendações</b>	<b>61</b>
8.1	Conclusão . . . . .	61
8.2	Recomendações para trabalhos futuros . . . . .	62
	<b>Referências Bibliográficas</b>	<b>63</b>
	<b>Apêndice A</b>	<b>66</b>

---

---

# CAPÍTULO 1

---

## INTRODUÇÃO

A concentração de um poluente no ar é o resultado final de vários fatores, que compreendem a emissão pelas fontes, suas interações físicas (diluição) e químicas (reações) na atmosfera. A movimentação do ar e as trocas de calor que ocorrem na atmosfera são capazes de dispersar os poluentes, misturando-os eficientemente a um grande volume de ar, o que contribui para a redução dos níveis de concentração do poluente na atmosfera. A capacidade de dispersão depende da topografia e das condições meteorológicas (LIRA, 2009).

Alguns poluentes atmosféricos são produzidos como resultado de reações químicas na atmosfera, por exemplo o ozônio troposférico que é formado na atmosfera como resultado de reações químicas envolvendo poluentes orgânicos, óxidos de nitrogênio, oxigênio e luz solar. Outro importante exemplo são os hidrocarbonetos que podem reagir com outras substâncias, tais como oxigênio, nitrogênio, cloro e enxofre e formar uma grande variedade de compostos derivados como aldeídos, ácidos, álcoois, éteres, cetonas e ésteres.

A resolução CONAMA nº 3 de 28/06/90, define poluição atmosférica qualquer forma de matéria ou energia com intensidade e em quantidade, concentração, tempo ou características em desacordo com os níveis estabelecidos, e que tornem ou pos-

sam tornar o ar impróprio, nocivo ou ofensivo à saúde, inconveniente ao bem-estar público, danoso aos materiais, à fauna e à flora ou prejudicial à segurança , ao uso e gozo da propriedade e às atividades normais da comunidade.

O impacto da poluição atmosférica sobre o bem-estar humano tem sido a principal motivação para o seu estudo e controle. A poluição atmosférica tem afetado a saúde da população, mesmo quando seus níveis encontram-se aquém do que determina a legislação vigente (MARTINS et al;2001).

Os padrões nacionais de qualidade do ar são baseados na resolução CONAMA *n*º 3 de 28/06/90 que estabelece uma média diária para as partículas totais em suspensão (*PTS*) de  $240 \mu\text{g}/\text{m}^3$ , para partículas inaláveis (*PM*<sub>10</sub>) uma média de  $150 \mu\text{g}/\text{m}^3$ , para o dióxido de enxofre (*SO*<sub>2</sub>) uma média de  $365 \mu\text{g}/\text{m}^3$ , para o dióxido de nitrogênio (*NO*<sub>2</sub>) uma média de 1 hora máxima de  $320 \mu\text{g}/\text{m}^3$ , para o monóxido de carbono *CO* uma média de 1 hora máxima de  $40.000 \mu\text{g}/\text{m}^3$  e para o ozônio (*O*<sub>3</sub>) uma média de 1 hora máxima de  $160 \mu\text{g}/\text{m}^3$ .

A Organização Mundial de Saúde (OMS) apresentou em 2005, novas diretrizes para o estabelecimento de padrões de qualidade do ar. A média diária recomendada para o *PM*<sub>10</sub> é de  $50 \mu\text{g}/\text{m}^3$ , para o *O*<sub>3</sub>  $100 \mu\text{g}/\text{m}^3$  a média de 8 hora, para o *SO*<sub>2</sub> a média diária é de  $20 \mu\text{g}/\text{m}^3$  e para 10 min  $500 \mu\text{g}/\text{m}^3$  e o *NO*<sub>2</sub> de  $200 \mu\text{g}/\text{m}^3$  média de 1 hora. Demonstrando um rigor maior do que os padrões nacionais atuais.

O monitoramento da qualidade do ar é realizado para determinar o nível de concentração dos poluentes presentes na atmosfera. Seus resultados não só permitem um acompanhamento sistemático da qualidade do ar na área monitorada, como também constituem elementos básicos para elaboração de diagnósticos da qualidade do ar, subsidiando ações governamentais para o controle das emissões. O Índice de Qualidade do Ar (IQA) é um parâmetro utilizado para transformar as concentrações medidas dos diversos poluentes em um único valor inteiro não negativo e adimensional que possibilita a comparação com os limites legais de concentração (padrões

de qualidade do ar) para os diversos poluentes. O IQA tem como objetivo principal proporcionar à população o entendimento sobre a qualidade do ar local, em relação a diversos poluentes atmosféricos amostrados nas estações de monitoramento. Sendo divulgados através de um boletim da qualidade do ar das últimas 24 horas. Segundo Lira (2009) o conhecimento prévio dos níveis dos poluentes na atmosfera de uma região pode ser útil para fornecer dados para ativar ações de emergência durante períodos de estagnação atmosférica, quando os níveis de poluentes na atmosfera possam representar risco à saúde pública.

Para esse conhecimento prévio, várias estratégias de modelagens têm sido desenvolvidas e otimizadas, no intuito de gerar previsões satisfatórias das concentrações dos poluentes em diversos cenários e conseqüentemente obter seus índices de qualidade do ar. A utilização de modelos de previsão é um artifício importante para conhecer os diversos comportamentos e características de determinados poluentes, podendo dessa forma, prever possíveis picos de concentração e possíveis índices de qualidade do ar indesejáveis. Para isso, existem duas classes de modelos, os experimentais e os matemáticos. Nesta última, têm-se os modelos determinísticos e os modelos estocásticos. Este estudo se concentra na classe de modelagem estocástica.

Muitas das séries temporais observadas são séries de valores inteiros não negativos e, em particular, séries de contagens. Os modelos usuais, quer lineares quer não lineares, para séries temporais não são neste caso adequados pois o produto de uma constante real por uma variável aleatória de valor inteiro produz uma variável aleatória real. Assim, McKenzie (1986, 1988) e Al-Osh e Alzaid (1987) recorreram à operação *thinning* binomial definida por Steutal e Van Harn (1979) para substituir a operação de multiplicação usual e propuseram o Modelo Auto Regressivo de valores Inteiros (INAR). O modelo INAR respeita a característica de contagem dos dados, retornando em suas simulações apenas valores inteiros não negativos, obtendo em suas modelagens e previsões valores que acompanham as séries de contagem. Os modelos INAR explicam as características da série apenas com os dados da própria série, baseado nas características probabilísticas dos mesmos.

O presente estudo baseia-se no uso do modelo INAR para a modelagem das séries temporais de índices de qualidade do ar dos poluentes investigados nesta dissertação, dando como resposta à modelagem, a previsão da qualidade do ar para determinado período e região. Não foram encontrados estudos envolvendo o modelagem INAR em poluição do ar.

O trabalho está dividido como segue. A próxima Seção mostra os objetivos do trabalho. A Seção 3 relata sobre a qualidade do ar, seus parâmetros e padrões e ainda apresenta os principais poluentes e seus efeitos, na Seção 4 é apresentado o modelo INAR e suas propriedades. A Seção 5 apresenta uma revisão bibliográfica do estudo. A Seção 6 mostra a metodologia de modelagem. As análises e os resultados da modelagem e previsão são apresentados na Seção 7. As conclusões da pesquisa e recomendações para trabalhos futuros encontram-se na Seção 8.

---

---

# CAPÍTULO 2

---

## OBJETIVO

### 2.1 Objetivo Geral

Modelar os índices de qualidade do ar e gerar previsões, através do uso do modelo Auto Regressivo de Valor Inteiro (INAR).

### 2.2 Objetivos Específicos

- Modelar a série de valores inteiros não negativos, dito série de contagem.
- Modelar a série de índices de qualidade do Ar dos principais poluentes atmosféricos da Região da Grande Vitória.
- Realizar previsões de índices de qualidade do ar;
- Uso do Critério  $AICC_{INAR}$  na escolha da melhor ordem para o modelo INAR(p)

---

---

## CAPÍTULO 3

---

### QUALIDADE DO AR

A poluição do ar tem sido, desde a primeira metade do século XX, um grave problema nos centros urbanos industrializados, com a presença cada vez maior dos automóveis, que vieram a somar com as indústrias, como fontes poluidoras. Episódios de poluição excessiva causaram aumento do número de mortes em algumas cidades da Europa e Estados Unidos. O primeiro episódio ocorreu em 1930, no vale de Meuse, Bélgica. Nos cinco primeiros dias do mês de dezembro, condições meteorológicas desfavoráveis, como a ausência de ventos, impediram a dispersão dos poluentes, que permaneceram estacionados sobre a região. Imediatamente foi registrado um aumento do número de doenças respiratórias e um excesso de mortes (60 mortes) até dois dias após o início do episódio. Alguns anos após, um episódio semelhante ocorreu durante os últimos cinco dias do mês de outubro de 1948 na cidade de Donora, Pensilvânia. Os produtos da combustão das indústrias locais permaneceram sobre a cidade devido à ocorrência de inversões térmicas que impediram a dispersão destes poluentes. Durante este período foram observadas 20 mortes ao invés das duas mortes esperadas normalmente em uma comunidade de 14.000 pessoas (BRAGA et al 2002).

Porém o mais clássico, e mais grave, dos episódios acerca dos efeitos deletérios dos poluentes do ar foi o acontecido em Londres. Durante o inverno de 1952, um episódio de inversão térmica impediu a dispersão de poluentes, gerados então pelas indústrias e pelos aquecedores domiciliares que utilizavam carvão como combustível, e uma nu-

vem, composta principalmente por material particulado e enxofre (em concentrações até nove vezes maiores do que a média de ambos), permaneceu estacionada sobre a cidade por aproximadamente três dias, levando a um aumento de 4.000 mortes em relação à média de óbitos em períodos semelhantes (BRAGA et al 2002).

### **3.1 Efeitos da Poluição Atmosférica**

A poluição atmosférica tem vários efeitos prejudiciais, diretos ou indiretos, sobre as plantas, os materiais, o meio ambiente e a saúde de seres humanos e animais. Esses efeitos causam perdas econômicas pelo aumento da ocorrência de algumas doenças, diminuem a produção agrícola, aceleram a taxa de corrosão dos metais, aumentam o custo de sua proteção e diminuem o tempo de vida dos edifícios, construções e monumentos históricos, entre outros (LIU e LIPTAK,1999).

Os efeitos dos poluentes atmosféricos variam em função do tempo de exposição e de suas concentrações. De maneira geral, os efeitos podem ser classificados como:(a) agudos- de caráter temporário, relacionado a exposição a altas concentrações de poluentes e os efeitos são imediatos; e (b) crônicos - de caráter permanente, relacionada a exposição a baixas concentrações de poluentes e os efeitos são a longo prazo.(LIRA,2009)

#### **Efeitos sobre as plantas**

Segundo Pedroso (2007) nas plantas, os poluentes prejudicam o processo químico. Danos na membrana celular, interferência no mecanismo de abertura e fechamento de estômatos e corrosão da cutícula das folhas e acículas são alguns dos efeitos dos poluentes químicos. Geralmente, os poluentes do ar que causam danos às plantas são gasosos, como os óxidos de nitrogênio, dióxido de enxofre, hidrocarbonetos e substâncias foto-oxidantes. O efeito é direto ou indireto sobre as plantas e depende de sua concentração e período de exposição. Quando as plantas estão sujeitas a altas concentrações de poluentes, sofrem danos agudos, com sintomas exterior-

mente visíveis: despigmentação da clorofila, descoloração das folhas, necrose de áreas de tecido e órgãos ou a morte. Com baixas concentrações de poluentes, não há, de início, nenhum envenenamento exteriormente visível. Mas mudanças químicas, bioquímicas, estruturais e funcionais podem ocorrer (entupimento dos estômatos, alterações na fisiologia da planta), além da susceptibilidade a pragas e doenças.

### **Efeitos sobre os materiais**

O principal efeito da poluição atmosférica sobre os metais é a corrosão, com eventual perda de material da superfície e alterações nas propriedades elétricas do metal. A corrosão é causada principalmente, por gases ácidos, em especial  $SO_2$  (BOUBEL et al, 1994).

O ataque aos materiais de construção não metálicos ocorre pela ação do  $SO_2$  que reage com os carbonatos na presença de umidade, formando sulfatos, mais solúveis, causando deterioração do material. O  $CO_2$ , na presença de umidade, forma o ácido carbônico que converte a pedra calcária em bicarbonato, que é solúvel em água e pode ser lixiviado pela chuva (BOUBEL et al, 1994).

O  $SO_2$  afeta couro e papel e provoca significativa deterioração e enfraquecimento. Esse poluente é absorvido e convertido em  $H_2SO_4$ , que ataca a estrutura desses materiais tornando-os frágeis (BOUBEL et al, 1994).

### **Efeitos sobre o clima**

Segundo Lira (2009) os efeitos da poluição do ar sobre o meio ambiente são caracterizados, em escala global, pela redução da camada de ozônio e pelo efeito estufa.

A camada de ozônio é uma camada de ar com alta concentração de  $O_3$ , localizada de 20 a 30 km acima da superfície da Terra, na estratosfera. Ela concentra em torno de 90% do  $O_3$  da atmosfera e age como um filtro natural que protege o planeta dos níveis indesejados de radiação ultravioleta provenientes do sol. Algumas das consequências da diminuição da concentração de  $O_3$  nessa camada são : danos ao homem- catarata, câncer, queimaduras; e danos à Natureza - à vegetação e agricultura, diminuindo a

capacidade de fotossíntese e crescimento das plantas. As substâncias denominadas clorofluorcarbonos (CFC) são as responsáveis em grande parte pela eliminação do  $O_3$  da estratosfera. Estas substâncias são muito estáveis quimicamente e permanecem na atmosfera por dezenas de anos (ALVARES JR et al,2002).

O efeito estufa está relacionado ao aumento de temperatura da Terra, provocado pela maior absorção da radiação infravermelha emitida pela superfície terrestre por ela refletida, em função do aumento da concentração de determinados gases que tem essa propriedade, tais com  $CO_2$ ,  $CH_4$ ,  $N_2O$  e CFC , chamados gases do efeito estufa. Os dois principais gases do efeito estufa ( $CO_2$  e  $CH_4$ ) são provenientes da queima de combustíveis fósseis, florestas e pastagens (ALVARES JR et al,2002).

### **Efeitos sobre a saúde humana**

A poluição atmosférica tem afetado a saúde da população, mesmo quando seus níveis encontram-se aquém do que determina a legislação vigente (MARTINS et al,2001). Serão apresentados aqui os poluentes do ar cujos padrões estão estabelecidos pela resolução CONAMA e que podem provocar agravos na saúde.

#### ***Monóxido de Carbono (CO):***

O CO é um gás tóxico, inodoro, incolor e fruto da combustão de diversos processos industriais, do escapamento de veículos automotores e do fumo do cigarro. Sua toxicidade se deve em parte à sua propriedade relacionada à afinidade pela heme da hemoglobina e da mioglobina. O CO, quando absorvido pelo sangue, forma a carboxihemoglobina, que por sua vez produz uma diminuição da oxihemoglobina e uma redução do transporte de oxigênio até os tecidos. O CO possui uma afinidade de até 300 vezes maior com a hemoglobina do que o oxigênio, o que favorece a hipoxemia em pessoas expostas. O CO foi associado a aumento na mortalidade por infarto cardíaco agudo entre as pessoas idosas (ESCAMILLA ,1999).

#### ***Óxidos de Nitrogênio ( $NO_x$ ):***

No grupo genérico dos óxidos de nitrogênio ( $NO_x$ ) estão incluídos o dióxido de nitrogênio ( $NO_2$ ), o monóxido de nitrogênio (NO) e os nitratos que se formam na combustão de gasolina, carbono e petróleo e em qualquer processo de combustão a

elevadas temperaturas. O  $NO_2$ , quando inalado, atinge as porções mais periféricas do pulmão devido à sua baixa solubilidade. Seu efeito tóxico está relacionado ao fato de ser um agente oxidante (CASTRO et al,2003).

***Dióxido de Enxofre ( $SO_2$ ):***

O  $SO_2$  é um gás amarelado, com odor característico de enxofre, e muito irritante quando em contato com superfícies úmidas, pois se transforma em ácido sulfúrico. O  $SO_2$  é um contaminante produzido por queima de combustível fóssil, escapamento de veículos automotores e subproduto de processos industriais. A intoxicação pelo  $SO_2$  pode causar irritação da mucosa respiratória desde a nasofaringe e a orofaringe até os alvéolos, levando a inflamação, hemorragia e necrose. A resposta fisiológica primária à inalação do  $SO_2$  é uma broncoconstrição reflexa e reversível (CASTRO,2003).

***Material Particulado (PM):***

O ar atmosférico contém partículas em suspensão que podem ser menores que  $10 \mu m$  de diâmetro. Partículas maiores que  $10 \mu m$  são efetivamente filtradas pelo nariz e pela nasofaringe, onde essas grandes partículas ficam depositadas e podem ser vistas em expectorações e/ou saliva. Partículas menores que  $10 \mu m$  de diâmetro ( $PM_{10}$ ) ficam retidas nas vias aéreas superiores e podem ser depositadas na árvore traqueobrônquica. As partículas menores que  $2,5 \mu m$  de diâmetro ( $PM_{2,5}$ ) depositam-se no brônquio terminal e nos alvéolos. As concentrações de  $PM_{2,5}$  representam, em geral, cerca de 45 a 65% da concentração de  $PM_{10}$ . A indústria e o transporte são as principais fontes poluidoras de material particulado (WHO, 2000).

***Ozônio ( $O_3$ )***

O ozônio ( $O_3$ ) se forma na atmosfera a partir da reação do oxigênio molecular ( $O_2$ ) com o oxigênio atômico ( $O$ ). O  $NO_2$ , através de uma reação fotoquímica, produz o oxigênio atômico. Portanto, a combinação com o  $O_2$  produz  $O_3$ . A formação de  $O_3$  é típica de áreas urbanas. Normalmente, é um poluente com uma concentração maior em ambientes externos do que em ambientes internos. O  $O_3$  é relativamente pouco solúvel em água e costuma atingir os alvéolos com mais facilidade, produzindo seus efeitos tóxicos nesta região. Os níveis de  $O_3$  iniciam sua elevação no período da manhã e têm o seu pico durante a tarde. Os níveis decaem durante o início da noite. O ozônio é um poluente oxidante, citotóxico (provoca lesão das células), que

atinge as porções mais distais das vias aéreas (CASTRO et al,2003).

## 3.2 Padrões de qualidade do ar

Diante desses novos episódios, ainda na década de 60, os Estados Unidos estabeleceram padrões de qualidade do ar, especificando os seis poluentes atmosféricos que seriam controlados, quais sejam: partículas totais em suspensão (PTS), dióxido de enxofre ( $SO_2$ ), monóxido de carbono ( $CO$ ), dióxido de nitrogênio ( $NO_2$ ), ozônio ( $O_3$ ) e chumbo ( $Pb$ ). A fim de efetivar esse controle, criou-se em 1970 a Agência de Proteção Ambiental norte-americana (U.S EPA). Várias medidas de controle foram, então, implantadas, visando não só atingir as fontes de emissão móveis, como também as estacionárias. Em 1990, foram conferidos à U.S EPA poderes para determinar os critérios técnicos de controle das substâncias tóxicas, com base nos seus efeitos à saúde.(LIRA ,2009)

Em 1976, uma comissão de países europeus (Comission of the European Communities CEC), estabeleceu padrões de qualidade do ar para  $SO_2$ ,  $CO$ ,  $NO_2$ , material particulado e oxidantes foto-químicos. Esses padrões foram sendo aprimorados ao longo dos anos, subsidiando as legislações dos diversos países europeus .

A Organização Mundial da Saúde (OMS) em 2005 publicou a revisão do seu guia de qualidade do ar, que é usado com objetivo de informar e dar base aos responsáveis pela elaboração das políticas de proteção e gestão da qualidade do ar em diferentes partes do mundo. Os padrões de qualidade do ar, segundo publicação da (OMS) em 2005, variam de acordo com a abordagem adotada para balancear riscos à saúde, viabilidade técnica, considerações econômicas e vários outros fatores políticos e sociais, que por sua vez, dependerão, entre outras coisas, do nível de desenvolvimento e da capacidade nacional de gerenciar a qualidade do ar. As diretrizes da OMS encontram-se na Tabela 3.1.

No Brasil, através da Portaria Normativa n.º 348 de 14/03/90 o IBAMA estabeleceu os padrões nacionais de qualidade do ar e os respectivos métodos de referência. Os

Tabela 3.1: Diretrizes da Organização Mundial de Saúde 2005

Poluentes	Tempo de amostragem	Concentração $\mu/m^3$
$SO_2$	24h	20
	10min	500
$O_3$	8h	100
$NO_2$	1h	200
	MAA	40
$PM_{10}$	24h	50
	MAA	20

padrões estabelecidos através dessa portaria foram submetidos ao CONAMA em 28/06/90 e transformados na Resolução CONAMA  $n^\circ$  03/90.

São estabelecidos dois tipos de padrões de qualidade do ar: os primários e os secundários:

- **Padrões primários de qualidade do ar:** As concentrações de poluentes que, ultrapassadas, poderão afetar a saúde da população. Podem ser entendidos como níveis máximos toleráveis de concentração de poluentes atmosféricos, constituindo-se em metas de curto e médio prazo.
- **Padrões secundários de qualidade do ar:** As concentrações de poluentes atmosféricos abaixo das quais se prevê o mínimo efeito adverso sobre o bem estar da população, assim como o mínimo dano à fauna e à flora, aos materiais e ao meio ambiente em geral. Podem ser entendidos como níveis desejados de concentração de poluentes, constituindo-se em meta de longo prazo. O intuito do estabelecimento de padrões secundários é criar uma base para uma política de prevenção da degradação da qualidade do ar.

Os parâmetros regulamentados são os seguintes : partículas totais em suspensão, fumaça, partículas inaláveis, dióxido de enxofre, monóxido de carbono, ozônio e dióxido de nitrogênio. Os padrões nacionais de qualidade do ar fixados na Resolução CONAMA  $n^\circ$  03 de 28/06/90 são apresentados na Tabela (3.2).

A mesma resolução estabelece ainda os critérios para episódios agudos de poluição do ar. Esses critérios são apresentados na Tabela (3.3).

Tabela 3.2: Padrões nacionais de qualidade do ar (Resolução CONAMA 003 de 28/06/1990)

Poluente	Tempo de Amostragem	Padrão Primário $\mu g/m^3$	Padrão Secundário $\mu g/m^3$	Método de Amostragem
PTS	24 horas <sup>1</sup>	240	150	separação/inercial filtração
	MGA <sup>2</sup>	80	60	
PM <sub>10</sub>	24 horas <sup>1</sup>	150	150	separação/inercial filtração
	MMA <sup>3</sup>	50	50	
SO <sub>2</sub>	24 horas	365	100	Infravermelho não dispersivo
	MAA <sup>3</sup>	80	40	
NO <sub>2</sub>	1 hora	320	190	Quimiluminescência
	MAA	100	100	
CO	1 hora	40.000/35 ppm	40.000/35 ppm	Pararosanilina
	8 horas	10.000 (9ppm)	10.000 (9ppm)	
O <sub>3</sub>	1 hora <sup>1</sup>	160	160	Quimiluminescência

(1) Não deve ser excedido mais que uma vez ao ano.

(2) Média geométrica anual.

(3) Média aritmética anual.

Tabela 3.3: Critérios para episódios agudos de poluição do ar

PARÂMETROS	ATENÇÃO	ALERTA	EMERGÊNCIA
<i>PTS</i> - 24 horas( $\mu g/m^3$ )	375	625	875
<i>PM</i> <sub>10</sub> - 24 horas( $\mu g/m^3$ )	250	420	500
<i>SO</i> <sub>2</sub> - 24 horas( $\mu g/m^3$ )	800	1.600	2.100
<i>SO</i> <sub>2</sub> <i>XPTS</i> - 24 horas( $\mu g/m^3$ )( $\mu g/m^3$ )	65.000	261.000	393.000
<i>NO</i> <sub>2</sub> - 1 hora( $\mu g/m^3$ )	1.130	2.260	3.000
<i>CO</i> - 8 horas( $\mu g/m^3$ )	17.000	34000	46.000
<i>O</i> <sub>3</sub> - 1 hora( $\mu g/m^3$ )	400	800	1.000
Fumaça-24horas( $\mu g/m^3$ )	250	420	500

No Espírito Santo, o Instituto Estadual de Meio Ambiente e recursos hídricos (IEMA) monitora a qualidade do ar da região da Grande Vitória tendo como objetivos principais medir a exposição horária da população na Região da Grande Vitória (RGV) aos poluentes: material particulado (PTS), partículas inaláveis com diâmetro menores que 10 micron metros (*PM*<sub>10</sub>), dióxido de enxofre (*SO*<sub>2</sub>), óxidos de nitrogênio (*NO*<sub>x</sub>), hidrocarboneto (*HC*) e ozônio (*O*<sub>3</sub>); avaliar a qualidade do ar na RGV de modo a verificar os níveis de concentrações de poluentes a fim de se comparar com os padrões de qualidade do Ar fixados pela Resolução CONAMA

*n*º 03 de 28/06/90 e fornecer subsídios para avaliar os efeitos da poluição na saúde humana, nos materiais, na fauna e na flora da região.(IEMA,2007).

### 3.3 Índice de qualidade do Ar

A classificação da qualidade do ar em diversas regiões em especial na Região da Grande Vitória, é baseada no cálculo de um índice de qualidade do ar, que é uma ferramenta matemática desenvolvida para simplificar o processo de divulgação da qualidade do ar. O índice é obtido através de uma função linear segmentada, onde os pontos de inflexão são os padrões de qualidade do ar. Desta função, que relaciona a concentração do poluente com o valor índice, resulta em número adimensional referido a uma escala com base nos padrões de qualidade do ar. Para cada poluente medido, é calculado um índice conforme a equação (3.1). Para efeito de divulgação, é utilizado o índice mais elevado dos poluentes medidos em cada estação. Portanto, a qualidade do ar de uma estação é determinada pelo pior caso. A qualificação do ar está associada com efeitos sobre a saúde humana. Depois de calculado o valor do índice, o ar recebe uma qualificação, conforme Tabela (3.4) .

Tabela 3.4: Faixas de concentração dos poluentes para o cálculo do IQA.

Classificação e Faixas do IQA		PTS Média (24h)	PM10 Média (24h)	SO2 Média (24h)	NO2 Média (1h)	O3 Média (1h)	CO Média (8h)
BOM	0-50	0-80	0-50	0-80	0-80	0-80	0-5.000
REGULAR	51-100	81-240	51-150	81-365	101-320	81-160	5.001-10.000
INADEQUADO	101-199	241-375	151-250	366-800	321-1130	161-200	10.001-17.000
MÁ	200-299	376-625	251-420	801-1600	1131-2260	201-800	17.001-34.000
PÉSSIMO	300-399	626-875	421-500	1601-2100	2261-3000	801-1000	34.001-46.000
CRÍTICA	≥400	≥876	≥501	≥2101	≥3001	≥1001	≥46.001

Baseados no processo de interpolação linear, obtemos a equação que converte os dados de concentração de um determinado poluente-p em IQA.

Sendo:

$I_p$ =índice do poluente-p a ser calculado;

$C_p$ =concentração do poluente-p;

$B_s$ =limite superior da concentração do poluente-p ,obtido na Tabela (3.4);

$B_i$ =limite inferior da concentração do poluente-p, obtido na Tabela (3.4);

$I_s$ =IQA correspondente a  $B_s$ , obtido na Tabela (3.4);

$I_i$ =IQA correspondente a  $B_i$ ,obtido na Tabela (3.4).

Temos:

$$\frac{I_p - I_i}{I_s - I_i} = \frac{C_p - B_i}{B_s - B_i} \Rightarrow I_p - I_i = \frac{I_s - I_i}{B_s - B_i}(C_p - B_i)$$

$$I_p = \frac{I_s - I_i}{B_s - B_i}(C_p - B_i) + I_i. \quad (3.1)$$

---

---

# CAPÍTULO 4

---

## MODELO AUTOREGRESSIVO DE VALORES INTEIROS E SUAS PROPRIEDADES

### 4.1 Modelo Auto-Regressivo de Valores Inteiros (INAR)

O modelo Auto-Regressivo de Valores Inteiros de ordem 1 (INAR(1)), apresentado por Al-Osh e Alzaid (1987), baseia-se no operador *thinning* binomial proposto por Steutel e Van Harn (1979). Du e Li (1991) generalizaram o modelo INAR para ordem  $p$ , isto é, INAR( $p$ ). O operador *thinning* é definido na Seção 4.1.2, a definição e as propriedades do modelo INAR(1) são dadas nas Seção 4.1.3, a definição e as propriedades do modelo INAR( $p$ ) são dadas na Seção 4.1.5 e o critério automático para a seleção da ordem do modelo INAR( $p$ ) é apresentado na Seção 4.2.2.

#### 4.1.1 Distribuição de Poisson

Segundo James (1981) e Meyer (2000) ao se estudar as distribuições discretas abordadas em séries temporais, a de Poisson é a mais observada na teoria e na prática. Uma variável aleatória  $\mathbf{X}$  possui distribuição Poisson com parâmetro  $\lambda$  se a dis-

tribuição de probabilidade de  $\mathbf{X}$  é dada por

$$\begin{aligned} P(\mathbf{X} = x) &= \frac{e^{-\lambda} \lambda^x}{x!}; & x = 0, 1, 2, \dots \\ &= 0 & ; \quad c.c. \end{aligned}$$

A distribuição de Poisson pode ser desenvolvida a partir da distribuição Binomial como se segue. Seja  $\mathbf{Y}$  uma variável aleatória com distribuição Binomial com parâmetros  $n$  (número de ensaios) e  $p$  (probabilidade de sucessos), onde sua distribuição de probabilidade é dada por

$$P(\mathbf{Y} = y) = \frac{n!}{y!(n-y)!} p^y (1-p)^{n-y}; \quad y = 0, 1, 2, \dots$$

Tomando  $np = b$  tem-se  $p = b/n$  e  $1-p = 1 - b/n = (n-b)/n$ . Daí:

$$\begin{aligned} P(\mathbf{Y} = y) &= \frac{n(n-1)(n-2)\dots(n-y+1)}{y!} \left[\frac{b}{n}\right]^y \left[\frac{n-b}{n}\right]^{n-y} \\ &= \frac{b^y}{y!} \left[1 \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \dots \left(1 - \frac{y-1}{n}\right)\right] \left[1 - \frac{b}{n}\right]^n \left[1 - \frac{b}{n}\right]^{-y}. \end{aligned}$$

Fazendo  $n \rightarrow \infty$  e  $p \rightarrow 0$  de modo que  $np = b$  permaneça fixo, tem-se que  $(1 - \frac{1}{n}) \rightarrow 1$ ,  $(1 - \frac{2}{n}) \rightarrow 1$ , ...,  $(1 - \frac{y-1}{n}) \rightarrow 1$ ,  $(1 - \frac{b}{n})^{-y} \rightarrow 1$ . Dado que  $(1 - \frac{b}{n})^n \rightarrow e^{-b}$  quando  $n \rightarrow \infty$  então

$$P(Y = y) = \frac{b^y e^{-b}}{y!},$$

que é a distribuição de Poisson com parâmetro  $b$ , isto é,  $Y \sim Po(b)$ .

Para verificar que a expressão acima é realmente uma distribuição de probabilidade, esta deve satisfazer  $\sum_{x=0}^{\infty} P(X = x) = 1$ .

De fato, como

$$\begin{aligned}
 P(X = 0) &= e^{-\lambda}; \\
 P(X = 1) &= \lambda e^{-\lambda}; \\
 P(X = 2) &= \frac{\lambda^2 e^{-\lambda}}{2!}; \\
 &\vdots \\
 P(X = n) &= \frac{\lambda^n e^{-\lambda}}{n!}; \\
 &\vdots
 \end{aligned}$$

então

$$\sum_{x=0}^{\infty} P(X = x) = e^{-\lambda} + \lambda e^{-\lambda} + \frac{\lambda^2 e^{-\lambda}}{2!} + \dots = e^{-\lambda} \left[ 1 + \lambda + \frac{\lambda^2}{2!} + \dots \right] = e^{-\lambda} e^{\lambda} = 1.$$

A função geratriz de momentos da distribuição de Poisson é dada por

$$\begin{aligned}
 M_x(t) &= E[e^{tx}] \\
 &= \sum_{x_i=0}^{\infty} e^{tx_i} p(x = x_i) \\
 &= \sum_{x_i=0}^{\infty} e^{tx_i} \frac{e^{-\lambda} \lambda^{x_i}}{x_i!} \\
 &= e^{-\lambda} + e^t \frac{e^{-\lambda} \lambda}{1!} + e^{2t} \frac{e^{-\lambda} \lambda^2}{2!} + e^{3t} \frac{e^{-\lambda} \lambda^3}{3!} + \dots \\
 &= e^{-\lambda} \left[ 1 + e^t \lambda + e^{2t} \frac{\lambda^2}{2!} + e^{3t} \frac{\lambda^3}{3!} + \dots \right] \\
 &= e^{-\lambda} \left[ 1 + (e^t \lambda)^1 + \frac{(e^t \lambda)^2}{2!} + \frac{(e^t \lambda)^3}{3!} + \dots \right] \\
 &= e^{-\lambda} e^{\lambda e^t} \\
 &= e^{\lambda e^t - \lambda} \\
 &= e^{\lambda(e^t - 1)}.
 \end{aligned}$$

Com a função geratriz de momentos, pode-se facilmente calcular os momentos da

distribuição da variável  $X$  em torno da origem. Estes são dados por:

$$E[X] = \frac{d}{dt} M_x(t) |_{t=0} = e^{\lambda(e^t-1)} \lambda e^t |_{t=0} = 1 \cdot \lambda \cdot 1 = \lambda.$$

$$E[X^2] = \frac{d^2}{dt^2} M_x(t) |_{t=0} = e^{\lambda(e^t-1)} (\lambda e^t)^2 + e^{\lambda(e^t-1)} \lambda e^t |_{t=0} = 1\lambda^2 + 1\lambda = \lambda + \lambda^2.$$

Daí

$$\text{var}[X] = E[X^2] - (E[X])^2 = \lambda + \lambda^2 - \lambda^2 = \lambda.$$

$$\begin{aligned} E[X^3] &= \frac{d^3}{dt^3} M_x(t) |_{t=0} = e^{\lambda(e^t-1)} (\lambda e^t)^3 + e^{\lambda(e^t-1)} 2\lambda e^t \lambda e^t + e^{\lambda(e^t-1)} (\lambda e^t)^2 + e^{\lambda(e^t-1)} \lambda e^t |_{t=0} \\ &= \lambda + 3\lambda^2 + \lambda^3. \end{aligned}$$

**Teorema 4.1.1.** *Sejam  $X_1, X_2, \dots, X_n$  variáveis aleatórias independentes com distribuição de Poisson de parâmetros  $\lambda_1, \lambda_2, \dots, \lambda_n$ , respectivamente. Seja  $Z = X_1 + X_2 + \dots + X_n$ , então  $Z$  tem distribuição de Poisson com parâmetro  $\lambda = \lambda_1 + \lambda_2 + \dots + \lambda_n$ .*

*Demonstração:*

A função geratriz de momentos de  $X_i$  é dada por

$$M_{X_i}(t) = e^{\lambda_i(e^t-1)}$$

Como

$$\begin{aligned} M_Z(t) &= E[e^{tZ}] = E[e^{t(X_1+\dots+X_n)}] = E[e^{tX_1} \cdot e^{tX_2} \cdot \dots \cdot e^{tX_n}] \\ &= E[e^{tX_1}] \cdot \dots \cdot E[e^{tX_n}] = M_{X_1}(t) \cdot M_{X_2}(t) \cdot \dots \cdot M_{X_n}(t), \end{aligned}$$

então,

$$M_Z(t) = e^{\lambda_1(e^t-1)} \cdot \dots \cdot e^{\lambda_n(e^t-1)} = e^{(\lambda_1+\dots+\lambda_n)(e^t-1)},$$

que é a função geratriz de momentos de uma variável aleatória com distribuição de Poisson de parâmetro  $\lambda = \lambda_1 + \lambda_2 + \dots + \lambda_n$ .  $\square$

### 4.1.2 Operador *thinning* binomial

A Definição 4.1.1 refere-se à versão geral do operador *thinning* binomial, proposto por Steutel e Van Harn (1979), onde  $Y_j$  é uma sequência de variáveis aleatórias de valor inteiro não-negativo i.i.d., independente de  $X$  tal que  $P(Y_j = 1) = 1 - P(Y_j = 0) = \alpha$ .

**Definição 4.1.1.** *Seja  $X$  variável aleatória de valor inteiro não-negativo e seja  $\alpha$  número real tal que  $\alpha \in [0,1]$ . O operador *thinning*, denotado por  $\circ$ , é definido como*

$$\alpha \circ X = \sum_{j=1}^X Y_j$$

onde  $Y_j$  é uma sequência de variáveis aleatórias de valor inteiro não-negativo independentes e identicamente distribuídas (i.i.d.), chamada série de contagem, independente de  $X$  e tal que  $E[Y_j] = \alpha$ ,  $\text{var}[Y_j] = \sigma^2$ ,  $E[Y_j^3] = \gamma$  e  $E[Y_j^4] = \kappa$ , isto é, momentos finitos.

Silva (2005) apresenta e demonstra as propriedades do operador *thinning*, sendo apresentadas na Proposição 4.1.1 e demonstradas no Apêndice A.

**Proposição 4.1.1.** *(Propriedades do operador *thinning*) Seja  $X_i (i = 1, \dots, 4)$ , uma variável aleatória identicamente distribuída e  $\alpha_i (i = 1, \dots, 4)$  uma constante real não-negativa. Considere a série de contagem,  $Y_{i,j} (j = 1, \dots, X_i)$ , com  $\alpha_i \circ X_i (i = 1, \dots, 4)$ , sendo mutuamente independente e identicamente distribuída, independente de  $X_i$  e tal que  $E[Y_{j,i}] = \alpha_i$  e  $\text{var}[Y_{j,i}] = \sigma_i^2$ ,  $E[Y_{j,i}^3] = \gamma_i$  e  $E[Y_{j,i}^4] = \kappa_i$ . Portanto,*

- (i)  $0 \circ X_1 = 0$ ,
- (ii)  $1 \circ X_1 = X_1$ ,
- (iii)  $\alpha_1 \circ (\alpha_2 \circ X_2) = (\alpha_1 \alpha_2) \circ X_2$ ,
- (iv)  $\alpha \circ (X_1 + X_2) = \alpha \circ X_1 + \alpha \circ X_2$

- (v)  $E[\alpha_1 \circ X_1] = \alpha_1 E[X_1]$ ,
- (vi)  $E[(\alpha_1 \circ X_1)X_2] = \alpha_1 E[X_1 X_2]$ ,
- (vii)  $E[(\alpha_1 \circ X_1)X_2 X_3] = \alpha_1 E[X_1 X_2 X_3]$ ,
- (viii)  $E[(\alpha_1 \circ X_1)^2] = \alpha_1^2 E[X_1^2] + \sigma_1^2 E[X_1]$ ,
- (ix)  $E[(\alpha_1 \circ X_1)^2 X_2] = \alpha_1^2 E[X_1^2 X_2] + \sigma_1^2 E[X_1 X_2]$ ,
- (x)  $E[(\alpha_1 \circ X_1)(\alpha_2 \circ X_2)] = \alpha_1 \alpha_2 E[X_1 X_2]$ ,
- (xi)  $E[(\alpha_1 \circ X_1)(\alpha_2 \circ X_2)X_3] = \alpha_1 \alpha_2 E[X_1 X_2 X_3]$ ,
- (xii)  $E[(\alpha_1 \circ X_1)^3] = \alpha_1^3 E[X_1^3] + 3\alpha_1 \sigma_1^2 E[X_1^2] + (\gamma_1 - 3\alpha_1 \sigma_1^2 - \alpha_1^3) E[X_1]$ ,
- (xiii)  $E[(\alpha_1 \circ X_1)^2(\alpha_2 \circ X_2)] = \alpha_1^2 \alpha_2 E[X_1^2 X_2] + \sigma_1^2 \alpha_2 E[X_1 X_2]$ ,
- (xiv)  $E[(\alpha_1 \circ X_1)(\alpha_2 \circ X_2)(\alpha_3 \circ X_3)] = \alpha_1 \alpha_2 \alpha_3 E[X_1 X_2 X_3]$ ,
- (xv)  $E[(\alpha_1 \circ X_1)^4] = \alpha_1^4 E[X_1^4] + 6\alpha_1^2 \sigma_1^2 E[X_1^3] + (3\sigma_1^4 - 12\sigma_1^2 \alpha_1^2 + 4\alpha_1 \gamma_1 - 4\alpha_1^4) E[X_1^2] + (\kappa_1 - 3\sigma_1^4 + 6\sigma_1^2 + \alpha_1^2 - 4\alpha_1 \gamma_1 + 3\alpha_1^4) E[X_1]$ ,
- (xvi)  $E[(\alpha_1 \circ X_1)^3 X_2] = \alpha_1^3 E[X_1^3 X_2] + 3\alpha_1 \sigma_1^2 E[X_1^2 X_2] + (\gamma_1 - 3\alpha_1 \sigma_1^2 - \alpha_1^3) E[X_1 X_2]$ ,
- (xvii)  $E[(\alpha_1 \circ X_1)(\alpha_2 \circ X_2)] = \sigma_1^2 \sigma_2^2 E[X_1 X_2] + \alpha_1^2 \alpha_2^2 E[X_1^2 X_2^2] + \alpha_1^2 \sigma_2^2 E[X_1^2 X_2] + \alpha_2^2 \sigma_1^2 E[X_1 X_2^2]$ ,
- (xviii)  $E[(\alpha_1 \circ X_1)^2 X_2 X_3] = \alpha_1^2 E[X_1^2 X_2 X_3] + \sigma_1^2 E[X_1 X_2 X_3]$ ,
- (xix)  $E[(\alpha_1 \circ X_1)X_2 X_3 X_4] = \alpha_1 E[X_2 X_3 X_4]$ ,
- (xx)  $E[(\alpha_1 \circ X_1)(\alpha_2 \circ X_2)^3] = \alpha_1 \alpha_2^3 E[X_1 X_2^3] + 3\alpha_1 \alpha_2 \sigma_2^2 E[X_1 X_2^2] + (\alpha_1 \gamma_2 - 3\alpha_1 \alpha_2 \sigma_2^2 - \alpha_1 \alpha_2^3) E[X_1 X_2]$ ,
- (xxi)  $E[(\alpha_1 \circ X_1)(\alpha_2 \circ X_2)(\alpha_3 \circ X_3)^2] = \alpha_1 \alpha_2 \alpha_3^2 E[X_1 X_2 X_3] + \alpha_1 \alpha_2 \sigma_3^2 E[X_1 X_2 X_3] - \alpha_1 \alpha_2 \alpha_3^2 E[X_1 X_2 X_3] + \alpha_1 \alpha_2 \alpha_3^2 E[X_1 X_2 X_3^2]$ ,
- (xxii)  $E[(\alpha_1 \circ X_1)(\alpha_2 \circ X_2)(\alpha_3 \circ X_3)(\alpha_4 \circ X_4)] = \alpha_1 \alpha_2 \alpha_3 \alpha_4 E[X_1 X_2 X_3 X_4]$ .

### 4.1.3 Modelo INAR(1)

O modelo Auto-Regressivo de Valores Inteiros não-negativos de ordem 1 (INAR(1)), introduzido por Al-Osh e Alzaid (1987), é expresso na forma:

$$X_t = \alpha \circ X_{t-1} + \epsilon_t, \quad (4.1)$$

onde  $\alpha \in [0,1]$  e  $\{\epsilon_t\}$  é uma sequência não correlacionada de variáveis aleatórias com valores inteiros não-negativos de média  $\mu$  e variância  $\sigma^2$  finita.

A distribuição marginal do modelo (4.1) pode ser expressa em termos da sequência  $\{\epsilon_t\}$  como

$$\begin{aligned} X_1 &= \epsilon_1 \\ X_2 &= \alpha \circ \epsilon_1 + \epsilon_2 \\ X_3 &= \alpha \circ (\alpha \circ \epsilon_1 + \epsilon_2) + \epsilon_3 \\ &= \alpha \circ (\alpha \circ \epsilon_1) + \alpha \circ \epsilon_2 + \epsilon_3 \\ &= \alpha^2 \circ \epsilon_1 + \alpha \circ \epsilon_2 + \epsilon_3 \\ &\vdots \\ X_n &\stackrel{d}{=} \sum_{j=0}^{\infty} \alpha^j \circ \epsilon_{n-j}. \end{aligned}$$

A média e a variância do processo  $X_t$  como definido em (4.1) são dadas por

$$\begin{aligned} E[X_t] &= E[\alpha \circ X_{t-1} + \epsilon_t] \\ &= \alpha E[X_{t-1}] + \mu \\ &= \alpha E[\alpha \circ X_{t-2} + \mu] + \mu \\ &= \alpha^2 E[X_{t-2}] + \alpha\mu + \mu \\ &= \alpha^t E[X_0] + \mu \sum_{j=0}^{t-1} \alpha^j \end{aligned}$$

e

$$\begin{aligned}
\text{var}[X_t] &= \text{var}[\alpha \circ X_{t-1} + \epsilon_t] \\
&= \text{var}[\alpha \circ X_{t-1}] + \text{var}[\epsilon_t] \\
&= E[(\alpha \circ X_{t-1})^2] - E[\alpha \circ X_{t-1}]^2 + E[\epsilon_t^2] - E[\epsilon_t]^2 \\
&= \alpha^2 E[X_{t-1}^2] + \text{var}[X_{t-1}]E[X_{t-1}] - \alpha^2 E[X_{t-1}]^2 + \sigma^2 + \mu^2 - \mu^2 \\
&= \alpha^2 [E[X_{t-1}^2] - E[X_{t-1}]^2] + \alpha(1 - \alpha)E[X_{t-1}] + \sigma^2 \\
&= \alpha^2 \text{var}[X_{t-1}] + \alpha(1 - \alpha)E[X_{t-1}] + \sigma^2 \\
&= \alpha^2 [\alpha^2 \text{var}[X_{t-2} + \alpha(1 - \alpha)E[X_{t-2}] + \sigma^2] + \alpha(1 - \alpha)E[X_{t-1}] + \sigma^2 \\
&= \alpha^4 \text{var}[X_{t-2}] + \alpha(1 - \alpha) [\alpha^3 E[X_{t-2}] + \alpha E[X_{t-1}]] + (\alpha^2 + 1)\sigma^2 \\
&\vdots \\
&= \alpha^{2t} \text{var}[X_0] + (1 - \alpha) \sum_{j=1}^t \alpha^{2j-1} E[X_{t-j}] + \sigma^2 \sum_{j=1}^t \alpha^{2(j-1)}.
\end{aligned}$$

A função de autocovariância é dada por

$$\begin{aligned}
\gamma(k) &= \text{Cov}(X_{t-k}, X_t) \\
&= \text{Cov}(X_{t-k}, \alpha^k \circ X_{t-k}) + \text{Cov}\left(X_{t-k}, \sum_{j=0}^{k-1} \alpha^j \circ \epsilon_{t-j}\right) \\
&= E[X_{t-k}(\alpha^k \circ X_{t-k})] - E[X_{t-k}]E[\alpha^k \circ X_{t-k}] + E\left[X_{t-k} \sum_{j=0}^{k-1} \alpha^j \circ \epsilon_{t-j}\right] - \\
&\quad - E[X_{t-k}]E\left[\sum_{j=0}^{k-1} \alpha^j \circ \epsilon_{t-j}\right] \\
&= \alpha^k E[X_{t-k}^2] - \alpha^k E[X_{t-k}]^2 + \sum_{j=0}^{k-1} \alpha^j E[X_{t-k}\epsilon_{t-j}] - \sum_{j=0}^{k-1} \alpha^j E[X_{t-k}]E[\epsilon_{t-j}] \\
&= \alpha^k \text{var}[X_{t-k}] + \sum_{j=0}^{k-1} \alpha^j \text{Cov}(X_{t-k}, \epsilon_{t-j}), \\
&= \alpha^k \gamma(0)
\end{aligned}$$

e a função de autocorrelação por

$$\rho(k) = \gamma(k)/\gamma(0) = \alpha^k.$$

Logo, o processo INAR(1) dado por  $X_t = \alpha \circ X_{t-1} + \epsilon_t$  é estacionário quando  $|\alpha| < 1$  e não estacionário quando  $|\alpha| = 1$ .

#### 4.1.4 Estimador de mínimos quadrados condicional

Seja  $\{X_t, t \in \mathbb{Z}\}$  definido como em (4.1) e suponha-se que  $\{\epsilon_t\}$  tenha distribuição de Poisson com parâmetro  $\lambda$ , então, tem-se

$$\begin{aligned}
 E[X_t|X_{t-1}] &= E[\alpha \circ X_{t-1} + \epsilon_t|X_{t-1}] \\
 &= E[\alpha \circ X_{t-1}] + E[\epsilon_t|X_{t-1}] \\
 &= E[Y_1 + \dots + Y_{X_{t-1}}|X_{t-1}] + E[\epsilon_t] \\
 &= E[Y_1] + \dots + E[Y_{X_{t-1}}] + \lambda \\
 &= \alpha + \dots + \alpha + \lambda \\
 &= \alpha X_{t-1} + \lambda.
 \end{aligned}$$

Sejam  $X_1, X_2, \dots, X_n$ ,  $n$  observações do processo  $\{X_t\}$ . O estimador de mínimos quadrados condicional de  $\alpha$  e  $\lambda$  são os valores que minimizam

$$L = \sum_{t=1}^n [X_t - (\alpha X_{t-1} + \lambda)]^2.$$

Derivando  $L$  em relação a  $\alpha$  e em relação a  $\lambda$  obtém-se

$$\partial L / \partial \alpha = -2 \sum_{t=1}^n (X_t - \hat{\alpha} X_{t-1} - \hat{\lambda}) X_{t-1} = 0, \quad (4.2)$$

$$\partial L / \partial \lambda = -2 \sum_{t=1}^n (X_t - \hat{\alpha} X_{t-1} - \hat{\lambda}) = 0. \quad (4.3)$$

Resolvendo (4.2) e (4.3) resulta que

$$\sum_{t=1}^n X_t - \hat{\alpha} \sum_{t=1}^n X_{t-1} - n\hat{\lambda} = 0.$$

Tem-se, então, que

$$n\hat{\lambda} = \sum_{t=1}^n X_t - \hat{\alpha} \sum_{t=1}^n X_{t-1}.$$

Portanto,

$$\hat{\lambda} = \frac{1}{n} \left( \sum_{t=1}^n X_t - \hat{\alpha} \sum_{t=1}^n X_{t-1} \right)$$

é o estimador de mínimos quadrados de  $\lambda$ , e

$$\sum_{t=1}^n X_t X_{t-1} - \hat{\alpha} \sum_{t=1}^n X_{t-1} X_{t-1} - \hat{\lambda} \sum_{t=1}^n X_{t-1} = 0.$$

Logo,

$$\hat{\alpha} \sum_{t=1}^n X_{t-1} X_{t-1} = \sum_{t=1}^n X_t X_{t-1} - \hat{\lambda} \sum_{t=1}^n X_{t-1}.$$

Então

$$\hat{\alpha} \sum_{t=1}^n X_{t-1} X_{t-1} = \sum_{t=1}^n X_t X_{t-1} - \frac{1}{n} \sum_{t=1}^n X_t \sum_{t=1}^n X_{t-1} + \frac{1}{n} \hat{\alpha} \left( \sum_{t=1}^n X_{t-1} \right)^2.$$

Portanto

$$\hat{\alpha} \left[ \sum_{t=1}^n X_{t-1} X_{t-1} - \frac{1}{n} \left( \sum_{t=1}^n X_{t-1} \right)^2 \right] = \sum_{t=1}^n X_t X_{t-1} - \frac{1}{n} \sum_{t=1}^n X_t \sum_{t=1}^n X_{t-1}.$$

Obtem-se que

$$\hat{\alpha} = \frac{\sum_{t=1}^n X_t X_{t-1} - \frac{1}{n} \sum_{t=1}^n X_t \sum_{t=1}^n X_{t-1}}{\sum_{t=1}^n X_{t-1}^2 - \frac{1}{n} \left( \sum_{t=1}^n X_{t-1} \right)^2},$$

é o estimador de mínimos quadrados de  $\alpha$ .

#### 4.1.5 Modelo INAR(p)

**Definição 4.1.2.** *Uma extensão natural para a ordem  $p$ , INAR(p), é:*

$$X_t = \alpha_1 \circ X_{t-1} + \dots + \alpha_p \circ X_{t-p} + \epsilon_t \quad (4.4)$$

onde  $\epsilon_t \in \mathbb{N}$  são aleatórias i.i.d, com média  $\mu_\epsilon$  e variância  $\sigma_\epsilon^2$  finitas,  $\alpha_i \in [0, 1], i = 1, \dots, p$ ,  $\alpha_p \neq 0$ , e as séries de contagem de  $\alpha_k \circ X_{t-k}$ ,  $k = 1, \dots, p$ , são mutualmente

*independentes e independentes de  $\epsilon_t$ . (DU e LI 1991)*

### Propriedades Gerais

Du e Li (1991), de posse das propriedades do operador *thinning*, apresentaram para o modelo INAR(p), a função de autocovariância na forma escalar, a função de autocovariância na forma vetorial, o estimador para a função de autocovariância, o estimador para a função de autocorrelação, a função espectral e os resíduos do modelo INAR(p).

*A Matriz de autocovariância  $R(k)$  na forma escalar:*

A matriz de autocovariância  $R(k)$ , satisfaz um conjunto de equações do tipo Yule-Walker,

$$\begin{cases} R(0) = V_p + \sum_{i=1}^p \alpha_i R(i) \\ R(k) = \sum_{i=1}^p \alpha_i R(i - k), k \geq 1 \end{cases}$$

*A Função de autocovariância  $R(k)$  na forma vetorial:*

$$R_p \alpha = \begin{bmatrix} R(0) & R(1) & \cdots & R(p) \\ R(1) & R(0) & \cdots & R(p-1) \\ \vdots & \vdots & \ddots & \vdots \\ R(p) & R(p-1) & \cdots & R(0) \end{bmatrix} \begin{bmatrix} -1 \\ \alpha_1 \\ \vdots \\ \alpha_p \end{bmatrix} = \begin{bmatrix} -V_p \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

sendo

$$V_p = \sigma_\epsilon^2 + \mu_X \sum_{i=1}^p \sigma_i^2 \quad (4.5)$$

Onde  $\sigma_\epsilon^2 = Var[\epsilon_t]$ ,  $\mu_X = E[X_t]$ ,  $\sigma_i^2$  é a variância da série de contagem envolvida no  $i$ -ésimo operador *thinning*,  $\alpha_i \circ X_{t-i}, i = 1, \dots, p$

*Função de autocovariância  $\hat{R}(k)$  amostral*

$$\hat{R}(k) = \frac{1}{N} \sum_{t=1}^{N-k} (X_t - \bar{X})(X_{t+k} - \bar{X}) \quad (4.6)$$

*Função de autocorrelação  $\hat{\rho}$  amostral*

$$\hat{\rho}(k) = \frac{\hat{R}(k)}{\hat{R}(0)}, 0 \leq k \leq N - 1 \quad (4.7)$$

No domínio de frequência temos a densidade espectral  $f(w)$  expressa por Silva e Oliveira (2004, 2005) por:

*Função de densidade espectral  $f(w)$*

$$f(w) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} R(k)e^{-i\omega k} = \frac{1}{2\pi} \frac{V_p}{|1 - \sum_{k=1}^p \alpha_k e^{-i\omega k}|^2}, -\pi \leq w \leq \pi \quad (4.8)$$

onde  $V_p$  está definido em (4.5).

*Os resíduos do processo INAR*

$$r_{N+1} = X_{N+1} - E[X_{N+1}|X_N, X_{N-1}, \dots] \quad (4.9)$$

Considerando  $(X_1, \dots, X_N)$  uma realização de um processo INAR (p).

*Estimação dos parâmetros do modelo INAR, baseados em equações do tipo Yule-Walker:*

Para a estimação dos parâmetros do modelo INAR,  $\alpha_1, \dots, \alpha_p$ , deve-se obter a solução do sistema de equações lineares abaixo:

$$\hat{R}_{p-1}\hat{\alpha} = \begin{bmatrix} \hat{R}(0) & \hat{R}(1) & \cdots & \hat{R}(p-1) \\ \hat{R}(1) & \hat{R}(0) & \cdots & \hat{R}(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{R}(p-1) & \hat{R}(p-2) & \cdots & \hat{R}(0) \end{bmatrix} \begin{bmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \vdots \\ \hat{\alpha}_p \end{bmatrix} = \begin{bmatrix} \hat{R}(1) \\ \hat{R}(2) \\ \vdots \\ \hat{R}(p) \end{bmatrix}$$

$$\hat{\mu}_\epsilon = \bar{X} \left( 1 - \sum_{i=1}^p \hat{\alpha}_i \right) \quad (4.10)$$

$$\hat{\sigma}_\epsilon^2 = \hat{V}_p - \bar{X} \sum_{i=1}^p \hat{\sigma}_i^2 \quad (4.11)$$

Considerando:

$$\hat{V}_p = \hat{R}(0) - \sum_{i=1}^p \hat{\alpha}_i \hat{R}(i) \quad (4.12)$$

Silva e Oliveira (2003) concluíram que a modelação efetiva de uma série de observações por modelos INAR depende da determinação da ordem do modelo a ser usado.

## 4.2 Critérios Automáticos de Seleção de Ordem

Segundo Zhang (1992) o procedimento para a escolha da ordem em modelos de regressão ou de séries temporais é escolher uma ordem  $k$  que minimiza um critério que pode ser escrito como uma função das observações (em geral, o erro quadrático médio da previsão a 1-passo ou a soma dos quadrados dos resíduos) mais um termo de penalização que depende do número de observações e da ordem do modelo a ser ajustado .

A escolha de uma ordem menor que a verdadeira provoca inconsistência na estimação dos parâmetros enquanto a escolha de uma ordem superior, conduz ao incremento da variância desses estimadores. Os critérios automáticos para seleção de ordem têm sido propostos tendo como objetivo equilibrar o risco da escolha de ordem menores ou maiores que a verdadeira. Este equilíbrio é feito através da atribuição de um custo ou penalização pela introdução de variáveis adicionais (SILVA e SILVA (2003)).

### 4.2.1 Critérios automáticos

#### AIC(p)

Dadas  $n$  observações independentes de um processo com vetor de parâmetros  $\theta$ , a proposta de Akaike (1974) consiste em considerar a verossimilhança  $f(x|\theta)$  para várias dimensões de  $\theta$ . Akaike (1974) mostra que, desde que a distribuição verdadeira pertença à família das distribuições aproximantes, i.e.,  $g(x) = f(x|\theta_0)$ , e sendo  $\hat{\theta}$  (estimador de máxima verossimilhança de  $\theta$ ) suficientemente próximo de  $\theta_0$ , tem-se que

$$AIC(p) = E[I(f(x|\theta_0), f(x|\theta))] \simeq -2 \log(\hat{\sigma}_\epsilon^2) + 2(p + 1) \quad (4.13)$$

onde  $\hat{\sigma}_\epsilon^2$  é o estimador de máxima verossimilhança da variância do ruído e  $p$  é a ordem do modelo aproximante. A ordem do modelo que melhor se ajusta às observações é o valor de  $p$  que minimiza AIC.

### AICC(p)

Versão corrigida do AIC, definido por HURVICH e TSAI (1989) :

$$AICC(k) = n \log(\hat{\sigma}_\epsilon^2) + n \frac{1 + k/n}{1 - (k + 2)/n}. \quad (4.14)$$

## 4.2.2 Critério automático para modelos INAR(p)

Segundo Silva e Oliveira (2003) o uso dos critérios expostos anteriormente para determinar a melhor ordem do modelo INAR a considerar não é correto, uma vez que a interpretação de  $\sigma_\epsilon^2$  como variância do erro de previsão a 1-passo ou, ainda, como variância dos resíduos para os modelos AR não tem analogia para os modelos INAR. Na construção do critério  $AICC_{inar}$  consideramos uma aproximação da função de verossimilhança através da função de densidade espectral, proposta por Whittle (1953) e usualmente designada por critério de Whittle, a fórmula de Kolmogorov adaptada ao processo INAR e as propriedades das equações de Yule-Walker, obtendo:

$$AICC_{inar}(k) = n \log(\hat{V}_k) + n \frac{1 + k/n}{1 - (k + 2)/n} \quad (4.15)$$

com

$$\hat{V}_k = \hat{\sigma}_\epsilon^2 + \bar{X} \left( \sum_{k=1}^p \hat{\beta}_k \right) \quad (4.16)$$

onde  $\hat{\beta}_k$  é o estimador da variância da série de contagem envolvida no k-ésimo operador *thinning*,  $\alpha_k \circ X_{t-k}$   $k = 1, \dots, p$ ,  $\bar{X}$  é a média amostral.

### 4.3 Previsão INAR

Após o término do processo de identificação, estimação dos parâmetros e verificação, o último passo é realizar as previsões de observações futuras dado que temos observado uma série até o tempo  $n$ , ou seja,  $x_n = (x_1, x_2, \dots, x_n)$ . Partindo-se da origem em  $n$ , e supondo que o objetivo é prever a série em um período futuro  $h$ ,  $\hat{X}_{n+h}$  representa a previsão para um período  $n+h$  feita em  $n$ . A previsão múltiplos passos para o período  $n+h$  é normalmente construída a partir de sucessivas previsões para os períodos  $n+1, n+2, \dots, n+h-1$  (MONTGOMERY, 1976). Neste procedimento, o valor de  $X_{n+h}$ , o qual não se conhece no tempo  $n$ , é substituído pela sua previsão  $\hat{X}_{n+h}$ . Portanto, a previsão se torna o cálculo do valor esperado de uma futura observação condicionada aos valores passados e ao valor presente da variável. Ou seja,  $\hat{X}_{n+h}$  é o valor previsto para um horizonte de  $h$  períodos de tempo futuro e  $n$  o período de origem da previsão, então,

$$\hat{X}_{n+h}|x_n = E(X_{n+h}|X_n) = \alpha^h \left[ X_n - \frac{\lambda}{1-\alpha} \right] + \frac{\lambda}{1-\alpha}, h = 1, 2, 3, \dots \quad (4.17)$$

Segundo Freeland (1998), a função de geração do momento de  $X_{n+h}$  dado  $X_n$  é :

$$\varphi_{X_{n+h}|X_n}(s) = [\alpha^h e^s + (1 - \alpha^h)]^{X_n} \exp \left[ \lambda \frac{1 - \alpha^h}{1 - \alpha} (e^s - 1) \right] \quad (4.18)$$

Expressão 4.18 mostra que a distribuição dos  $X_{n+h}|X_n$  é uma combinação de uma distribuição binomial com parâmetros  $\alpha^h$  e  $X_n$  e uma distribuição de Poisson com parâmetro  $\lambda(1 - \alpha^h)/(1 - \alpha)$ . Ou seja, a função de probabilidade de  $X_{n+h}|X_n$  é dada por

$$\begin{aligned} f(x_{n+h}|x_n) &= P(X_{n+h} = x_{n+h}|X_n = x_n) \\ &= \exp \left\{ -\lambda \frac{1 - \alpha^h}{1 - \alpha} \right\} \sum_{i=0}^{M_h} \frac{1}{x_{n+h-i}!} \times \left( \lambda \frac{1 - \alpha^h}{1 - \alpha} \right)^{x_{n+h-i}} \binom{x_n}{i} (\alpha^h)^i (1 - \alpha^h)^{x_n-i} \end{aligned} \quad (4.19)$$

Onde  $M_h = \min(X_{n+h}, X_n)$  e  $x_{n+h} = 0, 1, \dots$

A fim de obter previsões coerentes para  $X_{n+h}$  Freeland e McCabe (2003) sugerem a utilização do valor que minimiza o erro esperado absoluto, dada a amostra, i.e.,  $E \left[ \left| X_{n+h} - \hat{X}_{n+h} \right| \middle| X_n \right]$ . Assim, eles concluíram que  $\hat{X}_{n+h} = \hat{m}_{n+h}$  é a mediana do h-passo à frente da distribuição condicional  $f(x_{n+h}|x_n) = P(X_{n+h} = x_{n+h}|X_n = x_n)$ .

### 4.3.1 Intervalo de previsão

Um intervalo de confiança para o preditor  $\hat{X}_{n+h}$ , pode ser calculado através da função de probabilidade do h-passo à frente do erro de previsão, dada por:

$$e_{n+h}|x_n = X_{n+h}|x_n - \hat{X}_{n+h}|x_n \quad (4.20)$$

Substituindo  $\hat{X}_{n+h}|x_n$  dado por 4.17, obtemos

$$e_{n+h}|x_n = k - \alpha^h x_n - \lambda \frac{1 - \alpha^h}{1 - \alpha}, k = 0, 1, 2, \dots \quad (4.21)$$

Assim,

$$\begin{aligned} P \left( e_{n+h} = k - \alpha^h x_n - \lambda \frac{1 - \alpha^h}{1 - \alpha} \middle| x_n \right) &= \\ &= P(X_{n+h} = k | X_n = x_n) \end{aligned} \quad (4.22)$$

$$= \exp \left\{ -\lambda \frac{1 - \alpha^h}{1 - \alpha} \right\} \sum_{i=0}^{M_h} \frac{1}{(k-i)!} \left( \lambda \frac{1 - \alpha^h}{1 - \alpha} \right)^{k-i} \binom{x_n}{i} (\alpha^h)^i (1 - \alpha^h)^{x_n - i}$$

A partir da expressão 4.22, obtemos um intervalo de confiança para  $X_{n+h}$ .

$$\left( \hat{X}_{n+h} + e_{t_1}, \hat{X}_{n+h} + e_{t_2} \right), \quad (4.23)$$

onde  $\hat{X}_{n+h}$  é dada por 4.17,  $e_{t_1}$  é o maior valor de  $e_{n+h}$ , com  $P(e_{n+h} \leq e_{t_1}) \leq (1-\gamma)/2$  e  $e_{t_2}$  é o menor valor de  $e_{n+h}$ , com  $P(e_{n+h} \leq e_{t_2}) \geq (1+\gamma)/2$  ao nível  $\gamma$ .

---

---

## CAPÍTULO 5

---

### PREVISÃO DA QUALIDADE DO AR

A previsão da qualidade do ar pode ser utilizada para alertar previamente sobre a concentração de poluentes na atmosfera e permitir a adaptação de comportamentos adequados pela população e grupos de risco (crianças, idosos e pessoas com doenças respiratórias), podendo ainda fornecer às autoridades as informações para a preparação de planos de ação de curto prazo para a redução de emissões e gestão da qualidade do ar.

A previsão da qualidade do ar e seus índices podem ser efetuados através de diferentes metodologias destacando-se:

Os modelos numéricos, determinísticos, que simulam a dispersão e as transformações químicas dos poluentes na atmosfera, recorrendo a inventários de emissões de poluentes atmosféricos e a informação meteorológica.

Os modelos estatísticos, que se baseiam na identificação das relações entre as condições meteorológicas e as concentrações dos poluentes medidas nas estações de monitoramento de qualidade do ar através da análise do histórico de alguns anos. Podem também fazer uso de modelos probabilísticos que tem como base os dados das concentrações de cada poluente, metodologia usada nesta dissertação.

## 5.1 Modelos Estatísticos para a modelagem da qualidade do ar

Neto et al (2004) baseado em métodos estocásticos, realizou previsões da qualidade do ar para a cidade de Lisboa, Portugal. Os poluentes modelados foram  $O_3$  e o  $PM_{10}$  combinados com as variáveis meteorológicas. O período das análises foi de 3 anos (2001 à 2003). Os dados serviram para testar estatisticamente novos modelos usando a análise de regressão múltipla (RM) e análise de regressão em árvore (CART-Classification and Regression Tree). Para a modelagem e previsão das concentrações máximas de  $O_3$  foram usados dois modelos, um de regressão múltipla e outro de regressão em árvore+regressão múltipla. De igual forma para a modelagem e previsão das concentrações máximas de  $PM_{10}$  foram usados dois modelos, um de regressão múltipla e outro de regressão em árvore+regressão múltipla. A variância explicada pelos modelos foi superior a 72%. Para a validação dos modelos apresentados foi utilizado outro período de tempo e verificado que as variâncias explicadas pelos modelos foram ligeiramente inferiores a 72%, no entanto suficientemente significativos para demonstrar robustez da metodologia aplicada.

Agirre-Basurko et al. (2006) comparam três modelos para a modelagem e previsão da qualidade do ar da cidade de Bilbao, Espanha. Um modelo de Regressão Linear Múltipla (RLM) e dois modelos de rede perceptron ou Multilayer perceptron (MLP). Os modelos usaram variáveis meteorológicas (temperatura, umidade relativa, pressão, radiação, gradiente de temperatura, direção do vento, velocidade do vento) e fluxo de veículos como dados de entrada, no período de 1993 à 1994 e, como saída prevista pelos modelos, tem-se a concentração de  $O_3$  e  $NO_2$  com horizonte de previsão de 8 horas à frente. Para classificar o modelo de melhor desempenho foram utilizados cinco indicadores estatísticos, coeficiente de correlação, erro quadrático médio normalizado, considerando  $C_o$  =(casos observados) e  $C_p$  = (casos previstos)  $0.5 < C_o/C_p < 2$ , a Polarização Fracionaria (FP) e a Variância Fracionária (FV). Os modelos de rede perceptron obtiveram melhores resultados para a previsão das concentrações de  $O_3$  e  $NO_2$  quando comparados ao modelo de Regressão Linear

Múltipla. Quando comparados os desempenhos dos modelos de rede perceptron o que mais se destacou foi o modelo que considerou a sazonalidade da série das concentrações de  $O_3$  e  $NO_2$ .

Lira (2009) desenvolveu modelos empíricos para a modelagem e previsão da qualidade do ar da cidade de Uberlândia-MG. Vários modelos lineares, tais como Auto-Regressivo com entradas exógenas (ARX), Auto-Regressivo com média móvel e entradas exógenas (ARMAX), erro na saída (OE) e Box-Jenkins (BJ), bem como modelos baseados em redes neurais foram avaliados. Os modelos usaram variáveis meteorológicas (temperatura, umidade relativa, precipitação, direção do vento, velocidade do vento, nebulosidade e insolação) e fluxo veículos como dados de entrada, no período de 2003 à 2007 e, como saída prevista pelo modelo, tem-se a concentração de  $PM_{10}$  com horizonte de previsão de três dias à frente. Para a escolha da ordem de cada modelo foram utilizados três critérios: o critério de informação ( $AIC$ ), o erro final de predição ( $FPE$ ) e o critério de informações de Bayes ( $BIC$ ). Após a escolha da ordem dos modelos foi necessário classificar o modelo de melhor desempenho e para isso foram utilizados três indicadores estatísticos. Um deles é a raiz quadrada do erro médio quadrático ( $RMSE$ ), que mede os desvios entre os valores observados e os preditos pelos modelos. O coeficiente de determinação ( $R^2$ ), que informa quanto da variabilidade da variável observada é explicada pelo modelo. E por fim, uma média relativa de erro denominada índice de concordância ( $d$ ). As melhores estimativas foram obtidas pelo modelo BJ.

## 5.2 Modelos INAR

O modelo Auto-Regressivo de Valores Inteiros de ordem 1 (INAR(1)), apresentado por Al-Osh e Alzaid (1987), baseado no operador *thinning* binomial proposto por Steutel e Van Harn (1979), veio com o objetivo da modelagem de séries de contagem. Du e Li (1991) generalizaram o modelo INAR para ordem  $p$ , isto é, INAR( $p$ ). Gauthier e Latour (1994) generalizaram o conceito do operador *thinning* permitindo

que as séries de contagem sigam qualquer distribuição discreta. Em 1998, Latour apresenta a condição de estacionaridade do modelo INAR(p) sendo  $\sum_{k=1}^p \alpha_k < 1$ . Hellstron (2000) simula o poder do teste Dickey-Fuller em modelos INAR(1) obtendo resultados próximos da aplicação em modelos autoregressivo de ordem 1, AR(1). Em 2003, Silva e Silva apresentam uma ferramenta de seleção automática da ordem dos modelos INAR(p) o  $AICC_{INAR}$ . Silva e Oliveira (2004) investigaram momentos e cumulantes de grande ordem do modelo INAR(1) estendendo o estudo no contexto do domínio da frequência. Ferland et al (2006), com base no modelo INAR, estudaram o processo com erros GARCH. A modelagem INAR considera a característica de contagem dos dados, retornando em suas simulações apenas valores inteiros e positivos, obtendo em suas modelagens e previsões valores que acompanham as séries ditas de contagem. Os modelos INAR explicam as características da série apenas com os dados da própria série, baseado nas características probabilísticas dos mesmos.

McKenzie (1986, 1988) e Al-Osh e Alzaid (1987), descreveram em seus trabalhos que muitas das séries temporais observadas são séries de valores inteiros não negativos e, em particular, séries de contagens. Os modelos usuais, quer lineares quer não lineares, para séries temporais não são neste caso adequados pois o produto de uma constante real por uma variável aleatória de valor inteiro produz uma variável aleatória real. Assim, recorreram à operação *thinning* binomial para substituir a operação de multiplicação usual e propuseram os modelos INAR(1) para modelagem e previsão de séries de valores inteiros não negativos. Contrariando a modelagem tradicional que considera dados discretos como contínuos.

Silva e Silva (2003) consideraram a importância da escolha da melhor ordem-p para os modelos INAR(p) usando critérios estatísticos para a seleção automática da ordem. Os critérios automáticos têm como objetivo equilibrar o risco da escolha de uma ordem menor que a verdadeira, o que provoca inconsistência na estimação dos parâmetros, e o da escolha de uma ordem superior, que conduz ao incremento da variância desses estimadores. Este equilíbrio é feito através da atribuição de um

custo ou penalização pela introdução de variáveis adicionais. A idéia é, então, escolher a ordem  $k$  que minimiza um critério que pode ser escrito como uma função das observações (em geral, o erro quadrático médio da previsão a 1-passo ou a soma dos quadrados dos resíduos) mais um termo de penalização que depende do número de observações e da ordem do modelo a ajustar. Foram comparados dois critérios o primeiro  $AICC$  que é um critério tradicional e o segundo  $AICC_{INAR}$  que é um critério proposto para modelos INAR. Para verificar o desempenho dos critérios, foi calculada a frequência de seleção de ordem em 100 realizações de modelos INAR( $p$ ) com inovações de Poisson e operação *thinning* binomial, para diferentes ordens e valores dos parâmetros. Consideraram-se três dimensões de amostras:  $n = 50$ ,  $n = 100$  e  $n = 200$ . Em todas as condições geradas o critério  $AICC_{INAR}$  mostrou eficiência superior ao critério  $AICC$ .

Silva e Silva (2003) e Latour(1998) propuseram a modelagem da série de contagem de ataques epiléticos de um dado paciente. Estes dados consistem em valores inteiros não negativos sendo uma série de 121 observações. Latour baseado na análise da função de autocorrelação amostral e na função autocorrelação parcial amostral propõe um processo INAR(14). Silva e Silva (2003) ao aplicarem o critério  $AICC_{INAR}$  para selecionar a ordem do modelo INAR que melhor se adapta a este conjunto de dados. Com uma ordem máxima possível de 20, o valor mínimo atingido pelo critério foi de 105,67 para uma ordem  $p = 6$ , i.e., o critério seleciona um modelo INAR(6). Para comparar os dois modelos propostos, o INAR(14) e o INAR(6), calculou-se o valor do critério para o modelo proposto por Latour, obtendo-se (153,18), que é maior que o valor obtido pelo modelo INAR(6) (105,67), concluindo com base no critério  $AICC_{INAR}$  que o modelo INAR(6) é a melhor escolha para a modelagem.

Pavlopoulos e Karlis (2007) propuseram a modelagem de valores inteiros de taxas de chuva (mm/h) para uma região remota no mar da China ( $2^{\circ}S$ ,  $156^{\circ}E$ ) que cobre uma área de  $240 \times 240 \text{ km}^2$ . Os dados foram coletados via radar de precipitação acoplado em um navio. Na seleção da ordem dos modelos INAR, usou-se os tradi-

cionais critérios automáticos  $AIC$  e  $BIC$  proposto por Akaike. A melhor ordem para os modelos  $INAR(p)$  foi a ordem  $p=1$ .

Silva et al (2009) apresentaram duas metodologias para a previsão em modelos  $INAR$ . Os autores compararam a metodologia clássica com a metodologia Bayesiana com a finalidade de obter previsões coerentes para o modelo  $INAR$ . Obtiveram por meio de simulações previsões teóricas para as duas metodologias, em seguida foram apresentados os intervalos de confiança para cada metodologia. Nas simulações a metodologia Bayesiana obteve melhores resultados. Do ponto de vista prático foi modelado a série de trabalhadores envolvidos em acidentes por queimadura, para um conjunto de dados de 120 observações com previsão de até 6 dias, com seus respectivos intervalos de confiança. O modelo  $INAR(1)$  foi o escolhido para a modelagem e previsão dos dados e novamente a metodologia Bayesiana obteve melhores resultados.

### Resumo cronológico da Modelagem de séries de contagem

1979	<b>Steutel e Van Harn</b>	Apresentaram o operador <i>thinning</i> binomial
1987	<b>Al-Osh e Alzaid</b>	Apresentaram o modelo $INAR(1)$
1991	<b>Du e Li</b>	Generalizaram o modelo $INAR(p)$
1994	<b>Gauthier e Latour</b>	Generalizaram o conceito do operador <i>thinning</i>
1998	<b>Latour</b>	Condição de estacionaridade do modelo $INAR$
2000	<b>Hellstron</b>	Teste Dickey-Fuller em modelos $INAR(1)$
2003	<b>Silva e Silva</b>	Critério de seleção da ordem- $AICC_{INAR}$
2003	<b>Silva e Silva</b>	Modelaram a série de ataques epiléticos de um dado paciente
2007	<b>Pavlopoulos e Karlis</b>	Modelaram a série de Taxas de chuva
2009	<b>Silva et al</b>	Previsão tradicional <i>versus</i> Bayesiana para modelos $INAR$

Assim baseados nos trabalhos anteriores decidiu-se estudar o comportamento das séries de índices de qualidade do ar (valores inteiros não negativos) dos principais poluentes atmosféricos monitorados na Região da Grande Vitória. O presente trabalho utiliza técnicas para séries temporais de contagem afim de modelar o comportamento e prever esses índices, evidenciando o uso dos modelos  $INAR(p)$  para a modelagem e previsão, contribuindo na elaboração de políticas ambientais e de saúde.

---

---

# CAPÍTULO 6

---

## MATERIAIS E MÉTODOS

### 6.1 Região de estudo

Segundo o Instituto Brasileiro de Geografia e Estatística (IBGE 2007) a Região da Grande Vitória (RGV) é constituída de 1.454.016 habitantes, correspondente a 43% da população total do Estado do Espírito Santo, abrange uma área de 1.461  $km^2$ , sendo um dos principais pólos de desenvolvimento urbano e industrial do Estado.

### 6.2 Rede automática de monitoramento da qualidade do ar da RGV

A RGV possui uma Rede Automática de Monitoramento da Qualidade do Ar (RAMQAR) inaugurada em 06/06/2000, de propriedade do IEMA sendo gerenciada por este Instituto. A referida rede é distribuída em oito estações localizadas estrategicamente e com equipamentos de medição avançados. Os resultados do monitoramento são divulgados sob forma de índices, para uma melhor compreensão dos níveis de qualidade do ar, todos os dias úteis, via internet e disponibilizados para a imprensa local. As estações componentes da RAMQAR estão distribuídas nos municípios que compõem a Região da Grande Vitória, da seguinte forma : o município Serra com duas estações localizadas nas regiões de Laranjeiras e Carapina; o

município Vitória com três estações localizadas nas regiões de Jardim Camburi, Enseada do Suá e Centro de Vitória. O município de Vila Velha apresenta duas estações localizadas nas regiões do Ibes e Centro de Vila Velha e por último o município de Cariacica com uma estação em Cariacica. (Tabela 6.1 e Figura 6.1).

Tabela 6.1: Estações de monitoramento da qualidade do ar da RGV, coordenadas UTM em metros

Estação	RAMQAR	Coordenadas UTM
Laranjeiras	RAMQAR 01	X: 368.670, 000000 Y: 7.766.879,000000
Carapina	RAMQAR 02	X: 368.755, 000000 Y: 7.762.824,000000
Jardim Camburi	RAMQAR 03	X: 367.512, 000000 Y: 7.759.855,000000
Enseada do Suá	RAMQAR 04	X: 365.228, 000000 Y: 7.753.434,000000
Vitória - Centro	RAMQAR 05	X: 360.392, 000000 Y: 7.752.592,000000
Vila Velha - Ibes	RAMQAR 06	X: 362.433, 000000 Y: 7.749.409,000000
Vila Velha-Centro	RAMQAR 07	X: 364.933, 000000 Y: 7.750.925,000000
Cariacica	RAMQAR 08	X: 353.808, 000000 Y: 7.750.062,000000



Figura 6.1: Estações de monitoramento da qualidade do ar da RGV

A Rede Automática de Monitoramento da Qualidade do Ar da Grande Vitória (RAMQAR) permite que o IEMA quantifique e conheça o comportamento dos seguintes poluentes atmosféricos: Partículas Totais em Suspensão (*PTS*), média 24h; Partículas Inaláveis (*PM<sub>10</sub>*), média 24h; Ozônio (*O<sub>3</sub>*), média 1h; Óxidos de Nitrogênio (*NO<sub>x</sub>*), média 1h; Monóxido de Carbono (*CO*), média 8h e Hidrocarbonetos (*HC*). Para viabilizar a análise da qualidade do ar, realiza-se ainda, o monitoramento dos seguintes parâmetros meteorológicos: Direção dos ventos (*DV*) (direção predominante em 24h); Velocidade dos Ventos (*VV*), média 24h, unidade em *m/s*; Precipitação Pluviométrica (*PP*), precipitação em 24h, unidade em *mm* ; Umidade Relativa do Ar (*UR*) média 1h, unidade em %, Temperatura (*T*), média 1h, unidade em *°C*; Pressão Atmosférica (*P*), média 1h, unidade em *mbar* e Radiação Solar (*I*), média 1h, unidade em *W/m<sup>2</sup>*. Os parâmetros meteorológicos e poluentes monitorados em cada estação RAMQAR encontram-se na Tabela 6.2. Os métodos amostragem dos poluentes monitorados pela RAMQAR são apresentados na Tabela 6.3.

Tabela 6.2: Parâmetros meteorológicos e poluentes monitorados em cada estação RAMQAR

RAMQAR	<i>PTS</i>	<i>PM<sub>10</sub></i>	<i>SO<sub>2</sub></i>	<i>CO</i>	<i>NO<sub>2</sub></i>	<i>HC</i>	<i>O<sub>3</sub></i>	Meteorologia
RAMQAR 01	X	X	X	X	X		X	DV,VV,UR,PP,P,T,I
RAMQAR 02	X	X						
RAMQAR 03	X	X	X		X			DV,VV
RAMQAR 04	X	X	X	X	X	X	X	
RAMQAR 05		X	X	X	X	X		DV,VV
RAMQAR 06	X	X	X	X	X	X	X	
RAMQAR 07		X	X					DV,VV,T,UR
RAMQAR 08	X	X	X	X	X		X	

O período de análises compreendeu 01/01/07 à 19/03/07 com uma previsão para o dia 20/03/07. Na estação de Laranjeiras os poluentes investigados foram *CO* e o *NO<sub>2</sub>*, na de Carapina o *PM<sub>10</sub>*, na estação de Enseada do Suá o *PM<sub>10</sub>*, *CO* e o *NO<sub>2</sub>*, para a estação de Vitória-Centro foram investigados os poluentes *PM<sub>10</sub>* e *CO*. Nas estações de Vila Velha o *PM<sub>10</sub>* e o *SO<sub>2</sub>* foram analisados nas duas estações e por fim em Cariacica foram analisados os poluentes *SO<sub>2</sub>*, *NO<sub>2</sub>*, *O<sub>3</sub>* e *CO*.

Tabela 6.3: Método de amostragem

Poluente	Método de amostragem
$PTS$	Microbalança
$PM_{10}$	Microbalança
$SO_2$	Fluorescência por ultravioleta
$CO$	Absorção infravermelha
$NO_2$	Quimiluminescência
$O_3$	Absorção ultravioleta

### 6.3 Modelagem INAR

A transformação dos valores das concentrações médias dos poluentes investigados, registradas pelas estações de monitoramento da qualidade do ar em índices de qualidade do ar é o primeiro passo para a modelação efetiva das séries de índices de qualidade do ar. O índice é obtido através de uma função linear segmentada, onde os pontos de inflexão são os padrões de qualidade do ar. Desta função, que relaciona a concentração do poluente com o valor índice, resulta um número adimensional de valor inteiro não negativo referido a uma escala com base nos padrões de qualidade do ar. Para cada poluente medido, é calculado um índice conforme a equação (3.1). Os valores obtidos são arredondados para o inteiro mais próximo, servindo como dado de entrada do modelo. Na modelagem das séries de contagem de índices de qualidade do ar da RGV aplicou-se o modelo autoregressivo de valores inteiros INAR, apresentado no Capítulo 4. Na estimação dos parâmetros do modelo foi utilizado o *software* MATLAB 7.8. Para a análise e previsão dos índices foi utilizado o *software* R 2.9.

---

---

# CAPÍTULO 7

---

## RESULTADOS E ANÁLISES

Neste capítulo são apresentadas e discutidas as análises de modelagem e previsão dos índices de qualidade do ar dos principais poluentes atmosféricos da Região da Grande Vitória (RGV), Brasil, para um período de análises de 1 de janeiro de 2007 a 19 de março de 2007 e previsões para 20 de março de 2007 à 25 de março de 2007. Os valores dos índices de qualidade do ar a serem modelados são valores inteiros não negativos, ditos de contagem, sendo utilizado na modelagem desses dados o modelo auto regressivo de valores inteiros INAR(p). O modelo INAR retorna em suas simulações apenas valores inteiros não negativos, o que difere de outros modelos tradicionais que retornam em suas simulações valores reais. Na Seção 7.1 são feitas as análises descritivas das séries de índices de qualidade do ar, na Seção 7.2 são escolhidas as ordens dos modelos e os respectivos valores  $AICC_{INAR}$ , na Seção 7.3 são discutidos e apresentados os ajustes dos modelos INAR(p) e seus parâmetros. Na Seção 7.4 são apresentados os gráficos de modelagem e as tabelas com as previsões para o dia 20/03/07 e as previsões para os dias 21/03/07 à 25/03/07, juntamente com as discussões finais sobre a modelagem e ajuste de cada modelo.

### 7.1 Análise descritiva dos dados

As tabelas e gráficos apresentados a seguir têm como objetivo resumir os dados de índices de qualidade do ar investigados na presente dissertação e respectivas análises descritivas mais relevantes. Os poluentes analisados, as respectivas estações de mo-

nitoramento e toda a estatística descritiva dos dados analisados encontram-se na Tabela 7.1. Para cada poluente foram analisados 78 dados com máximas chegando a 54 e mínimas 1. Avaliou-se também se os poluentes de cada estação respeitaram os padrões Nacionais de qualidade do ar, tabela 3.2, as diretrizes da Organização Mundial de Saúde (OMS 2005), Tabela 3.1, e as classificações da qualidade do ar, Tabela 3.4.

Para a estação de Laranjeiras o poluente que apresentou maior risco à saúde humana conforme resolução CONAMA 03/90, foi o  $NO_2$  com média 16.2 (desvio padrão igual a 5.9) e máxima 37, e na estação de Carapina o poluente  $PM_{10}$  com média 23.7 (desvio padrão igual a 6.4) e máxima de 43. Nas duas estações os padrões nacionais de qualidade do ar, e as diretrizes da OMS, foram respeitados e ainda baseado na classificação da qualidade do ar todos os poluentes investigados nestas estações obtiveram classificação BOA.

Para a estação de Enseada do Suá o poluente que apresentou maior risco à saúde humana conforme resolução CONAMA 03/90, foi o  $PM_{10}$  com média 28.6 (desvio padrão igual a 8.1) e máxima de 54. Para a visualização da distribuição dos dados dessa série foi utilizado o *box-plot* (Figura 7.1), sendo destacado três valores acima do permitido pelas diretrizes da OMS 2005 cuja concentração média em 24h não pode exceder  $50 \mu/m^3$  e ainda baseado na classificação da qualidade do ar foram obtidos três classificações consideradas REGULAR.

Na estação de Vitória-Centro o  $PM_{10}$  novamente se destaca com média 23.8 (desvio padrão igual a 4.8) e máxima 36. Na estação os padrões nacionais de qualidade do ar e as diretrizes da Organização Mundial de Saúde, foram respeitados e ainda baseado na classificação da qualidade do ar todos os poluentes investigados nesta estação obtiveram classificação BOA.

Para a estação de VilaVelha-Centro o destaque foi o  $SO_2$  cuja média foi de 10.4 (desvio padrão igual a 4.4) e máxima de 26. As concentrações correspondentes aos

índices dessa série encontram-se dentro dos padrões nacionais de qualidade do ar, mas quando a análise é feita tendo como base as diretrizes da OMS, que prevê para o  $SO_2$  uma média 24h de  $20 \mu/m^3$ , tem-se 16 observações que excederam essa diretriz, mas mesmo havendo 16 observações excedentes, a classificação da qualidade do ar foi considerada BOA para todos os poluentes investigados nesta estação.

No caso da estação Vila Velha-Ibes os dois poluentes investigados merecem destaque. O primeiro o  $PM_{10}$  de média 26.7 (desvio padrão igual a 8.1) e máxima de 51. Para a visualização da distribuição dos dados dessa série foi utilizado o *box-plot* (figura 7.2), sendo destacado um valor acima do permitido pelas diretrizes da OMS, cuja concentração média em 24h não pode exceder  $50 \mu/m^3$ , e uma classificação da qualidade do ar considerada REGULAR. O segundo foi o  $SO_2$  cuja média foi de 7.3 (desvio padrão igual a 4.4) e máxima 26. As concentrações correspondentes aos índices dessa série encontram-se dentro dos padrões nacionais de qualidade do ar, mas quando a análise é feita tendo como base as diretrizes da OMS, que prevê para o  $SO_2$  uma média 24h de  $20 \mu/m^3$ , tem-se 8 observações de excedem essa diretriz, mas mesmo havendo 8 observações excedentes, a classificação da qualidade do ar foi considerada BOA para todos os poluentes investigados nesta estação.

Na estação de Cariacica o poluente  $O_3$  de média 23.2 (desvio padrão igual a 8.2) e máxima de 50 cujo valor é o limite superior da faixa considerada BOA, Tabela 3.4. E o mesmo em termos de concentração não excedeu as diretrizes da OMS que prevê para o  $O_3$  uma média 8h de  $100 \mu/m^3$ .

O resumo dos poluentes que excederam as diretrizes da OMS 2005 e as respectivas estações de monitoramento encontram-se na Tabela 7.2

Tabela 7.1: Estatísticas descritivas dos índices de qualidade do ar dos principais poluentes das estações de monitoramento da qualidade do ar da RGV, no período de 01/01/07 à 19/03/07. (78 dias)

ESTAÇÃO	POLUENTE	n	média	mínimo	Percentis			máximo
					25	50	75	
LARANJEIRAS	<i>CO</i>	78	5.6	4	5	6	6	10
	<i>NO<sub>2</sub></i>	78	16.2	9	12	14.5	19	37
CARAPINA	<i>PM<sub>10</sub></i>	78	23.7	10	19	23	27	43
ENSEADA DO SUÁ	<i>PM<sub>10</sub></i>	78	28.6	10	23	27	33	54
	<i>CO</i>	78	6.6	3	5	6	8	20
	<i>NO<sub>2</sub></i>	78	20.4	12	17	20	24	35
VITÓRIA-CENTRO	<i>PM<sub>10</sub></i>	78	23.8	8	21	23	27	36
	<i>CO</i>	78	17.2	7	11	17	22	30
VV-IBES	<i>PM<sub>10</sub></i>	78	26.7	10	21.25	26	32	51
	<i>SO<sub>2</sub></i>	78	7.3	3	4	6	9.75	26
VV-CENTRO	<i>PM<sub>10</sub></i>	78	22.1	10	17	20	26	46
	<i>SO<sub>2</sub></i>	78	10.4	6	8	9	11	26
CARIACICA	<i>SO<sub>2</sub></i>	78	2.0	1	1	2	3	5
	<i>NO<sub>2</sub></i>	78	20.8	6	14.25	20	26.75	38
	<i>O<sub>3</sub></i>	78	23.2	9	17	23	27	50
	<i>CO</i>	78	4.9	2	4	5	6	8

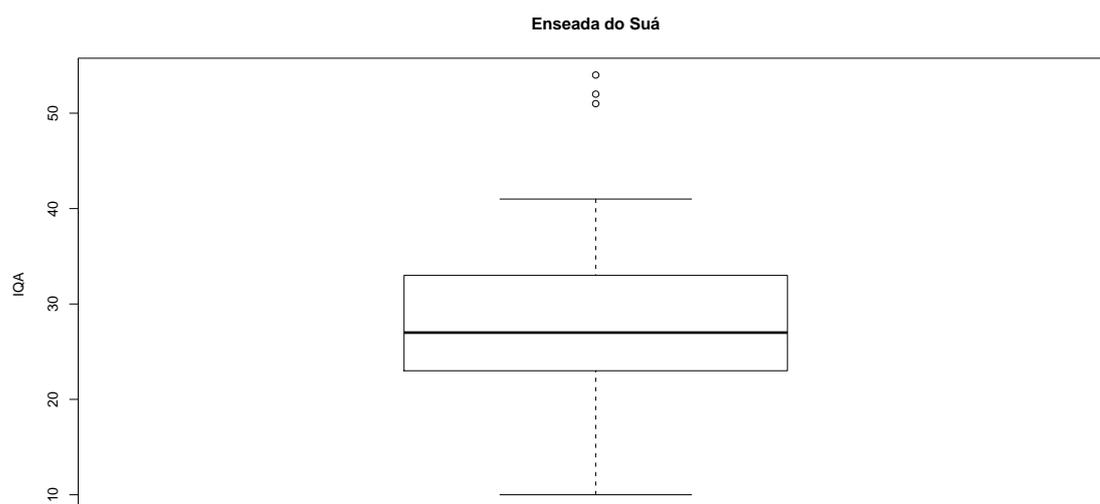


Figura 7.1: Box Plot *PM<sub>10</sub>* para Enseada do Suá

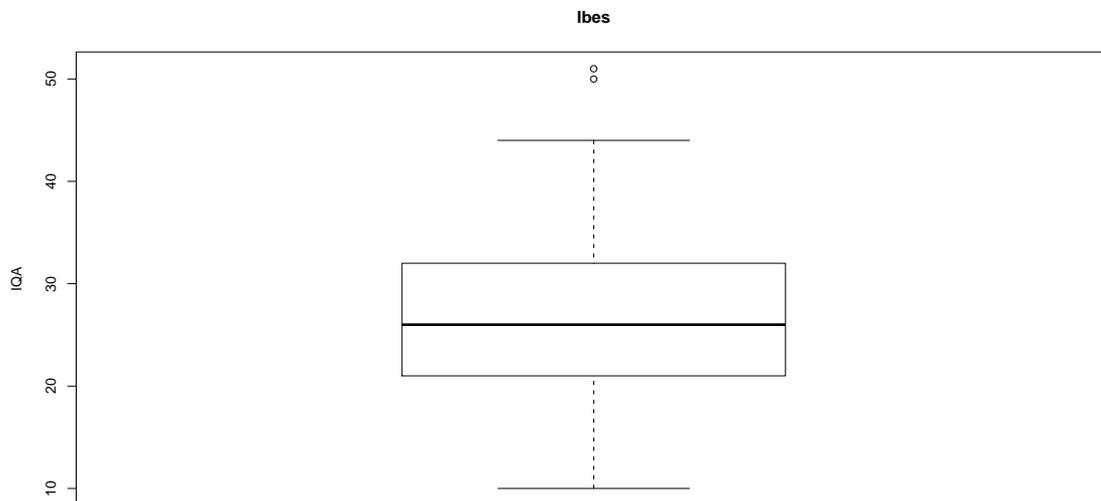


Figura 7.2: Box Plot  $PM_{10}$  para Ibes

Tabela 7.2: Poluentes que excederam as diretrizes da OMS 2005

Estação	Poluente	$n^{\circ}$ de excedentes
Enseada do Suá	$PM_{10}$	2
VilaVelha-Ibes	$PM_{10}$	1
	$SO_2$	8
VilaVelha-Centro	$SO_2$	16

## 7.2 Seleção da ordem dos modelos INAR(p)

Segundo Silva e Silva(2003) a modelação efetiva de uma série de observações por modelos INAR depende da determinação da ordem do modelo a usar. Para cada poluente investigado nesta dissertação foram propostas sete ordens sendo que para a seleção da ordem propõe-se o critério automático  $AICC_{INAR}$ , conforme metodologia descrita na Seção 4.2. A idéia é escolher a ordem  $k$ ,  $k = 1, \dots, p$  que minimiza o critério  $AICC_{INAR}$ .

Na estação de Laranjeiras, para as séries de índices de qualidade do ar dos poluentes  $CO$  e  $NO_2$  as ordens escolhidas foram as mesmas,  $p=1$ , os respectivos valores do  $AICC_{INAR}$  foram 123.52 e 346.59. Na estação de Carapina a ordem escolhida para a série de índices de qualidade do ar do poluente  $PM_{10}$  foi  $p=2$  e o valor do  $AICC_{INAR}$  é 361.79. Na estação de Enseada do Suá, para as séries de índices de qualidade do ar dos os poluentes  $PM_{10}$ ,  $CO$  e  $NO_2$  as ordens escolhidas foram as mesmas,  $p=1$ , os respectivos valores do  $AICC_{INAR}$  foram 400.90, 229.85 e 320.53. Para a estação Vitória-Centro a escolha da ordem para série do poluente  $PM_{10}$  foi  $p=2$  com  $AICC_{INAR}$  igual a 323.90 e para o poluente  $CO$  a ordem foi  $p=7$  com  $AICC_{INAR}$  igual a 358.44. A ordem das séries de índices de qualidade do ar dos poluentes  $PM_{10}$  e  $SO_2$  da Estação de Vila Velha-IBES foram as mesmas,  $p=1$ , os respectivos valores do  $AICC_{INAR}$  foram 392.32 e 299.46. Na estação de Vila Velha-Centro, para as séries de índices de qualidade do ar dos poluentes  $PM_{10}$  e  $SO_2$  as ordens escolhidas foram,  $p=1$  e  $p=7$  os respectivos valores do  $AICC_{INAR}$  foram 379.72 e 280.58. Para a estação de Cariacica os poluentes  $SO_2$ ,  $NO_2$ ,  $O_3$  e  $CO$  obtiveram as mesmas ordens,  $p=1$ , os respectivos valores do  $AICC_{INAR}$  foram 86.15, 394.90, 384.48 e 153.35. A Tabela 7.3 resume as ordens e os respectivos valores  $AICC_{INAR}$  destacados em negrito.

Tabela 7.3: Seleção da ordem do modelo INAR(p) baseados no critério automático  $AICC_{INAR}$ , período das análises 01/01/07 à 19/03/07.

ESTAÇÃO/POLUENTE	ORDENS						
	p=1	p=2	p=3	p=4	p=5	p=6	p=7
<b>LARANJEIRAS</b>							
<i>CO</i>	<b>123.52</b>	126.56	128.49	131.88	134.87	136.82	139.19
<i>NO<sub>2</sub></i>	<b>346.59</b>	349.66	352.03	355.28	357.59	361.09	363.78
<b>CARAPINA</b>							
<i>PM<sub>10</sub></i>	362.07	<b>361.79</b>	364.92	368.74	372.09	375.34	381.83
<b>ENSEADA DO SUÁ</b>							
<i>PM<sub>10</sub></i>	<b>400.90</b>	408.97	410.28	411.96	414.42	415.53	418.14
<i>CO</i>	<b>229.85</b>	231.60	232.75	236.13	239.41	242.12	245.86
<i>NO<sub>2</sub></i>	<b>320.53</b>	325.53	322.49	325.53	328.69	329.20	335.14
<b>VITÓRIA-CENTRO</b>							
<i>PM<sub>10</sub></i>	325.14	<b>323.90</b>	327.78	331.70	335.45	337.00	377.95
<i>CO</i>	364.07	366.75	367.21	376.71	369.94	361.23	<b>358.44</b>
<b>VV-IBES</b>							
<i>PM<sub>10</sub></i>	<b>392.32</b>	392.67	403.26	398.36	404.01	405.49	407.62
<i>SO<sub>2</sub></i>	<b>299.46</b>	302.33	305.50	308.37	310.26	311.85	310.61
<b>VV-CENTRO</b>							
<i>PM<sub>10</sub></i>	<b>379.72</b>	381.07	392.85	387.44	396.95	389.49	391.35
<i>SO<sub>2</sub></i>	284.29	297.41	290.59	293.98	296.26	283.75	<b>280.58</b>
<b>CARIACICA</b>							
<i>SO<sub>2</sub></i>	<b>86.15</b>	89.35	92.63	95.95	98.61	91.08	87.48
<i>NO<sub>2</sub></i>	<b>394.90</b>	395.41	397.47	398.18	401.61	411.50	417.64
<i>O<sub>3</sub></i>	<b>384.48</b>	396.66	394.60	396.25	395.71	396.55	400.37
<i>CO</i>	<b>153.35</b>	154.07	156.58	157.49	159.08	161.55	161.39

### 7.3 Modelos Ajustados

Após a determinação da ordem dos modelos verifica-se a não correlação dos resíduos dos modelos, e para isso foi utilizado o teste Box-Pierce. Na tabela 7.4 todo valor-p do teste Box-Pierce para os modelos propostos foram maiores que 0.05 ou seja os resíduos são não correlacionados em todos os casos. Depois de verificada a não correlação dos resíduos do modelo estima-se os parâmetros de cada modelo. Os estimadores dos parâmetros dos modelos INAR(p) escolhidos, são obtidos através do critério proposta por Whittle (1953) e usualmente designada por critério de Whittle considerando as seguintes restrições:  $0 < \alpha_i < 1$ ;  $i=1, \dots, p$ ,  $\sum_1^p \alpha_i < 1$  e  $0 < \lambda < 60$ . Conforme tabela 7.4 todos os parâmetros estão dentro do especificado pelo critério de Whittle (SILVA e SILVA 2004). Segundo Latour (1998) a condição de estacionariedade de um processo INAR é confirmada quando  $\sum_1^p \alpha_i < 1$ , logo todas as séries geradas são estacionárias.

Tabela 7.4: Modelos INAR(p) ajustados

ESTAÇÃO	POLUENTE	ORDEM	$AICC_{INAR}$	$\lambda$	$\sum_1^p \alpha$	BOX-PIERCE Valor-p
LARANJEIRAS	$CO$	1	123.52	4.4923	0.2108	0.9523
	$NO_2$	1	346.59	9.5943	0.4079	0.4733
CARAPINA	$PM_{10}$	2	361.79	17.0456	0.2805	0.5550
ENSEADA DO SUÁ	$PM_{10}$	1	400.90	18.8655	0.3410	0.5318
	$CO$	1	229.85	5.2430	0.2105	0.5465
	$NO_2$	1	320.53	12.6337	0.3818	0.5565
VITÓRIA-CENTRO	$PM_{10}$	2	323.90	18.4466	0.2269	0.9412
	$CO$	7	358.44	1.8870	0.3952	0.5102
VV-IBES	$PM_{10}$	1	392.32	16.1203	0.3978	0.5162
	$SO_2$	1	299.46	4.2009	0.4311	0.4930
VV-CENTRO	$PM_{10}$	1	379.72	12.8390	0.4191	0.3044
	$SO_2$	7	284.29	1.2070	0.8842	0.4229
CARIACICA	$SO_2$	1	86.15	1.0400	0.4651	0.4659
	$NO_2$	1	394.90	10.9937	0.4723	0.4449
	$O_3$	1	384.48	10.9179	0.5308	0.3639
	$CO$	1	153.35	3.0570	0.3774	0.4780

## 7.4 Modelagem e previsão das séries dos índices de qualidade do ar

Nesta seção são apresentados os gráficos dos índices de qualidade do ar dos poluentes investigados nesta dissertação, Figuras 7.3 a 7.18. Cada figura possui seis gráficos dispostos em três linhas e duas colunas. Na primeira coluna primeira linha de cada figura encontra-se o gráfico da série IQA. Na primeira coluna segunda linha de cada figura encontra-se o gráfico da função de autocorrelação (acf) da série IQA e na primeira coluna terceira linha a função de autocorrelação parcial (pacf) da série IQA. Na segunda coluna primeira linha de cada figura encontra-se o gráfico da modelagem INAR destacado em vermelho. Na segunda coluna segunda linha de cada figura encontra-se o gráfico da função de autocorrelação (acf) do modelo INAR proposto para modelar a série IQA e na segunda coluna terceira linha a função de autocorrelação parcial (pacf) do modelo INAR proposto para modelar a série IQA. Na avaliação da modelagem INAR destaca-se dois pontos, o primeiro a não correlação dos resíduos, e o segundo ponto a semelhança das estruturas de autocorrelação da série o do modelo.

Conforme Tabela 7.4, em todas as estações de monitoramento da qualidade do ar os resíduos dos modelos são não correlacionados ou seja os resíduos não apresentam uma estrutura de dependência, o que nos garante que toda a dependência das séries estudadas nesta dissertação foram explicadas pelos modelos. A verificação da não correlação dos resíduos dos modelos teve como base o teste *BOX-PIERCE* cujo valor-p foram satisfatórios e variando de 0.3044 à 0.9523.

A verificação da semelhança das estruturas de autocorrelação dos dados e dos modelos é realizada com base nos gráficos de acf e pacf das séries e dos modelos. Em todos os casos foram obtidas consideráveis semelhanças estruturais, satisfazendo com isso os dois pontos de verificação.

Para a previsão, cada figura destacou a previsão para o dia 20/03/07. O resumo das previsões do dia 20/03/07 encontram-se na Tabela 7.5. As previsões dos índices de qualidade do ar para os dias 21/03/07 à 25/03/07 encontram-se na Tabela 7.6.

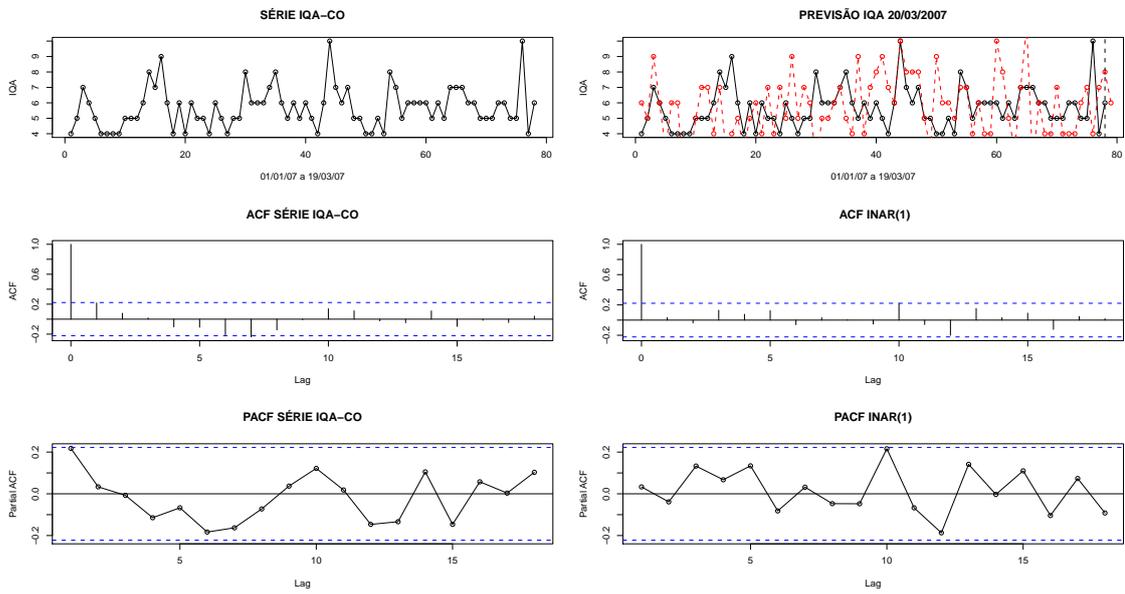


Figura 7.3: Modelagem da Série IQA-CO Laranjeiras

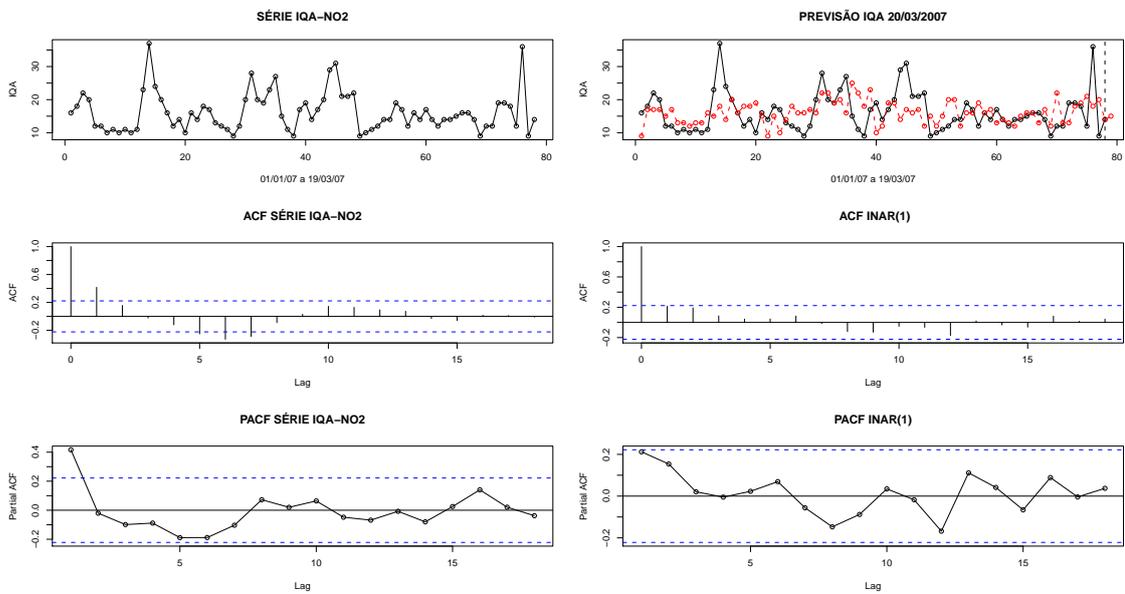


Figura 7.4: Modelagem da Série IQA-NO<sub>2</sub> Laranjeiras

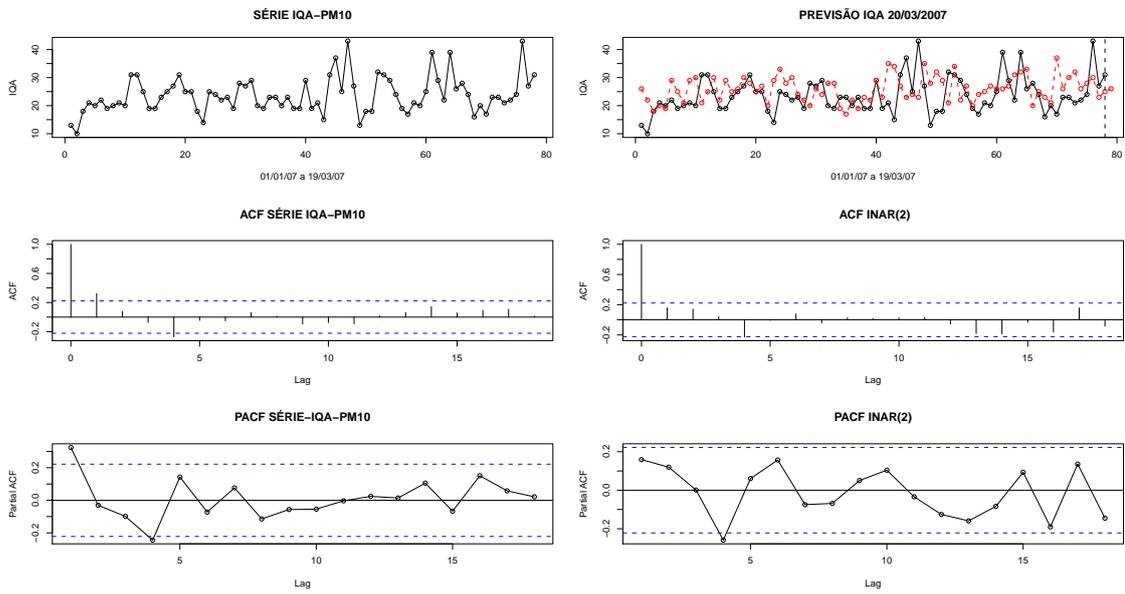


Figura 7.5: Modelagem da Série IQA- $PM_{10}$  Carapina

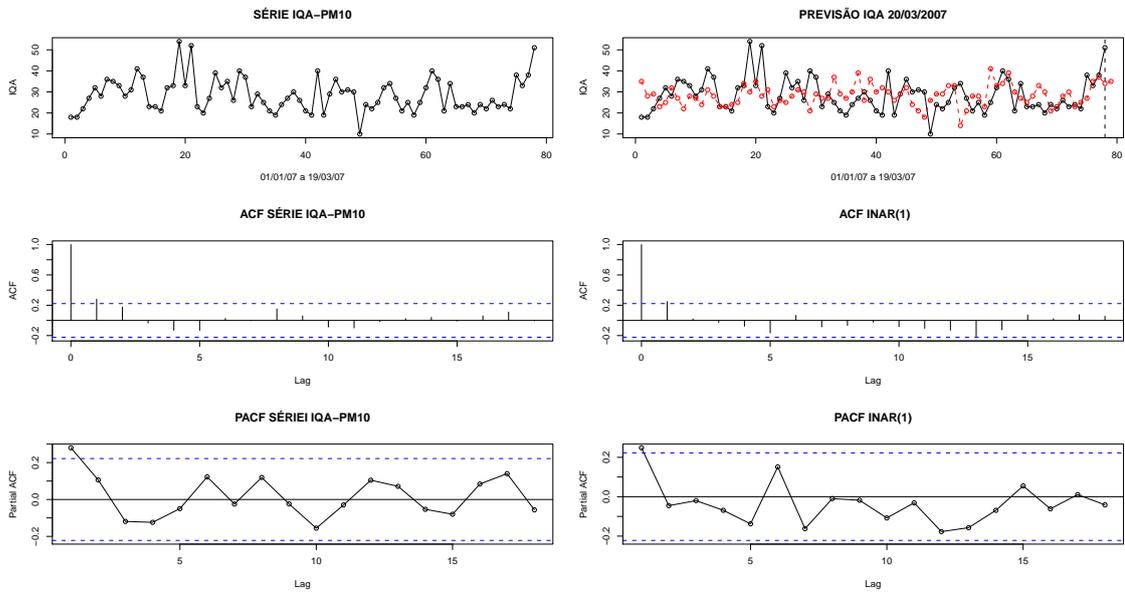


Figura 7.6: Modelagem da Série IQA- $PM_{10}$  Enseada do Suá

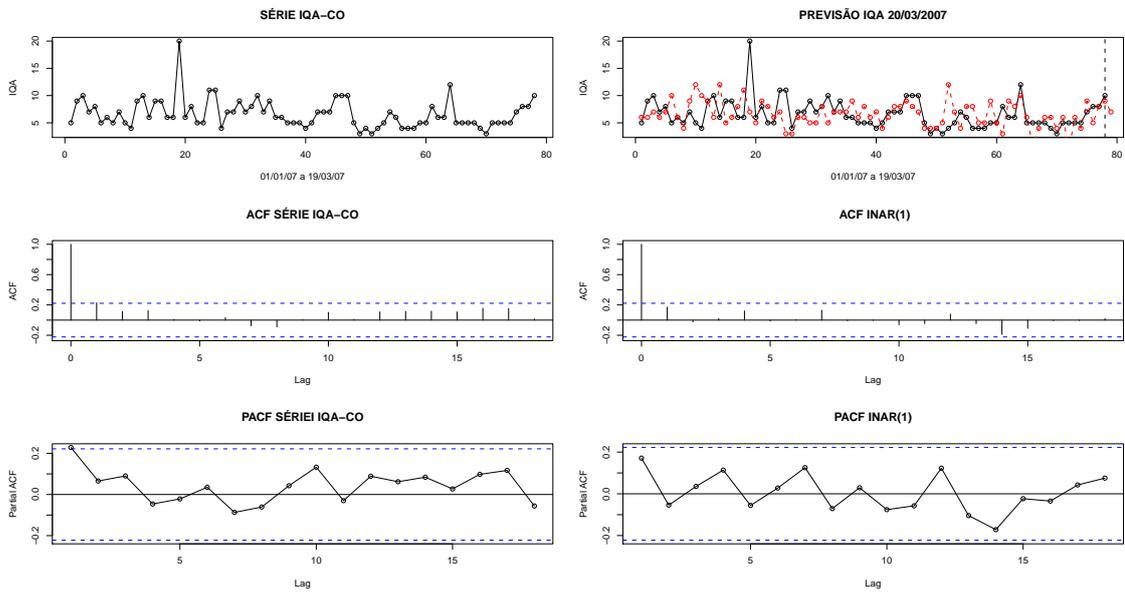


Figura 7.7: Modelagem da Série IQA-CO Enseada do Suá

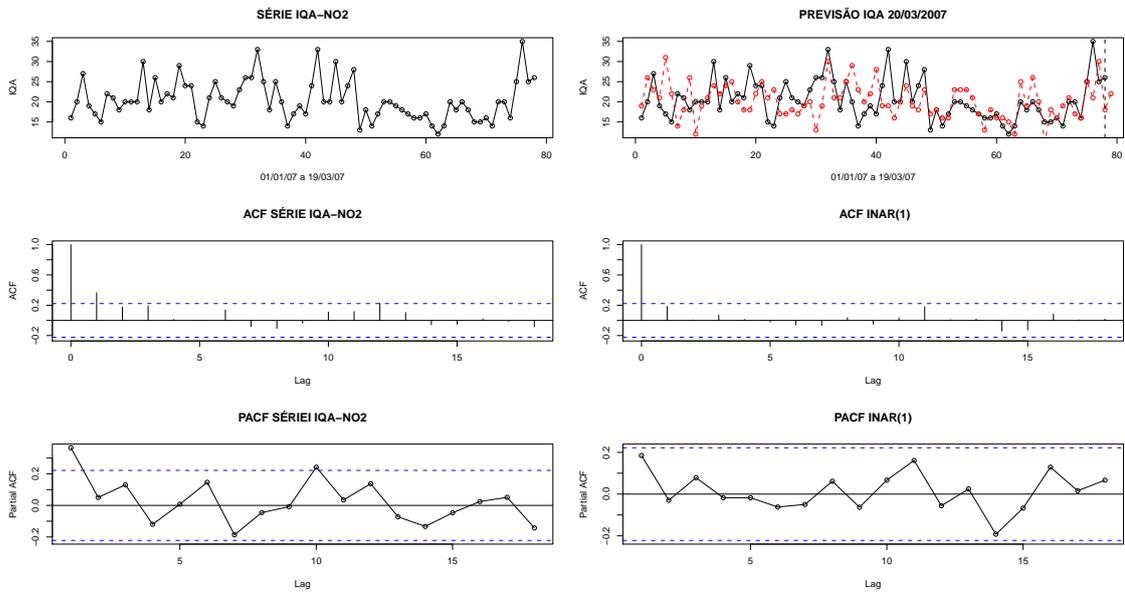


Figura 7.8: Modelagem da Série IQA-NO<sub>2</sub> Enseada do Suá

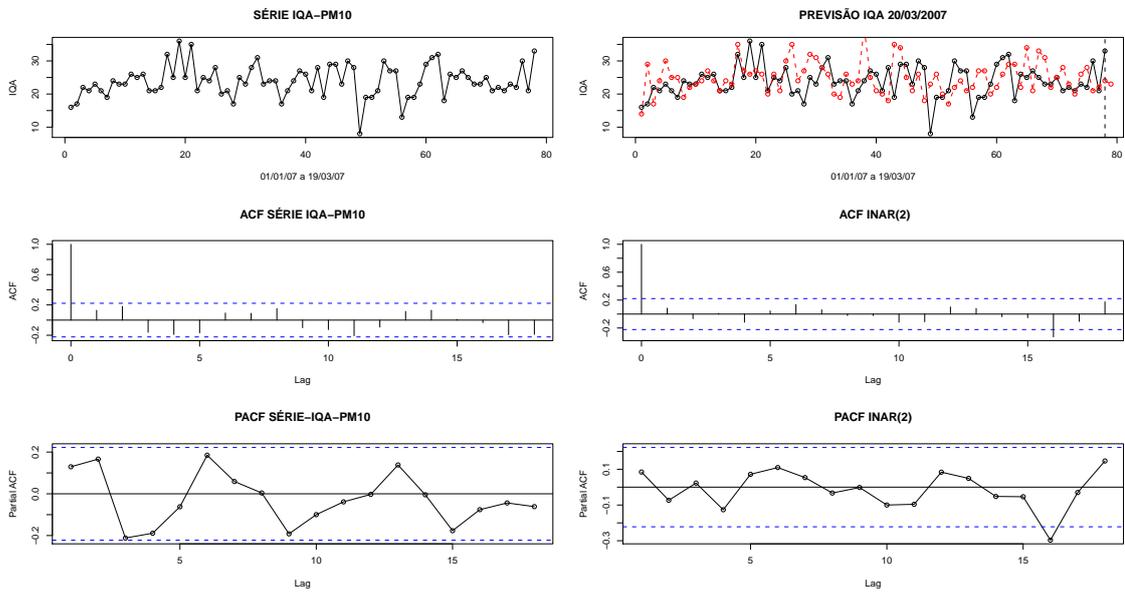


Figura 7.9: Modelagem da Série  $IQA-PM_{10}$  Vitória -Centro

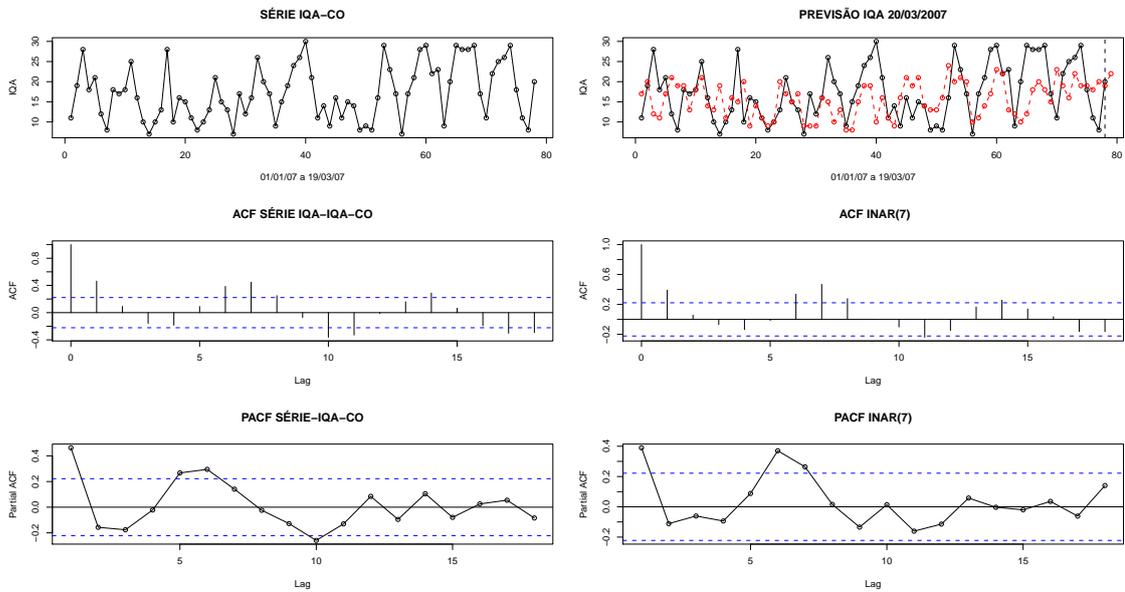


Figura 7.10: Modelagem da Série  $IQA-CO$  Vitória -Centro

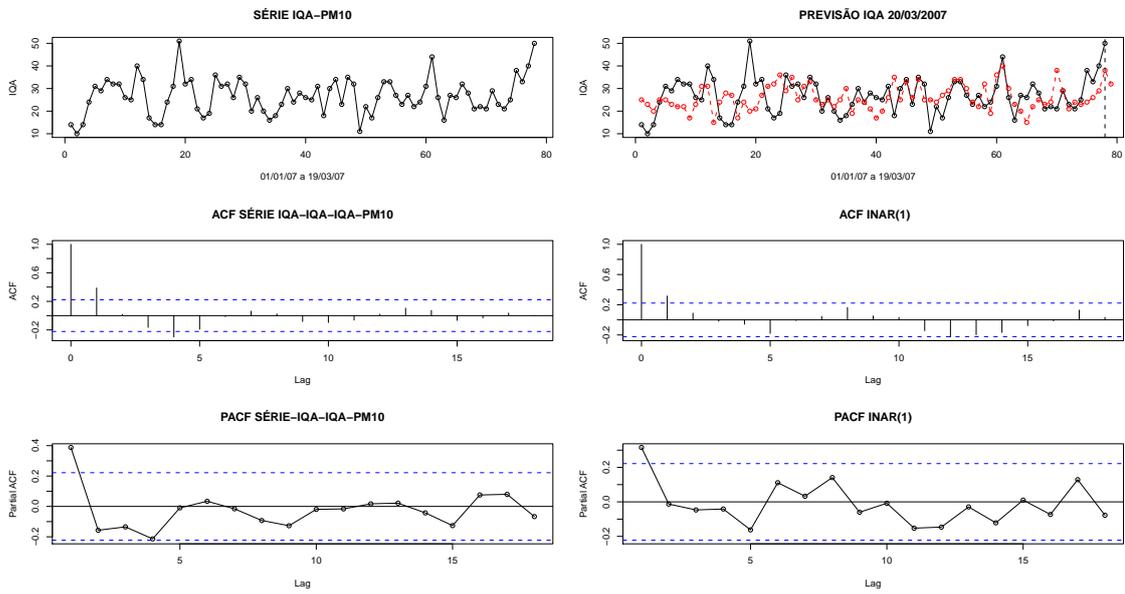


Figura 7.11: Modelagem da Série IQA- $PM_{10}$  VV-Ibes

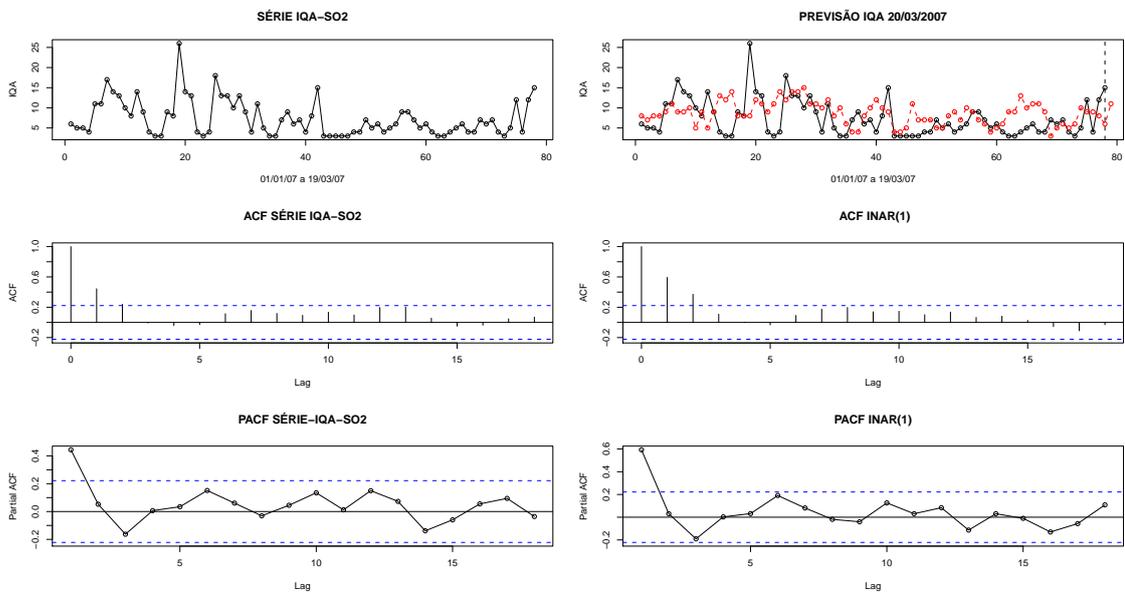


Figura 7.12: Modelagem da Série IQA- $SO_2$  VV-Ibes

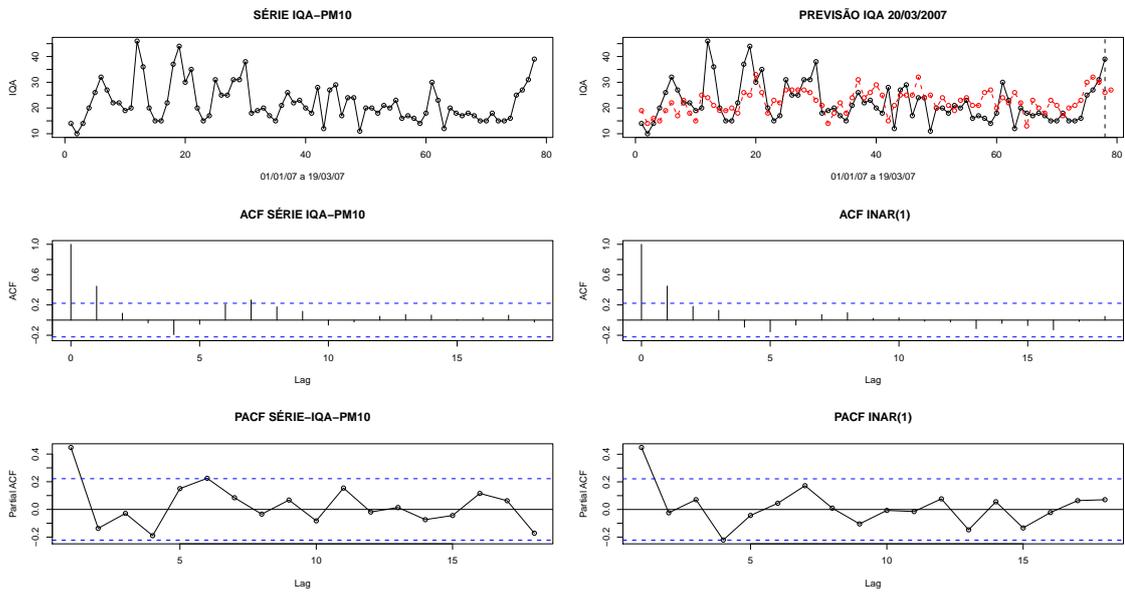


Figura 7.13: Modelagem da Série IQA- $PM_{10}$  VV-Centro

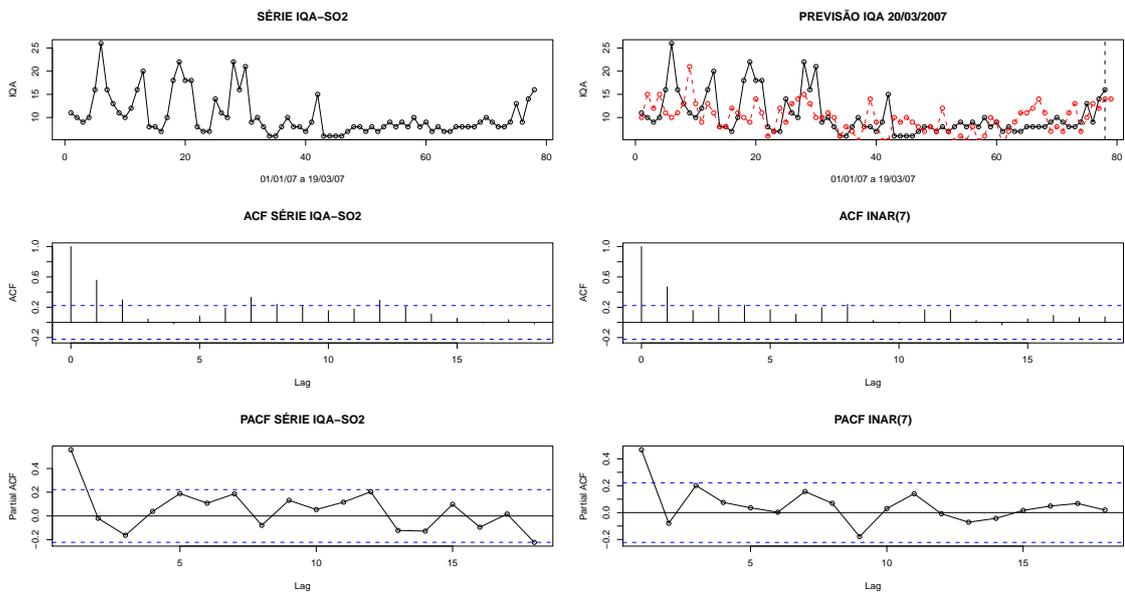


Figura 7.14: Modelagem da Série IQA- $SO_2$  VV-Centro

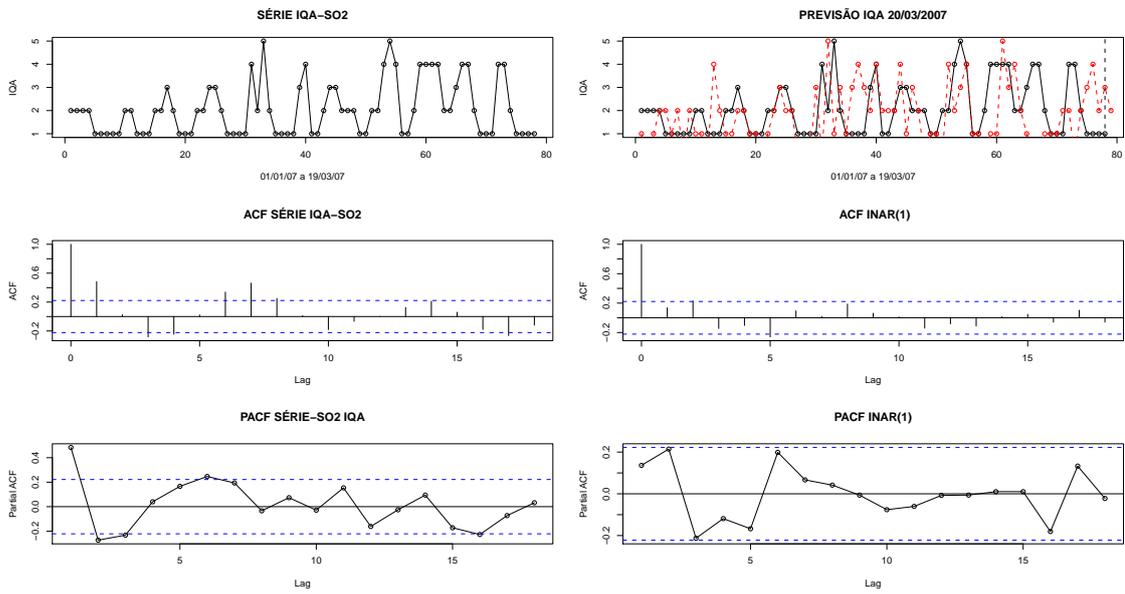


Figura 7.15: Modelagem da Série IQA- $SO_2$  Cariacica

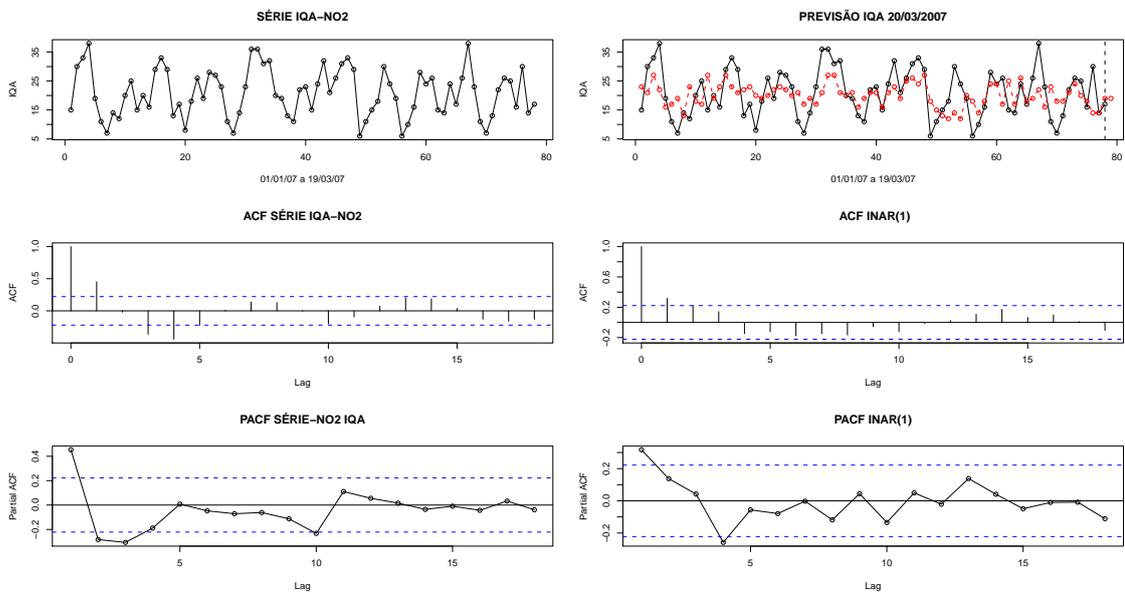


Figura 7.16: Modelagem da Série IQA- $NO_2$  Cariacica

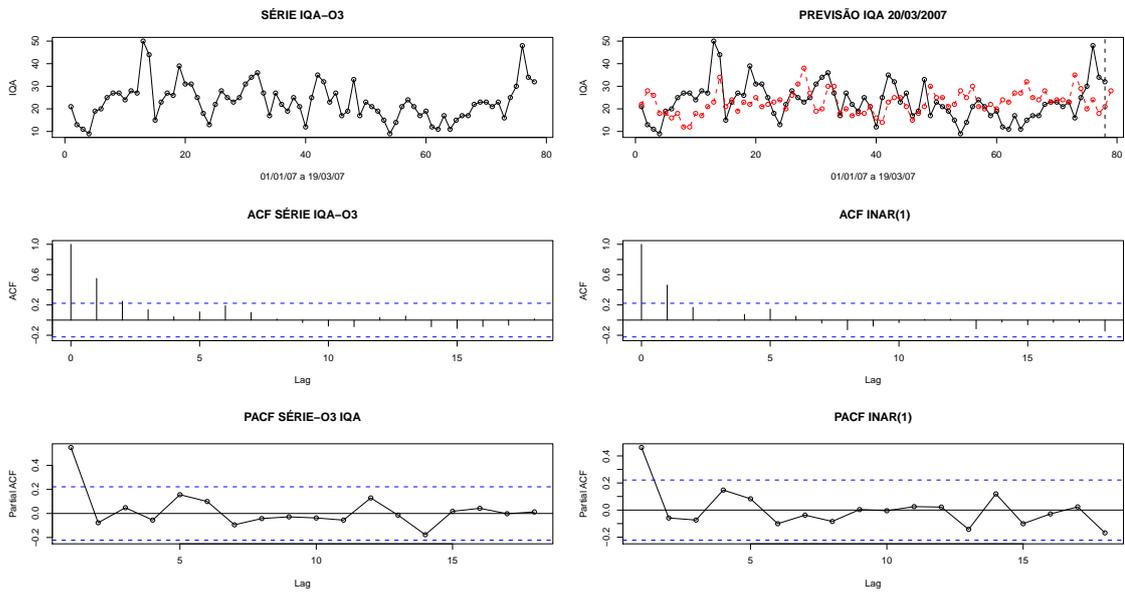


Figura 7.17: Modelagem da Série IQA- $O_3$  Cariacica

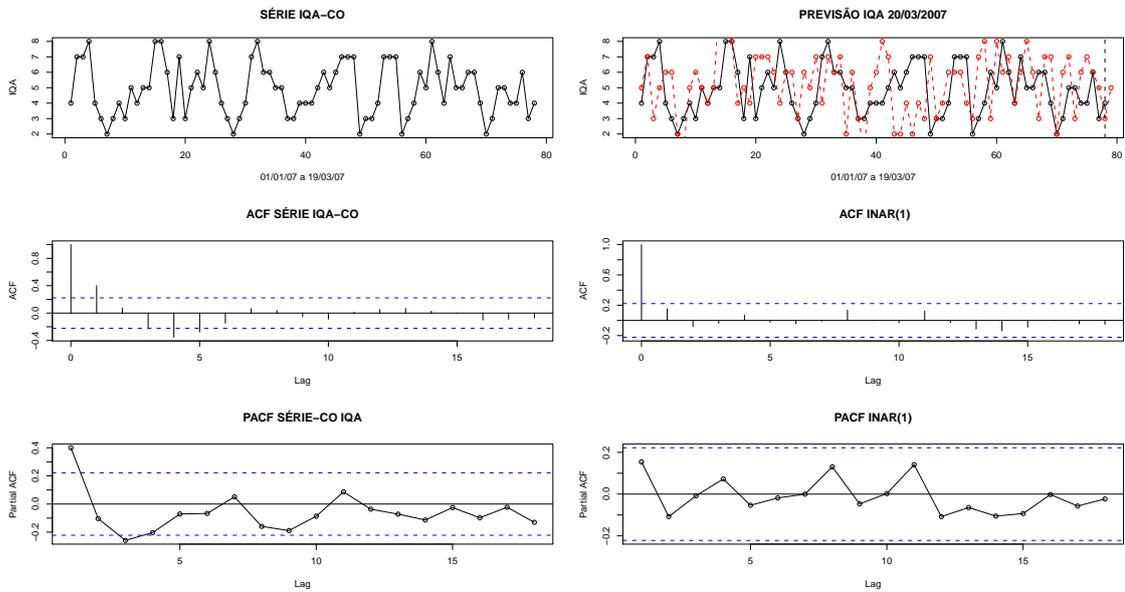


Figura 7.18: Modelagem da Série IQA- $CO$  Cariacica

A previsão dos índices de qualidade do ar do dia 20/03/07 dos poluentes atmosféricos investigados nessa dissertação e seus respectivos intervalos de confiança (IC), estão representados na Tabela 7.5. Todas as previsões resultantes encontram-se dentro do padrão de qualidade do ar considerada **BOA** sendo que a qualidade do ar real (QA REAL) e a qualidade do ar prevista (QA PREVISTA) foram as mesmas para todos as previsões do dia 20/03/07. No entanto baseados nas diretrizes da OMS 2005, a previsão do poluente  $SO_2$  da estação Vila Velha-Centro no dia 20/03/07 excedeu em termos de concentração ( $\mu g/m^3$ ) o valor de  $20 \mu g/m^3$  para uma média de 24h, resultado esse que levanta a discussão da reformulação dos padrões Nacionais que datam de 1990.

Tabela 7.5: Previsão da qualidade do ar-(20/03/07)

ESTAÇÃO	POLUENTE	IQA	IQA	IC	QA	QA
		REAL	PREVISTO		REAL	PREVISTA
LARANJEIRAS	$CO$	5	6	(3.17 , 8.34)	BOA	BOA
	$NO_2$	15	15	(4.60 , 26.96)	BOA	BOA
CARAPINA	$PM_{10}$	31	29	(14.02 , 38.09)	BOA	BOA
ENSEADA DO SUÁ	$PM_{10}$	37	35	(19.51 , 50.28)	BOA	BOA
	$CO$	12	7	(2.34 , 12.47)	BOA	BOA
	$NO_2$	30	22	(13.36 , 31.57)	BOA	BOA
VITÓRIA CENTRO	$PM_{10}$	26	23	(14.23 , 32.51)	BOA	BOA
	$CO$	22	22	(11.54 , 33.27)	BOA	BOA
VV IBES	$PM_{10}$	19	32	(17.42 , 46.20)	BOA	BOA
	$SO_2$	16	11	(2.83 , 18.69)	BOA	BOA
VV CENTRO	$PM_{10}$	39	27	(14.47 , 40.47)	BOA	BOA
	$SO_2$	13	14	(3.66 , 20.72)	BOA	BOA
CARIACICA	$SO_2$	1	2	(0.64 , 4.32)	BOA	BOA
	$NO_2$	26	19	(5.60 , 32.33)	BOA	BOA
	$O_3$	33	28	(14.40 , 41.72)	BOA	BOA
	$CO$	5	5	(1.57 , 7.52)	BOA	BOA

As previsões para os dias 21/03/07 à 25/03/07 encontram-se na Tabela 7.5. As previsões para os respectivos dias tendem para a média dos dados de cada poluente. As médias dos poluentes encontram-se na Tabela 7.1.

Tabela 7.6: Previsões da qualidade do ar-(21/03/07) à (25/03/07)

ESTAÇÃO		PREVISÕES				
		21/03	22/03	23/03	24/03	25/03
LARANJEIRAS	<i>CO</i>	5	6	5	6	6
	<i>NO<sub>2</sub></i>	16	15	16	16	16
CARAPINA	<i>PM<sub>10</sub></i>	26	26	24	24	24
ENSEADA DO SUÁ	<i>PM<sub>10</sub></i>	32	26	27	28	28
	<i>CO</i>	8	5	7	7	7
	<i>NO<sub>2</sub></i>	21	21	21	20	20
VITÓRIA-CENTRO	<i>PM<sub>10</sub></i>	25	23	22	23	23
	<i>CO</i>	25	21	15	17	17
IBES	<i>PM<sub>10</sub></i>	24	21	20	24	27
	<i>SO<sub>2</sub></i>	9	8	8	8	7
VILA VELHA- CENTRO	<i>PM<sub>10</sub></i>	21	21	20	22	26
	<i>SO<sub>2</sub></i>	12	12	11	11	11
CARIACICA	<i>SO<sub>2</sub></i>	3	2	2	1	1
	<i>NO<sub>2</sub></i>	20	23	23	22	21
	<i>O<sub>3</sub></i>	26	25	24	24	24
	<i>CO</i>	6	6	5	5	5

---

---

# CAPÍTULO 8

---

## CONCLUSÃO E RECOMENDAÇÕES

### 8.1 Conclusão

A presente dissertação apresentou um estudo de modelagem e previsão de índices de qualidade do ar da Região da Grande Vitória (RGV), considerando séries de contagem, dos principais poluentes atmosféricos monitorados pelas estações de qualidade do ar da RGV.

Os modelos INAR(p) utilizados neste trabalho consideraram a característica de contagem dos dados, retornando em suas simulações apenas valores inteiros e positivos, obtendo em suas modelagens e previsões valores que acompanharam as séries de contagem. Os modelos INAR explicaram as características das séries apenas com os dados da própria série, baseados nas características probabilísticas dos mesmos.

A escolha das ordens para os modelos INAR(p) foram baseadas no critério  $AICC_{INAR}$ . Estas escolhas foram consideradas satisfatórias tendo em vista a qualidade dos ajustes dos modelos. Foram gerados para o dia 20/03/07 previsões de índices de qualidade do ar dos poluentes investigados. Baseados nas previsões e nas classificações da qualidade do ar do dia 20/03/07, todos os poluentes obtiveram classificação considerada BOA, destacando apenas a previsão do poluente  $SO_2$  da estação de Vila Velha-Centro o qual excedeu em termos de concentração as diretrizes da OMS. As previsões para os dias 21/03/07 à 25/03/07 tenderam para a média de cada poluente.

## 8.2 Recomendações para trabalhos futuros

- Efetuar um estudo de comparação das técnicas de seleção da ordem para modelos INAR envolvendo o critério automático  $AICC_{INAR}$  e os tradicionais critérios automáticos presentes na literatura, tendo como base de dados os índices de qualidade do ar.
- Efetuar um estudo de comparação entre as metodologias de previsão clássicas e a metodologia de previsão Bayesiana para modelos INAR
- Implementação sazonal na modelagem INAR.

A importância de realizarmos estudos que consideram as características de contagem dos dados nos remete as modificações das estruturas das estatísticas convencionais de modelagem. Comumente as estatísticas tradicionais retornam em suas simulações valores reais que deixam de considerar as características de contagem dos dados analisados. Na modelagem INAR temos a resposta pertencente aos inteiros não negativos, respeitando e preservando as características de contagem dos dados.

---

## REFERÊNCIAS BIBLIOGRÁFICAS

- AGIRRE-BASURKO, E., IBARRA-BERASTEGI, G. e MADARIAGA, I. (2006). Regression and multilayer perceptron-based models to forecast hourly  $O_3$  and  $NO_2$  levels in the Bilbao area. *Environmental Modelling e Software*, 21, 430:446.
- AKAIKE, H.(1974). A new look at the statistical model identification. *IEEE Trans. Auto.Control*, AC-19, 716:723.
- AL-OSH, M. A. e ALZAID, A. A.(1987). First-order integer valued autoregressive (INAR(1)) process. *Journal of Time Series Analysis*, 8, 261:75.
- ÁLVARES JR., O. M.;LACAVA, C. I. V.; FERNANDES, P. S. (2002). Emissões atmosféricas. *SENAI/DN*.
- BOUBEL, R. et al. (1994). Fundamentals of air pollution. *San Diego, CA: Academic Press*, 3.
- BRAGA, A. et al.(2002). Poluição atmosférica e seus efeitos na saúde humana. In: Sustentabilidade na geração e uso de energia no Brasil: os próximos 20 anos, Campinas. Disponível em [www.cgu.rei.unicamp.br/energia2020/papers/paperSaldiva.pdf](http://www.cgu.rei.unicamp.br/energia2020/papers/paperSaldiva.pdf).
- CASTRO, H. A., GOUVEIA, N., ESCAMILLA-CEJUDO, J. A. (2003). Questões metodológicas para a investigação dos efeitos da poluição do ar na saúde. *Rev. Bras. Epidemiol.*,6,2.
- CONAMA.(1990). Resolução CONAMA 03, de 28 de junho de 1990.
- DU, J.-G. e LI, Y. (1991). The integer-valued autoregressive (INAR(p)) model. *Journal of Time Series Analysis*, 12,129:42.
- ESCAMILLA, J. Á. (1999). Air Pollution and Fatal Acute Myocardial Infarction in Mexico City: A Case-crossover study: Tese Ph.D. Baltimore: School

of Hygiene and Public Health, Johns Hopkins University.

FERLAND, R., LATOUR, A. e ORAICHI, D. (2006). Integer-Valued Garch Process. *Journal of Time Series Analysis*, 27, 6, 923:942.

GAUTHIER, G. e LATOUR, A. (1994). Convergence Forte des Estimateurs des Parametres d'un Processus GENAR(p). *Ann. Sci. Math. Quebec*, 18, 49:71.

HELLSTRÖM, J. (2001). Unit root testing in integer-valued AR(1) models. *Economics Letters*, 70, 9:14.

HURVICH, C. e TSAI, C.L. (1989) Regression and time series model selection in small samples. *Biometrika*, 76, 297:307.

IEMA. (2007). Relatório da qualidade do ar no estado do Espírito Santo, 2007.

JAMES, B.R. (1981). *Probabilidade: um curso em nível intermediário*. Impa, Rio de Janeiro.

LATOUR, A. (1998) Existence and stochastic structure of a non-negative integer-valued autoregressive process. *J. time ser. Anal*, 19, 439:455.

LIRA, T, S. (2009) Modelagem e previsão da qualidade do ar na Cidade de Uberlândia-MG. Tese de Doutorado, Minas Gerais: Programa de Pós-Graduação em Engenharia Química: Universidade Federal de Uberlândia, Faculdade de Engenharia Química.

LIU, D. H. F.; LIPTAK, B. G. (1999). *Environmental engineers handbook*. Boca Raton: CRC Press LLC.

MARTINS, L. et al. (2001). Relação entre poluição atmosférica e atendimentos por infecção de vias aéreas superiores no município de São Paulo: avaliação do rodízio de veículos. *Revista Brasileira de Epidemiologia*, 4, 220:229.

MCKENZIE, E. (1986) Autoregressive moving-average process with negative-binomial and geometric marginal distributions. *Adv. Appl. Probab.*, 18, 679:705.

MCKENZIE, E. (1988) Some ARMA models for dependent sequences of Poisson counts. *Adv. Appl. Probab.*, 20, 822:835.

MEYER, P. (2000). *Probabilidade. Aplicações à Estatística*. LTC, Rio de Janeiro.

NETO, J., TORRES, P., FERREIRA, F. (2004). Previsão da Qualidade do Ar para Lisboa - a Abordagem estatística. Actas da 8<sup>a</sup> Conferência Nacional de Ambiente, Lisboa, 27-29 Outubro 2004.

- PAVLOPOULOS, H. e KARLIS, D. (2007). INAR(1) modeling of overdispersed count series with an environmental application. *Environmetrics*, 19,369:393.
- PEDROSO, A. N. V. (2007). Poluentes Atmosféricos e Plantas Bioindicadoras. *Instituto de Botânica, São Paulo*.
- SILVA, I. e OLIVEIRA, M.E.(2003). Selecção de ordem em modelos INAR. Actas do *X Congresso Anual da SPE*, Portugal, 627:638.
- SILVA, I. e OLIVEIRA, M.E. (2004).Estimação de Whittle para réplicas de processos INAR. Actas do *XI Congresso Anual da SPE*, Portugal, 767:778.
- SILVA, I. (2005). Contributions to the Analysis of discrete-valued time series. Tese de Doutorado, Portugal: Programa de pós graduação em Matemática Aplicada: Faculdade de Ciências da Universidade do Porto.
- SILVA, I. e OLIVEIRA, M.E.(2004). Selecção de ordem em modelos INAR. Actas do *X Congresso Anual da SPE*, Portugal.
- SILVA, N; PEREIRA, I; SILVA, M. E. (2009). Forecasting in INAR(1) model. *Statistical Journal*, 7, 1, 119:134.
- STEUTAL, F. e VAN HARN, K. (1979). Discrete analogues of self-decomposability and stability. *Ann. Prob.* 7, 893:9.
- WHITTLE, P. (1953) The analysis of multiple stationary time series. *J. R. Stat. Soc., B*, 15, 125:139.
- WHO, (2006) Air quality guidelines: global update 2005.
- WHO, (2000). Guidelines for air quality.
- ZHANG, P. (1992) On the distributional properties of model selection criteria. *J. Am. Stat.Assoc.*, Vol. 87, pp. 732-7.

---

# APÊNDICE A

## Propriedades do operador thinning

Esta seção apresenta a demonstração de algumas das propriedades do operador *thinning*.

(i)  $0 * X_1 = \sum_{j=1}^{X_1} Y_{j,1}$  tal que  $P(Y_{j,1} = 1) = \alpha = 0 \Rightarrow P(Y_{j,1} = 0) = 1 - \alpha = 1$

Logo

$$0 * X_1 = \sum_{j=1}^{X_1} Y_{j,1} = \sum_{j=1}^{X_1} 0 = 0$$

(ii)  $1 * X_1 = \sum_{j=1}^{X_1} Y_{j,1}$  tal que  $P(Y_{j,1} = 1) = \alpha = 1$

Logo

$$1 * X_1 = \sum_{j=1}^{X_1} Y_{j,1} = \sum_{j=1}^{X_1} 1 = X_1$$

(iii)  $\alpha_1 * (\alpha_2 * X_2) = (\alpha_1 \alpha_2) * X_2$

(iv)

$$\begin{aligned} E[\alpha_1 * X_1] &= E[E[(\alpha_1 * X_1)|X_1]] \\ &= E \left[ E \left[ \sum_{j=1}^{X_1} Y_{j,1} | X_1 \right] \right] \\ &= E \left[ \sum_{j=1}^{X_1} E[Y_{j,1} | X_1] \right] \\ &= E \left[ \sum_{j=1}^{X_1} E[Y_{j,1}] \right] \\ &= E \left[ \sum_{j=1}^{X_1} \alpha_1 \right] \\ &= E[\alpha_1 X_1] \\ &= \alpha_1 E[X_1] \end{aligned}$$

(v)

$$\begin{aligned} E[(\alpha_1 * X_1)X_2] &= E \left[ \left( \sum_{j=1}^{X_1} Y_{j,1} \right) X_2 \right] \\ &= E[(Y_{1,1} + Y_{2,1} + \dots + Y_{X_1,1})X_2] \\ &= E[Y_{1,1}X_2 + Y_{2,1}X_2 + \dots + Y_{X_1,1}X_2] \\ &= E[Y_{1,1}]E[X_2] + E[Y_{2,1}]E[X_2] + \dots + E[Y_{X_1,1}]E[X_2] \\ &= E[X_2][E[Y_{1,1}] + E[Y_{2,1}] + \dots + E[Y_{X_1,1}]] \\ &= E[X_2][E[Y_{1,1} + Y_{2,1} + \dots + Y_{X_1,1}]] \\ &= E[X_2]E[\alpha_1 * X_1] \\ &= \alpha_1 E[X_1]E[X_2] \\ &= \alpha_1 E[X_1 X_2] \end{aligned}$$

(vi)

$$\begin{aligned} E[(\alpha_1 * X_1)X_2X_3] &= E \left[ \left( \sum_{j=1}^{X_1} Y_{j,1} \right) X_2X_3 \right] \\ &= E \left[ \left( \sum_{j=1}^{X_1} Y_{j,1} \right) \right] E[X_2X_3] \\ &= E[\alpha_1 * X_1] E[X_2X_3] \\ &= \alpha_1 E[X_1] E[X_2X_3] \\ &= \alpha_1 E[X_1X_2X_3] \end{aligned}$$

(vii)

$$\begin{aligned} E[(\alpha_1 * X_1)^2] &= E [ E [ (\alpha_1 * X_1)^2 | X_1 ] ] \\ &= E \left[ E \left[ \left( \sum_{i=1}^{X_1} Y_{i,1}^2 + \sum_{j=1}^{X_1} \sum_{\substack{k=1 \\ j \neq k}}^{X_1} Y_{j,1} Y_{k,1} \right) | X_1 \right] \right] \\ &= E \left[ \sum_{i=1}^{X_1} E[Y_{i,1}^2 | X_1] + \sum_{j=1}^{X_1} \sum_{\substack{k=1 \\ j \neq k}}^{X_1} E[Y_{j,1} Y_{k,1} | X_1] \right] \\ &= E \left[ \sum_{i=1}^{X_1} E[Y_{i,1}^2] + \sum_{j=1}^{X_1} \sum_{\substack{k=1 \\ j \neq k}}^{X_1} E[Y_{j,1} Y_{k,1}] \right] \\ &= E[X_1(\sigma_1^2 + \alpha_1^2) + X_1(X_1 - 1)(\alpha_1\alpha_1)] \\ &= E[\sigma_1^2 X_1 + \alpha_1^2 X_1^2] \\ &= \sigma_1^2 E[X_1] + \alpha_1^2 E[X_1^2] \end{aligned}$$

(viii)

$$\begin{aligned} E[(\alpha_1 * X_1)^2 X_2] &= E \left[ \left( \sum_{i=1}^{X_1} Y_{i,1}^2 + \sum_{j=1}^{X_1} \sum_{\substack{k=1 \\ j \neq k}}^{X_1} Y_{j,1} Y_{k,1} \right) X_2 \right] \\ &= E \left[ \left( \sum_{i=1}^{X_1} Y_{i,1}^2 + \sum_{j=1}^{X_1} \sum_{\substack{k=1 \\ j \neq k}}^{X_1} Y_{j,1} Y_{k,1} \right) \right] E[X_2] \\ &= (\sigma_1^2 E[X_1] + \alpha_1^2 E[X_1^2]) E[X_2] \\ &= \sigma_1^2 E[X_1 X_2] + \alpha_1^2 E[X_1^2 X_2] \end{aligned}$$

(ix)

$$\begin{aligned} E[(\alpha_1 * X_1)(\alpha_2 * X_2)] &= E[E[(\alpha_1 * X_1)(\alpha_2 * X_2) | X_1, X_2]] \\ &= E \left[ E \left[ \left( \sum_{i=1}^{X_1} Y_{i,1} \right) \left( \sum_{j=1}^{X_2} Y_{j,2} \right) | X_1, X_2 \right] \right] \\ &= E \left[ \sum_{i=1}^{X_1} \sum_{j=1}^{X_2} E[Y_{i,1} Y_{j,2} | X_1, X_2] \right] \\ &= E \left[ \sum_{i=1}^{X_1} \sum_{j=1}^{X_2} E[Y_{i,1} Y_{j,2}] \right] \\ &= E[X_1 X_2 \alpha_1 \alpha_2] \\ &= \alpha_1 \alpha_2 E[X_1 X_2] \end{aligned}$$

(x)

$$\begin{aligned} E[(\alpha_1 * X_1)(\alpha_2 * X_2) X_3] &= E \left[ \left( \sum_{i=1}^{X_1} Y_{i,1} \right) \left( \sum_{j=1}^{X_2} Y_{j,2} \right) X_3 \right] \\ &= E \left[ \left( \sum_{i=1}^{X_1} Y_{i,1} \right) \left( \sum_{j=1}^{X_2} Y_{j,2} \right) \right] E[X_3] \\ &= E[(\alpha_1 * X_1)(\alpha_2 * X_2)] E[X_3] \\ &= \alpha_1 \alpha_2 E[X_1 X_2] E[X_3] \\ &= \alpha_1 \alpha_2 E[X_1 X_2 X_3] \end{aligned}$$

(xi)  $E[(\alpha_1 * X_1)^3] =$

$$E[E[\sum_{i=1}^{X_1} Y_{i,1}^3 + 3 \sum_{i=1}^{X_1} \sum_{j=1, i \neq j}^{X_1} Y_{i,1}^2 Y_{j,1} + \sum_{i=1}^{X_1} \sum_{j=1, i \neq j}^{X_1} \sum_{k=1, i \neq k, j \neq k}^{X_1} Y_{i,1} Y_{j,1} Y_{k,1}] | X_1] =$$

$$E[\sum_{i=1}^{X_1} E[Y_{i,1}^3 | X_1] + 3 \sum_{i=1}^{X_1} \sum_{j=1, i \neq j}^{X_1} E[Y_{i,1}^2 Y_{j,1} | X_1] + \sum_{i=1}^{X_1} \sum_{j=1, i \neq j}^{X_1} \sum_{k=1, i \neq k, j \neq k}^{X_1} E[Y_{i,1} Y_{j,1} Y_{k,1}]] =$$

$$E[\sum_{i=1}^{X_1} E[Y_{i,1}^3] + 3 \sum_{i=1}^{X_1} \sum_{j=1, i \neq j}^{X_1} E[Y_{i,1}^2 Y_{j,1}] + \sum_{i=1}^{X_1} \sum_{j=1, i \neq j}^{X_1} \sum_{k=1, i \neq k, j \neq k}^{X_1} E[Y_{i,1} Y_{j,1} Y_{k,1}]] =$$

$$E[\gamma_1 X_1 + 3X_1(X_1 - 1)(\sigma_1^2 + \alpha_1^2)(\alpha_1) + X_1(X_1 - 1)(X_1 - 2)\alpha_1\alpha_1\alpha_1] =$$

$$\gamma_1 E[X_1] + 3\sigma_1^2\alpha_1 E[X_1^2] + 3\alpha_1^3 E[X_1^2] - 3\sigma_1^2\alpha_1 E[X_1] - 3\alpha_1^3 E[X_1] + \alpha_1^3 E[X_1^3] -$$

$$3\alpha_1^3 E[X_1^2] + 2\alpha_1^3 E[X_1] =$$

$$\alpha_1^3 E[X_1^3] + 3\alpha_1\sigma_1^2 E[X_1^2] + (\gamma_1 - 3\alpha_1\sigma_1^2 - \alpha_1^3) E[X_1]$$

(xii)  $E[(\alpha_1 * X_1)^2(\alpha_2 * X_2)] =$

$$E[E[(\alpha_1 * X_1)^2(\alpha_2 * X_2) | X_1, X_2]] =$$

$$E[E[(\sum_{i=1}^{X_1} Y_{i,1}^2 + \sum_{i=1}^{X_1} \sum_{j=1, i \neq j}^{X_1} Y_{i,1} Y_{j,1})(\sum_{k=1}^{X_2} Y_{k,2}) | X_1, X_2]] =$$

$$E[E[\sum_{i=1}^{X_1} \sum_{j=1}^{X_2} Y_{i,1}^2 Y_{j,2} | X_1, X_2] + E[\sum_{i=1}^{X_1} \sum_{j=1, i \neq j}^{X_1} \sum_{k=1}^{X_2} Y_{i,1} Y_{j,1} Y_{k,2} | X_1, X_2]] =$$

$$E[\sum_{i=1}^{X_1} \sum_{j=1}^{X_2} E[Y_{i,1}^2 Y_{j,2} | X_1, X_2] + \sum_{i=1}^{X_1} \sum_{j=1, i \neq j}^{X_1} \sum_{k=1}^{X_2} E[Y_{i,1} Y_{j,1} Y_{k,2} | X_1, X_2]] =$$

$$E[\sum_{i=1}^{X_1} \sum_{j=1}^{X_2} E[Y_{i,1}^2 Y_{j,2}] + \sum_{i=1}^{X_1} \sum_{j=1, i \neq j}^{X_1} \sum_{k=1}^{X_2} E[Y_{i,1} Y_{j,1} Y_{k,2}]] =$$

$$E[X_1 X_2 (\sigma_1^2 + \alpha_1^2) \alpha_2 + X_1 (X_1 - 1) X_2 \alpha_1^2 \alpha_2] =$$

$$(\sigma_1^2 \alpha_2 + \alpha_1^2 \alpha_2) E[X_1 X_2] - \alpha_1^2 \alpha_2 E[X_1 X_2] + \alpha_1^2 \alpha_2 E[X_1^2 X_2] =$$

$$\alpha_1^2 \alpha_2 E[X_1^2 X_2] + \sigma_1^2 \alpha_2 E[X_1 X_2]$$

(xiii)  $E[(\alpha_1 * X_1)(\alpha_2 * X_2)(\alpha_3 * X_3)] =$

$$E[E[(\sum_{i=1}^{X_1} Y_{i,1})(\sum_{j=2}^{X_2} Y_{j,2})(\sum_{k=3}^{X_3} Y_{k,3}) | X_1, X_2, X_3]] =$$

$$E[(\sum_{i=1}^{X_1})(\sum_{j=2}^{X_2})(\sum_{k=3}^{X_3}) E[Y_{i,1} Y_{j,2} Y_{k,3} | X_1, X_2, X_3]] =$$

$$E[(\sum_{i=1}^{X_1})(\sum_{j=2}^{X_2})(\sum_{k=3}^{X_3}) E[Y_{i,1} Y_{j,2} Y_{k,3}]] =$$

$$E[X_1 X_2 X_3 \alpha_1 \alpha_2 \alpha_3] =$$

$$\alpha_1 \alpha_2 \alpha_3 E[X_1 X_2 X_3]$$

(xv)  $E[(\alpha_1 * X_1)^3 X_2] =$

$$E[(\sum_{i=1}^{X_1} Y_{i,1}^3 + 3 \sum_{i=1}^{X_1} \sum_{j=1, i \neq j}^{X_1} Y_{i,1}^2 Y_{j,1} + \sum_{i=1}^{X_1} \sum_{j=1, i \neq j}^{X_1} \sum_{k=1, i \neq k, j \neq k}^{X_1} Y_{i,1} Y_{j,1} Y_{k,1}) X_2] =$$

$$E[\sum_{i=1}^{X_1} Y_{i,1}^3 + 3 \sum_{i=1}^{X_1} \sum_{j=1, i \neq j}^{X_1} Y_{i,1}^2 Y_{j,1} + \sum_{i=1}^{X_1} \sum_{j=1, i \neq j}^{X_1} \sum_{k=1, i \neq k, j \neq k}^{X_1} Y_{i,1} Y_{j,1} Y_{k,1}] E[X_2] =$$

$$\begin{aligned} E[(\alpha_1 * X_1)^3] E[X_2] &= \\ (\alpha_1^3 E[X_1^3] + 3\alpha_1 \sigma_1^2 E[X_1^2] + (\gamma_1 - 3\alpha_1 \sigma_1^2 - \alpha_1^3) E[X_1]) E[X_2] &= \\ \alpha_1^3 E[X_1^3 X_2] + 3\alpha_1 \sigma_1^2 E[X_1^2 X_2] + (\gamma_1 - 3\alpha_1 \sigma_1^2 - \alpha_1^3) E[X_1 X_2] & \end{aligned}$$

$$\begin{aligned} \text{(xvii)} \quad E[(\alpha_1 * X_1)^2 X_2 X_3] &= \\ E[(\alpha_1 * X_1)^2] E[X_2 X_3] &= \\ (\alpha_1^2 E[X_1^2] + \sigma_1^2 E[X_1]) E[X_2 X_3] &= \\ \alpha_1^2 E[X_1^2 X_2 X_3] + \sigma_1^2 E[X_1 X_2 X_3] & \end{aligned}$$

$$\begin{aligned} \text{(xviii)} \quad E[(\alpha_1 * X_1) X_2 X_3 X_4] &= \\ E[(\alpha_1 * X_1)] E[X_2 X_3 X_4] &= \\ \alpha_1 E[X_2 X_3 X_4] & \end{aligned}$$

$$\begin{aligned} \text{(xxi)} \quad E[(\alpha_1 * X_1)(\alpha_2 * X_2)(\alpha_3 * X_3)(\alpha_4 * X_4)] &= \\ E[E[(\sum_{i=1}^{X_1} Y_{i,1})(\sum_{j=1}^{X_2} Y_{j,2})(\sum_{k=1}^{X_3} Y_{k,3})(\sum_{l=1}^{X_4} Y_{l,4}) | X_1, X_2, X_3, X_4]] &= \\ E[\sum_{i=1}^{X_1} \sum_{j=1}^{X_2} \sum_{k=1}^{X_3} \sum_{l=1}^{X_4} E[Y_{i,1} Y_{j,2} Y_{k,3} Y_{l,4} | X_1, X_2, X_3, X_4]] &= \\ E[\sum_{i=1}^{X_1} \sum_{j=1}^{X_2} \sum_{k=1}^{X_3} \sum_{l=1}^{X_4} E[Y_{i,1} Y_{j,2} Y_{k,3} Y_{l,4}]] &= \\ E[X_1 X_2 X_3 X_4 \alpha_1 \alpha_2 \alpha_3 \alpha_4] &= \\ \alpha_1 \alpha_2 \alpha_3 \alpha_4 E[X_1 X_2 X_3 X_4] & \end{aligned}$$