

Algoritmos de Previsão Utilizando Modelos de Estados: Aplicações em Teletráfego

Edson Luiz Ursini ^{de} π/γ

Orientador: Prof. Dr. Michel Daoud [Yacoub] t.

Co-orientador: Prof. Dr. Wagner Caradori do [Amaral] 1952 - t

FEF

Este exemplar corresponde à redação final da tese defendida por Edson Luiz Ursini e aprovada pela Comissão Julgadora em 28/10/1994.

Michel Daoud
Orientador

Tese apresentada à Universidade Estadual de Campinas como requisito parcial para a obtenção do título de Doutor em Engenharia Elétrica

45583

RESUMO

A expansão da rede de telecomunicações exige investimentos vultosos e deve ser cuidadosamente planejada. Esta expansão leva em conta a demanda histórica do tráfego e utiliza algoritmos de previsão para projetar o tráfego para o curto, médio ou longo prazo. A escolha do algoritmo de previsão apropriado deve levar em consideração, entre outras, 1) a precisão requerida para o erro entre o valor projetado e o valor observado, 2) a facilidade de automatização do algoritmo em vista do grande número de rotas de tráfego existentes, e 3) a estabilidade dos valores projetados n passos à frente.

Este trabalho originou-se da necessidade de obtenção de projeções de valores de tráfego (ou, equivalentemente, do número de terminais) para atender ao planejamento e à administração de tráfego.

A maioria dos requisitos citados para a escolha dos algoritmos de previsão é preenchido pelo modelo de Espaço de Estados com o estado estimado pelo Filtro de Kalman. Isto se deve a que o modelo: 1) permite associar as características físicas das séries temporais às variáveis de estado, 2) pode ser simplificado e, conseqüentemente, automatizado, e 3) tem uma concepção intrinsecamente recursiva. De fato, a utilização deste modelo tem sido objeto da grande maioria dos recentes trabalhos de teletráfego onde a previsão de tráfego é o alvo principal. A aplicação do modelo de Espaço de Estados, para este caso, recai num problema de filtragem não-linear, onde tanto o estado quanto a matriz de transição de estado — relacionados numa única equação — devem ser estimados simultaneamente. Isto tem sido resolvido por meio de uma técnica de relaxação: o algoritmo *Expectation-Maximization*, *EM*. No entanto, apesar de produzir resultados de acordo com as especificações, esse algoritmo requer um elevado esforço computacional. Além disso, em alguns casos, como por exemplo na mudança brusca da dinâmica do sistema, o algoritmo *EM* pode desempenhar relativamente pior que um algoritmo mais simples.

Propõe-se, neste trabalho, o Algoritmo de Projeção Seqüencial Robusto - APS Robusto, que, às custas de um desempenho ligeiramente pior que aquele obtido pela aplicação do *EM*, é substancialmente mais simples de se implementar além de apresentar uma grande robustez, caracterizada pela sua pouca sensibilidade a grandes variações dos seus parâmetros. Esta robustez foi comprovada por testes efetuados com dados de campo de 40 rotas de tráfego.

Com o intuito de se reunirem as características de precisão de erro do *EM*, da robustez do APS Robusto, e poder atuar em situações de perturbações não-estacionárias, propõe-se o Algoritmo de Projeção Seqüencial Adaptativo Modificado - APSAM. Os testes realizados mostram que esse algoritmo supera o desempenho do *EM*, requer menos carga de processamento, além de ser capaz de se adaptar às variações nos valores dos parâmetros do modelo, sendo apropriado para aplicações em tempo real.

Aos meus filhos, Gabriel e Giovana

V.	Et.		
TUMBO BC/	21.108		
PROC.	286/94		
C	<input type="checkbox"/>	D	<input checked="" type="checkbox"/>
PREÇO	CR\$ 800,00		
DATA	06/05/94		
Nº CPD			

CM-00058259-2

Agradecimentos

A Alessandro Girolami, que iniciou o estudo dos modelos de previsão que resultaram nesta tese, e sempre incentivou e auxiliou com sugestões no decorrer dos trabalhos;

A Jorge Moreira de Souza, que me recebeu na Seção de Qualidade de Sistemas, e pelo seu apoio aos trabalhos de pesquisa;

A Michel Daoud Yacoub, meu orientador, pela sua paciência na revisão dos textos e sugestões de melhoria;

A Wagner Caradori do Amaral, meu co-orientador, que contribuiu com diversas sugestões para a melhoria dos modelos;

A todos os colegas da Seção de Qualidade de Sistemas, que, de uma forma ou de outra, contribuíram para a realização deste trabalho. Em particular, a Lourival Aparecido dos Reis, que possibilitou sua conclusão;

Aos colegas profissionais do setor de telecomunicações, de várias empresas, públicas e privadas, com os quais tive a satisfação de trabalhar e aprender. Eles sempre estiveram abertos a contribuições ou a discussão de várias questões, e seus trabalhos e experiência foram utilizados na elaboração desta tese;

Ao pessoal da Área de Documentação do CPqD, que elaborou as figuras deste trabalho utilizando o equipamento PCAD.

Conteúdo

1	Introdução	1
1.1	Dados de Tráfego	1
1.2	Estrutura da Rede	3
1.3	Métodos Existentes versus Métodos Propostos	3
1.4	Cronologia dos Métodos de Estimção Linear	4
1.5	Breve Descrição dos Capítulos	5
1.6	Contribuições do Trabalho	5
2	Revisão da Literatura	8
2.1	Introdução	8
2.2	Classificação dos Modelos Para Previsão de Séries Temporais	10
2.3	Modelos Estáticos (de erro)	13
2.4	Modelos Dinâmicos	22
2.4.1	Modelos Entrada×Saída	22
2.4.2	Considerações Gerais sobre Aspectos Importantes na Aplicação dos Modelos Dinâmicos (Particularmente os Entrada×Saída)	25
2.4.3	Modelos de Espaço de Estados	33
2.5	Vantagens e Desvantagens dos Modelos Apresentados	35
2.6	Comentários Finais e Conclusões	37
3	Representação por Espaço de Estados com o Filtro de Kalman	41
3.1	Introdução	41
3.2	O Modelo de Espaço de Estados Formado por Decomposição Aditiva da Série Temporal	43
3.3	Os Modelos (H_j, Φ_j, Γ_j)	45
3.3.1	Modelo de Tendência (H_1, Φ_1, Γ_1)	45
3.3.2	Modelos para o Componente Sazonal (H_2, Φ_2, Γ_2)	47
3.3.3	Efeito de Variação Diário $(H_{3k}, \Phi_3, \Gamma_3)$	50
3.3.4	Modelo para Tendência Estocástica (Auto-regressivo) (H_4, Φ_4, Γ_4)	51
3.3.5	Ilustração do Modelo	51
3.4	Estimção de Estado com o Filtro de Kalman	53
3.4.1	Equações do Filtro de Kalman	53
3.5	Estimção Conjunta do Estado e dos Parâmetros	57
3.5.1	Estimção de Máxima Verossimilhança do Algoritmo <i>EM</i>	58
3.6	Crterios para Avaliação de Desempenho	62
3.7	Aplicação do Algoritmo <i>EM</i>	64
3.8	Otimidade, Sensibilidade e Robustecimento	71

3.9	Aplicações ao Problema de Projeção de Tráfego	75
3.10	Sumário e Conclusões	77
4	Algoritmo de Projeção Seqüencial Robusto: Uma Proposta	78
4.1	Introdução	78
4.2	O Modelo do Algoritmo de Filtragem do APS	80
4.2.1	Características Gerais	80
4.2.2	Iniciação	84
4.2.3	Seqüência de Ganhos e Robustecimento	86
4.3	Influência da Matriz de Transição de Estado no Desempenho do Filtro	87
4.4	Reiniciação e Robustecimento	89
4.4.1	Intervalo de Confiança para Detecção de Observações Duvidosas (<i>Outliers</i>)	90
4.4.2	Correção da Carga Projetada n Passos à Frente	91
4.4.3	Exemplo de Aplicação	91
4.5	Conclusões	93
5	Algoritmo de Projeção Seqüencial Adaptativo Modificado: Uma Proposta	94
5.1	Introdução	94
5.2	Estimação das Covariâncias dos Erros Q_k e R_k	95
5.2.1	Estimação da Covariância do Erro de Modelagem (Q_k)	97
5.2.2	Estimação da Covariância do Erro de Observação (R_k)	102
5.2.3	Testes Realizados e Considerações	102
5.3	Estimação da Matriz de Transição de Estado (Φ_k)	105
5.3.1	Equações para a Estimação da Matriz de Transição de Estado	105
5.3.2	Testes Realizados e Considerações	107
5.4	Iniciação	108
5.5	Reiniciação e Robustecimento	110
5.5.1	Intervalo de Confiança para Reiniciação e Robustecimento	110
5.6	Aplicações em Tempo Real	111
5.6.1	Fator de Tendência de Laplace	112
5.6.2	Verificação se a Seqüência de Inovações é Descorrelacionada	113
5.6.3	Testes Realizados e Considerações	114
5.7	Conclusões	121
6	Conclusões e Trabalhos Futuros	124
A	Extensão do Algoritmo Adaptativo para Multivariáveis	134
A.1	Submodelo de Espaço de Estados para os Elementos de Q_k	134
A.2	Estimação do Estado do Subsistema (Subfiltro de Kalman)	136
A.3	Robustecimento	138
A.4	Exemplo de Obtenção das Matrizes para um Modelo Bivariável	138
B	Comparação entre as Equações de Fahmeier e de Chow (ou de Abraham e Ledoiter)	141
C	Estabilidade e Estacionariedade do Modelo de Espaço de Estados	143

D	Avaliação da Equação do Ganho do Filtro de Kalman e Equação de Ricatti Discreta	145
D.1	Avaliação do Ganho do Filtro de Kalman	145
D.2	Dedução da Equação de Ricatti Discreta	146
E	Dedução das Equações do Suavizador de Kalman (Kalman smoother)	147
E.1	Dedução da Equação do Estado Suavizado	148
E.2	Dedução da Equação da Covariância do Erro de Estimção do Suavizador .	149
E.3	Simplificações de Devine	150
E.4	Dedução da Covariância de Transição de Estado do Suavizador	150
F	Desenvolvimento das Relações entre Critérios Empregados na Avaliação de Desempenho dos Algoritmos de Previsão	152
G	Aplicação do Algoritmo EM aos Modelos da Seção 3.3	153
H	Exemplo de Obtenção das Matrizes do Modelo de Decomposição Aditiva da Série Temporal	163
I	Considerações sobre a Posição do Estado Verdadeiro (Não-observável) no Funcionamento do Filtro de Kalman	167
J	Diferentes Formas da Equação da Covariância do Erro de Estimção (\bar{P}_k)	168
K	Geração de Séries Temporais pelo Método de Monte Carlo	169
K.1	Introdução	169
K.1.1	Geração da Distribuição Gaussiana	169
K.1.2	Distribuição Normal Contaminada	169
K.1.3	Estimção da Correlação Cruzada entre Duas Variáveis Aleatórias Gaussianas	169
K.1.4	Estimção de \bar{X}_0 e \bar{P}_0	170
K.2	Geração das Observações	170
K.3	Exemplo de Aplicação	171
L	Lista de Símbolos	172

Capítulo 1

Introdução

As previsões dos valores de tráfego (*traffic forecasting*) são parte integrante do Planejamento da Rede incluindo o Planejamento da Expansão, da Administração de Tráfego, da Gerência de Rede e do Planejamento Financeiro das empresas que operam no setor de telecomunicações.

O papel das projeções de tráfego no planejamento da rede é prover dados para dimensionamento e otimização constituindo previsões a longo prazo, da ordem de dois a dez anos. Para a administração de tráfego e alocação de recursos requer-se a projeção da ordem de alguns meses à frente, até um ano. Para as atividades de gerência de rede há a necessidade de aquisição e projeção de dados quase em tempo real, da ordem de minutos.

A Telesp (Telecomunicações de São Paulo S. A.) possui atualmente em torno de 10000 rotas de tráfego que precisam ser continuamente observadas e projetadas. Estudos da Bell System para os USA [23] indicam que uma rota de tráfego possui um ciclo de vida variando de um a dez anos ¹, sendo típicos os valores de três a cinco anos (três a cinco observações anuais). Portanto, dois fatores são cruciais para o desenvolvimento dos algoritmos de projeção: a necessidade de se produzir um grande número de projeções em um sistema totalmente automatizado e a típica carência de dados em cada série individual.

Para a correta provisão de recursos é necessário que se conheça o tráfego médio cursado pela rede, sendo este o insumo dos algoritmos de projeção.

1.1 Dados de Tráfego

Dois tipos fundamentais de tráfego devem ser considerados para efeito de previsão/planejamento: o tráfego obtido por meio de medições diretamente nos circuitos e o tráfego obtido como subproduto do sistema de tarifação (que é o chamado tráfego **ponto-a-ponto** porque mede o período de ocupação entre duas localidades ou áreas de origem/destino diferentes). Na Telesp, a coleta de dados da estatística **ponto-a-ponto** permite variância dos erros da ordem dos cometidos pela medição direta dos circuitos. O procedimento para a obtenção dos dados **ponto-a-ponto** se baseia na observação das direções de tráfego durante dez dias

¹Esse período de duração é devido a mudanças: na estrutura da rede, de numeração de encaminhamento, etc.

consecutivos, 24 h por dia.

Existem dois fatores que afetam as séries temporais com dados de tráfego: A falta de dados suficientes e a falta de confiabilidade dos dados disponíveis. A falta de confiabilidade não corresponde apenas à variância do erro devido ao método de medição, mas também às condições da rede. Em situações de congestionamento observam-se circuitos sobrecarregados mas não se sabe qual o tráfego verdadeiro oferecido à rota (o tráfego oferecido dá origem à carga básica [10] que será projetada).

Para se obter mais confiabilidade nos dados, mesmo em condições de congestionamento, empregam-se duas técnicas para a determinação do tráfego oferecido:

- Utiliza-se a fórmula de Erlang invertida (que aqui se denomina “ técnica de campo ”). Ao invés de se calcular a perda (grau de serviço) a partir do tráfego e do número de circuitos, obtém-se o tráfego oferecido [57]. No entanto, sem a consideração das chamadas que são repetidas quando a taxa de chamadas não-completadas é elevada, esse procedimento apresenta resultados inadequados;
- Utilizam-se modelos de chamadas repetidas, onde estima-se uma função para a perseverança do usuário (com que intensidade o usuário está disposto a realizar tentativas até obter sucesso) e utiliza-se também um parâmetro indicador da qualidade da rede (quantas tentativas, em média, são feitas na rede para que se obtenha sucesso). Atualmente a rede brasileira tem um índice de sucesso de 40% a 50 %, ou seja, quatro a cinco chamadas completadas a cada dez realizadas. Aplicando-se um modelo de chamadas repetidas [58]² [10], obtém-se o número de intenções de chamadas e daí o tráfego oferecido.

Para a consideração do período em que os circuitos devem ser observados, utiliza-se o conceito de Hora de Maior Movimento, HMM. A HMM corresponde ao período de 60 minutos consecutivos onde se observa o maior valor médio da intensidade de tráfego.

Para se utilizar dados de tráfego mensais, define-se o Valor Representativo Mensal, VRM. Em cada mês utilizam-se medições do tráfego cursado durante cinco dias consecutivos, durante a HMM (ou hora representativa mensal, atualmente entre 10 e 11 horas da manhã), e define-se o VRM do mês como o segundo maior valor diário. O Valor Representativo Anual, VRA, é definido como o segundo maior VRM.

Além das séries de tráfego, as projeções são feitas com número de terminais, geralmente quando não se dispõe de dados de tráfego ou o horizonte de projeção é muito distante. Em outras tecnologias, como por exemplo, nas redes de pacotes, utiliza-se o número de segmentos de informação.

A complexidade da rede também é um fator que deve ser considerado, e isto será comentado a seguir.

²A recomendação E.501 do CCITT (atual ITU-TS) propõe uma “ técnica de campo ” como descrita acima, baseada nos valores de tráfego medido nos circuitos e se utiliza do modelo de chamadas repetidas para se obter um fator de correção da distorção causada no caso da aplicação da técnica em rotas congestionadas. Em [58], propõe-se um procedimento mais detalhado usando o mesmo modelo de chamadas repetidas (Anexo A da E.501) e como fonte de dados as estatísticas **ponto-a-ponto** e de chamadas repetidas como o sistema DDD-X da Embratel, que consiste na separação do tipo de ocorrência da chamada efetuada: OK (completada), NR (não-responde), LO (assinante do destino ocupado), CO1 (congestionamento na origem), CO3 (congestionamento na rota) e CO2 (congestionamento no destino).

1.2 Estrutura da Rede

A rede interurbana nacional obedece a uma estrutura hierarquizada onde se permite que as centrais de trânsito de uma determinada classe se liguem apenas a outras centrais de sua classe ou a classes acima da sua. Por exemplo, a maior classe, a central de trânsito internacional, recebe ligações apenas das centrais de classe I ou de outros trânsitos internacionais. Por sua vez, as de classe I recebem de outras classe I ou das de classe II. Os tipos de rotas interurbanas³ do Estado de São Paulo, administradas pela Telesp, são ⁴:

- Rotas de primeira escolha (que têm preferência de encaminhamento);
- Rotas de segunda escolha ou alternativas (entre Trânsitos classe II);
- Rotas finais (local para trânsito e vice-versa e Trânsito classe III para classe II e vice-versa).

Como exemplo de tráfego local, nas grandes regiões metropolitanas como a cidade de São Paulo, as ligações diretas entre as centrais telefônicas só se realizam quando há alto interesse de tráfego (rotas de primeira escolha direta, de alta perda), caso contrário são encaminhadas para centrais denominadas Tandem (rotas alternativas) que concentram pequenas quantidades de tráfego de muitas regiões, formando um volume de tráfego significativo, que é enviado ao destino como rota final, possibilitando economia devido à maior utilização dos circuitos.

Há estudos para a utilização de redes não-hierarquizadas no futuro (e algumas já em funcionamento) e também para cursar uma mistura de tipos diferentes de tráfego, tais como: voz, dados, imagem, etc.

As distribuições dos tráfegos que cursam essas diferentes rotas são bastante diferentes e isso deve ser levado em conta não somente pelos métodos de medição como também pelos métodos de projeção.

1.3 Métodos Existentes versus Métodos Propostos

Muitos dos métodos de projeção ainda hoje empregados pelas empresas operadoras utilizam-se de algoritmos de extrapolação e do ajuste de funções que foram originalmente projetados para cálculos manuais e que estavam disponíveis antes da disseminação do uso de computadores e do advento das modernas técnicas de estimação.

De maneira geral, os modelos de projeção são classificados em **estáticos**, que não levam em conta o valor de suas saídas passadas, e **dinâmicos**, que as levam em conta (Capítulo 2). Este trabalho enfocará principalmente os modelos dinâmicos (considerados integrantes das modernas técnicas de estimação), e entre os dinâmicos, particularmente os modelos de Espaço de Estados [27] (que foram aplicados pela primeira vez por David e Pack [22]). Esta escolha, além das razões descritas no Capítulo 2, se deve também à tendência observada nos trabalhos de projeção de tráfego mais recentes.

³O tráfego interurbano é mais rentável que o tráfego local (em uma mesma cidade), portanto é a ocupação principal das empresas operadoras dos serviços de telecomunicações.

⁴Em [45], o Algoritmo de Projeção Seqüencial Robusto proposto por Moreland [12], foi aplicado a 111 rotas de tráfego e fez-se o diagrama dos erros obtidos por tipo de rota. Nesse artigo, considerou-se cinco tipos de rotas de tráfego, aglutinadas aqui em três tipos.

Os seguintes países podem ser citados como sendo aqueles que já publicaram trabalhos utilizando modelos de Espaço de Estados para problemas de Teletráfego: Estados Unidos, França, Portugal, Japão e Holanda. Todos os trabalhos desses países mostraram a utilização de dados práticos colhidos de suas redes. Particularmente a França, com quem o contato foi mais estreito, explicou que os modelos de Espaço de Estados são utilizados em sua rede para projeções de tráfego. Para projeções dos valores dos números de chamadas efetuadas, devido a serem processos estocásticos melhor comportados, a administração francesa utiliza-se de modelos de projeção mais simples.

1.4 Cronologia dos Métodos de Estimação Linear

A seguir uma breve cronologia histórica da evolução dos métodos de estimação linear:

- 1632** Galileu, com o método dos mínimos quadrados em estudos de astronomia;
- 1795** Gauss, com mínimos quadrados ponderados que foi publicado por Legendre pela primeira vez em 1805;
- 1923** Wittaker, métodos de suavização;
- 1930-1940** Kolmogorov e Krein, processos estocásticos de tempo discreto, e Wiener, processos estocásticos contínuos (filtros ótimos no domínio da frequência);
- 1947** Levinson, filtragem de Wiener para o caso discreto;
- 1960** Kalman, filtros ótimos no domínio do tempo, ideais para implementação em computadores digitais por serem intrinsecamente recursivos;
- 1960 - 1970** Kailath, reformulou a teoria de Kalman;
- 1970** Box e Jenkins, modelos lineares na área de controle;
- 1971** Harrison e Stevens, aplicação do Filtro de Kalman na engenharia de controle;
- 1979** David e Pack, introdução do Filtro de Kalman para projeções na área de telecomunicações.

Além destes, trabalhos de outras áreas influenciaram o estágio atual dos métodos disponíveis:

- No campo econômico e financeiro: os modelos clássicos de Holt (1957) e Winters (1960, efeito sazonal), Brown (1963, suavização exponencial) e Harrison (1965, modelo sazonal expresso em termos de harmônicas);
- No campo do processamento digital de sinais, o desenvolvimento de algoritmos adaptativos: Leadrow e Hopf (1959, algoritmo gradiente ou *Least Mean Squares*); Godard (1974, algoritmo dos mínimos quadrados ou aplicação da teoria de Kalman); Fakoner e Lyung (1978), Carayamanis et al (1983), e Cioffi e Kailath (1984, mínimos quadrados rápidos); Mori et al (1977, filtro treliça);

- No campo da análise espectral: Yule (1927) e Burg (1967, modelo auto-regressivo); Griffiths (1975, modelo auto-regressivo adaptativo).

Os filtros lineares ótimos têm duas origens: uma ligada à área das comunicações e devida a Wiener e outra ligada à área de controle e devida a Kalman.

1.5 Breve Descrição dos Capítulos

O Capítulo 2 apresenta uma revisão dos modelos existentes aplicados às projeções de tráfego, dividindo-os em Estáticos e Dinâmicos, analisa suas vantagens e desvantagens, e termina por sugerir a utilização dos modelos dinâmicos e, particularmente, o modelo de Espaço de Estados. O Capítulo 3 descreve os modelos aditivos de Espaço de Estados, mostra as equações do Filtro de Kalman e do Suavizador de Kalman e compara o desempenho dos dois tipos de suavizador. Ao final, propõe a implementação de dois algoritmos: o Algoritmo de Projeção Seqüencial Robusto - APS Robusto, e o Algoritmo de Projeção Seqüencial Adaptativo Modificado - APSAM. Além disso, sugere que a Técnica de Relaxação utilizada para se estimarem os parâmetros, maximizando recursivamente a função de verossimilhança, continue a ser utilizada como ferramenta de comparação e suporte. O Capítulo 4 apresenta o APS Robusto proposto com suas duas principais vantagens: estabilidade na projeção a n passos e menor dependência em relação ao valor verdadeiro do estado inicial. O Capítulo 5 discute o APSAM com a estimação completa da matriz da covariância do erro de modelagem, Q_k (ruído no estado), por meio de um modelo de estados adicional, cujo estado é estimado por um outro Filtro de Kalman, denominado subfiltro; propõe equações adicionais para a estimação recursiva da matriz da covariância do erro de observação, R_k , e também propõe a estimação recursiva da matriz de transição de estado, Φ . O Apêndice A apresenta uma extensão deste algoritmo para multivariáveis. Uma série de outros apêndices, de B a K, é introduzida ao final com o objetivo de tornar a leitura do corpo da tese mais suave, visando eliminar partes mais carregadas de equações do texto principal.

1.6 Contribuições do Trabalho

Aqui são listadas as contribuições do trabalho, por capítulo:

- Capítulo 2
 - Breve descrição dos modelos de projeção utilizados para projeção de rotas de tráfego, discutindo suas vantagens e desvantagens;
 - Justificativa sobre a preferência dos modelos dinâmicos em relação aos estáticos. Entre os modelos dinâmicos, considerações sobre a utilização dos modelos de Espaço de Estados sobre os modelos Entrada×Saída.
- Capítulo 3
 - Proposta de vários tipos de modelo de Espaço de Estados aditivos, combinando os desenvolvidos com características gerais e os desenvolvidos com características específicas para projeção de tráfego, e sugestão de nova variante de modelo aditivo com parametrização dos elementos da matriz de observação, H ;

- Exemplo de obtenção das matrizes do modelo aditivo (Apêndice H);
- Aplicação do suavizador com Jota [20, 21] e implementação do programa do suavizador de De Jong [30] (Apêndice E) com proposta para estimação conjunta do estado inicial, \bar{X}_0 , e da covariância do erro de estimação inicial, \bar{P}_0 , e aplicação a vários modelos;
- Comparação de desempenho entre o algoritmo com Jota e o de De Jong;
- Proposta e exemplos de aplicação de estimação de partes da matriz de transição de estado (na literatura há um exemplo escalar apenas);
- Desenvolvimento (Apêndice F) dos critérios empregados na avaliação de desempenho dos algoritmos de previsão;
- Aplicação do algoritmo *EM* a diversos modelos da Seção 3.3 (Apêndice G);
- Discussão sobre critérios gerais de estabilidade e estacionariedade do modelo de Espaço de Estados (Apêndice C);
- Apresentação e verificação de diferentes formas da covariância do erro de estimação, \bar{P}_k (Apêndice J).

• Capítulo 4

- Desenvolvimento de um Algoritmo de Projeção Sequencial Robusto (APS Robusto) com as seguintes contribuições:
 - * Comparação de diversos tipos de ganho robusto, proposta de utilização de ganho fixo, e estimação do ganho utilizando o critério da mediana;
 - * Proposta de estimação da matriz de transição de estado, Φ , pelo critério da mediana;
 - * Introdução de um intervalo de confiança para detecção de *outliers* ou eventos determinísticos não-previstos e para estimação de erro de previsão a n passos;
 - * Testes do APS Robusto com dados de campo de 40 rotas de tráfego.

• Capítulo 5

- Desenvolvimento de um Algoritmo de Projeção Sequencial Adaptativo Modificado (APSAM) com as seguintes contribuições:
 - * Proposta da estimação completa da matriz da covariância Q_k (e não apenas diagonal) por meio de um Subfiltro de Kalman;
 - * Proposta de um modelo com diferentes elementos para a matriz de observação H e testes de sensibilidade em relação aos diferentes valores utilizados;
 - * Considerações sobre a posição do estado verdadeiro, não-observável ⁵ (Apêndice I);
 - * Utilização de equações adicionais para estimação da matriz da covariância R_k e inclusão (Apêndice B) de considerações sobre a equação mais adequada para se estimar R_k , bem como sobre sua velocidade de convergência;

⁵Este conceito também foi utilizado nos Capítulos 4 e 5.

- * Proposta de utilização de um conjunto de equações adicionais para se estimar a matriz de transição de estado Φ ;
- * Proposta de um intervalo de confiança para reiniciação e robustecimento;
- * Diversos testes, em cada um dos pontos dos itens anteriores, com dados de campo;
- * Testes do APSAM com dados de campo de sete rotas de tráfego e uma rota gerada por Monte Carlo (Apêndice K) com as condições de tráfego do funcionamento em tempo real para Gerência de Rede;
- * Desenvolvimento de um modelo para multivariáveis, incluindo também um intervalo de confiança para reiniciação e robustecimento (Apêndice A).

- Apêndices

- Todos os apêndices foram desenvolvidos pelo autor, excetuando-se o Apêndice C que é uma contribuição que corresponde a vários trabalhos de onde foram retirados aspectos importantes relativos a este trabalho de tese, o Apêndice E que é uma organização das notas de aula do professor Shumway [30] e o Apêndice L que é apenas uma lista dos principais símbolos utilizados no trabalho.

Capítulo 2

Revisão da Literatura

Este capítulo tem por finalidade fazer uma revisão na literatura existente a respeito de modelos para previsão de séries temporais relacionados ao tráfego telefônico (teletráfego). Os modelos são classificados em dois tipos: estáticos e dinâmicos. Os modelos dinâmicos dependem dos valores de suas saídas passadas e os estáticos não. Para cada um dos tipos, foram abordados os procedimentos mais usuais com suas características básicas. A escolha do modelo depende da precisão requerida, do comprimento do período de projeção, da disponibilidade de dados históricos e facilidades de computação, do nível de agregação e da clara descrição das variáveis que caracterizam o tráfego. De uma maneira geral, os modelos dinâmicos, com o mesmo número de parâmetros que os estáticos, apresentam maior precisão porque têm informação sobre a dinâmica do sistema. Embora os modelos Entrada×Saída e Espaço de Estados tenham desempenho semelhante, o capítulo conclui pelo uso da representação por Espaço de Estados, principalmente devido às facilidades para a identificação física do modelo e para a projeção n passos à frente.

2.1 Introdução

Nesta seção serão introduzidos alguns conceitos básicos utilizados ao longo deste trabalho. São eles: 1) Série Temporal, 2) Previsão, Projeção e Predição, 3) Identificação e Modelagem, e 4) Especificação do Modelo de Projeção.

1) Série Temporal

Uma série temporal é um conjunto de observações ordenadas, em geral, no tempo. Em alguns casos, a ordenação pode utilizar outras dimensões como, por exemplo, profundidade, altura, etc. Embora existam séries temporais discretas e contínuas, os registros das observações, geralmente, se dão de forma discreta, mesmo para as séries contínuas (por exemplo, valor da temperatura tomado de hora em hora). Assim estas séries também podem ter um tratamento discreto.

2) Previsão, Projeção e Predição

Se uma série temporal Y_k possui valores até o instante de tempo k e deseja-se prevê-la para o instante de tempo $k + n$, diz-se que k é a origem e Y_{k+n}^* é a previsão da série no instante $k + n$ (Figura 2.1).

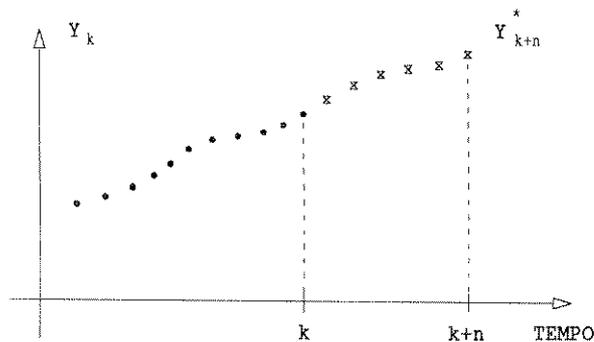


Figura 2.1: Série Temporal com Previsões de Origem k e Comprimento n do Período de Previsão

No decorrer do texto as palavras **previsão** — que etimologicamente significa “ ver algo antes que exista ” —, **projeção** e **predição** serão usadas indistintamente, dependendo da circunstância em que pareçam mais adequadas¹, para indicar o que, supostamente, deverá existir no futuro.

3) Identificação e Modelagem

Entende-se por **identificação** a determinação de um **modelo** do sistema representado por seus aspectos essenciais, de uma forma adequada para a sua utilização [1].

“ Um modelo é uma representação de um objeto, sistema ou idéia em alguma forma outra que não a da própria entidade ” [2]. Um modelo de um objeto ou de um sistema pode ser uma reprodução exata do objeto ou do sistema (embora executado em material ou escala diferentes), ou pode ser uma abstração das propriedades do objeto ou do sistema. Os modelos de previsão de séries temporais constituem-se em uma classe de modelos [3]. A forma do modelo é determinada pelo tipo do sistema em consideração, como por exemplo, estático, dinâmico, linear, não-linear, estocástico, determinístico, discreto, contínuo, etc. [4]. Existem duas linhas de ação que podem ser relacionadas entre si para a determinação do modelo no problema de identificação [4]:

- i) A **análise físico-matemática** - *model building* em que o modelo é obtido através da descrição do comportamento físico do sistema [5];
- ii) A **identificação experimental** em que se utiliza o fato do processo estar em funcionamento para, através de observações, construir um modelo paramétrico do sistema [6, 5].

Tendo em vista os objetivos e especificações do sistema, deve-se obter um modelo que represente seus aspectos fundamentais.

Neste trabalho, em geral, estuda-se o problema de identificação experimental dado que a aplicação em mente é a projeção de séries temporais ligadas aos valores experimentais

¹Por exemplo, nas equações do Filtro de Kalman é mais usual se usar **predição**, por outro lado, aos valores futuros das rotas de tráfego costuma-se usar **projeção**, etc.

(medições) do tráfego telefônico. Quando for apresentado o modelo de Espaço de Estados, também serão utilizadas as características físicas do sistema, o que representa uma vantagem significativa em relação aos demais modelos.

4) Especificação do Modelo de Projeção

Devido a que os investimentos nos sistemas telefônicos, decorrentes de estudos de projeções de tráfego, podem requerer a alocação de vários milhões de dólares anualmente, é importante que as projeções possuam determinadas propriedades estatísticas. Por exemplo, as projeções devem ser não-polarizadas (não-viesadas), ou seja, não ser consistentemente altas ou baixas.

A projeção da carga de tráfego (a estimativa da carga futura de tráfego) é a base para o planejamento e a administração da rede de troncos das empresas operadoras. Uma projeção deve possuir dois atributos principais: estabilidade² e precisão. Esses atributos são normalmente conflitantes no sentido que uma maior precisão da projeção pode resultar em estabilidade mais pobre, e vice-versa [22, 23]. Dentre os objetivos e requisitos que norteiam a seleção do modelo, ressaltam-se:

- a) Uma clara descrição das variáveis descrevendo o tráfego, como por exemplo, o tráfego oferecido à rede nos períodos de ocupação;
- b) O tamanho do período de projeção;
- c) Disponibilidades de dados históricos e facilidades de computação;
- d) Nível de agregação — os dados podem ser originados de medições **ponto-a-ponto** ou **área a área** correspondentes à agregação de várias rotas de tráfego;
- e) Precisão das projeções;
- f) Estabilidade das projeções.

O modelo que melhor se adequar a estas condições, dentro de um enfoque custo×eficácia, terá maiores chances de ser escolhido.

2.2 Classificação dos Modelos Para Previsão de Séries Temporais

De uma forma geral, uma série temporal pode ser escrita como:

$$Y_k = f_k + v_k \quad (2.1)$$

onde Y_k é o valor da saída observada do sistema, f_k é o sinal e v_k é uma perturbação não-mensurável (ruído)³.

²O Capítulo 3 aborda os critérios para estabilidade das projeções e discute sua estreita relação com a projeção n passos à frente. O Apêndice C enfoca a estabilidade dos algoritmos baseados no modelo de Espaço de Estados com o estado estimado pelo Filtro de Kalman, de uma maneira geral.

³Na maioria dos modelos enfocados neste trabalho, v_k é um ruído branco Gaussiano de média nula. Nos modelos multivariáveis, a perturbação é um vetor que representa-se por V_k .

Para a análise de séries temporais existem duas abordagens predominantes. Na primeira, a análise é feita no domínio do tempo e os modelos propostos são, em geral, paramétricos. Na segunda, a análise é realizada no domínio da frequência e os modelos propostos são, em geral, não-paramétricos.

Nas séries ligadas ao tráfego telefônico (e em muitas séries econômicas) os eventos ocorrem em uma seqüência temporal que torna a análise no domínio do tempo mais adequada.

Os modelos para previsão de tráfego, no domínio do tempo, podem ser classificados como **estáticos (ou de erro) e dinâmicos**.

Nos modelos estáticos, o sinal f_k depende somente de entradas externas (em outras palavras, pode-se dizer que os modelos estáticos não possuem entradas internas defasadas). Nos modelos dinâmicos, o sinal f_k depende dos valores passados de Y_k .⁴ Devido ao fato dos modelos dinâmicos estarem relacionados ao conceito de **estado**, uma breve descrição é dada a seguir, quase exatamente (a menos de tradução e pessoa verbal) como em [63]:

“ O estado é alguma representação suficientemente completa do sistema para permitir prever, com base nas entradas, exatamente que saídas ocorrerão, e também atualizar o próprio valor do estado ”. Por exemplo, comparando os sistemas físicos (contínuos) da passagem de uma corrente por um resistor e por um capacitor. Como o resistor não possui memória, sua saída é unicamente determinada pelo valor da entrada naquele instante. Entretanto, o capacitor requer uma explícita descrição da variável de estado a fim de que se saiba qual será o valor de sua saída (é necessário o conhecimento de sua tensão inicial).

1) Modelos Estáticos

No caso geral, a função f_k pode ser escrita como em (2.2):

$$f_k = g(\mu_k, T_k, S_k) \tag{2.2}$$

onde o componente μ_k representa o valor de um nível constante da série, ou que pode variar lentamente com o tempo, T_k representa a ciclotendência e S_k é o componente sazonal. Os componentes de $g(\cdot)$ podem ser do tipo aditivo ou multiplicativo.

De uma maneira geral, os modelos de regressão linear múltipla podem ser representados por:

$$g(\mu_k, T_k, S_k) = \beta_0 + \beta_1 X_{1k} + \dots + \beta_l X_{lk} \tag{2.3}$$

onde

- $\beta_i, \quad i = 0, 1, \dots, l$ são os parâmetros a serem estimados;
- $X_{ik}, \quad i = 1, 2, \dots, l$ são os valores das entradas externas no instante k .

Quando os modelos são aditivos, uma maneira de se modelar a soma $\mu_k + T_k$ é por meio de um polinômio de grau l como em (2.4) [7]:

$$\mu_k + T_k = \beta_0 + \beta_1 k^1 + \dots + \beta_l k^l \tag{2.4}$$

⁴Os modelos *MA - moving average* (será apresentado mais à frente, neste capítulo), que pertencem à classe de modelos Entrada×Saída, são uma exceção, já que não têm sua saída dependente de valores de saídas passadas; porém, podem ser considerados dinâmicos pois seus pólos são localizados na origem.

Neste caso, tem-se uma tendência polinomial determinística de grau l , de modo que o componente sistemático se mova suave, lenta e progressivamente no tempo.

O componente S_k pode ser modelado como um polinômio harmônico ou por meio da série de Fourier (combinação linear de senos e cossenos com coeficientes constantes) como em (2.5) [16]:

$$S_k = \sum_{j=1}^L [a_j \cos \frac{2\pi}{L}jk + b_j \sin \frac{2\pi}{L}jk] \quad (2.5)$$

onde L é o período de S_k .

Além dos modelos de regressão existem os procedimentos de suavização. Pode-se dizer que os procedimentos de suavização apenas trabalham com os dados, procurando reduzir sua variância para serem utilizados em um modelo de regressão, por exemplo. Aos modelos de regressão que trabalham ao mesmo tempo com os dados suavizados, dá-se o nome de modelos mistos.

Os modelos aqui denominados extrapolativos são um outro tipo de modelo estático, em que não se faz nenhuma consideração sobre o padrão da série temporal, e a projeção é feita multiplicando-se o valor da medida de tráfego atual por um fator de projeção da tendência de crescimento das centrais telefônicas [22].

Existem ainda os procedimentos estáticos que trabalham no ajuste de projeção individual ou agregada e são utilizados para matrizes de tráfego. As projeções que são ajustadas por esses procedimentos podem ter sido originadas por qualquer método de projeção.

Nas aplicações em projeção de tráfego telefônico, os coeficientes de (2.3) são determinados minimizando-se o erro entre a demanda real e a demanda estimada para o referido tráfego.

2) Modelos Dinâmicos

Dois tipos de abordagem sobre modelos dinâmicos serão estudadas: os modelos **Entrada×Saída** e os modelos de **Espaço de Estados**.

Os modelos **Entrada×Saída** podem ser representados por:

$$\Phi(B)Y_k = \Theta(B)V_k + \xi(B)U_k \quad (2.6)$$

onde

$$\Phi(B) = 1 + \phi_1 B + \dots + \phi_p B^p \quad (2.7)$$

$$\Theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q, \quad q \leq p \quad (2.8)$$

$$\xi(B) = 1 + \xi_1 B + \dots + \xi_t B^t, \quad t < p \quad (2.9)$$

Nas equações de (2.6) a (2.9), $\Phi(B)$ é um polinômio em B , de ordem p , $\Theta(B)$ é um polinômio em B de ordem $q \leq p$, $\xi(B)$ é um polinômio em B de ordem inferior a p , Y_k é o valor da saída no instante k , V_k é um ruído branco Gaussiano de média nula e variância

σ_V^2 constante, U_k é uma seqüência de entradas e B é o operador deslocamento para trás (*backward*), tal que $BY_k \triangleq Y_{k-1}$.

Os modelos de Espaço de Estados de tempo discreto podem ser representados na sua forma geral por:

$$X_{k+1} = f(X_k, U_k, W_k) \quad (2.10)$$

$$Y_k = g(X_k, V_k) \quad (2.11)$$

Se o modelo for linear:

$$X_{k+1} = \Phi X_k + LU_k + \Gamma W_k \quad (2.12)$$

$$Y_k = HX_k + V_k \quad (2.13)$$

Nas equações de (2.10) a (2.13):

- Y_k é o vetor de saídas (observações) do sistema;
- X_k é um vetor $n \times 1$ de variáveis de estado (não-observáveis) no instante k ;
- L é uma matriz de entradas $n \times l$;
- U_k é um vetor $l \times 1$ de entradas;
- Φ é a matriz de transição de estado $n \times n$;
- Γ é uma matriz de entradas $n \times r$;
- W_k é um vetor de erros de modelagem $r \times 1$;
- H é chamada matriz de observação $m \times n$;
- V_k é um vetor dos erros de medição $m \times 1$.

2.3 Modelos Estáticos (de erro)

Dependendo do valor atribuído à função f_k , descrita em (2.2), é possível obterem-se diferentes modelos para a projeção de tráfego.

Para os modelos de regressão linear múltipla, os coeficientes β_i de (2.3) são calculados de forma a minimizar D tal que:

$$D = \sum_{k=1}^N w^{N-k} [Y_k - \beta_0 - \beta_1 X_{1k} - \dots - \beta_l X_{lk}]^2 \quad (2.14)$$

onde w é a constante de suavização. A solução de mínimos quadrados de (2.14) é dada por (2.15):

$$\hat{B} = (X^T X)^{-1} X^T Y \quad (2.15)$$

onde $X^T X$ é não-singular e \hat{B} é o estimador de

$$B = [\beta_0 \quad \beta_1 \quad \cdots \quad \beta_l]^T$$

e

$$Y = [Y_1 \quad Y_2 \quad \cdots \quad Y_N]^T$$

e

$$X = [X_1 \quad X_2 \quad \cdots \quad X_N]^T, \quad k = 1, \dots, N$$

e

$$X_k = [1 \quad X_{1k} \quad X_{2k} \quad \cdots \quad X_{lk}]^T$$

Se as variáveis X_{ik} forem correlacionadas, então o valor dos coeficientes pode ser estimado levando-se em conta essas correlações [9]. O número de observações N deve ser maior que l . Se as perturbações não são ruído branco Gaussiano, utiliza-se o método dos mínimos quadrados generalizados para se obterem estimativas não-polarizadas dos coeficientes β_i [34]. Também é possível estimarem-se os coeficientes por uma forma alternativa, usando os polinômios ortogonais [7]. Para verificar a existência da regressão, existe um teste que utiliza a distribuição \mathcal{F} de Snedecor [49], ou quando a regressão possui apenas os parâmetros β_0 e β_1 , o teste de existência é feito pela distribuição t de Student [49]. O ajuste desses modelos é medido pelo coeficiente de explicação ou determinação R^2 (R é o coeficiente de correlação múltipla) [62, 49]. O coeficiente de explicação, R^2 , é expresso pela relação entre a variação esperada (erro quadrático entre os valores estimados pelo modelo e o valor médio observado) e a variação total (erro quadrático entre os valores observados e seu valor médio) [49].

Os procedimentos de suavização — que de fato apenas provêem uma maneira de reduzir a variância dos dados — têm uma ou mais constantes de alisamento, para cuja determinação procede-se à minimização de (2.16), onde:

$$D = \sum_{k=1}^N (Y_k - Y_k^*)^2 \quad (2.16)$$

Pode-se verificar que (2.16) é equivalente a (2.14), para $w = 1$ e utilizando-se Y_k^* como sendo a projeção da saída no instante k a partir dos dados $k - 1$. Existem vários outros critérios para a determinação das constantes de alisamento como, por exemplo, aquele que leva em conta o tipo de autocorrelação entre os dados e o custo da previsão [7].

Os procedimentos extrapolativos projetam as cargas de tráfego por um modelo da forma $Y_{k+n}^* = Y_k p$, onde Y_k é a medida do grupo de troncos no ano corrente, Y_{k+n}^* é a carga futura estimada no ano $k + n$ e p é um fator de projeção da tendência de crescimento das centrais. As medidas das cargas correntes são obtidas de uma estatística **ponto-a-ponto** ou **área-a-área**, que são um subproduto do sistema de tarifação, ou por meio de medição direta do grupo de troncos. Os procedimentos extrapolativos, geralmente, trabalham com dados agregados.

Os métodos que trabalham com matrizes de tráfego (podem ser utilizados para ajuste de projeções individuais ou agregadas) utilizam as seguintes técnicas para se obter a matriz

de tráfego (a essa forma de obtenção das matrizes de tráfego, dá-se o nome de fechamento)⁵ [10]:

- O método de **Kruithof**, que utiliza a última matriz de tráfego conhecida e usa projeções das linhas e das colunas para construir a nova matriz, por meio de um procedimento iterativo;
- O método de **Kruithof estendido** que, além das projeções das linhas e das colunas, usa também as projeções do tráfego **ponto-a-ponto** para obter consistência com a soma das linhas e colunas;
- O método dos **mínimos quadrados ponderados** que é baseado na suposição de que as projeções da soma das linhas, da soma das colunas e **ponto-a-ponto** são incertas, dando-se, assim, pesos diferentes a essas várias projeções.

Os modelos estáticos podem ainda ser divididos em:

- 1) Simples;
- 2) Aditivos;
- 3) Multiplicativos;
- 4) Gerais.

A seguir serão apresentados os tipos de modelos mais comumente utilizados na literatura.

1) Modelos Simples

- Suponha $g(\mu_k, T_k, S_k) = \mu_k$, e $T_k = S_k = 0$. Para este caso tem-se:

– Média Móvel Simples (MMS)

A técnica da MMS calcula a média aritmética das observações mais recentes, isto é:

$$\hat{\mu}_k = \frac{Y_k + Y_{k-1} + \dots + Y_{k-r+1}}{r} \quad (2.17)$$

onde r é a constante de alisamento e $\hat{\mu}_k$ é o estimador de μ_k . Um grande valor da constante de alisamento, r , faz com que a projeção acompanhe lentamente o valor do nível μ_k ; um pequeno valor implica em uma reação mais rápida. A projeção dos valores futuros é igual à última média móvel calculada:

$$Y_{k+n|k}^* = \hat{\mu}_k \quad (2.18)$$

onde $Y_{k+n|k}^*$ é a projeção n passos à frente a partir do instante k . Na sua forma recursiva:

$$Y_{k+n|k}^* = Y_{k+n|k-1}^* + \frac{Y_k - Y_{k-r}}{r} \quad (2.19)$$

$$\forall n > 0, n = 1, 2, 3, \dots$$

⁵Quando a quantidade de dados para projetar a matriz de tráfego é insuficiente, Zolfaghari [42] sugere a utilização de um modelo baseado nas distâncias entre os centros de comutação, chamado de Modelo Gravitacional de Tráfego.

– Alisamento Exponencial Simples (AES)

O AES é definido por:

$$\hat{\mu}_k = \alpha Y_k + (1 - \alpha)\hat{\mu}_{(k-1)} \quad (2.20)$$

onde $\hat{\mu}_0 = Y_1$, $k = 1, \dots, N$ e $0 < \alpha < 1$ é a constante de alisamento. O último valor exponencialmente alisado fornece a projeção a n passos:

$$Y_{k+n|k}^* = \hat{\mu}_k \quad (2.21)$$

ou na sua forma recursiva:

$$Y_{k+n|k}^* = \alpha Y_k + (1 - \alpha)Y_{k+n|k-1}^* \quad (2.22)$$

$\forall n > 0, n = 1, 2, 3, \dots$

É fácil notar que o modelo MMS é um caso particular do AES (para $\alpha = \frac{2}{r-1}$), que é definido de maneira recursiva.

- Suponha $g(\mu_k, T_k, S_k) = S_k$, e $\mu_k = T_k = 0$. Para este caso tem-se:

– Polinômios Harmônicos

O modelo dos polinômios harmônicos (Seção 2.2) modela S_k como em (2.5).

2) Modelos Aditivos

- Suponha $g(\mu_k, T_k, S_k) = \mu_k + T_k$, e $S_k = 0$

– Ajuste de Tendência Polinomial

Este modelo de regressão é um caso particular de (2.3) com $X_k = k^i$, $i = 1, \dots, l$:

$$\mu_k + T_k = \beta_0 + \beta_1 k + \dots + \beta_l k^l \quad (2.23)$$

– Procedimento de Holt

O procedimento de Holt é descrito por:

$$\hat{\mu}_k = AY_k + (1 - A)(\hat{\mu}_{k-1} + \hat{T}_{k-1}), \quad 0 < A < 1$$

e

$$\hat{T}_k = C(\hat{\mu}_k - \hat{\mu}_{k-1}) + (1 - C)\hat{T}_{k-1}, \quad 0 < C < 1 \quad (2.24)$$

onde A e C são as constante de alisamento e \hat{T}_k é a estimativa da tendência. A projeção de Y_k a n passos é:

$$Y_{k+n|k}^* = \hat{\mu}_k + n\hat{T}_k \quad (2.25)$$

$\forall n > 0, n = 1, 2, 3, \dots$

– Alisamento Exponencial Duplo (AED)

Este também é um caso de ajuste polinomial com suavização. Se a série possuir tendência linear:

$$\mu_k + T_k = b_1 + b_2 k, \quad k = 1, \dots, N \quad (2.26)$$

O AED calcula um valor exponencialmente alisado pela segunda vez:

$$\hat{\mu}_{T_k} = \alpha \hat{\mu}_{T_k} + (1 - \alpha) \hat{\mu}_{T_k} \quad (2.27)$$

onde $\hat{\mu}_{T_k}$ é a estimativa de $\mu_k + T_k$ e $\hat{\mu}_{T_1} = Y_1$, onde $\hat{\mu}_{T_k}$ é dado por $\hat{\mu}$ de (2.20). A projeção a n passos a partir da k -ésima observação é:

$$Y_{k+n|k}^* = \hat{a}_{1,k} + \hat{b}_{2,k} n \quad (2.28)$$

onde pode-se demonstrar que [8]: $\hat{a}_{1,k} = 2\hat{\mu}_{T_k} - \hat{\mu}_{T_k}$ e $\hat{b}_{2,k} = \frac{\alpha}{1-\alpha}(\hat{\mu}_{T_k} - \hat{\mu}_{T_k})$

• Suponha $g(\mu_k, T_k, S_k) = \mu_k + T_k + S_k$. Para este caso tem-se:

– Tendência Polinomial + Polinômio Harmônico

O modelo de tendência polinomial + polinômio harmônico combina (2.23) e (2.5).

– Alisamento Exponencial Geral - Método de Brown - AEG

É o modelo mais geral entre os casos particulares de regressão linear múltipla estudados. Em (2.14), o valor da constante de suavização w pode variar entre 0 e 1, sendo esta a única diferença em relação ao modelo anterior. O fato de $w < 1$ implica que a função tem validade em um período de tempo limitado. O valor dos coeficientes, ao contrário do que acontece nas maneiras tradicionais de calculá-los nas regressões lineares múltiplas, pode ser atualizado a cada nova observação, evitando-se a inversão de matriz. Para se estimar a constante de suavização, podem, por exemplo, ser utilizados os critérios citados na Seção 2.3. No caso em que o número de observações, N , tende a infinito, além da atualização dos coeficientes a cada nova observação, este procedimento permite a construção de intervalos de confiança para a projeção a n passos ($Y_{k+n|k}^*$) [7].

– Procedimento de Holt Winters Aditivo - HWA

Existem três equações de alisamento representando sazonalidade, nível e tendência, respectivamente, dadas por:

$$\begin{aligned} \hat{S}_k &= D(Y_k - \hat{\mu}_k) + (1 - D)\hat{S}_{k-s}, \quad 0 < D < 1 \\ \hat{\mu}_k &= A(Y_k - \hat{S}_{k-s}) + (1 - A)(\hat{\mu}_{k-1} + \hat{T}_{k-1}), \quad 0 < A < 1 \\ \hat{T}_k &= C(Y_k - \hat{\mu}_{k-1}) + (1 - C)\hat{T}_{k-1}, \quad 0 < C < 1 \end{aligned} \quad (2.29)$$

onde A , B e C são as constantes de alisamento, \hat{S} é a estimativa da sazonalidade e s é o período sazonal. A projeção a n passos é:

$$\begin{aligned} Y_{k+n|k}^* &= \hat{\mu}_k + n\hat{T}_k + \hat{S}_{k+n-s}, \quad n = 1, 2, \dots, s \\ Y_{k+n|k}^* &= \hat{\mu}_k + n\hat{T}_k + \hat{S}_{k+n-2s}, \quad n = s + 1, \dots, 2s \\ &\vdots \end{aligned} \quad (2.30)$$

3) Modelos Multiplicativos

• Suponha $g(\mu_k, T_k, S_k) = \mu_k S_k + T_k$. Para este caso tem-se:

– Procedimento de Holt Winters Multiplicativo - HWM

As três equações de alisamento representando sazonalidade, nível e tendência, respectivamente, são:

$$\begin{aligned}\hat{S}_k &= D \frac{Y_k}{\hat{\mu}_k} + (1 - D) \hat{S}_{k-s}, \quad 0 < D < 1, \quad k = s + 1, \dots, N \\ \hat{\mu}_k &= A \frac{Y_k}{\hat{S}_{k-s}} + (1 - A)(\hat{Y}_{k-1} + \hat{T}_{k-1}), \quad 0 < A < 1, \quad k = s + 1, \dots, N \\ \hat{T}_k &= C(\hat{\mu}_k - \hat{\mu}_{k-1}) + (1 - C)\hat{T}_{k-1}, \quad 0 < C < 1, \quad k = s + 1, \dots, N\end{aligned}\quad (2.31)$$

onde A , B e C são as constantes de alisamento e s é o período sazonal. A projeção a n passos é:

$$\begin{aligned}Y_{k+n|k}^* &= (\hat{\mu}_k + n\hat{T}_k)\hat{S}_{k+n-s}, \quad n = 1, 2, \dots, s \\ Y_{k+n|k}^* &= (\hat{\mu}_k + n\hat{T}_k)\hat{S}_{k+n-2s}, \quad n = s + 1, \dots, 2s \\ &\vdots\end{aligned}\quad (2.32)$$

Os valores iniciais para as equações dos procedimentos sazonais de HW, tanto para o aditivo quanto para o multiplicativo, são:

$$\hat{S}_j = \frac{Y_j}{\frac{1}{s} \sum_{i=1}^s Y_i}, \quad j = 1, 2, \dots, s$$

$$\hat{\mu}_k = \frac{1}{s} \sum_{i=1}^s Y_i, \quad k = 1, 2, \dots, s$$

$$\hat{T}_s = 0 \quad (2.33)$$

– Tendência Polinomial \times Polinômio Harmônico

Este modelo corresponde ao produto dos primeiros termos de (2.23) — geralmente apenas β_0 representando μ_k — com (2.5), representando S_k . Este resultado é somado aos demais termos de (2.23) (β_0 já foi retirado), que representam T_k . O tratamento deste modelo é análogo aos outros de regressão já apresentados.

• Suponha $g(\mu_k, T_k, S_k) = (\mu_k + T_k)S_k$. Para este caso tem-se:

– Tendência Polinomial \times Polinômio harmônico

Este modelo corresponde ao produto de (2.23) por (2.5). O tratamento deste modelo já foi visto para os outros modelos de regressão apresentados.

4) Modelos Gerais

• Suponha $g(\mu_k, T_k, S_k)$ geral. Neste caso geral não há uma função explícita para modelar μ_k, T_k e S_k . Como exemplos, podem-se citar:

– Ajuste de Funções

Seja um modelo exponencial dado por

$$f_k = ae^{bk} \quad (2.34)$$

Aplicando-se logaritmo em (2.34), tem-se:

$$\ln f_k = \ln a + bk$$

ou

$$Z_k = \beta_0 + \beta_1 X_k + v_k$$

onde $Z_k = \ln f_k$, $\beta_0 = \ln a$, $\beta_1 = b$, $X_k = k$ e v_k é o erro. De maneira análoga, outros modelos de funções podem ser usados para projeção de tráfego:

* Linear: $f_k = a + bk$

* Parabólica: $f_k = a + bk + ck^2$

* Logística: $f_k = \frac{M}{1+ae^{bk}}$

* Gompertz: $f_k = Ma^{bk}$

onde f_k é o valor do tráfego no instante k ; a , b e c são os parâmetros a serem estimados; M é um parâmetro que descreve o nível de saturação para as curvas Logística e Gompertz.

A Figura 2.2 mostra os vários tipos de curvas mencionados [10]. Na abscissa são mostrados os anos (de 1980 até o ano 2000) e na ordenada o número de terminais telefônicos necessários em cada ano, obtidos por meio de cinco tipos de funções: Gompertz, Logística, Exponencial, Parabólica e Linear. O ano de 1980 é tomado como referência com o valor 100. As funções Logística e Gompertz diferem das demais devido a possuírem um nível de saturação M , e devem ser utilizadas quando se tem alguma previsão desse ponto. A função a ser adotada deve ser aquela com o coeficiente R^2 mais próximo a um, o que indica melhor ajuste às observações disponíveis. No entanto, ajustes adequados às observações não implicam em projeções adequadas. Por exemplo, se o número de terminais no ano 2000 estiver próximo a 20000 (modelo exponencial) mas for adotado o modelo linear, o erro será significativo.

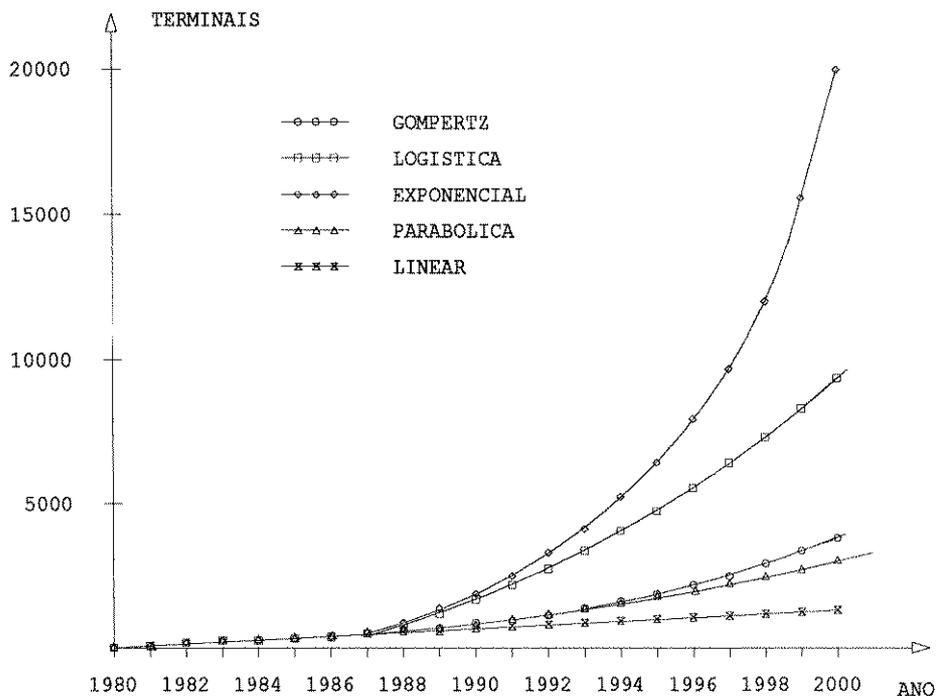


Figura 2.2: Tipos de Funções para Ajustar Dados de Tráfego

– Modelos Econométricos (ou com Variáveis Explicativas)

Como exemplo de modelo econométrico, seja a demanda de chamadas (o número de chamadas é facilmente transformado em tráfego) em um determinado país relacionado com sua densidade de telefones pela seguinte função [53]:

$$C = aD^b e^{v_k} \quad (2.35)$$

onde

- * C é a demanda em chamadas por habitante por ano;
- * D é a densidade em terminais por 100 habitantes;
- * a é uma constante de escala;
- * b é a elasticidade com relação à densidade de terminais;
- * v_k é uma variável aleatória que serve para compensar as variáveis não colocadas explicitamente no modelo, $v_k \sim \mathcal{N}(0, \sigma_v^2)$.

Aplicando-se logaritmos, recai-se nos casos anteriores. Desta forma este modelo econométrico pode ser tratado como um modelo de regressão (a diferença importante é a existência de variáveis explicativas ou relacionadas a alguma teoria. Mesmo modelos simples como este exemplo podem ser bastante eficazes). Por exemplo, levantou-se a função relacionando-se a densidade de telefones com o número de chamadas, cujos dados foram retirados da revista da Siemens [55] que possui dados de densidade de chamadas e densidade de telefones para todos os países do mundo, desenvolvidos ou em desenvolvimento. A esses dados procurou-se ajustar os modelos linear, exponencial e potencial ($f_k = ak^b$). O melhor coeficiente de explicação ocorreu para o modelo exponencial. Desta

forma, a equação com os parâmetros estimados é:

$$C = \hat{a}D^{\hat{b}} \quad (2.36)$$

com os dados disponíveis, obteve-se $\hat{a} = 29,84$, $\hat{b} = 0,804$ e o coeficiente de explicação $R^2 = 0,76$.

• Procedimentos Extrapolativos

Os procedimentos extrapolativos são bastante simples. Quando existem poucos dados, ou quando o horizonte de projeção é muito longo (acima de 3 a 4 anos), é melhor fazerem-se projeções com métodos mais simples. A utilização conjunta dos procedimentos extrapolativos com o modelo de ajuste da matriz de tráfego produz resultados dentro das especificações a longo prazo. Os modelos extrapolativos podem cometer grandes erros com relação às projeções individuais, porque o fator de crescimento agregado é médio e global. Além disso, devido à quantidade limitada de dados em que as medidas são baseadas, podem ocorrer erros estatísticos significativos, pois os modelos dependem do tamanho da carga de tráfego e do tipo de sistema de medição utilizado. Como exemplo, se as taxas de crescimento da central de origem, da central de destino e da área de destino são denotadas por O, D e T , respectivamente, então p pode ser selecionado de fórmulas tais como $\frac{OD}{T}$, $\frac{O+D}{2}$, $O\sqrt{D}$, O ou D ; $Y_{k+n|k}^*$ é calculado como $Y_k p$ [22]. No Brasil, a Telesp (Telecomunicações de São Paulo S.A.) utiliza-se de fórmulas propostas no trabalho de Rapp [54]⁶.

• Projeção de Matrizes de Tráfego

Como exemplo de um modelo desse tipo, pode-se citar o método dos mínimos quadrados ponderados. Seja a equação da forma quadrática:

$$D = \sum_{i,j} a_{ij}(c_{ij} - D_{ij})^2 + \sum_i b_i(c_i - D_i)^2 + \sum_j (c_j - D_j)^2 \quad (2.37)$$

onde $\{a_{ij}\}$, $\{b_i\}$ e $\{c_j\}$ são constantes escolhidas ou pesos, $\{c_{ij}\}$, $\{c_i\}$ e $\{c_j\}$ são as projeções dos tráfegos **ponto-a-ponto**, da soma das linhas e da soma das colunas,

⁶Existem três equações propostas por Rapp:

$$A_{jl} = A_{jl}^o \frac{N_j \frac{N_l}{N_l^o} + N_l \frac{N_j}{N_j^o}}{N_j + N_l} \quad (a), \quad A_{jl} = A_{jl}^o \frac{N_j^2 \frac{N_l}{N_l^o} + N_l^2 \frac{N_j}{N_j^o}}{N_j^2 + N_l^2} \quad (b) \quad \text{e} \quad A_{jl} = A_{jl}^o \frac{N_j N_l M^o}{N_j^o N_l^o M} \quad (c)$$

onde:

- A_{jl}^o é o tráfego atual entre as centrais j e l (Y_k na equação geral);
- A_{jl} é o tráfego futuro (Y_{k+n}^* na equação geral);

As outras variáveis compõem o valor de p na equação geral:

- N_j^o , N_l^o são os números atuais de assinantes nas áreas j e l ;
- N_j , N_l são os números futuros de assinantes nas áreas j e l ;
- M_o , M_l são o número total de assinantes da área, atual e futuro, respectivamente.

A equação (a) é utilizada pela Telesp para o tráfego interurbano, a (c) para o tráfego local, e a (b) não é utilizada.

respectivamente. As projeções de mínimos quadrados ponderados $\{D_{ij}\}$ são obtidas de modo a minimizar D :

$$\min D(D_{ij}), \text{ sujeito a } \begin{cases} D_{.j} = \sum_i D_{ij} & j = 1, 2, \dots \\ D_{.i} = \sum_j D_{ij} & i = 1, 2, \dots \end{cases} \quad (2.38)$$

Os pesos podem ser escolhidos como o inverso das variâncias das projeções. Uma forma de se estimar o desvio padrão das projeções é compará-las com valores anteriores, já conhecidos, e se calcular a raiz do erro quadrático médio [10].

A Tabela 2.1 mostra os diversos tipos de modelos estáticos em função dos efeitos modelados⁷.

Tabela 2.1 - Tipos de Modelos Estáticos (de Erro)

Tipo Efeito Modelado	Regressão Linear Múltipla						Extrapolativ. Procedim. Extrapol.
	Tendência Polinomial	Polinômio Harmônico	Tend. Polinom. + Polin. Harm.	Tend. Polinom. x Polin. Harm.	Econômétrico	Ajuste de Funções	
$S_k, \mu_k = T_k = 0$		X					
$\mu_k + T_k, S_k = 0$	X						
$\mu_k + T_k + S_k$			X				
$\mu_k S_k + T_k$				X			
$(\mu_k + T_k)S_k$				X			
$g(\mu_k, T_k, S_k)$					X	X	X
Tipo Efeito Modelado	Suavização				Mistos		
	Média Móvel Simples - MMS	Alisam. Expon. Simples - AES	Procedim. de Holt - ALB	Holt Winters Multiplicativo - HWM	Holt Winters Aditivo - HWA	Alisam. Expon. Duplo - AED	Alisam. Expon. Geral - AEG
$\mu_k, T_k = S_k = 0$	X	X					
$\mu_k + T_k, S_k = 0$			X			X	
$\mu_k + T_k + S_k$					X		X
$\mu_k S_k + T_k$				X			

2.4 Modelos Dinâmicos

Como foi apresentado na Seção 2.2, os modelos dinâmicos são classificados em modelos do tipo Entrada×Saída e Espaço de Estados.

2.4.1 Modelos Entrada×Saída

- Modelos *ARMA* (*Autoregressive Moving Average*)

Os modelos *ARMA* são descritos por uma equação a diferenças conforme (2.39):

$$Y_k = \phi_1 Y_{k-1} + \dots + \phi_p Y_{k-p} + V_k - \theta_1 V_{k-1} - \theta_2 V_{k-2} - \dots - \theta_q V_{k-q} \quad (2.39)$$

onde Y_k são os valores das saídas (observações), V_k é uma perturbação branca de média nula, ϕ_i e θ_i são os valores dos parâmetros a serem estimados. Em forma compacta:

$$\Phi(B)Y_k = \Theta(B)V_k \quad (2.40)$$

⁷Os modelos de projeção de matrizes de tráfego não aparecem na tabela, porém, são classificados como modelos estáticos e podem trabalhar com projeções fornecidas por qualquer outro modelo, estático ou dinâmico.

onde:

$$\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad (2.41)$$

$$\Theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q \quad (2.42)$$

Note que (2.39) é a equação (2.6) com $\xi(B) = 0$. No caso em que o processo possui um nível constante, pode-se incluir este nível na série temporal ou trabalhar com a média subtraída do mesmo (o termo constante em um modelo *ARMA* corresponde à tendência polinomial apresentada nos modelos estáticos).

Quando $\Phi(B) = 1$, o modelo corresponde ao processo *MA* - *Moving Average* e quando $\Theta(B) = 1$, o modelo corresponde ao processo *AR* - *Auto-Regressivo*.

Pode-se demonstrar que um processo *AR* será estacionário se todas as raízes da equação característica, $\Phi(B) = 0$, estiverem fora do círculo unitário [7]. Pode-se inverter o processo *MA* se todas as raízes da equação característica, $\Theta(B) = 0$, estiverem fora do círculo unitário (se ao invés de B fosse usado B^{-1} como operador deslocamento para trás, então as raízes deveriam estar dentro do círculo unitário). Qualquer processo estocástico estacionário no sentido amplo, ou fracamente estacionário, pode ser representado tanto por um processo *MA* puro ou por um processo *AR* puro. Portanto, um processo *ARMA* será estacionário se as raízes de $\Phi(B) = 0$ estiverem todas fora do círculo unitário e pode-se inverter esse processo se as raízes de $\Theta(B) = 0$ estiverem fora do círculo unitário.

- Modelos *ARIMA* (*Autoregressive Integrated Moving Average*)

Se a série temporal Y_k tornar-se estacionária, ao se tomar um número finito de diferenças, d , ela é chamada não-estacionária homogênea. Tomar a diferença de uma série temporal, Y_k , significa fazer:

$$\Delta Y_k = Y_k - Y_{k-1} = (1 - B)Y_k, \quad \text{com } BY_k = Y_{k-1}$$

e a d -ésima diferença é:

$$\Delta^d Y_k = \Delta(\Delta^{d-1} Y_k)$$

Se a série $Z_k = \Delta^d Y_k$ for estacionária, então é possível representar Z_k por um modelo *ARMA*(p, q), ou:

$$\Phi(B)Z_k = \Theta(B)V_k$$

Diz-se que Y_k segue um modelo denominado auto-regressivo integral e médias móveis, ou *ARIMA*(p, d, q):

$$\Phi(B)\Delta^d Y_k = \Theta(B)V_k \quad (2.43)$$

Também é possível escrever (2.43) como:

$$\Omega(B)Y_k = \Theta(B)V_k \quad (2.44)$$

onde:

$$\Omega(B) = \Phi(B)\Delta^d = \Theta(B)(1 - B)^d$$

$\Omega(B)$ é um operador auto-regressivo não-estacionário, de ordem $p + d$, com d raízes iguais a um, sobre o círculo unitário, e p raízes fora do círculo unitário se o sistema é estável.

- Modelo Sazonal *ARMA*

Em analogia com a descrição dos modelos *ARMA*, pode-se dizer que os modelos *ARMA* Sazonais são descritos por:

$$Y_k = \phi_1^s Y_{k-1} + \dots + \phi_P^s Y_{k-P} + \alpha_k - \theta_1^s \alpha_{k-1} - \theta_2^s \alpha_{k-2} - \dots - \theta_Q^s \alpha_{k-Q} \quad (2.45)$$

ou, em forma compacta:

$$\Phi(B^s)Y_k = \Theta(B^s)\alpha_k \quad (2.46)$$

onde α_k é uma perturbação não-branca, representado por um processo *ARIMA*,

$$\Phi(B^s) = 1 - \phi_1^s B^s - \phi_2^s B^{2s} - \dots - \phi_P^s B^{Ps}$$

é o operador auto-regressivo sazonal de ordem P e

$$\Theta(B^s) = 1 - \theta_1^s B^s - \theta_2^s B^{2s} - \dots - \theta_Q^s B^{Qs}$$

é o operador médias móveis sazonal de ordem Q , s é o período sazonal (geralmente igual a 12 meses) e B^s é o deslocamento de um período sazonal para trás, da forma:

$$\Delta^s Y_k = Y_k - Y_{k-s} = (1 - B^s)Y_k$$

ou seja, $\Delta^s = 1 - B^s$.

Em geral, α_k pode ser representado por um processo *ARIMA* como em (2.43):

$$\Phi(B)\Delta^d \alpha_k = \Theta(B)V_k$$

onde V_k é ruído branco Gaussiano de média nula.

A equação (2.46) se torna então:

$$\Phi(B)\Phi(B^s)\Delta^d Y_k = \Theta(B)\Theta(B^s)V_k \quad (2.47)$$

Este modelo é chamado multiplicativo e é gerado por um *ARIMA*(p, d, q) e um *ARMA*(P, Q) e representa-se $(p, d, q) \times (P, Q)$ [20].

- Modelo Sazonal *ARIMA*

Se a série temporal sazonal Z_k se tornar estacionária ao se tomar um número finito de diferenças, D , então (2.46) pode ser escrita como:

$$\Phi(B)\Phi(B^s)\Delta^d Z_k = \Theta(B)\Theta(B^s)V_k \quad (2.48)$$

Desta maneira, a equação (2.47) se torna agora:

$$\Phi(B)\Phi(B^s)\Delta^d \Delta^D Y_k = \Theta(B)\Theta(B^s)V_k \quad (2.49)$$

Este modelo é chamado multiplicativo, gerado por dois processos *ARIMA* e representado por $(p, d, q) \times (P, D, Q)$ [20].

- Modelos *AR* Sazonais

Uma outra maneira de modelar S_k ⁸ para os modelos dinâmicos é por meio da consideração que a sazonalidade é um fenômeno periódico, isto é, $S_k \simeq S_{k+pL}$, para algum inteiro positivo L (tamanho do período sazonal) e para todo inteiro p . A partir desta consideração, podem-se formular modelos em que $\sum_{i=0}^{L-1} = 0$ (e também outras variantes que tem coeficientes inteiros, mas com diferentes valores para possibilitar a modelagem de séries crescentes ou decrescentes, etc.). Devido a terem sua saída futura em função de suas saídas passadas, os modelos *AR* Sazonais são modelos dinâmicos, porém, não obedecem às condições gerais de estacionariedade ou mesmo de possibilidade de inversão de um modelo *AR* e além disso têm seus coeficientes definidos a priori. Como exemplos, podem-se citar:

- $\sum_{i=0}^{L-1} S_{k-i} = V_k$;
- $\Delta^{s \times m} S_k = V_k$, onde $\Delta^s = S_k - S_{k-s}$;
- $(1 - \sum_{i=1}^{L-1} B^i)^2 S_k = V_k$, onde $B^i S_k = S_{k-i}$. Este caso, devido a prover amostras do sinal em diferentes ciclos consecutivos, tem condições de modelar sinais periódicos crescentes ou decrescentes [19].

2.4.2 Considerações Gerais sobre Aspectos Importantes na Aplicação dos Modelos Dinâmicos (Particularmente os Entrada×Saída)

Serão abordados aqui diversos aspectos fundamentais relacionados aos modelos de previsão. A ênfase dada aos modelos dinâmicos se deve a que esses modelos possuem algumas vantagens em relação aos estáticos⁹, porém, requerem uma maior elaboração no tratamento estatístico. Este item, especificamente, abordará mais os aspectos relacionados com os modelos Entrada×Saída e o item seguinte e o Capítulo 3 abordarão os modelos de Espaço de Estados e complementarão com conceitos específicos, quando for o caso.

⁸No presente caso, a suposição é que a série temporal seja constituída apenas por sazonalidade, ou seja, o valor observado $Y_k = S_k$.

⁹Geralmente com mesma quantidade de observações possibilitam previsões mais precisas. Vide outras conclusões no final do capítulo.

- Séries Não-Estacionárias

O tratamento dos dados nos modelos Entrada×Saída requerem que a série seja estacionária (vide uma forma de tratamento de não-estacionariedade em [15]).

A utilização de uma forma sistemática de transformação sobre os dados a fim de se obter estacionariedade é proposta por Box-Cox, que a partir da média aritmética (\bar{Y}) e da amplitude (A) dos dados indica a transformação a ser aplicada [7].

A transformação a ser utilizada é:

$$Y_k^\lambda = \begin{cases} \frac{Y_k^\lambda - c}{\lambda}, & \lambda \neq 0 \\ \log Y_k, & \lambda = 0 \end{cases}$$

onde o parâmetro λ exprime a forma de transformação recomendada para os dados. Por exemplo, se $\lambda = \frac{1}{2}$, utiliza-se $Z_k = \sqrt{Y_k}$. O parâmetro c corresponde a um nível constante que pode ser subtraído da série. A amplitude A é calculada como:

$$w = \text{máx } Y_k - \text{mín } Y_k, \quad k = 1, \dots, N$$

Entrando com o par (w, \bar{Y}) na Figura 2.3 de [7], encontra-se um valor para o parâmetro λ . Após os dados transformados, se ainda não foi obtida estacionariedade ¹⁰ procede-se à diferenciação até que esta seja alcançada ¹¹. Para o modelo *ARIMA*, o número de vezes em que a série foi diferenciada é o valor de d ¹². Achar o número de diferenças para a série se tornar estacionária, d , é o primeiro passo do processo de identificação. Após a série se tornar estacionária, o próximo passo é a determinação da ordem do modelo *ARMA* resultante.

- Determinação da Autocorrelação Amostral

A função de autocorrelação amostral desempenha um papel primordial no estudo das propriedades dos modelos dinâmicos, e é calculada como:

$$\rho_j = \frac{\nu_j}{\nu_0}$$

onde

$$\nu_j = \frac{1}{N-1} \sum_{k=1}^{N-j} (Y_k - \bar{Y})(Y_{k+j} - \bar{Y})$$

e

$$\bar{Y} = \frac{1}{N} \sum_{k=1}^N Y_k \quad (2.50)$$

As quantidades ν_j são as autocovariâncias do processo Y_k .

¹⁰Para se testar se a série é estacionária verifica-se a função de autocorrelação amostral dos dados, que deve decrescer rapidamente para zero conforme aumenta o *lag* das amostras [7].

¹¹Ao se tomar d observações para que a série alcance estacionariedade, restam $n - d = N$ observações de um total de n observações.

¹²Se a série for *ARIMA* sazonal, com dados mensais, então é preciso diferenciá-la para períodos de 12 meses e depois diferenciá-la mês a mês; o primeiro número de diferenças é o valor de D e o do segundo é d .

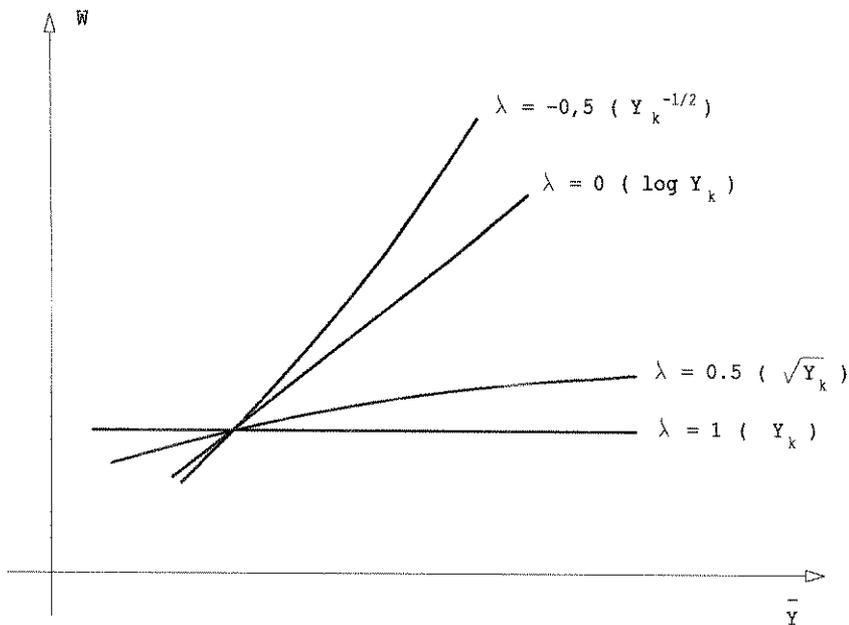


Figura 2.3: Transformação de Box-Cox

• Identificação (determinação da ordem do modelo) ¹³

O próximo passo é a determinação ¹⁴ da ordem do modelo Entrada×Saída. Para esse fim existem vários procedimentos. Entre eles podem-se citar:

a) Função de Autocorrelação e Autocorrelação Parcial

Sejam a autocorrelação, ρ_j , e a autocorrelação parcial, $\hat{\phi}_{jj}$ — a autocorrelação parcial é obtida a partir da matriz de funções de autocorrelação e corresponde à estimação do último coeficiente de um modelo $AR(p)$ [7]: Se $\hat{\phi}_{jj} \sim \mathcal{N}(0, \frac{1}{N})$, $j > p$, o modelo identificado é do tipo $AR(p)$. Se $\rho_j \sim \mathcal{N}(0, \frac{1}{N}(1 + 2\sum_{i=1}^q \rho_i^2))$, $j > q$, o modelo identificado é do tipo $MA(q)$.

Se ao mesmo tempo os cortes forem em $AR(p)$ e em $MA(q)$, estará identificado o modelo $ARMA(p, q)$.

Esse método pode identificar modelos diferentes a cada amostragem, principalmente quando se dispõe de poucas observações, sendo este um aspecto negativo.

b) Procedimento AIC

O procedimento *Akaike Information Criterion* - AIC procura o modelo cuja ordem minimize o valor da equação [7]:

$$AIC(p, q) = \log \hat{\sigma}_V^2 + 2 \frac{(p + q)}{N} \quad (2.51)$$

¹³O conceito de identificação abordado no início do Capítulo 2 é mais amplo englobando não apenas a determinação da ordem ou da estrutura, mas também os aspectos essenciais do modelo.

¹⁴O processo de identificação supõe uma estimativa inicial dos parâmetros. No item seguinte (estimação), será abordada a questão das estimativas iniciais.

onde $\hat{\sigma}_v^2$ é a estimativa da variância dos resíduos, N é o número de observações e p e q são as ordens procuradas.

O que se faz é procurar o mínimo AIC para $0 \leq p \leq PM$ e $0 \leq q \leq QM$, onde PM e QM são valores máximos admissíveis. Um aspecto que vale ressaltar é que o AIC pode ser considerado também um instrumento de verificação da adequação do modelo porque quanto menor o seu valor, mais ajustado o modelo estará aos dados. No Capítulo 3 será mostrada a aplicação do critério AIC ao modelo de Espaço de Estados.

c) Outros Tipos de Procedimento

Alguns outros tipos de procedimento existentes na literatura são citados em seguida [7]:

- Com pequenas alterações no critério AIC , Akaike, e depois Rissanen e Schwarz sugeriram o critério BIC . Também com pequenas alterações no AIC , Hannan e Quinn sugeriram outro procedimento similar;
- Um outro critério, também sugerido por Akaike, é o FPE - *Final Prediction Error*, que minimiza uma equação que é função dos parâmetros estimados e da função de autocorrelação da série temporal;
- Gray, Kelley e McIntire apresentam um procedimento para a identificação dos modelos em função das funções de autocorrelação. Nesta mesma linha encontra-se o procedimento de Nerlove, Grether e Carvalho;
- Cleveland introduziu a análise da função de autocorrelação inversa, que é definida em termos do recíproco do espectro da série, para auxiliar no processo de identificação;
- Para se estimar a ordem de um processo $AR(p)$, existem os procedimentos: CAT - *Criterion Autoregressive Transfer Function*, de Parzen, e o procedimento de Anderson;
- Como outros procedimentos gerais, existem os de: Gray, Kelley e McIntire, o de Nerlove, Grether e Carvalho.

Testes realizados com vários métodos, por exemplo (FPE , AIC , CAT), para a identificação de modelos AR mostraram que os critérios identificam modelos diferentes dependendo da ordem verdadeira do modelo e do tamanho da amostra [7].

• Estimação dos parâmetros

Após o modelo provisoriamente identificado, o próximo passo é a estimação de seus parâmetros (neste item, mais aplicável aos modelos Entrada×Saída). Entre os procedimentos de estimacão são destacados os seguintes:

a) Procedimentos para Estimativas Iniciais dos Parâmetros

As estimativas iniciais dos parâmetros são necessárias para os procedimentos de identificação e também podem servir de valores iniciais para outros tipos de procedimentos de estimacão mais elaborados (exemplo, estimativa por Máxima Verossimilhança).

I) Processo $AR(p)$

No caso dos processos $AR(p)$, a estimativa inicial dos parâmetros é realizada pelo procedimento de Yule-Walker. Em um processo AR , tem-se:

$$\begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{p-1} \\ \rho_1 & 1 & \cdots & \rho_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p-1} & \rho_{p-2} & \cdots & 1 \end{bmatrix} \cdot \begin{bmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \\ \vdots \\ \hat{\phi}_p \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{bmatrix} \quad (2.52)$$

onde ρ_j são as autocorrelações do processo Y_k conforme (2.50). Resolvendo esse sistema, são encontrados os valores iniciais dos coeficientes $\{\hat{\phi}_j\}$. Uma maneira prática de se calcular esses coeficientes de forma recursiva, para ordens sucessivas de $AR(p)$ é a utilização do procedimento de Levinson-Durbin. Para se encontrar a estimativa inicial de σ_V^2 (a variância da perturbação), resolve-se:

$$\hat{\sigma}_V^2 = \nu_0 [1 - \hat{\phi}_1 \rho_1 - \cdots - \hat{\phi}_p \rho_p]$$

onde ν_0 é como em (2.50).

II) Processo $MA(q)$

Para se encontrar estimativas iniciais dos $\hat{\theta}_j$ para o processo $MA(q)$, resolve-se a função de autocorrelação do processo $MA(q)$ conforme [7]. Em seguida, para se achar a estimativa de σ_V^2 , resolve-se:

$$\nu_0 = \hat{\sigma}_V^2 (1 + \hat{\theta}_1^2 + \cdots + \hat{\theta}_q^2)$$

III) Processo $ARMA(p, q)$

Obtém-se a estimativa para $\{\hat{\phi}_j\}$ resolvendo:

$$\rho_j = \hat{\phi}_1 \rho_{j-1} + \cdots + \hat{\phi}_p \rho_{j-p}, \quad j = q+1, \dots, q+p$$

Depois, usando a função de autocovariância do processo $ARMA(p, q)$, (2.50), é possível se obter $\{\hat{\theta}_j\}$ e $\hat{\sigma}_V^2$.

b) Método dos Mínimos Quadrados

Este método é geral e também pode ser aplicado aos modelos de Espaço de Estados. Seja a representação linear discreta:

$$Y_k = \bar{U}^T \Upsilon + V_k \quad (2.53)$$

onde \bar{U} contém todas as entradas do sistema e Υ contém todos os parâmetros desconhecidos, que podem ser os valores de Φ , coeficientes do polinômio relativo às

perturbações V_k (de Θ), ou de ξ , coeficientes do polinômio relativo aos controles U_k , de (2.6). Para N observações, tem-se:

$$\underline{Y}(N) \triangleq \begin{bmatrix} Y_0 \\ Y_1 \\ \dots \\ Y_{N-1} \end{bmatrix} = \begin{bmatrix} \bar{U}_0^T \\ \bar{U}_1^T \\ \dots \\ \bar{U}_{N-1}^T \end{bmatrix} \Upsilon + \begin{bmatrix} V_0 \\ V_1 \\ \dots \\ V_{N-1} \end{bmatrix} \quad (2.54)$$

Se $E[V_k V_k^T] = \sigma_v^2 W$, então a solução de (2.54) será:

$$\hat{\Upsilon} = [\bar{U}^T W^{-1} \bar{U}]^{-1} \bar{U}^T W^{-1} \underline{Y}(N) \quad (2.55)$$

se \bar{U} for não-singular, onde:

$$\bar{U} = [\bar{U}_0^T \quad \bar{U}_1^T \quad \dots \quad \bar{U}_{N-1}^T]$$

Se W for a matriz identidade, (2.55) será a solução de mínimos quadrados (que já foi apresentada para a solução dos coeficientes da regressão linear múltipla (2.15)), e se W for uma outra matriz possível de ser invertida — W é utilizada sempre que as perturbações não são Gaussianas — (2.55) será a solução de mínimos quadrados generalizados (também denominado de mínimos quadrados ponderados ou estimador de Markov).

Os estimadores de mínimos quadrados e os estimadores de mínimos quadrados generalizados são da classe *BLUE - best linear unbiased estimators*, que significa serem não-polarizados e de variância mínima.

c) Método da Máxima Verossimilhança

Neste caso, abordará o modelo Entrada×Saída. No Capítulo 3 abordará o modelo de Espaço de Estados. Dado um modelo *ARIMA*(p, d, q), seja $\Xi(\Phi, \Theta, \sigma_v^2)$ seu conjunto de $p+q+1$ parâmetros ($p+q+2$ se a série tiver um nível constante, μ , desconhecido), onde (maiores detalhes podem ser encontrados em /citemor):

$$\Phi = (\phi_1, \dots, \phi_p)$$

$$\Theta = (\theta_1, \dots, \theta_q)$$

e σ_v^2 é a variância do ruído. A função de verossimilhança será:

$$L = (\Xi | Y_1, \dots, Y_N)$$

E os estimadores de máxima verossimilhança de Ξ serão os valores que maximizam $\log L$.

Sob a suposição de que o processo $V_k \sim \mathcal{N}(0, \sigma_v^2)$, os estimadores de máxima verossimilhança serão equivalentes aos estimadores de mínimos quadrados [7, 34].

Seja $Z_k = \Delta^d Y_k$, resultando o modelo $ARMA(p, q)$ estacionário e possível de ser invertido. Então pode-se escrever:

$$V_k = Z_k - \phi_1 Z_{k-1} - \dots - \phi_p Z_{k-p} + \theta_1 V_{k-1} + \dots + \theta_q V_{k-p} \quad (2.56)$$

A função densidade conjunta dos V_k , $k = 1, \dots, N$ será:

$$L(V_1, \dots, V_N) = (2\pi)^{-\frac{n}{2}} (\sigma_V)^{-n} e^{-\sum_{k=1}^N \frac{V_k^2}{2\sigma_V^2}} \quad (2.57)$$

Também pode-se escrever $\log L$, a partir de (2.56) e (2.57), como:

$$\log L = -N \log \sigma_V - \frac{S(\eta)}{2\sigma_V^2} \quad (2.58)$$

onde $\eta = (\Phi, \Theta)$. Maximizar $\log L$ equivale a minimizar $S(\eta)$.

Existem dois procedimentos de máxima verossimilhança:

I) Procedimento Condicional

Neste caso, usam-se valores iniciais considerados adequados, fixando-se p valores de Z_k , chamados Z^* , e q valores de V_k , chamados V^* , e minimiza-se [7]:

$$S(\eta) = \sum_{k=1}^N V_k^2(\eta | Z, Z^*, V^*) \quad (2.59)$$

onde Z corresponde às demais $N - p$ observações disponíveis.

II) Procedimento Incondicional

$S(\eta)$ será:

$$S(\eta) = \sum_{k=-\infty}^N (E[V_k | \eta, Z])^2 \quad (2.60)$$

Neste caso Z corresponde às N observações. O procedimento incondicional requer que se use a forma *backward* do filtro para se estimarem valores passados:

$$\Phi(F)Z_k = \Theta(F)e_k$$

onde F é o operador translação para o futuro, $FY_k = Y_{k+1}$, e e_k é um ruído branco Gaussiano com mesma variância que V_k .

– Como considerações adicionais:

- a) Ao se derivar (2.58) em relação a cada parâmetro, obtém-se um sistema de equações não-lineares nos parâmetros. Uma outra forma de se calcular esses parâmetros é linearizar (2.58) em termos da série de Taylor ao redor de algum valor inicial η_0 ;

b) Para poucos parâmetros, existe um procedimento simples (chamado soma dos quadrados), em que se fixam diferentes valores para os parâmetros e se procura minimizar $S(\eta)$;

c) Existem procedimentos que calculam recursivamente os parâmetros, usando alguma técnica de cálculo numérico.

- Verificação (validação do modelo)

Após o modelo identificado e com os parâmetros estimados, é preciso verificar sua adequação. Entre os procedimentos para esse fim destacam-se:

a) Teste de Autocorrelação Residual e Correlação Cruzada

Estimados Θ e Φ , as quantidades r_k são os resíduos estimados, ou seja, a diferença entre os valores previstos na projeção a um passo e os valores observados. Esses resíduos devem ter um padrão branco (ser não-correlacionados) e sua função de autocorrelação pode ser calculada por (2.50) com $\bar{Y} = 0$ e com $Y_k = r_k$. Considerando ρ_j assintoticamente normal, a comparação de ρ_j com os limites $\pm \frac{2}{\sqrt{N}}$ fornece uma indicação adequada do modelo de erro. Para modelos com ordens pequenas (um ou dois), este limite pode ser bem menor que $\frac{2}{\sqrt{N}}$ [7].

Para testar a autocorrelação, também existe o teste de Box-Pierce, que, apesar de não detectar o padrão branco, pode indicar se os valores autocorrelacionados são muito altos [7].

Se o modelo é adequado, r_k e Y_{k-j} devem ser independentes para $j \geq 1$, logo a covariância entre r_k e Y_{k-j} deve ser nula. Para verificar isso, pode-se estudar a função de correlação cruzada:

$$\hat{c}c_j = \frac{\sum_{k=j+1}^N r_k (Y_{k-j} - \bar{Y})}{[\sum_{k=1}^N r_k^2 \sum_{k=1}^N (Y_k - \bar{Y})^2]^{1/2}}, \quad j = 1, 2, \dots$$

$\hat{c}c$ deve estar dentro dos limites $\pm \frac{2}{\sqrt{N}}$, exceto para j pequeno.

A inspeção dos $\hat{c}c_j$ pode sugerir a inclusão de novos termos auto-regressivos no modelo e a inspeção dos ρ_j , quando os $\hat{c}c_j$ não se mostrarem mais significantes, pode sugerir a inclusão de novos termos de médias móveis no modelo.

- Previsão para os modelos Entrada×Saída

Como exemplo de projeção n passos à frente com um modelo *ARIMA*, pode-se ter:

$$Y_{k+n}^* = \hat{\xi}_1[Y_{k+n-1}] + \dots + \hat{\xi}_{p+d}[Y_{k+n-p-d}] - \hat{\theta}_1[V_{k+n-1}] - \dots - \hat{\theta}_q[V_{k+n-q}] + [V_{k+n}], \quad n \geq 1 \quad (2.61)$$

onde $\hat{\xi}_i$ são os coeficientes estimados de $\Omega(B)$, $\hat{\theta}_i$ são os coeficientes estimados de $\Theta(B)$ e

$$[Y_{k+n}] = E[Y_{k+n} | Y_k, Y_{k-1}, \dots]$$

$$[Y_{k+j}] = Y_{k+j}^*, \quad j > 0$$

$$\begin{aligned} [Y_{k+j}] &= Y_{k+j}, \quad j \leq 0 \\ [V_{k+j}] &= 0, \quad j > 0 \\ [V_{k+j}] &= e_{k+j}, \quad j \leq 0 \end{aligned}$$

onde e_{k+j} é o resíduo entre a projeção a um passo e o valor observado.

Pode-se mostrar que a variância do erro de projeção a n passos é dada por [7]:

$$VAR[e_{k+n}] = (1 + \varsigma_1^2 + \dots + \varsigma_{n-1}^2)\sigma_V^2 \quad (2.62)$$

onde

$$e_{k+n} = Y_{k+n} - Y_{k+n}^*$$

e $\{\varsigma_i\}$ são os $n - 1$ primeiros coeficientes de um modelo MA de ordem infinita, equivalente ao modelo *ARIMA* [7].

Se $V_k \sim \mathcal{N}(0, \sigma_V^2)$ para Y_k, Y_{k-1}, \dots , então Y_{k+n} será $\mathcal{N}(Y_{k+n}^*, VAR[e_{k+n}])$. Um intervalo de confiança com coeficiente ν é dado por:

$$Y_{k+n}^* \pm z_\nu (VAR[e_{k+n}])^{\frac{1}{2}} \quad (2.63)$$

onde z_ν é obtido da distribuição normal padrão.

2.4.3 Modelos de Espaço de Estados

Todos os modelos, estáticos ou dinâmicos, possuem as fases de identificação (no sentido de se conhecer a estrutura do modelo), estimação e verificação, mesmo que elas estejam implícitas. Nos modelos dinâmicos, em geral, essas fases são tratadas de maneira mais formal. Vários conceitos que serão abordados nesta subseção já foram apresentados na subseção anterior e outros serão apresentados em detalhes no Capítulo 3.

- Identificação (determinação da estrutura do modelo)

Quando, além de tendência, os dados de tráfego apresentam também um padrão sazonal, pode-se usar um modelo de Espaço de Estados formado pela superposição de um modelo de crescimento linear (tendência) com um modelo sazonal (No Capítulo 3 será mostrado um modelo aditivo formado pela superposição de vários efeitos).

O modelo de Espaço de Estados, para aplicações no tráfego telefônico, permite facilidades no processo de identificação, já que é possível associar significado físico às variáveis componentes do vetor de estado. Como exemplo, uma atribuição comum é associar uma variável de estado ao nível da carga de tráfego e outra ao seu incremento de crescimento [12, 26]. Ainda com relação à identificação, vários estudos mostram que o modelo de crescimento linear (tendência) representa adequadamente a maior parte dos casos de projeção de tráfego com base de dados mensal ou anual [12, 13, 22]. Para os modelos de Espaço de Estados, a conclusão da fase de identificação (determinação da ordem do modelo) se dá em conjunto com a fase de estimação dos parâmetros, para escolher um entre alguns modelos previamente identificados e nos quais se utilizam os parâmetros estimados. No Capítulo 3 utiliza-se o critério *AIC* para definir o melhor modelo entre os vários pré-selecionados.

- Estimação

Devido à maior facilidade de associação das propriedades físicas, que caracterizam a demanda de tráfego, com a representação por Espaço de Estados, é possível se conhecer a dimensão da matriz de transição de estado, Φ . Desconhece-se, porém, os valores de seus componentes. Como há um problema de estimação conjunta de estado, X_k , e matriz de transição, Φ , caracteriza-se um problema de filtragem não-linear, cuja solução será proposta no Capítulo 3 através do uso de uma Técnica de Relaxação que maximiza iterativamente a função de verossimilhança.

O modelo de Espaço de Estados necessita de um estimador para o estado e uma solução usual, quando a matriz de transição é conhecida, é a utilização do Filtro de Kalman. Sob as condições formuladas na Secção 2.2, determina-se o valor de \bar{X}_k (filtragem), utilizando-se:

$$\bar{X}_k = X_k^* + K_k(Y_k - HX_k^*)$$

O ganho do filtro é calculado de tal forma que:

$$\bar{P}_k = E[(\bar{X}_k - X_k)(\bar{X}_k - X_k)^T]$$

seja mínimo, onde

- \bar{X}_k é a estimativa do vetor de estado no instante k (filtragem);
- X_k^* é o valor do estado estimado um passo à frente a partir do instante $k - 1$ (predição);
- K_k é a matriz do ganho do Filtro de Kalman e pode ser calculada recursivamente;
- \bar{P}_k é a covariância do erro de estimação;
- X_k é o vetor de variáveis de estado verdadeiras (não-observável);

K_k é função da covariância do erro de estimação, \bar{P}_k , da covariância do erro de predição, P_k^* , da covariância do erro de modelagem, Q_k , e da covariância do erro de observação, R_k (supondo Φ e H invariantes com o tempo). $Q_k = E[W_k W_k^T]$ e $R_k = E[V_k V_k^T]$ são matrizes definidas não-negativas para que o filtro seja estável (Apêndice C). Se as matrizes de transição de estado, Φ , e de observação, H , forem invariantes no tempo, o valor de regime do ganho do filtro será constante e corresponderá à solução de estado estacionário da equação de Ricatti discreta (Capítulo 4 e Apêndice D).

A covariância correspondente a X_k^* é $P_k^* = E[(X_k^* - X_k)(X_k^* - X_k)^T]$, a covariância do erro de predição.

- Verificação (validação do modelo)

A verificação do funcionamento correto do algoritmo pode ser feita pela observação da seqüência de inovações (ou resíduos), $r_k = Y_k - HX_k^*$. Se a seqüência for decorrelacionada no tempo (ruído branco), será uma indicação do correto funcionamento do filtro. Os testes de Autocorrelação Residual e Correlação Cruzada, apresentados para os modelos Entrada×Saída, também podem ser aplicados aos resíduos dos modelos de Espaço de Estados. O Capítulo 3 apresenta vários critérios utilizados para verificação de desempenho.

- Previsão

O modelo de Espaço de Estados permite também facilidades para a projeção do estado n passos à frente, bastando fazer:

$$X_{k+n}^* = \Phi^n \bar{X}_k \quad (2.64)$$

onde X_{k+n}^* é o estimador do estado, dadas as observações Y_1, Y_2, \dots, Y_k e \bar{X}_k é o valor do estado estimado no instante k . Com a suposição de que os ruídos são Gaussianos, pode-se construir um intervalo de confiança com coeficiente ν para a previsão a n passos [18]:

$$Y_{k+n}^* \pm z_\nu (V_{Y_{k+n}^*})^{\frac{1}{2}} \quad (2.65)$$

onde

$$V_{Y_{k+n}^*} = R_k + H M_n H^T, \quad M_n = \Phi^n \bar{P}_k (\Phi^T)^n + \sum_{j=0}^{n-1} \Phi^j Q_k (\Phi^T)^j$$

e z_ν é obtido da distribuição normal padrão.

2.5 Vantagens e Desvantagens dos Modelos Apresentados

A seguir são listadas algumas vantagens e desvantagens dos modelos citados:

- Os modelos de regressão independem de observações faltantes e não necessitam trabalhar com séries temporais equiespaçadas. Os modelos de Espaço de Estados também trabalham com observações faltantes (sem a necessidade de qualquer tratamento aos dados observados) [21];
- Os modelos de regressão que trabalham com variáveis explicativas, como exemplo, os modelos econométricos, ainda possuem as seguintes características adicionais:
 - Ponderam as observações;
 - Incorporam os efeitos de observações atrasadas;
 - Podem eliminar ou atenuar as observações aberrantes (*outliers*);
 - Podem incorporar mudanças nos fatores que afetam suas variáveis;
 - Além disso, a fonte de erros de previsão é mais fácil de ser encontrada devido à identificação explícita de sua estrutura, o que não ocorre com os modelos de ajuste de funções.

Os modelos dinâmicos multivariáveis, que trabalham com diversas variáveis sob observação, e portanto com diversas séries correlacionadas, também têm condições de identificar fontes de erro estruturais devido às relações entre as variáveis observadas;

- Os modelos estáticos, geralmente, requerem menos observações que as necessárias para outros modelos de séries temporais (exemplo, modelo Entrada \times Saída que requer várias observações para o processo de determinação de sua ordem) [10, 7, 12];

- Em geral, o grau de especialização do pessoal para aplicação e análise dos resultados dos modelos estáticos pode ser menor que o grau de especialização para aplicação dos modelos dinâmicos;
- Os procedimentos mais simples, caso de vários tipos de modelos estáticos e alguns modelos dinâmicos simplificados, são aplicáveis mesmo quando se dispõe de um pequeno número de observações;
- Para os modelos estáticos que utilizam polinômios para ajustar a tendência, ou harmônicos para ajustar a sazonalidade, é difícil determinar-se o número de termos polinomiais para ajustar a tendência ou a quantos harmônicos ajustar a sazonalidade. Na prática, supõe-se vários polinômios ou harmônicos cabíveis e escolhe-se o de menor erro de projeção;
- É difícil a determinação da(s) constante(s) de suavização. Para séries que apresentam maior complexidade, a constante de suavização deveria ser grande o suficiente para permitir que um maior número de observações participe do alisamento. Por outro lado, para melhor adaptação à variação nas observações, é preciso dar peso maior às medições mais recentes. As duas condições são incompatíveis e requerem que se estabeleça um compromisso entre elas;
- No caso dos procedimentos mais complexos, existem dificuldades ou impossibilidades de se estudarem as propriedades estatísticas dos valores projetados e, conseqüentemente, de se determinarem intervalos de confiança para as previsões;
- Existem modelos que são intrinsecamente recursivos, como exemplo, o modelo do Espaço de Estados com o Filtro de Kalman, o que permite o processamento seqüencial das observações. Outros tipos de modelos podem ser tornados recursivos, como exemplo, o Filtro de Wiener;
- Os modelos estáticos requerem um maior número de parâmetros para atingir a mesma precisão dos modelos dinâmicos, pelo fato de não trabalharem com suas saídas passadas. Segundo Morettin, os modelos de regressão, em geral, se ajustam melhor aos dados, mas os modelos dinâmicos produzem melhores previsões [7];
- Embora as etapas de identificação (determinação da ordem), estimação e verificação estejam presentes qualquer que seja o modelo (estático ou dinâmico), nos modelos dinâmicos, em geral, há mais formalização teórica disponível, o que permite maior confiança na aplicação e nos resultados obtidos pelo modelo escolhido;
- Os modelos dinâmicos descrevem séries temporais onde a dinâmica do processo é incorporada aos procedimentos, o que permite que, para alcançar precisão semelhante, os modelos estáticos demandem a utilização de um maior número de parâmetros;
- Geralmente, os modelos dinâmicos são mais complexos e requerem do usuário um conhecimento formal mais sofisticado. Além disso, são freqüentes os tipos em que o usuário tem necessidade de programas sofisticados (*softwares*) de computador, nem sempre disponíveis;

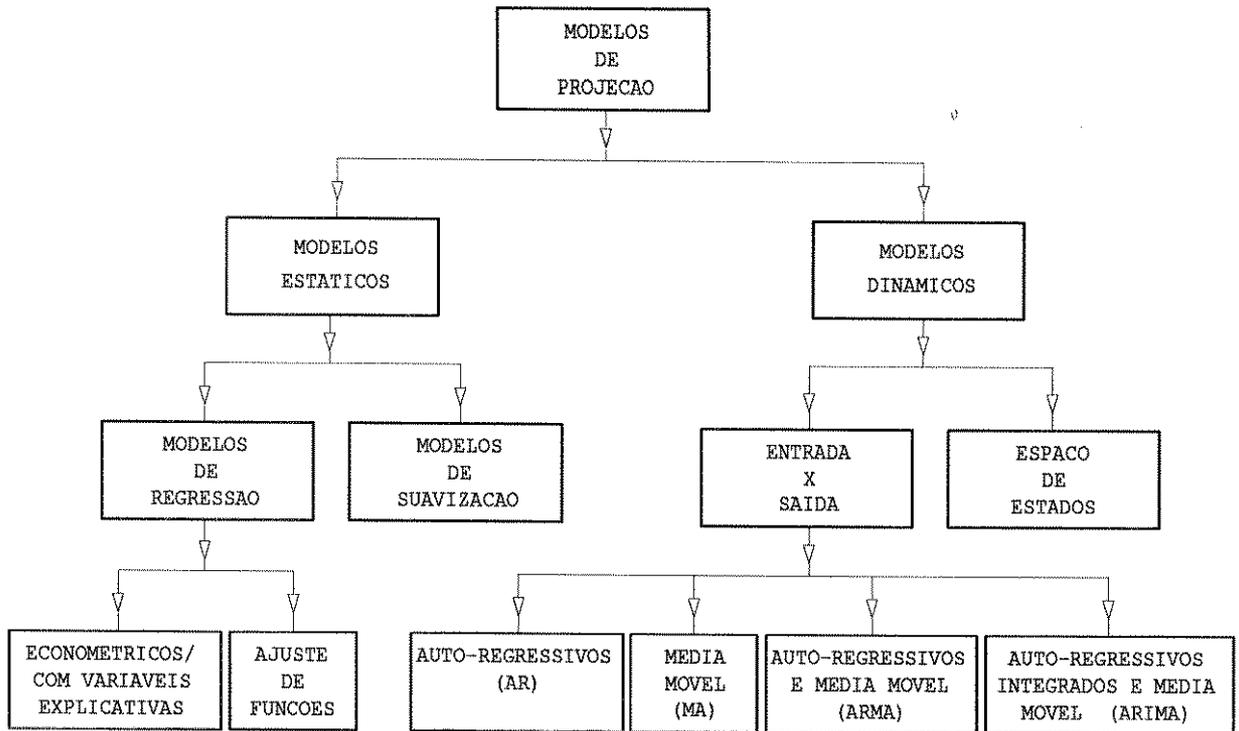


Figura 2.4: Modelos de Projeção de Tráfego

- A classe dos modelos estáticos possui, em geral, modelos mais simples, como os apenas de suavização, os de ajuste de funções ou os de regressão simples; e nos casos de se possuir poucas observações, geralmente os modelos mais simples fornecem melhores resultados.

2.6 Comentários Finais e Conclusões

Todos os modelos apresentados têm suas vantagens e desvantagens dependendo da condição em que são utilizados. A definição final sobre o modelo a ser empregado depende dos requisitos de desempenho, tais como precisão, estabilidade das projeções, facilidades de computação, etc. A Figura 2.4 mostra a classificação para os modelos de projeção de tráfego apresentados no decorrer do presente capítulo com a finalidade de ilustrar os comentários finais.

Os modelos de suavização¹⁵ podem modelar a série temporal como modelos independentes; no problema de teletráfego, porém, se prestam mais ao papel de coadjuvantes dos modelos de regressão ou dos modelos dinâmicos e têm como objetivo principal a redução da variância dos dados amostrados. Por outro lado, sempre é possível tratar os resíduos de um modelo de regressão por meio de um modelo dinâmico. Pela conceituação deste tra-

¹⁵No Capítulo 3 será apresentado um modelo de suavização dinâmico para o modelo de Espaço de Estados.

balho, o modelo completo se tornará dinâmico (como exemplo, os resíduos de um modelo econométrico que podem ser tratados por meio de um modelo auto-regressivo)¹⁶.

Algumas considerações que devem orientar a escolha do modelo são:

- Um modelo mais complexo tende a apresentar maior precisão. Porém, este aumento de complexidade acarreta um processamento adicional devido ao aumento das covariâncias envolvidas, o que implica em adicionar outras fontes de erro ao modelo. Os testes que serão apresentados no Capítulo 3 indicam que modelos mais complexos requerem um número maior de observações que os modelos mais simples. Modelos mais complexos são mais sensíveis à utilização de parâmetros desconhecidos [24, 12, 45];
- Quando o propósito do modelo é a estabilidade dos fatores que influenciam a variável em questão (fatores sujeitos a mudanças), ou quando a causalidade (relação de causa e efeito entre as variáveis) é primordial, os modelos econométricos (ou que trabalham com variáveis explicativas) são indicados porque podem manipular mudanças estruturais. Neste caso, os modelos que trabalham apenas com a série histórica dos dados desvinculada da estrutura que os gerou, não seriam adequados (com exceção, evidentemente, dos modelos multivariáveis em que as várias variáveis sob observação são correlacionadas);
- Quando existem poucos dados (ou não são confiáveis), os modelos Entrada×Saída não são adequados porque requerem entre 60 a 100 observações para a fase de determinação da estrutura [22, 14]. Nesse caso, um modelo que leve em conta a estrutura do sistema é mais eficaz porque reduz o número de observações necessárias para a fase de identificação;
- Quando a estrutura que gera os dados é difícil de ser identificada, então não há outra escolha senão a de empregar um modelo que é baseado nos dados históricos da variável de interesse. Sob esse aspecto, os modelos dinâmicos, pelo fato de trabalharem com valores passados das observações e das perturbações, geralmente requerem um menor número de parâmetros a serem estimados para alcançar a mesma precisão.

Do ponto de vista de desempenho, os modelos de Espaço de Estados e Entrada × Saída são essencialmente equivalentes, e tendo-se a estrutura de um é possível se obter a estrutura equivalente do outro [22].

Nos próximos capítulos, a continuidade do trabalho será centrada na representação Espaço de Estados, porque:

- A identificação dos modelos de Espaço de Estados pode ser baseada também no sentido físico do problema e não apenas no conjunto de observações disponíveis (exemplo: um modelo com dois estados, um relacionado ao nível da carga de tráfego e outro associado ao seu incremento de crescimento), o que permite que um menor número de observações possa ser usado na fase de identificação;

¹⁶Com relação aos modelos extrapolativos — que podem ser escritos como uma função, da mesma forma que os modelos de regressão —, e aos modelos que trabalham com a matriz de tráfego — que dependem sempre de algum dos outros modelos de projeção citados no capítulo (esses modelos, geralmente, são os procedimentos extrapolativos) —, por simplicidade foram incluídos junto aos modelos de regressão na Figura 2.4.

- Versatilidade em se formular intuitivamente modelos aditivos (ver Capítulo 3) envolvendo tendência, sazonalidade e componentes irregulares. A matriz de transição de estado pode ser fixada para representar tendência e sazonalidade e ter apenas uma pequena parte a ser estimada (no caso do modelo aditivo, a matriz de transição de estado completa é particionada em várias submatrizes que, em sua maioria, podem ser fixadas);
- No caso dos modelos Entrada×Saída, a cada conjunto de dados de rotas é possível identificar um modelo diferente (exemplo: $AR(1)$, $ARIMA(1,1,1)$, $AR(2)$, etc.) e só então se passa a estimação de seus parâmetros. Os modelos de Espaço de Estados, mesmo podendo ter matrizes de transição de estado diferentes para cada rota, mantêm a ordem das matrizes e o problema, para um conjunto grande de rotas (por exemplo, as que só possuem tendência, as que possuem tendência e sazonalidade, etc.) se torna apenas de estimação dos novos elementos da matriz de transição de estados e dos demais parâmetros do modelo;
- Os modelos Entrada×Saída requerem um certo tamanho (número de observações) para a etapa de identificação (determinação da estrutura), mas os modelos de Espaço de Estados podem utilizar qualquer tamanho;
- Os modelos de Espaço de Estados podem utilizar informações adicionais além dos dados históricos utilizados;
- Os modelos Entrada×Saída requerem que o processo estocástico que gerou os dados seja estacionário (ou tornado estacionário, vide técnicas do Capítulo 2). Os modelos de Espaço de Estados trabalham com qualquer processo gerador dos dados;
- Com o modelo de Espaço de Estados, é simples fazer a projeção n passos à frente, bastando fazer $Y_{k+n}^* = H\Phi^n \bar{X}_k$;
- Existem modelos de Espaço de Estados multivariáveis que podem trabalhar com séries com observações faltantes (por exemplo, bastando mudar valores da matriz de observação H), sem a necessidade de suprir as faltas de dados com interpolações, etc., o que não ocorre com os modelos Entrada×Saída [30];
- De acordo com Cole [43], o modelo de Espaço de Estados com o Filtro de Kalman representa um papel fundamental na busca de um arcabouço de metodologia unificada para os modelos de previsão. Todos os modelos apresentados podem ser representados na forma de Espaço de Estados;
- O modelo de Espaço de Estados com o estado estimado pelo Filtro de Kalman, além de ser um preditor do estado futuro, ainda tem intrinsecamente incorporado ao seu funcionamento um estimador (o Filtro de Kalman), o que possibilita que se adapte às variações no padrão das observações, ou seja, possui parâmetros variantes que são as covariâncias do erro de predição e de estimação, recalculadas a cada nova observação [34]. No entanto, esse grau de adaptação é relativo, já que o próprio Filtro de Kalman também requer que seus parâmetros sejam corretamente estimados¹⁷;

¹⁷Alguns autores [34, 59] denominam as covariâncias do erro de estimação e do erro de predição, do modelo de Espaço de Estados com o estado estimado pelo Filtro de Kalman, de parâmetros variantes que

- Os modelos de Espaço de Estados simplificados (portanto menos precisos, mas também menos sensíveis a mudanças de dados ou de parâmetros) são capazes de realizar projeções a mais longo prazo (diferentemente do modelo Entrada× Saída que é mais adequado para o curto prazo), sem cometer erros grosseiros;
- A tendência em se utilizar o modelo de Espaço de Estados é mostrada na maioria dos trabalhos recentes de projeção de rotas de tráfego [61].

são recalculados a cada nova observação. Neste texto, os parâmetros referidos são: As covariâncias do erro de observação e de modelagem, a matriz de transição de estado, a estimação inicial do estado e a covariância do erro de estimação inicial (Capítulo 3).

Capítulo 3

Representação por Espaço de Estados com o Filtro de Kalman

As abordagens de projeção baseadas no modelo de Espaço de Estados são revistas de maneira crítica, sendo exploradas suas características. Como resultado deste trabalho, este capítulo propõe um procedimento de projeção de tráfego formado pela adição de características desejáveis dos vários trabalhos pesquisados. Especial atenção é dada aos modelos construídos por decomposição aditiva da série temporal com a consideração de aspectos relacionados à identificação física do modelo, devido à sua forma intuitiva. O estado do modelo é estimado pelo Filtro de Kalman e a necessidade de se estimarem conjuntamente o estado e a matriz de transição de estado caracteriza um problema de filtragem não-linear. Neste capítulo utiliza-se como solução para este problema uma técnica de relaxação baseada na maximização da função de verossimilhança, denominada *EM* (*Expectation-Maximization*). Observa-se que um algoritmo mais simples (e robusto) tem melhor desempenho, com relação ao erro médio quadrático entre o valor da previsão a um passo e o valor observado, do que o algoritmo *EM*. Abordam-se, assim, as questões de otimalidade \times sensibilidade. Dependendo do tipo de aplicação, recomenda-se a utilização de um algoritmo robusto ou de um algoritmo adaptativo — ambos incorporando a estimação da matriz de transição de estado —, cujos desenvolvimentos serão detalhados nos Capítulos 4 e 5.

3.1 Introdução

As abordagens de projeção baseadas no modelo de Espaço de Estados são revistas de maneira crítica, sendo estudadas suas características desejáveis. Como resultado desse trabalho, neste capítulo propõe-se uma forma original de procedimento para as fases de **identificação**, **estimação** e **verificação**. A proposta baseia-se na seleção dos aspectos relevantes para os algoritmos de projeção de tráfego. Desta forma, características desejáveis de vários trabalhos são combinadas para formar uma nova alternativa. As propostas para as fases de **identificação**, detalhada nas Seções 3.2 e 3.3, **estimação**, abordada nas Seções 3.4 e 3.5, e **verificação**, comentada na Seção 3.6, são descritas como se segue:

1) Identificação (estrutura do modelo)

As propostas de Kitagawa, Kitagawa e Gersch, Shumway e Stoffer, Shumway [50, 19, 20, 21], baseadas na decomposição aditiva da série temporal, são preferidas porque supõem

padrões gerais e possuem formulação independente. Além disso, consideraram-se os modelos específicos baseados na interpretação física das variáveis de estado (e.g., o modelo com duas variáveis de estado, representando nível e incremento de crescimento) [12, 26].

O interesse do presente trabalho é a obtenção de projeções n passos à frente, $Y_{k+n}^* = H\Phi^n \bar{X}_k$. Portanto, também se propõe o modelo representando tendência com $H = \begin{bmatrix} \gamma & 1 - \gamma \end{bmatrix}$, onde γ é um parâmetro variando de zero a um (Seção 3.3).

2) Estimação

Para a estimação do estado do modelo de Espaço de Estados adota-se, como usual, o Filtro de Kalman. Todavia, a utilização do Filtro requer que as covariâncias do ruído de observação (ou covariância do erro de observação), R_k , e do ruído do sistema (ou do erro de modelagem), Q_k , sejam conhecidas, o que raramente ocorre na prática. Além disso, quando a matriz de transição de estado, Φ_k , é desconhecida, torna-se necessário estimá-la. A estimação simultânea do estado e da matriz Φ_k caracteriza um problema de filtragem não-linear. Devido ao caráter recursivo do filtro, há também um problema de iniciação que é crítico quando a quantidade de observações é pequena, ou seja, devem ser estimados o valor do estado inicial, cujo estimador é \bar{X}_0 , e a respectiva covariância do erro de estimação inicial, \bar{P}_0 .

Para resolver o problema dos parâmetros desconhecidos, incluindo o problema de filtragem não-linear, existem os algoritmos para Filtragem Não-linear [24], e o método de Relaxação baseado na maximização da função de verossimilhança, de Dempster et al [31].

Para projeções do tráfego telefônico, os trabalhos de Tomé e Cunha, e de Chemouil e Garnier utilizam o método de relaxação baseado na maximização da função de verossimilhança para estimarem os parâmetros desconhecidos [13, 17]¹. A técnica de relaxação *EM* de Dempster et al [31, 20, 21] será utilizada neste trabalho e produz resultados dentro das especificações desejadas [45].

O algoritmo *EM* utiliza a técnica da máxima verossimilhança para se estimarem os parâmetros desconhecidos, e as equações do Filtro de Kalman e do Suavizador de Kalman (*Kalman Smoother*) para se estimar o estado. O algoritmo *EM* não permite a estimação simultânea de \bar{X}_0 , e da respectiva imprecisão inicial, \bar{P}_0 , sendo necessário que se fixe um dos dois. Neste trabalho propõe-se um método alternativo para a estimação conjunta de \bar{X}_0 e \bar{P}_0 . A equação de estimação de \bar{P}_0 será detalhada no Capítulo 4 pelo fato de ser empregada no desenvolvimento do Algoritmo Robusto proposto nesse capítulo.

Na implementação do programa de um novo algoritmo proposto por Shumway para o Suavizador de Kalman, que diminui a necessidade de inversão de matrizes (Apêndice E), propõem-se também métodos para se estimarem partes de Φ (submatrizes) enquanto outras permanecem fixas.

Apresenta-se também um estudo comparativo do desempenho do suavizador ao se utilizar o algoritmo do Jota [31], ou as recursões de De Jong [30] para o modelo univariável (esses dois algoritmos serão detalhados no decorrer do capítulo). A comparação é feita por meio do ajuste de funções (Capítulo 2) e compara o tempo de processamento em função do número de estados.

3) Verificação

¹O método da maximização a posteriori da função de verossimilhança também é utilizado por Abe e Saito [64]

A validação do modelo é apresentada na Seção 3.6, como “ Critérios para Avaliação de Desempenho ”, onde se utilizam os critérios usuais (tais como Erro Quadrático Médio (*EQM*), etc.), voltados para os valores projetados, e não apenas para os estados estimados². Nesta seção aborda-se a descrição do procedimento *AIC* que, além de ser uma ferramenta de verificação do ajuste dos modelos aos dados, também serve para definir a escolha do modelo mais adequado, ou seja, encerra a fase de identificação (determinação da estrutura).

4) Outras questões

Posteriormente, terminadas as questões de identificação, estimação e verificação, na Seção 3.7 são realizados os testes com vários modelos propostos (testes adicionais são mostrados no Apêndice G). Também é nesta seção que se mostra que um modelo mais sofisticado, em alguns casos apresenta resultados piores que algoritmos mais simples. Quando as projeções são n passos à frente, esse fato torna-se evidente³.

Com parâmetros inadequados, verifica-se a divergência do filtro quando a sequência de ganhos do filtro tende a zero. Um filtro com ganhos constantes, por exemplo, elimina esse problema [45, 12]. A Seção 3.8 enfoca as questões de “ Otimalidade, Sensibilidade e Robustecimento ”, terminando por propor algoritmos mais simples, mas que podem evitar erros grosseiros.

A Seção 3.9 discute como os filtros apresentados nas seções anteriores são utilizados em diversas aplicações em teletráfego.

Finalmente, a Seção 3.10 apresenta o sumário e as conclusões.

3.2 O Modelo de Espaço de Estados Formado por Decomposição Aditiva da Série Temporal

Conforme discutido na Seção 2.2, uma série temporal pode ser descrita por:

$$Y_k = \mu_k + T_k + S_k + V_k \quad (3.1)$$

No problema do tráfego telefônico, a média e a tendência também podem ser decompostas. A média μ_k tem uma parcela contida em T_k ou em S_k e uma parcela que corresponde às flutuações diárias, denominada V_{D_k} . A tendência T_k de fato contém uma parte do tipo tendência polinomial determinística e uma parte formada por um processo do tipo *ARMA*. Para que não se introduzam mais símbolos, a primeira parte da tendência será denominada T_k (ou simplesmente tendência) e a segunda, T_{E_k} (ou tendência estocástica). Desta forma, uma série univariável pode ser descrita como⁴:

$$Y_k = T_k + S_k + T_{E_k} + V_{D_k} + V_k \quad (3.2)$$

onde

²O Apêndice F apresenta o desenvolvimento de algumas relações entre os critérios propostos.

³No entanto, mesmo para a projeção um passo à frente nota-se um aumento no Erro Quadrático Médio, que pode ser verificado pelos resultados do Apêndice G.

⁴Os modelos com perturbações autocorrelacionadas podem ser tratados por meio do filtro com o estado estendido [28], onde se recai no modelo de perturbação Gaussiana.

- Y_k é o valor da observação no instante k ;
- T_k é o valor da tendência da série no instante k ;
- S_k é o valor do componente sazonal no instante k ;
- T_{E_k} é o valor da tendência estocástica, geralmente representado por um processo auto-regressivo de ordem p ;
- V_{D_k} é o valor do componente de variação diário, representado por uma regressão que converge a valores fixos;
- V_k é uma variável aleatória Gaussiana, de média zero e covariância R_k .

O estado do modelo terá quatro componentes:

$$X_k = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix} \quad (3.3)$$

onde X_1 ⁵ corresponde a T_k , X_2 corresponde a S_k , X_3 corresponde a V_{D_k} e X_4 corresponde a T_{E_k} .

Como foi apresentado no Capítulo 2, o modelo de Espaço de Estados genérico para as observações Y_k , é dado por (2.13) e (2.12):

$$Y_k = H_k X_k + V_k$$

e

$$X_k = \Phi X_{k-1} + \Gamma W_k$$

onde considerou-se $U_k = 0$. A introdução dos efeitos de (3.2) em (2.13) e (2.12), conduz a (3.4) e (3.5), [19, 20]:

$$Y_k = [H_1 \ H_2 \ H_{3k} \ H_4] X_k + V_k \quad (3.4)$$

onde

$$X_k = \begin{bmatrix} \Phi_1 & 0 & 0 & 0 \\ 0 & \Phi_2 & 0 & 0 \\ 0 & 0 & \Phi_3 & 0 \\ 0 & 0 & 0 & \Phi_4 \end{bmatrix} X_{k-1} + \begin{bmatrix} \Gamma_1 & 0 & 0 & 0 \\ 0 & \Gamma_2 & 0 & 0 \\ 0 & 0 & \Gamma_3 & 0 \\ 0 & 0 & 0 & \Gamma_4 \end{bmatrix} W_k \quad (3.5)$$

Em (3.4), a matriz H_3 , representada por H_{3k} , é variante no tempo porque se trata de uma regressão de efeitos fixos em que os valores de H_3 mudam a cada nova observação (haverá detalhamento na próxima seção).

O modelo de Espaço de Estados (H, Φ, Γ) é obtido a partir das matrizes componentes (H_j, Φ_j, Γ_j) $j = 1, \dots, 4$. Essas matrizes modelam, respectivamente, os efeitos de tendência, sazonalidade, efeito de variação diário e tendência estocástica.

A seguir, discute-se a metodologia para a obtenção dessas matrizes.

⁵Para efeito de simplificação, onde for possível, não se utilizará o índice k .

3.3 Os Modelos (H_j, Φ_j, Γ_j)

A presente seção apresenta os modelos com os efeitos mais comumente observados em séries temporais relacionadas ao tráfego telefônico. Além de modelos desenvolvidos, em princípio, para finalidades gerais combinados com modelos específicos para aplicações de tráfego (vide exemplos de aplicação no Apêndice G), propõe-se um modelo com variações dos elementos da matriz de observação, $H = [\gamma \ 1 - \gamma]$. A seção termina com uma ilustração da combinação de modelos aditivos.

3.3.1 Modelo de Tendência (H_1, Φ_1, Γ_1)

O componente de tendência pode ter duas formas: a) **tendência polinomial local** e b) **crescimento linear**.

a) O componente de tendência polinomial satisfaz à seguinte equação a diferenças:

$$\Delta^m T_k = w_{1k}; \quad w_{1k} \sim \mathcal{N}(0, q_1) \quad (3.6)$$

onde $\{w_{1k}\}$ é uma seqüência iid (de variáveis aleatórias independentes e identicamente distribuídas) e Δ denota o operador diferença, definido por $\Delta T_k = T_k - T_{k-1}$.

Para $m = 1, 2, 3$, os valores da tendência das correspondentes matrizes H_1, Φ_1, Γ_1 e os componentes do vetor de estado, (X_1) , são:

$$m = 1: \quad T_k = T_{k-1} + w_{1k}, \quad X_1 = T_k, \quad \text{logo } \Phi_1 = [1], \quad \Gamma_1 = [1], \quad H_1 = [1] \quad (3.7)$$

$$m = 2: \quad T_k = 2T_{k-1} - T_{k-2} + w_{1k}, \quad X_1 = \begin{bmatrix} T_k \\ T_{k-1} \end{bmatrix}, \quad \text{logo}$$

$$\Phi_1 = \begin{bmatrix} 2 & -1 \\ 1 & 0 \end{bmatrix}, \quad \Gamma_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad H_1^T = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad (3.8)$$

$$m = 3: \quad T_k = 3T_{k-1} - 3T_{k-2} + T_{k-3} + w_{1k}, \quad X_1 = \begin{bmatrix} T_k \\ T_{k-1} \\ T_{k-2} \end{bmatrix}, \quad \text{logo}$$

$$\Phi_1 = \begin{bmatrix} 3 & -3 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad \Gamma_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad H_1^T = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad (3.9)$$

b) No caso de crescimento linear, a tendência no instante k , T_k , é representada pela adição do nível, T_{k-1} , com o incremento de crescimento, \dot{T}_{k-1} (o incremento de crescimento é o produto do nível pela taxa de crescimento da rota de tráfego):

$$T_k = T_{k-1} + \dot{T}_{k-1} + w_{11k} \quad (3.10)$$

e

$$\dot{T}_k = \dot{T}_{k-1} + w_{21k} \quad (3.11)$$

Supondo que

$$W_{1k} = \begin{bmatrix} w_{11k} \\ w_{21k} \end{bmatrix} \sim \mathcal{N}(0, Q_1)$$

e $\{w_{11k}\}$ e $\{w_{21k}\}$ são seqüências iid, \dot{T}_k é o incremento de crescimento no instante k ,

$$X_1 = \begin{bmatrix} T_k \\ \dot{T}_k \end{bmatrix}, \text{ logo } \Phi_1 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \Gamma_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, H_1^T = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Esse segundo tipo tem sido bastante usado para modelar a tendência do tráfego telefônico porque incorpora uma interpretação física, segundo a qual o nível futuro corresponderá ao nível presente adicionado com o incremento de crescimento em um período. Por outro lado, outra suposição é que o incremento de crescimento permanece constante.

Essas duas considerações correspondem à matriz de transição de estados $\Phi_1 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, [13, 17, 12, 18, 26].

Considerando duas variantes do item b), tem-se:

b.1) Uma extensão deste modelo utiliza a matriz de transição de estado estimada de acordo com os dados disponíveis. Neste caso, permanece a interpretação física de que a carga de tráfego é formada pela soma do nível com o incremento de crescimento, mas ponderados pelos coeficientes de Φ_1 . As demais considerações do item anterior também se aplicam a este modelo.

b.2) Existem casos em que H deve ser representada por $H = [\gamma \ 1 - \gamma]$, onde γ é um fator de ponderação, $0 < \gamma < 1$. Este tipo de representação se justifica porque se requer H diferente de $\begin{bmatrix} 1 & 0 \end{bmatrix}$ ⁶. Continuam válidas as condições gerais de b e da variante b.1, mas:

- T_k não será o valor do nível, mas o valor do nível real dividido por γ , i.e., $T_k = \frac{\text{nível}}{\gamma}$;
- \dot{T}_k não será o valor do incremento de crescimento, mas o valor do incremento real dividido por $1 - \gamma$, i.e., $\dot{T}_k = \frac{\text{incremento}}{1 - \gamma}$;
- Para este modelo, os valores limites de γ (zero e um) não fazem sentido porque provocam indeterminações. No entanto, na prática, o valor mais utilizado para γ é um (é o próprio modelo do item b ou este modelo com $\gamma \rightarrow 1$), que significa que a primeira variável de estado é o próprio nível da carga de tráfego.

Com as considerações acima, (2.13) é reescrita como:

$$Y_k = HX_k + V_k = \gamma T_k + (1 - \gamma)\dot{T}_k + V_k = \text{nível} + \text{incremento} + V_k \quad (3.12)$$

⁶O que importa é o valor da projeção a n passos, que é dado por $H\Phi^n \bar{X}_k$ e a coerência do produto HX_k^* em relação ao valor observado, conforme será descrito a seguir. No desenvolvimento do algoritmo Adaptativo, que será descrito no Capítulo 5, é importante que H não possua elementos nulos porque o valor de H é utilizado para a estimação de Q_k .

3.3.2 Modelos para o Componente Sazonal (H_2, Φ_2, Γ_2)

Diz-se que uma série temporal exibe sazonalidade se as observações equiespaçadas no tempo por algum intervalo fixo possuem comportamento semelhante. A seguir são descritas algumas representações do componente sazonal:

- **Amostras sazonais** (Seção 2.4 para os modelos *AR* sazonais), em que amostras equiespaçadas de um sinal durante o período sazonal são igualadas a zero, ou seja, o sinal se repete a cada ciclo e a soma dos valores positivos e negativos se anula. Iguala-se esta soma de valores positivos e negativos a uma perturbação branca de média nula, w_{2k} , que corresponde à modelagem do componente sazonal do estado, isto é:

$$\sum_{i=0}^{L-1} S_{k-i} = w_{2k} \quad (3.13)$$

onde $w_{2k} \sim \mathcal{N}(0, q_2)$, L é o número de divisões do período sazonal (e.g., $L = 4$ e $L = 12$, para dados trimestrais e mensais, respectivamente) e $\{w_{2k}\}$ é uma seqüência iid. Alternativamente, (3.13) é escrita como:

$$S_k = - \sum_{i=1}^{L-1} B^i S_k + w_{2k} \quad (3.14)$$

onde B^i é o operador translação para o passado, definido por $B^i S_k \triangleq S_{k-i}$.

Por exemplo para dados trimestrais, definindo:

$$X_2 = \begin{bmatrix} S_k \\ S_{k-1} \\ S_{k-2} \end{bmatrix}, \text{ tem-se } \Phi_2 = \begin{bmatrix} -1 & -1 & -1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \Gamma_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, H_2^T = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

Este modelo é usado para efeitos sazonais regulares, onde não há períodos de crescimento ou decrescimento do nível sazonal [19]. X_2 tem dimensão $L - 1$, de acordo com a equação (3.14).

- **Analogia com tendência polinomial** (Seção 3.3):

$$\Delta_y^m S_k = w_{2k} \quad (3.15)$$

onde $\Delta S_k = S_k - S_{k-1}$ e para $m = 1$ tem-se:

$$S_k = S_{k-L} + w_{2k} \quad (3.16)$$

onde L é o período sazonal e w_{2k} é como definido anteriormente. Para dados trimestrais, definindo:

$$X_2 = \begin{bmatrix} S_k \\ S_{k-1} \\ S_{k-2} \\ S_{k-3} \end{bmatrix}, \text{ tem-se } \Phi_2 = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \Gamma_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, H_2^T = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Neste caso, X_2 tem dimensão L .

- **Componentes sazonais crescentes ou decrescentes.** Este modelo utiliza amostras de dois ciclos sazonais consecutivos para permitir a inclusão de padrões crescentes ou decrescentes em relação ao ciclo anterior (Seção 2.4), [19]:

$$(1 - \sum_{i=1}^{L-1} B^i)^2 S_k = w_k \quad (3.17)$$

onde w_{2k} é como apresentado anteriormente. Para dados trimestrais, definindo:

$$X_2 = \begin{bmatrix} S_k \\ S_{k-1} \\ S_{k-2} \\ S_{k-3} \\ S_{k-4} \\ S_{k-5} \end{bmatrix}, \text{ tem-se } \Phi_2 = \begin{bmatrix} 2 & 1 & 0 & -3 & -2 & -1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}, \Gamma_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, H_2^T = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Para esta representação, X_2 tem dimensão $2(L-1)$.

- **Representação por série de Fourier** [26]. A sazonalidade é representada por uma série de Fourier (Seção 2.2), equação (3.18):

$$S_k = \sum_{j=0}^{\lfloor \frac{L}{2} \rfloor} (a_j \cos \frac{2\pi}{L} jk + b_j \sin \frac{2\pi}{L} jk) + w_{2k} \quad (3.18)$$

onde $\lfloor x \rfloor$ é o maior inteiro menor ou igual a x . Na verdade, tem-se $(\lfloor \frac{L}{2} + 1 \rfloor)2 = L+2$ coeficientes no caso par e $L+1$ no caso ímpar; porém, como b_0 e $b_{\frac{L}{2}}$ são nulos, usa-se apenas L coeficientes (no caso ímpar $b_{\frac{L}{2}}$ não existe).

Também escreve-se S_k como:

$$S_k = \alpha_1 + \sum_{j=1}^{\lfloor \frac{L-1}{2} \rfloor} (\alpha_{2j} \cos \frac{2\pi}{L} jk + \alpha_{2j+1} \sin \frac{2\pi}{L} jk) + \alpha_L (-1)^k \quad (3.19)$$

onde $\alpha_L (-1)^k = 0$, se L é ímpar. A primeira maneira de se expressar a sazonalidade, equação (3.18), é semelhante à forma geral para padrões sazonais, (2.5). As duas equações se equivalem com:

$$\alpha_{2j} = a_j \text{ e } \alpha_{2j+1} = b_j, \text{ para } 0 < j < \frac{L}{2}$$

onde $\alpha_1 = a_0$ e $\alpha_L = a_{\frac{L}{2}}$ para L par, e os coeficientes são dados por:

$$\begin{aligned} \alpha_1 &= \frac{1}{L} \sum_{k=1}^L S_k \\ \alpha_{2j} &= \frac{2}{L} \sum_{k=1}^L S_k \cos\left(\frac{2\pi}{L} jk\right) \\ \alpha_{2j+1} &= \frac{2}{L} \sum_{k=1}^L S_k \sin\left(\frac{2\pi}{L} jk\right) \end{aligned}$$

$$\alpha_L = \frac{1}{L} \sum_{k=1}^L S_k (-1)^k \quad (3.20)$$

Se S_k possuir média nula, $\alpha_1 = 0$.

Para obter o cálculo recursivo dos coeficientes, Szlag [26] desenvolveu a equação (3.19) em um instante genérico de tempo k e depois para o instante seguinte $k + 1$. Para transitar de k para $k + 1$, resultam as seguintes equações [26]:

$$\begin{aligned} \alpha_{1k} &= \alpha_{1(k+1)} \\ \begin{bmatrix} \alpha_{2jk} \\ \alpha_{(2j+1)k} \end{bmatrix} &= \phi_j^{-1} \begin{bmatrix} \alpha_{2j(k+1)} \\ \alpha_{(2j+1)(k+1)} \end{bmatrix} \\ \alpha_{Lk} &= (-1)\alpha_{L(k+1)} \end{aligned} \quad (3.21)$$

onde

$$\phi_j^{-1} = \begin{bmatrix} \cos \frac{2\pi}{L} j & -\sin \frac{2\pi}{L} j \\ \sin \frac{2\pi}{L} j & \cos \frac{2\pi}{L} j \end{bmatrix}$$

Definindo o estado X_2 como sendo os coeficientes do modelo sazonal, $X_2 = [\alpha_{1k}, \dots, \alpha_{Lk}]^T$, tem-se:

$$X_{2k} = \Phi_2^{-1} X_{2(k+1)}$$

ou

$$X_{2(k+1)} = \Phi_2 X_{2k} \quad (3.22)$$

onde

$$\Phi_2^{-1} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & \phi_1^{-1} & 0 & \dots & 0 \\ 0 & 0 & \phi_2^{-1} & \dots & 0 \\ 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \dots & -1 \end{bmatrix} \quad (3.23)$$

A utilização do índice k tornou-se primordial. Nos exemplos anteriores, o produto $H_j X_j$ fornecia o valor que seria adicionado à equação geral (3.2). Neste caso S_k é formada pelos coeficientes de (3.19) desenvolvida por Szlag em dois instantes de tempo consecutivos e, por imposição, múltiplos de $\frac{\pi}{2}$, de tal forma que os termos onde ocorrem as funções seno se anulam e os termos onde ocorrem as funções cosseno se tornam iguais a um, resultando:

$$S_k = \alpha_{1k} + \sum_j \alpha_{2jk} + \alpha_{Lk}$$

O produto $H_2 X_2$ fornecerá o valor S_k . Então o vetor L dimensional $H_2 = [1 \ 1 \ 0 \ 1 \ \dots \ 0]$

Como exemplo, para dados trimestrais:

$$\Phi_2^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \frac{\pi}{2} & -\sin \frac{\pi}{2} & 0 \\ 0 & \sin \frac{\pi}{2} & \cos \frac{\pi}{2} & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}$$

onde

$$\Phi_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}, \Gamma_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, H_2^T = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \end{bmatrix}$$

Desta forma, X_2 possui dimensão L . Segundo Szlag [26], resultados empíricos indicam que o modelo trigonométrico propicia, em algumas situações, resultados melhores que os modelos anteriores.

3.3.3 Efeito de Variação Diário (H_{3k}, Φ_3, Γ_3)

Neste modelo H_3 pode variar com o tempo. Devido aos diferentes números de dias da semana por mês, Kitagawa e Gersch [19] propuseram sua utilização para melhorar a previsão de modelos sazonais (por exemplo, existem meses que possuem mais domingos, outros possuem mais segundas-feiras, etc.). A introdução do efeito de variação diário no modelo de Espaço de Estados corresponde a remover da série temporal a flutuação devido a essas variações. Para um período fixo de observações, ajusta-se a variação diária por meio de uma regressão sobre coeficientes fixos. Dado que a flutuação da variação diária é semanal, tem-se que:

$$\sum_{i=1}^7 \gamma_{ik} = 0. \text{ Assim } \gamma_{7k} = -\sum_{i=1}^6 \gamma_{ik} \quad (3.24)$$

onde γ_{ik} denota o fator do efeito de variação diário dos i -ésimos dias da semana no mês (trimestre, etc) k . Conceitualmente, γ_{ik} é um sinal periódico da mesma forma que o sazonal, porém, representa apenas uma flutuação semanal sobre um período maior (mês, trimestre, etc.) que vai ser observado.

Se d_{ik}^* denotar o número de i -ésimos dias da semana (segundas, terças, etc.) no k -ésimo mês (ano, trimestre, etc.), escreve-se o efeito de variação diário como:

$$V_{Dk} = \sum_{i=1}^7 \gamma_{ik} d_{ik}^* \quad (3.25)$$

Fazendo $d_{ik} = d_{ik}^* - d_{7k}^*$, tem-se:

$$V_{Dk} = \sum_{i=1}^6 \gamma_{ik} d_{ik}^* + \gamma_{7k} d_{7k}^* = \sum_{i=1}^6 \gamma_{ik} d_{ik}^* - \sum_{i=1}^6 \gamma_{ik} d_{7k}^* = \sum_{i=1}^6 \gamma_{ik} (d_{ik}^* - d_{7k}^*) = \sum_{i=1}^6 \gamma_{ik} d_{ik}$$

Como $V_{Dk} = H_{3k} X_{3k}$, definem-se:

$$H_{3k} \triangleq \begin{bmatrix} d_{1k} \\ d_{2k} \\ d_{3k} \\ d_{4k} \\ d_{5k} \\ d_{6k} \end{bmatrix}, X_{3k} \triangleq \begin{bmatrix} \gamma_{1k} \\ \gamma_{2k} \\ \gamma_{3k} \\ \gamma_{4k} \\ \gamma_{5k} \\ \gamma_{6k} \end{bmatrix}$$

A equação a diferenças, não-perturbada, para a variação diária é ⁷:

$$\gamma_{ik} = \gamma_{i(k-1)} \quad (3.26)$$

Para dados mensais, tem-se:

$$\Phi_3 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} = I_6, \Gamma_3 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Para dados trimestrais, d_{ik}^* denota o número de i -ésimos dias da semana no k -ésimo trimestre, e pelo fato de serem 3 meses, o efeito diário é menos significativo.

3.3.4 Modelo para Tendência Estocástica (Auto-regressivo) (H_4, Φ_4, Γ_4)

Neste modelo, supõe-se que o componente estocástico estacionário, T_{E_k} , satisfaz um modelo auto-regressivo de ordem p . Isto é:

$$T_{E_k} = \phi_1 T_{E_{k-1}} + \dots + \phi_p T_{E_{k-p}} + w_{4k} \quad (3.27)$$

onde $w_{4k} \sim \mathcal{N}(0, q_4)$ e $\{w_{4k}\}$ é uma seqüência iid. Definindo:

$$X_{4k} = \begin{bmatrix} T_{E_k} \\ T_{E_{k-1}} \\ \vdots \\ T_{E_{k-p+1}} \end{bmatrix}$$

tem-se:

$$\Phi_4 = \begin{bmatrix} \phi_1 & \phi_2 & \dots & \phi_p \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & \dots & 1 & 0 \end{bmatrix}, \Gamma_4 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, H_4^T = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

3.3.5 Ilustração do Modelo

O exemplo a seguir, para dados mensais, possui tendência (crescimento linear), sazonalidade, tendência estocástica (AR de ordem 2) e efeito de variação diário (o Apêndice H

⁷Isto equivale à solução recursiva de uma regressão linear do tipo de (2.3) da Seção 2.3, com $\beta_i = \gamma_{ik}$ e $d_{ik} = X_{ik}$. Esta forma recursiva de solução da regressão linear múltipla é utilizada nos algoritmos de filtragem adaptativa (mínimos quadrados recursivos) e neste caso o modelo é um mero estimador de parâmetros (sem ruído no estado, ou seja, Q_k é nula) [11].

mostra um detalhamento da obtenção dessas matrizes):

$$X_k = \begin{bmatrix} T_k \\ \dot{T}_k \\ \hline T_{E_k} \\ T_{E_{k-1}} \\ \hline S_k \\ S_{k-1} \\ \vdots \\ S_{k-10} \\ \hline \gamma_{1k} \\ \vdots \\ \gamma_{6k} \end{bmatrix} = \begin{bmatrix} 1 & 1 & | & 0 & 0 & | & 0 & \dots & 0 & | & 0 & \dots & 0 \\ 0 & 1 & | & 0 & 0 & | & 0 & \dots & 0 & | & 0 & \dots & 0 \\ \hline 0 & 0 & | & \phi_1 & \phi_2 & | & 0 & \dots & 0 & | & 0 & \dots & 0 \\ 0 & 0 & | & 1 & 0 & | & 0 & \dots & 0 & | & 0 & \dots & 0 \\ \hline 0 & 0 & | & 0 & 0 & | & -1 & \dots & -1 & | & 0 & \dots & 0 \\ 0 & 0 & | & 0 & 0 & | & 1 & \dots & 0 & | & 0 & \dots & 0 \\ \hline \dots & \dots & | & \dots & \dots & | & \dots & \dots & \dots & | & \dots & \dots & \dots \\ 0 & 0 & | & 0 & 0 & | & & \ddots & \vdots & | & 0 & \dots & 0 \\ 0 & 0 & | & 0 & 0 & | & & & 1 & 0 & | & 0 & \dots & 0 \\ \hline 0 & 0 & | & 0 & 0 & | & 0 & \dots & 0 & | & 1 & \dots & 0 \\ \hline \dots & \dots & | & \dots & \dots & | & \dots & \dots & \dots & | & \dots & \dots & \dots \\ 0 & 0 & | & 0 & 0 & | & 0 & \dots & 0 & | & 0 & \dots & 0 \\ 0 & 0 & | & 0 & 0 & | & 0 & \dots & 0 & | & 0 & \dots & 1 \end{bmatrix} X_{k-1} +$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 \end{bmatrix} W_k$$

(3.28)

$$Y_k = [1 \ 0 \ 1 \ 0 \ 1 \ 0 \ \dots \ 0 \ d_{1k} \ \dots \ d_{6k}] X_k + V_k \quad (3.29)$$

onde

$$W_k = \begin{bmatrix} w_{11k} \\ w_{21k} \\ w_{2k} \\ w_{4k} \end{bmatrix}, \begin{bmatrix} W_k \\ V_k \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} Q_1 & \underline{0} & \underline{0} & \underline{0} \\ \underline{0}^T & q_2 & 0 & 0 \\ \underline{0}^T & 0 & q_4 & 0 \\ \underline{0}^T & 0 & 0 & R_k \end{bmatrix} \right)$$

onde

$$Q_1 = \begin{bmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{bmatrix}$$

e $\underline{0}$ é um vetor nulo de dimensão 2×1

Embora a seção tenha propósitos gerais, na prática os padrões de tendência e sazonalidade são suficientes para descrever o comportamento da série temporal do tráfego telefônico [13, 18, 17, 12, 26]. Isto também tem sido confirmado pelos diversos resultados que se obtiveram com dados de dezenas de rotas de tráfego ([45] e Apêndice G).

3.4 Estimação de Estado com o Filtro de Kalman

O Filtro de Kalman é um estimador ótimo⁸ que minimiza a variância do erro entre o estado estimado e o estado real. A presente seção mostrará as equações do Filtro de Kalman para a **filtragem** e para a **predição**. Como definido no Capítulo 2, **Filtragem** se refere à estimação do vetor de estado atual, baseada em todas as observações passadas e **Predição** se refere à estimação do estado em um instante futuro a partir do valor presente do estado estimado. A seguir discute-se a estimação do estado dadas N observações passadas, ou seja, os procedimentos de suavização (*smoothing*).

Nesta seção supõe-se que as condições iniciais e os demais parâmetros necessários ao funcionamento do filtro são conhecidos, ou seja, conhece-se \bar{X}_0 , \bar{P}_0 , R_k , Q_k , e Φ (a seção seguinte abordará a estimação desses parâmetros).

3.4.1 Equações do Filtro de Kalman

Serão apresentadas as equações do Filtro de Kalman e do Suavizador de Kalman.

- Filtro de Kalman

Na literatura existem diversas maneiras de deduzir as equações do Filtro de Kalman [24, 18, 28]. As equações do filtro são divididas em: equações de **estimação (filtragem ou atualização)** e equações de **predição** [25].

– As equações da filtragem são:

$$\bar{X}_k = X_k^* + K_k r_k \quad (3.30)$$

onde \bar{X}_k é o valor estimado do estado (filtrado), X_k^* é o valor da projeção do estado (predição) para o instante k a partir do instante $k - 1$, K_k é o ganho e r_k é o termo de inovação. Estes dois últimos parâmetros são calculados por:

$$r_k = Y_k - H X_k^* \quad (3.31)$$

e

$$K_k = P_k^* H^T (\sigma_k^2)^{-1} \quad (3.32)$$

onde P_k^* é a covariância do erro de predição, Y_k é o valor da observação no instante k , σ_k^2 é a covariância do termo de inovação e H é a matriz de observação. A covariância σ_k^2 é dada por:

$$\sigma_k^2 = H P_k^* H^T + R_k \quad (3.33)$$

onde $R_k = E[V_k V_k^T]$ é a covariância do erro de observação e V_k é um vetor de ruídos brancos gaussianos de média nula (perturbações nas observações). A covariância do erro de estimação (filtragem) é dada por:

$$\bar{P}_k = (I - K_k H) P_k^* \quad (3.34)$$

⁸Aqui os estimadores de interesse são os estimadores **não-polarizados** (aqueles cujo valor esperado é o mesmo que a quantidade sendo estimada), de **variância mínima** (cuja variância do erro é menor ou igual a qualquer outro estimador não-polarizado) e **consistentes** (que convergem ao valor verdadeiro da grandeza estimada, o vetor do estado, conforme o número de observações aumenta).

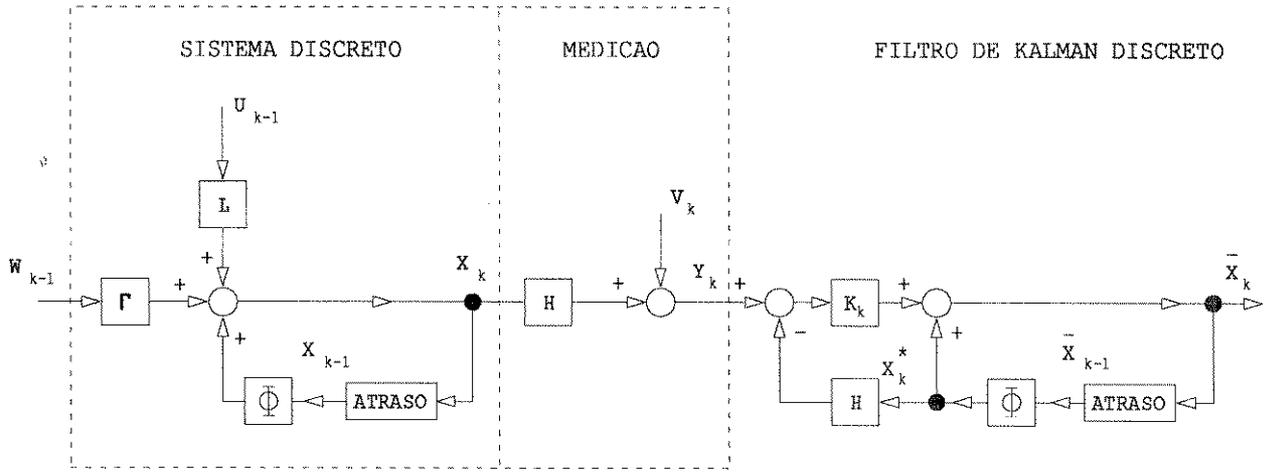


Figura 3.1: Filtro de Kalman Discreto

– As equações da predição são:

$$X_{k+1}^* = \Phi \bar{X}_k \quad (3.35)$$

e

$$P_{k+1}^* = \Phi \bar{P}_k \Phi^T + Q_k \quad (3.36)$$

onde $Q_k = E[W_k W_k^T]$ é a covariância do erro de modelagem, W_k é um vetor de perturbações das variáveis de estado (ruídos brancos gaussianos de média nula) e Φ é a matriz de transição de estado.

A Figura 3.1 mostra o modelo do sistema e o Filtro de Kalman discreto.

• Suavizador de Kalman

A suavização é uma forma de processamento de um conjunto de N dados, em tempo não real, que se utiliza de todas as observações entre $1 \leq k \leq N$ para estimar o estado de um sistema em um instante k entre 1 e N . O estado suavizado possui variância reduzida quando comparada à variância do estado filtrado. Desta forma, a utilização do suavizador também propicia a redução da imprecisão na estimação dos parâmetros. O suavizador a ser apresentado é o de intervalo fixo.

As equações do suavizador que se utilizam do ganho para trás, J_k , são [20, 21]:

$$\bar{X}_{k-1}^N = \bar{X}_{k-1} + J_{k-1}(\bar{X}_k^N - \Phi \bar{X}_{k-1}) \quad (3.37)$$

$$J_{k-1} = \bar{P}_{k-1} \Phi^T (P_k^*)^{-1} \quad (3.38)$$

$$\bar{P}_{k-1}^N = \bar{P}_{k-1} + J_{k-1}(\bar{P}_k^N - P_k^*) J_{k-1}^T \quad (3.39)$$

$$\bar{P}_{k-1,k-1} = \bar{P}_{k-1} J_{k-2}^T + J_{k-1}(\bar{P}_{k,k-1}^N - \Phi \bar{P}_{k-1}) J_{k-2}^T \quad (3.40)$$

onde

$$\bar{P}_{N,N-1}^N = (I - K_k H) \Phi \bar{P}_{N-1}$$

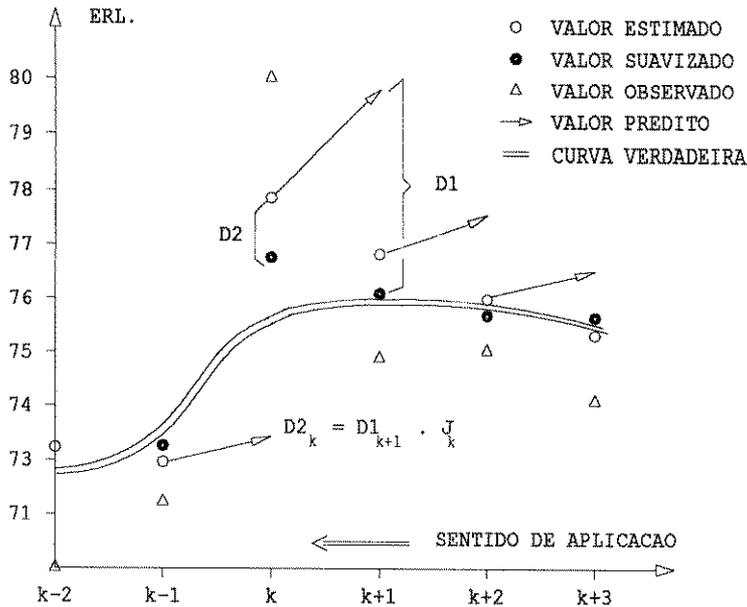


Figura 3.2: Funcionamento do Suavizador

onde $E[X_k | Y_1, Y_2, \dots, Y_N] = \bar{X}_k^N$, $E[(X_k - \bar{X}_k^N)(X_k - \bar{X}_k^N)^T | Y_1, Y_2, \dots, Y_N] \triangleq \bar{P}_k^N$, e $E[(X_k - \bar{X}_k^N)(X_{k-1} - \bar{X}_{k-1}^N)^T | Y_1, Y_2, \dots, Y_N] \triangleq \bar{P}_{k-1, k}^N$.

A Figura 3.2 [45] ilustra o desempenho do suavizador com as diferenças relativas entre o estado suavizado (\bar{X}_k^N), o estado filtrado (\bar{X}_k) e a curva verdadeira (não-observável). Nessa figura, as bolas pretas representam os estados suavizados, as bolas brancas representam os estados filtrados, as setas indicam o estado projetado um passo à frente e os triângulos representam os valores observados⁹. Pode-se notar que o estado suavizado, por utilizar uma maior quantidade de informação, tem a sua variância do erro de estimação significativamente diminuída quando comparado ao estado filtrado, com maior proximidade do estado suavizado em relação ao estado real. A variável J representa uma espécie de ganho para trás, e propicia um ajuste maior dos valores observados à curva verdadeira. As quantidades $d2$ e $d1$ correspondem a $\bar{X}_{k-1}^N - \bar{X}_{k-1}$ e $\bar{X}_k^N - \Phi \bar{X}_{k-1}$ da equação (3.37) e são uma representação do funcionamento do suavizador no modelo escalar. Esta figura foi obtida de uma aplicação do algoritmo *EM* sobre uma rota de tráfego do Estado de São Paulo (Marília 33 para a central trânsito, denominada T44). Os dados da rota de tráfego da Figura 3.2 correspondem à ampliação do detalhe retangular mostrado nas Figuras 3.4 e 3.5 deste capítulo, e também correspondem à Figura 4.3 do Capítulo 4.

Para evitar a inversão da matriz P_k^* de (3.38), implementa-se o programa do suavizador com a seguinte forma (recursões de De Jong e simplificações de Devine, que

⁹O estado filtrado (bolas brancas) fica entre o estado projetado a um passo (setas) e os valores observados (triângulos). O Apêndice I mostra, para um modelo univariável com dois estados, que a curva verdadeira (não-observável) se posiciona entre o valor da projeção a um passo e o valor observado.

são apresentadas no Apêndice E) [30]:

$$L_k = \Phi(I - K_k H) \quad (3.41)$$

$$Z_{k-1} = H^T(\sigma_k^2)^{-1} r_k + L_k^T Z_k, \quad Z_N = 0 \quad (3.42)$$

$$\bar{X}_k^N = X_k^* + P_k^* Z_{k-1} \quad (3.43)$$

$$S_{k-1} = H^T(\sigma_k^2)^{-1} H + L_k^T S_k L_k, \quad S_N = 0 \quad (3.44)$$

$$\bar{P}_k^N = P_k^* S_{k-1} P_k^* \quad (3.45)$$

Também:

$$\bar{X}_0^N = \bar{X}_0 + \bar{P}_0 \Phi^T Z_0$$

$$\bar{P}_0^N = \bar{P}_0 \Phi^T S_0 \Phi \bar{P}_0$$

e a covariância de transição de estado:

$$\bar{P}_{k,k-1}^N = (I - P_k^* S_{k-1}) L_{k-1} P_{k-1}^* \quad (3.46)$$

com:

$$\bar{P}_{1,0}^N = (I - P_1^* S_0) \Phi \bar{P}_0$$

Para efeito de comparação entre o tempo de execução ¹⁰ dos dois suavizadores, tomaram-se sete observações da Rota Bauru trânsito para Marília trânsito, T4244, (outros exemplos estão disponíveis no Apêndice G) e testaram-se diversos modelos compostos por um a cinco estados, para observação escalar. Três modelos de regressão foram utilizados (linear, exponencial e potencial) para o ajuste do tempo de execução com o número de variáveis de estado. O modelo que melhor se adequou foi do tipo potencial:

$$T = ab^E$$

onde T é o tempo de execução em minutos e E é o número de estados. A título de ilustração, o teste em uma calculadora HP28S (com outra Rota, T6362, Jaboticabal Trânsito para Araraquara Trânsito) para o suavizador utilizando J , resultou $a = 6,099$, $b = 1,291$ e o coeficiente de explicação $R^2 = 0,999$. Para o suavizador de De Jong, $a = 6,323$, $b = 1,265$ e o coeficiente de explicação $R^2 = 0,998$ (também neste caso, o modelo potencial foi o que melhor se adequou). Como se pode notar, os coeficientes de explicação próximos a um indicam uma relação praticamente determinística (isto tem que ser encarado com reservas porque foram utilizadas apenas cinco observações). Alguns comentários a respeito desses resultados são os seguintes:

- O suavizador com J aumenta a duração de sua execução conforme aumentam o número de estados devido à necessidade de inversão da matriz P_k^* ;
- Devido à forma de implementação dos programas, o suavizador com J executa um *loop* a mais que o suavizador de De Jong ($N + 1$ *loops*). No entanto, para N maior que quinze observações, esse processamento adicional se torna pouco significativo. O suavizador de De Jong, embora execute N *loops*, precisa executar

¹⁰Com relação ao código gerado, a diferença básica é que o suavizador de De Jong, no modelo univariável, não requer a utilização da rotina para inversão de matrizes.

algumas instruções adicionais devido a cálculos separados para \bar{X}_0^N , \bar{P}_0^N e $\bar{P}_{1,0}^N$. Isso justifica o fato do suavizador com J requerer menos tempo de execução para um ou dois estados (para sete observações);

- Para os dados disponíveis (acima de quinze observações) e modelos com um ou dois estados, os dois algoritmos são equivalentes do ponto de vista do tempo de execução (do ponto de vista da estimação dos parâmetros eles também se mostraram equivalentes, vide Apêndice G);

3.5 Estimação Conjunta do Estado e dos Parâmetros

Os procedimentos para estimação dos parâmetros do Filtro de Kalman são necessários porque, na prática, eles raramente são conhecidos. Para as aplicações deste trabalho, a matriz de transição de estado, Φ , é conhecida em parte, mas para a projeção n passos à frente é preciso que ela seja estimada com precisão.

Da equação do sistema, dada por (2.13), pode-se notar que o sistema é linear quando Φ é conhecido. Portanto, o desconhecimento simultâneo do estado X_k e da matriz de transição de estado, Φ , caracteriza um problema de filtragem não-linear.

Para se trabalhar com o problema de filtragem não-linear pode-se basear, de maneira geral, em três abordagens:

- Aproximar as funções não-lineares das equações do modelo por meio do **desenvolvimento em série de Taylor** de primeira ordem, segunda ordem, etc.;
- Utilizar estimação não-linear por **mínimos quadrados generalizados**. Esta técnica provê um ajuste aos dados observados de um modo mais determinístico que estocástico, onde o valor do estado estimado minimiza um índice de desempenho determinístico (é uma variante de (2.55), onde a minimização é em relação ao estado estimado, \bar{X}_k , e não em relação aos coeficientes) [24]. Neste caso não é necessário o conhecimento dos modelos estatísticos dos ruídos dos processos, W_k e V_k ;
- Utilizar a aproximação das funções não-lineares que descrevem o funcionamento do filtro por meio de **linearização estatística**, que ajusta polinômios às observações, por meio da técnica de mínimos quadrados (o caso linear, corresponde a interpolar uma reta aos trechos não-lineares).

No caso da projeção de rotas de tráfego telefônico existem as seguintes particularidades:

- Considera-se, na aplicação das técnicas de relaxação, a não-linearidade apenas na equação do sistema e não na equação da observação. A não-linearidade presente na equação da observação é suposta como devida à ocorrência de observações anômalas, *outliers*, e neste trabalho propõe-se seu tratamento por meio de algoritmos para detecção de observações aberrantes, *outliers*, que serão abordados nos Capítulos 4 e 5;
- Se a primeira ou segunda aproximações acima forem adotadas, haverá o problema do desconhecimento da função não-linear que representa a equação do sistema. Se a terceira aproximação for adotada, será necessário o conhecimento da função densidade de probabilidade que descreve a variável de estado (embora, em geral, considere-se que a variável de estado é Gaussiana mesmo nestes casos não-lineares);

- Além desse problema de filtragem não-linear, que envolve a estimação do estado, \bar{X}_k , e da matriz de transição de estado, Φ , também é preciso se estimarem os demais parâmetros do filtro: R_k , Q_k , \bar{P}_0 , e \bar{X}_0 .

Como será mostrado na Seção 3.7, após os parâmetros estimados, o filtro linear passa a operar adequadamente. Portanto, a questão da não-linearidade se restringe à fase de estimação dos parâmetros (para operação em tempo real, que será abordada no Capítulo 5, deve-se também considerar perturbações não-estacionárias, e portanto, parâmetros variantes no tempo).

Uma outra abordagem foi então proposta, tendo em vista esses aspectos. A técnica da **relaxação** (algoritmo *EM*), como descrita a seguir, foi utilizada para a estimação conjunta do estado, X_k , da matriz de transição de estado, Φ , e dos demais parâmetros (R_k , Q_k , \bar{P}_0 , e \bar{X}_0), e os resultados obtidos apresentaram erros inferiores ao dos algoritmos do Filtro de Kalman em que os parâmetros não foram estimados dessa maneira (Seção 3.7). Nesta proposta de aplicação do algoritmo *EM*, \bar{X}_0 e \bar{P}_0 são estimados simultaneamente, o que é fundamental quando se dispõe de poucas observações. O algoritmo de relaxação utilizado neste trabalho é um procedimento realizado em tempo não real, sobre um conjunto fixo de N valores medidos. Inicialmente os parâmetros são supostos conhecidos e estima-se o estado por meio de um procedimento de suavização. Em seguida, usando-se os valores do estado estimado pelo Filtro de Kalman, calculam-se os parâmetros. Com os parâmetros estimados, reestima-se o estado e prossegue-se dessa forma até que haja a convergência do algoritmo.

Existem diversos algoritmos para esse tipo de estimação. Devido às suas propriedades, o estimador de máxima verossimilhança é o mais comumente utilizado na literatura [36]. Este estimador, combinado com o procedimento de suavização, é denominado *EM Expectation-Maximization*. A Seção 3.7 mostra que os resultados são adequados e se aplicam às projeções de tráfego telefônico. Além disso, a seção faz uma consideração sobre a maneira de se estimarem partes (submatrizes) da matriz de transição de estado Φ . É interessante ressaltar que no trabalho de Shumway [20] só há um exemplo escalar.

3.5.1 Estimação de Máxima Verossimilhança do Algoritmo *EM*

Considerando que o logaritmo da distribuição conjunta dos erros dos valores iniciais, de modelagem e de observação, dados os estados X_0, X_1, \dots, X_N e as observações Y_1, Y_2, \dots, Y_N , é o logaritmo da distribuição conjunta de três Gaussianas [21], então:

$$\begin{aligned} \log L = & -\frac{1}{2} \log |\bar{P}_0| - \frac{1}{2} (X_0 - \bar{X}_0)^T \bar{P}_0^{-1} (X_0 - \bar{X}_0) \\ & - \frac{N}{2} \log |Q| - \frac{1}{2} \sum_{k=1}^N (X_k - \Phi X_{k-1})^T Q^{-1} (X_k - \Phi X_{k-1}) \\ & - \frac{N}{2} \log |R| - \frac{1}{2} \sum_{k=1}^N (Y_k - H X_k)^T R^{-1} (Y_k - H X_k) \end{aligned} \quad (3.47)$$

A primeira linha de (3.47) é o logaritmo da distribuição Gaussiana $\mathcal{N}(\bar{X}_0, \bar{P}_0)$, que corresponde à distribuição do erro do estado inicial. A segunda linha de (3.47) é o logaritmo da distribuição conjunta de N variáveis aleatórias, cada uma delas com distribuição $\mathcal{N}(0, Q)$, que corresponde ao erro de modelagem (equação do sistema) em cada observação. A terceira linha de (3.47) é o logaritmo da distribuição conjunta de N variáveis aleatórias,

cada uma delas com distribuição $\mathcal{N}(0, R)$, que corresponde ao erro de observação (equação da observação) em cada medição. A distribuição conjunta das distribuições dessas três linhas equaciona todas as fontes de erro do filtro. Nos três termos existe um termo constante e igual a 2π , que foi suprimido. A seguir, maximiza-se a função log da verossimilhança ($\log L$) com relação aos parâmetros $\bar{X}_0, \bar{P}_0, \Phi, R$ e Q .¹¹

Devido a que o logaritmo da verossimilhança depende da série de dados não-observáveis, $X_k, k = 0, 1, \dots, N$, maximiza-se a função (3.47) condicionada à série observada Y_1, Y_2, \dots, Y_N . Em outras palavras, definem-se os parâmetros estimados na r -ésima iteração como os valores $\bar{X}_0, \bar{P}_0, \Phi, R$ e Q que maximizam:

$$G(\bar{X}_0, \bar{P}_0, \Phi, Q, R) = E(\log L | Y_1, Y_2, \dots, Y_N) \quad (3.48)$$

onde a esperança condicional é relativa a uma densidade contendo os r -ésimos valores iterados $\bar{X}_0(r), \bar{P}_0(r), \Phi(r), Q(r)$ e $R(r)$.

O algoritmo é dividido em duas etapas:

- Maximização da função de verossimilhança para cálculo dos parâmetros do modelo (M);
- Estimação de estado (E).

A maximização de $\log L$ é realizada utilizando-se um procedimento iterativo, com as equações da Seção 3.4, até se obter o ponto de máximo.

Substituindo (3.47) em (3.48) é necessário calcularem-se \bar{X}_k^N, \bar{P}_k^N e $\bar{P}_{k-1,k}^N$. Estes valores são fornecidos pelas equações do filtro, (3.43), (3.45) e (3.46).

Desenvolvendo (3.47) para posterior utilização em (3.48), tem-se que o lado direito da primeira linha de (3.47) é:

$$-\frac{1}{2}(X_0 - \bar{X}_0)^T \bar{P}_0^{-1} (X_0 - \bar{X}_0) = -\frac{1}{2} TR \{ \bar{P}_0^{-1} (X_0 - \bar{X}_0) (X_0 - \bar{X}_0)^T \} \quad (3.49)$$

onde TR é o traço da matriz.

Somando-se e subtraindo-se \bar{X}_0^N de cada elemento entre parênteses e aplicando-se a esperança condicional, obtém-se:

$$\begin{aligned} & -\frac{1}{2} TR \{ \bar{P}_0^{-1} E[(X_0 - \bar{X}_0 - \bar{X}_0^N + \bar{X}_0^N)(X_0 - \bar{X}_0 - \bar{X}_0^N + \bar{X}_0^N)^T | Y_1, \dots, Y_N] \} \\ & = -\frac{1}{2} TR \{ \bar{P}_0^{-1} [\bar{P}_0^N + (\bar{X}_0^N - \bar{X}_0)(\bar{X}_0^N - \bar{X}_0)^T] \}, \quad E[X_0 - \bar{X}_0^N] = 0. \end{aligned} \quad (3.50)$$

De maneira análoga podem ser deduzidos outros termos de (3.48), conforme se segue:

$$\begin{aligned} G(\bar{X}_0, \bar{P}_0, \Phi, Q, R) & = -\frac{1}{2} \log |\bar{P}_0| - \frac{1}{2} TR \{ \bar{P}_0^{-1} [\bar{P}_0^N + (\bar{X}_0^N - \bar{X}_0)(\bar{X}_0^N - \bar{X}_0)^T] \} \\ & - \frac{N}{2} \log |Q| - \frac{1}{2} TR [Q^{-1} (C - B\Phi^T - \Phi B^T + \Phi A\Phi^T)] - \frac{N}{2} \log |R| - \\ & \frac{1}{2} TR \{ R^{-1} \sum_{k=1}^N [(Y_k - H\bar{X}_k^N)(Y_k - H\bar{X}_k^N)^T + H\bar{P}_k^N H^T] \} \end{aligned} \quad (3.51)$$

¹¹Como em cada iteração os valores das covariâncias são fixos, não se utilizou o índice k .

onde

$$A = \sum_{k=1}^N [\bar{P}_{k-1}^N + \bar{X}_{k-1}^N (\bar{X}_{k-1}^N)^T] \quad (3.52)$$

$$B = \sum_{k=1}^N [\bar{P}_{k,k-1}^N + \bar{X}_k^N (\bar{X}_{k-1}^N)^T] \quad (3.53)$$

$$C = \sum_{k=1}^N [\bar{P}_k^N + \bar{X}_k^N (\bar{X}_k^N)^T] \quad (3.54)$$

Os valores de \bar{X}_k^N , \bar{P}_k^N e $\bar{P}_{k,k-1}^N$ são computados, a cada iteração, utilizando-se o Filtro de Kalman, com os valores fixados dos parâmetros $\bar{X}_0(r)$, $\Phi(r)$, $Q(r)$ e $R(r)$, correspondentes à iteração r .

A maximização da função de verossimilhança é dada por:

$$\Phi(r+1) = BA^{-1} \quad (3.55)$$

$$Q(r+1) = \frac{(C - BA^{-1}B^T)}{N} \quad (3.56)$$

$$R(r+1) = \sum_{k=1}^N \frac{[(Y_k - H\bar{X}_k^N)(Y_k - H\bar{X}_k^N)^T + H\bar{P}_k^N H^T]}{N} \quad (3.57)$$

e

$$\bar{X}_0(r+1) = \bar{X}_0^N \quad (3.58)$$

Pode-se verificar que isto equivale a impor o traço da matriz identidade, $TR[I]$.

Devido a não ser possível se estimarem simultaneamente \bar{P}_0 e X_0 porque a equação (3.50) e as equações do suavizador formam um sistema indeterminado¹², é conveniente que se fixe um dos parâmetros (fixa-se \bar{P}_0 , sobre o qual há estudos relativos a aplicações em projeções do tráfego telefônico e se calcula \bar{X}_0 utilizando-se (3.58)). Moreland [12] propõe um valor \bar{P}_0 para o modelo com dois estados representando tendência. O detalhamento dessa proposta será feito no capítulo seguinte, que abordará o Algoritmo de Projeção Sequencial Robusto proposto.

Existem formas alternativas de se estimar \bar{P}_0 (vide a iniciação do Algoritmo de Projeção Sequencial Adaptativo Modificado, APSAM, Capítulo 5).

A Figura 3.3 ilustra o esquema do funcionamento do algoritmo *EM*. Inicialmente começa-se com o melhor conjunto de parâmetros disponível¹³ e roda-se o Filtro de Kalman, da primeira até a enésima observação. A cada observação processada guardam-se os valores calculados X_k^* , P_k^* e L_k se o suavizador utilizado é o De Jong e guarda-se \bar{X}_k , \bar{P}_k e P_k^* se é o suavizador com J que vai ser utilizado. Após terminado o processamento das N observações, o suavizador entra em ação processando os N conjuntos de valores armazenados na memória, em ordem inversa, do enésimo ao primeiro. Para se utilizar a mesma área de memória, esse conjunto de valores armazenados é substituído, após sua utilização, por \bar{X}_k^N , \bar{P}_k^N e $\bar{P}_{k,k-1}^N$. Depois de terminado o ciclo completo, estimam-se os parâmetros utilizando-se (3.55) a (3.58). Quando o número de iterações é completado, termina-se a execução, caso contrário inicia-se um novo ciclo.

¹²O valor máximo de (3.50) é obtido com o traço da matriz identidade porque corresponde ao máximo da função logaritmo.

¹³No entanto, após algumas iterações, o algoritmo *EM* obtém resultados adequados, mesmo com valores iniciais incorretos (desde que haja um número significativo de observações).

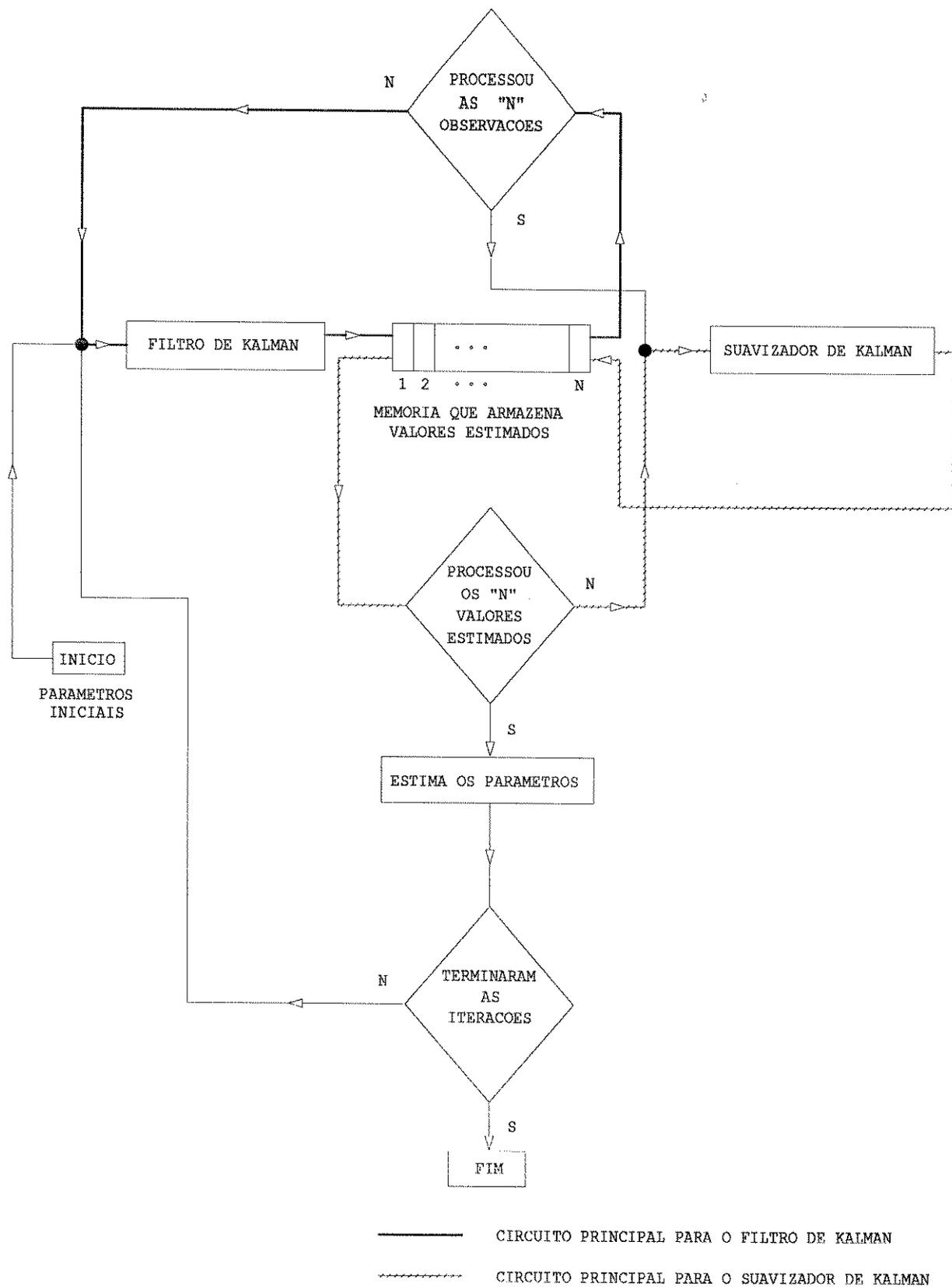


Figura 3.3: Funcionamento do Algoritmo *EM*

Para a verificação da convergência, a utilização de (3.47) é de fato trabalhosa. Existe uma maneira alternativa de se escrever a função de verossimilhança, por meio da série das inovações, que apresenta maior eficiência computacional porque seus valores já estão calculados para as equações do filtro (cálculo para a frente). O logaritmo da distribuição conjunta de N distribuições Gaussianas (de N termos de inovação ou resíduos), cada uma normalmente distribuída (um fator constante 2π foi cancelado): $\sim \mathcal{N}(0, HP_k^* H^T + R_k)$ é escrito como [13, 20, 32]:

$$\log L = -\frac{1}{2} \sum_{k=1}^N \log |HP_k^* H^T + R_k| - \frac{1}{2} \sum_{k=1}^N (Y_k - HX_k^*)(HP_k^* H^T + R_k)^{-1}(Y_k - HX_k^*) \quad (3.59)$$

Observações finais:

- Quando se querem estimar apenas algumas submatrizes da matriz de transição de estado Φ (Seção 3.2) e não a matriz completa, calculam-se essas submatrizes Φ_i desconhecidas utilizando $B_i A_i^{-1}$, onde B_i e A_i são as submatrizes das matrizes B e A , correspondentes à submatriz Φ_i . Uma vez calculadas, essas submatrizes são introduzidas na matriz Φ juntamente com as demais que são fixas. Em seguida, $Q(r+1)$ ¹⁴ é calculada como:

$$Q(r+1) = \frac{C - B\Phi^T - \Phi B^T + \Phi A\Phi^T}{N} \quad (3.60)$$

- Dempster et al [31] demonstraram a convergência do algoritmo *EM* para situações bastante gerais, inclusive para a família exponencial (que inclui a distribuição Gaussiana).

3.6 Critérios para Avaliação de Desempenho

Na literatura são utilizados diversos critérios para avaliação do desempenho dos algoritmos, e.g., [23, 26]. Relacionando com os mais comumente utilizados, voltados às previsões, podem-se citar (o Apêndice F mostra o relacionamento entre alguns critérios listados):

- Erro Quadrático Médio (*EQM*), que é definido como:

$$EQM = E[(Y_{k+n}^* - Y_{k+n})(Y_{k+n}^* - Y_{k+n})^T] \quad (3.61)$$

onde $Y_{k+n}^* = H\Phi^n \bar{X}_k$ é a projeção do estado estimado n passos à frente. Às vezes utiliza-se também o *REQM*, ou seja, a raiz quadrada do *EQM*. Este critério é o mais utilizado nas comparações entre algoritmos. O *EQM* percentual, para a projeção a n passos é estimado como:

$$EQM(\%) = \frac{1}{N} \sum_{i=1}^N \frac{(Y_{i+n}^* - Y_{i+n})(Y_{i+n}^* - Y_{i+n})^T}{(Y_{i+n})(Y_{i+n})^T}$$

¹⁴Em todos os lugares em que está escrito Q deveria estar $\Gamma Q \Gamma^T$, o que só não foi feito por simplicidade. Na aplicação do algoritmo, é importante que se mantenham nulos todos os elementos da matriz Q que não serão estimados.

Mesmo quando $N = 1$, em que não há significado estatístico como esperança matemática, se utilizará a denominação EQM (ou $REQM$).

A estimação dos critérios dos critérios EMP , EST e VP , que serão abordados a seguir, ocorre de forma similar.

- Erro Médio da Projeção, que é definido como:

$$EMP = E[Y_{k+n}^* - Y_{k+n}] \quad (3.62)$$

- Estabilidade da Projeção, que é definida como:

$$EST = E[(Y_{k+n}^* - Y_{k+n|k-1}^*)(Y_{k+n}^* - Y_{k+n|k-1}^*)^T] \quad (3.63)$$

Os testes realizados mostraram que os resultados apresentando melhor estabilidade¹⁵ também apresentam melhores projeções n passos à frente, ou seja, basta verificar o critério de estabilidade das projeções ou o EQM da projeção n passos à frente. Isto se explica porque tanto o critério de estabilidade como a projeção a n passos utilizam potências da matriz de transição de estado.

- Variância da Projeção, que é definida como:

$$VP = E[(Y_{k+n}^* - E[Y_{k+n}^*])(Y_{k+n}^* - E[Y_{k+n}^*])^T] \quad (3.64)$$

Pode-se mostrar que $EQM = EMP(EMP)^T + VP + R_{k+n}$ (Apêndice F). Mas, em geral, a covariância R_{k+n} é estimada com precisão e se admite $R_{k+n} = R = \text{constante}$. Portanto, é suficiente testarem-se dois entre os três critérios.

- Correlação Branca do Erro de Previsão

Se o algoritmo estiver funcionando adequadamente, o erro da previsão um passo à frente (em relação ao valor observado), ou resíduo, deve possuir um padrão descorrelacionado.

- AIC Critério de Informação de Akaike

Para se determinar um entre vários modelos pré-selecionados (por exemplo um apenas com tendência, outro com tendência e sazonalidade), utiliza-se aquele com o menor AIC - *Akaike Information Criterion*, onde [20, 19]:

$$AIC = -2 \log L + 2(n^{\circ} \text{ de parâmetros}) \quad (3.65)$$

O $\log L$ é dado por (3.59)¹⁶, e o n° de parâmetros corresponde ao tamanho de X_k somado ao número de elementos desconhecidos de Q_k , R_k e Φ . Como exemplo, um modelo sazonal para dados mensais, do tipo dado por (3.14), possui 11 estados, uma

¹⁵O Apêndice C aborda a questão da estabilidade do Filtro de Kalman de uma maneira geral, e não apenas a estabilidade das projeções.

¹⁶Por inspeção de (3.59) nota-se que essa equação corresponde a um erro quadrático normalizado para a projeção um passo à frente.

variável de perturbação da equação do sistema e uma variável de perturbação da equação de observação (corresponde a um elemento desconhecido em cada uma das respectivas covariâncias), o que significa $2 \times (11 + 1 + 1) = 26$ adicionado ao *AIC*. Para ser viável, a inclusão de sazonalidade em um modelo aditivo deve implicar na redução mínima de 26 ao valor de $-2 \log L$. O apêndice G apresenta vários testes em diversos tipos de modelos, nos quais se utiliza o critério *AIC*.

O desempenho do filtro deve ser analisado após o período transitório (como exceções, existem filtros, geralmente mais simplificados, tais como os filtros robustos, que são projetados para trabalhar com poucas observações que englobam basicamente o período transiente) [12, 45]. Seja m o período do transiente. Para $k \leq m$, o ganho do filtro, em geral, é fortemente influenciado pela relação \bar{P}_0/R e para $k > m$, o ganho do filtro depende da relação Q/R (Apêndice D). Os testes efetuados neste trabalho, em cinco rotas de tráfego, mostraram que o ganho se estabiliza em torno da décima observação¹⁷.

3.7 Aplicação do Algoritmo *EM*

O desempenho do filtro depende do conhecimento de seus parâmetros \bar{X}_0 , \bar{P}_0 , R , Q , e Φ .

Quando os parâmetros são aproximadamente invariantes no tempo, mas desconhecidos, o algoritmo *EM* estima-os com precisão, conforme os resultados práticos obtidos por Girolami e Ursini em quarenta rotas de tráfego para o modelo com dois estados [45].

O algoritmo *EM* foi aplicado aos modelos da Seção 3.3, o que se encontra detalhado no Apêndice G. Pode-se verificar neste apêndice que a estimação de Φ na projeção a n passos é de ressaltada importância. Os testes realizados ilustram a convergência dos parâmetros em cinco rotas de tráfego e na série temporal IPI (Índice de Produto Industrial) de Morettin [7]¹⁸. Esses testes se utilizam do *AIC - Akaike Information Criterion* (Seção 3.6) com o objetivo de comparar diferentes modelos da Seção 3.3.

Testes do Apêndice G

Os testes do Apêndice G possibilitam várias considerações e conclusões:

- No teste da primeira rota de tráfego, Marília 33 para a Central Trânsito, T44, com 24 observações disponíveis, o modelo de tendência polinomial com dois estados, (3.8), teste d), apresenta o menor valor para o *AIC* (igual a 53,88). No entanto, o modelo que apresenta o menor fator $-2 \log L$ (igual a 44,56) é o modelo dado por (3.10) e (3.11), variante 1 (com Φ também estimado), teste c). Isso indica que apesar do modelo de tendência polinomial ser o melhor pelo critério *AIC*, por ter menos parâmetros a serem estimados, com o aumento do número de observações essa situação muda porque a parcela do critério *AIC*, relativa ao número de parâmetros (equação (3.65)), é fixa¹⁹;

¹⁷Para verificação dessa estabilização os testes foram feitos observando-se a norma euclidiana do ganho no instante k , comparada com a norma euclidiana do ganho em $k + 1$.

¹⁸Por exemplo, as rotas de tráfego de São Paulo para Santos e vice-versa e São Paulo para Caraguatatuba e vice-versa, tipicamente sazonais, não foram testadas porque não foram obtidos seus dados. Devido a isso, a utilização da série IPI é fundamental.

¹⁹Além disso, outros testes no decorrer do capítulo mostram que o modelo dado por (3.10) e (3.11), variante 1 (com Φ também estimado) possui a melhor projeção a n passos (maior estabilidade da projeção).

- O teste da primeira rota de tráfego também evidencia que o modelo aditivo com uma parte de tendência dada por (3.10) e (3.11) e uma parte de um modelo $AR2$, dado por (3.27), pode ser descartado porque além de ter o maior número de parâmetros entre os modelos testados, não foi o que apresentou o menor valor para o fator $-2\log L$. Portanto, nem mesmo com o aumento do número de observações²⁰ este modelo terá melhor desempenho que os demais;
- Ainda com relação à primeira rota, o fato de se utilizar o suavizador de De Jong ou o suavizador com J , não trouxe diferença significativa, embora o suavizador de De Jong tenha apresentado um resultado ligeiramente melhor;
- A primeira rota também é testada para um conjunto de dados quadrimestrais (média aritmética dos dados de quatro meses), estimando-se os parâmetros com o uso de oito observações disponíveis. Neste caso, o modelo que apresenta o menor valor para o critério AIC é o modelo mais simples, dado por (3.7). Devido ao pequeno número de observações, o fator relativo ao número de parâmetros (equação (3.65)) terá um peso relativo maior. Por exemplo, a diferença percentual do valor AIC ao se utilizar o modelo dado por (3.10) e (3.11), variante 1 (com Φ também estimado), em relação ao modelo dado por (3.10) e (3.11) aumentou de 12% para 24% ao se passar de 24 para 8 observações disponíveis;
- Com relação à segunda rota testada, Botucatu Trânsito para Bauru Trânsito, T4942, também com 24 observações disponíveis, a utilização do suavizador com J acarreta um aumento no valor do fator $-2\log L$, indicando ainda que este fator possui convergência mais lenta nesse caso;
- Ainda com relação à segunda rota, o modelo aditivo formado por uma parte de tendência dada por (3.10) e (3.11) com Φ fixo e um modelo $AR2$, também mostra que deve ser descartado. O motivo do descarte é que o modelo possui uma número maior de parâmetros que os outros modelos comparados e não possui o menor fator $-2\log L$;
- A terceira rota testada, Bauru Trânsito para Botucatu Trânsito, T4249, com 24 observações disponíveis, da mesma forma que a primeira, também apresenta um valor AIC 12% maior ao se comparar o modelo dado por (3.10) e (3.11) com e sem a matriz Φ estimada;
- A quarta e quinta rotas testadas, Bauru Trânsito para Marília Trânsito, T4244, e Marília Trânsito para Bauru Trânsito, T4442, com 24 observações disponíveis em cada uma, apenas são incluídas na fase de teste para verificação da aplicabilidade do algoritmo e escolhe-se para ambas o modelo dado por (3.10) e (3.11), variante 1 (com Φ também estimado);

²⁰Evidente que o número de observações influencia a quantidade de informações presentes nos dados. Por exemplo, dados com periodicidade anual não incorporam o efeito sazonal, que só pode ser percebido em observações de meses consecutivos. Desta forma, evidentemente, dados mensais podem requerer um modelo mais complexo para modelar também a sazonalidade. Por outro lado, para dados em tempo muito curto (quase tempo real), por exemplo, o efeito sazonal vai corresponder a uma tendência localizada e não requer modelagem como efeito sazonal, apenas como tendência. Também modelos mais simplificados, que, às custas de uma menor precisão, são mais robustos (Capítulo 4), também não requerem muitos estados (ou muitos efeitos modelados).

- A série IPI (Índice de Produto Industrial) é testada para diversos modelos para sete observações quadrimestrais (média aritmética de quatro observações mensais) disponíveis. Entre os modelos testados, o modelo mais simples (devido a possuir o menor número de parâmetros a serem estimados), dado por (3.10) e (3.11), variante 1 (com Φ também estimado) é o que teve o menor *AIC*;
- Os testes de sazonalidade para a série IPI indicam, de fato, a ocorrência do efeito sazonal. Isto pode ser verificado pela comparação do teste a) com o teste c), em que há uma redução do fator $-2 \log L$ de 44,18 para 40,20 e também do teste b) com o teste d), em que há uma redução de 46,78 para 42,20. Essas reduções no valor do fator $-2 \log L$ indicam que o modelo sazonal melhora a precisão das projeções e deve ser empregado para um número maior de observações;
- Os testes do efeito de variação diário para a série IPI também indicam a ocorrência deste efeito, porém, de maneira mais fraca. Isto pode ser verificado pela comparação do teste d) (que possui apenas sazonalidade) com o teste f) (que possui sazonalidade e efeito de variação diário), em que há uma redução do fator $-2 \log L$ de 42,50 para 42,12 e do teste c) com o teste e), em que há uma redução de 40,20 para 39,28. Embora notada a existência do efeito, devido ao fato das observações, neste caso, serem quadrimestrais, a influência dos diferentes dias da semana no quadrimestre é pequena e não compensa a utilização deste modelo mesmo com um número maior de observações. Se as observações são mensais, este efeito é mais significativo (diferentes números de dias da semana nos diversos meses);
- Outro aspecto a ser comentado é o da sensibilidade do modelo, devido a se trabalhar com apenas sete observações, em relação aos parâmetros iniciais. Os testes a), c) e d), para a série IPI, utilizam modelos com elementos unitários (dígitos um) na matriz de covariância do erro de modelagem (os não nulos) e os testes b), d) e f) utilizam 200 para os elementos da matriz de covariância do erro de modelagem. Todos os outros parâmetros utilizados são iguais em cada par de testes. O teste b) em relação ao a) teve um valor do fator $-2 \log L$ (e do valor *AIC* porque cada par de testes possui o mesmo número de parâmetros) aumentado de 44,18 para 46,78; o teste d) em relação ao c) teve um aumento de 40,20 para 42,50 e o teste f) em relação ao e) teve um aumento de 39,28 para 42,12. Isto evidencia a necessidade de um melhor conjunto de parâmetros iniciais quando se possuem poucas observações;
- Os dois tipos de suavizadores utilizados, com *J* e De Jong, nem sempre convergem a um ponto estacionário. Algumas vezes, após atingir um ponto de máximo valor do fator $\log L$, o suavizador começa a diminuir novamente (tornar-se mais negativo). No entanto, os valores obtidos continuam próximos do ponto máximo (mudanças no valor decimal).

A Projeção n Passos à Frente

Diversos testes foram realizados com a projeção um passo à frente. A Figura 3.4 mostra os valores observados, filtrados e projetados a um passo da Rota T44, Marília 33 para a Central Trânsito, para o modelo com dois estados dado por (3.10) e (3.11). Para realizar a filtragem e a projeção supôs-se $T_0 = Y_0$ (o valor do nível igual à primeira observação), $\dot{T}_0 = 0,015Y_0$ (a taxa de crescimento suposta 1,5% do valor do nível), $Q_{11} = R$, $Q_{22} = R$,

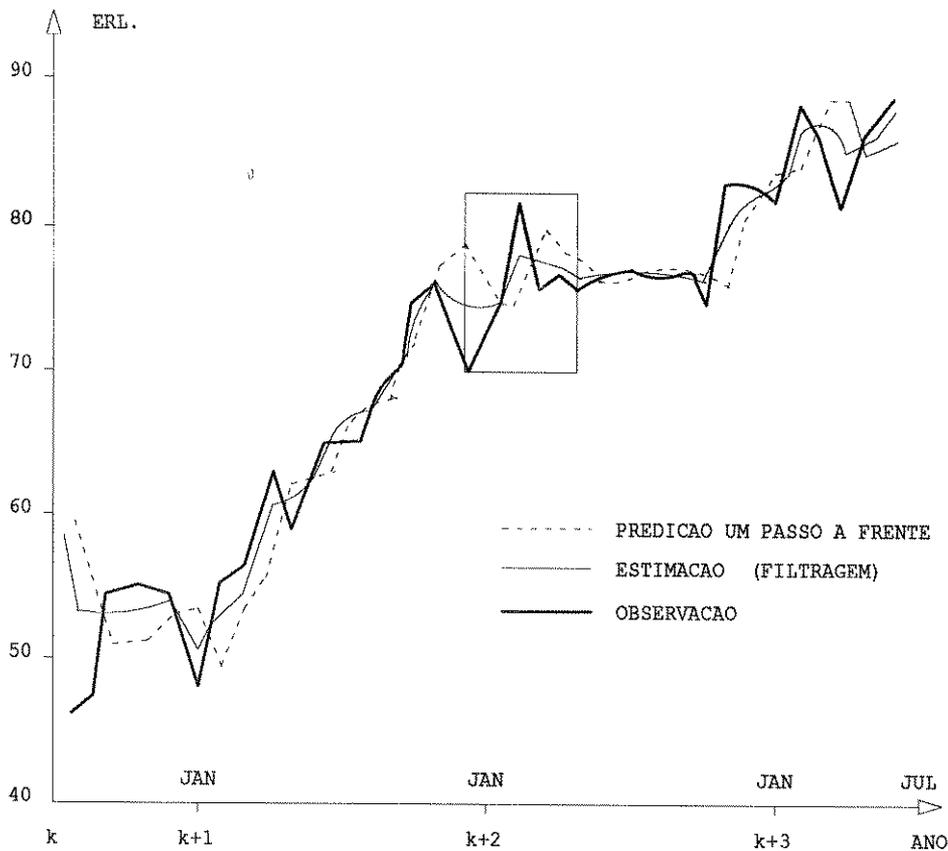


Figura 3.4: Projeção a um Passo

R estimado de acordo com Moreland (será detalhado no capítulo seguinte) e $\Phi = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ [44].

Pela Figura 3.4 ²¹ pode-se verificar que os erros de projeção a um passo são pequenos e estão em torno de 5%. Essa tendência de pequenos erros permanece mesmo para outros valores dos parâmetros, conforme [45, 51, 52]. Outras 111 rotas de tráfego também foram testadas utilizando-se um conjunto fixo de parâmetros para todas elas ²² [45, 51, 52], e em menos que 10% dos casos o erro supera 10%, mas se o horizonte de projeção for aumentado, os erros começam a ser significativos.

A Figura 3.5 mostra as projeções a um e a doze passos à frente da Rota Marília 33 para a Central Trânsito, T44, em duas situações:

- Parâmetros estimados da mesma forma que na situação anterior (projeção a um passo);

²¹O retângulo da Figura 3.4 indica uma região em que várias considerações foram feitas sobre dados de campo, como a figura 3.2 que mostra o funcionamento do suavizador com J (Seção 3.4) e a figura 4.3 do Capítulo 4.

²²No capítulo seguinte, ao ser abordado o Algoritmo de Projeção Sequencial Robusto, esta questão voltará a ser enfocada.

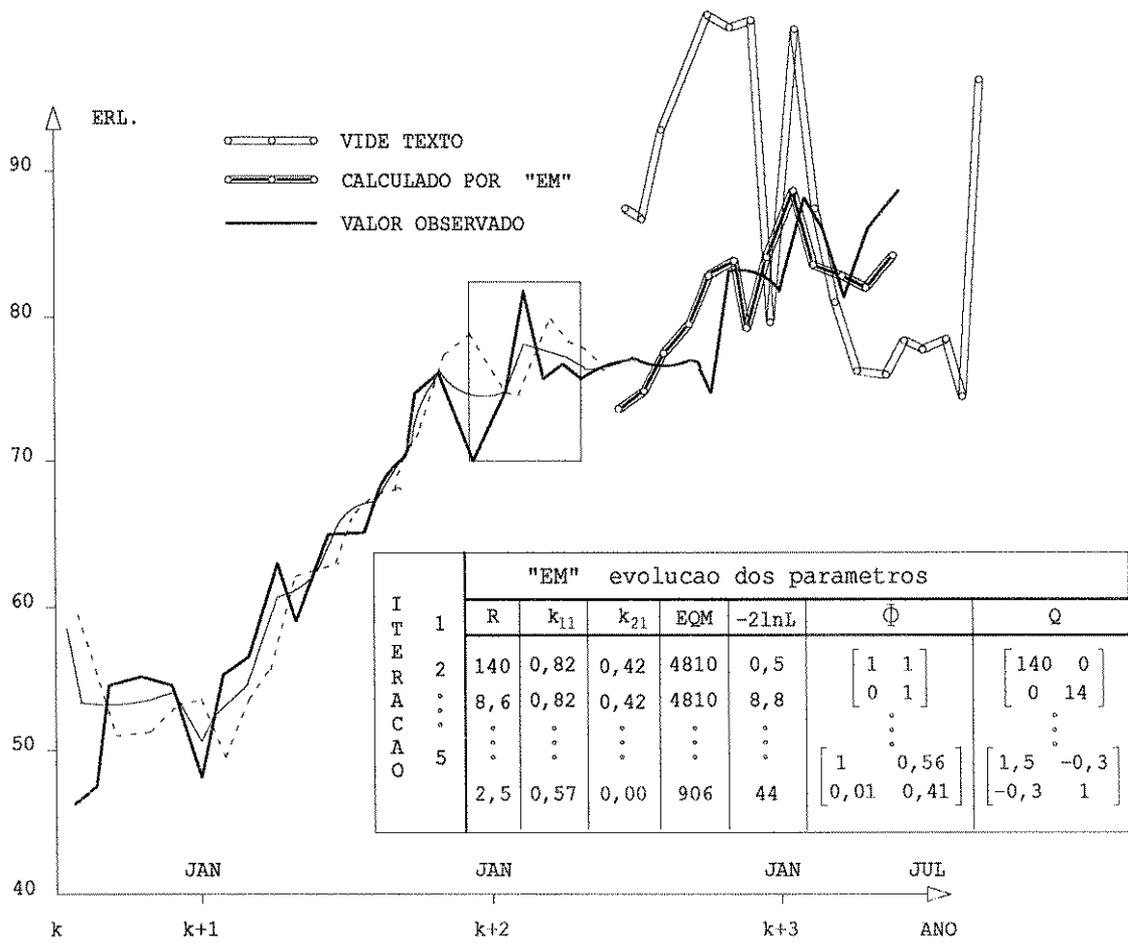


Figura 3.5: Projção Doze Passos à Frente

- Parâmetros estimados pelo procedimento *EM*.

Além disso, a Figura 3.5 (que é uma continuação da figura 3.4) também fornece a evolução do valor dos parâmetros com o número de iterações, bem como o valor do erro quadrático médio, *EQM*, dado por (3.61) e do valor de $-2 \log L$, dado por (3.59). Nesta figura fica evidente que a projeção a doze passos, utilizando o algoritmo *EM*, apresenta valores mais precisos e estáveis (não apenas mais precisos, mas também com flutuações bem menores em relação à média). A Figura 3.6 complementa a informação contida na Figura 3.5 com relação à evolução dos parâmetros. A Figura 3.6 (a) mostra que os *EQMs* da projeção a um passo, para as Rotas Bauru trânsito para Marília trânsito, T4244, e Marília trânsito para Bauru trânsito, T4442, decrescem conforme aumenta o número de iterações. Nota-se, no entanto, que a partir da 5ª iteração a taxa de decrescimento diminui. Na Figura 3.6 (b), a título de exemplo, é mostrada a evolução do valor estimado da covariância do erro de observação. A Figura 3.6 (c) mostra que o valor do incremento de crescimento (\hat{T}_k) também estabiliza após a 5ª iteração. Esse fato é importante porque a projeção n passos à frente, ao utilizar um valor suavizado de incremento e um valor de Φ também específico para cada rota, acaba fazendo com que $Y_{k+n}^* = H\Phi^n \bar{X}_k$ seja mais estável e mais preciso, conforme a Figura 3.5.

Portanto, a projeção 12 passos à frente, para o primeiro caso mencionado apresentou uma variância do erro de projeção bem maior que para a aplicação do algoritmo *EM*.²³ A projeção 12 passos à frente é calculada como $Y_{k+12}^* = H\Phi^{12} \bar{X}_k$, e como os erros da projeção a um passo não são elevados, deduz-se que esses erros são provocados pela utilização de uma matriz de transição de estado fixa (não estimada pelo conjunto dos dados), o que não ocorre com o algoritmo *EM*.

Após a aplicação do algoritmo *EM*, com os dados das N primeiras observações disponíveis e a convergência dos parâmetros, a aplicação do Filtro de Kalman é realizada após a $(N + 1)$ -ésima observação com os valores fixos dos parâmetros obtidos. Quando a dinâmica do sistema variar com o tempo (ou as perturbações não forem estacionárias), o algoritmo *EM*, que para o problema de projeção de tráfego seria a aproximação mais próxima do ótimo, pode produzir maiores erros que aqueles fornecidos por algoritmos mais simples.

Para ilustrar este fato, a Tabela 3.1 mostra os valores percentuais do erro quadrático médio da projeção doze passos à frente da Rota T6362, em que os resultados da coluna da esquerda foram obtidos com a aplicação do algoritmo *EM* e os resultados da coluna direita através da aplicação de um algoritmo de projeção sequencial robusto (e mais simplificado)²⁴, proposto por Moreland [12]. O fato de se utilizar projeções para comparação a partir da vigésima observação é porque o algoritmo *EM* necessita das primeiras observações (no caso 19) para calcular os parâmetros. Como pode-se notar, a média aritmética do *REQM* do Algoritmo Robusto é menor do que a do algoritmo *EM*, mas o mais importante é que o seu desvio-padrão é aproximadamente a metade do desvio-padrão do *EM*, indicando previsões mais estáveis. Além disso, o Algoritmo Robusto não requer a utilização de valores iniciais para estimar seus parâmetros (Capítulo 4, onde será proposto o Algoritmo de Projeção Sequencial Robusto).

Este fato justifica, em alguns casos, a utilização de algoritmos mais simples que o *EM*.

²³Foram usadas 43 observações mensais, sendo 31 para se estimarem os parâmetros e 12 para se verificarem os resultados.

²⁴O conceito de robustez será explicado nas seções seguintes.

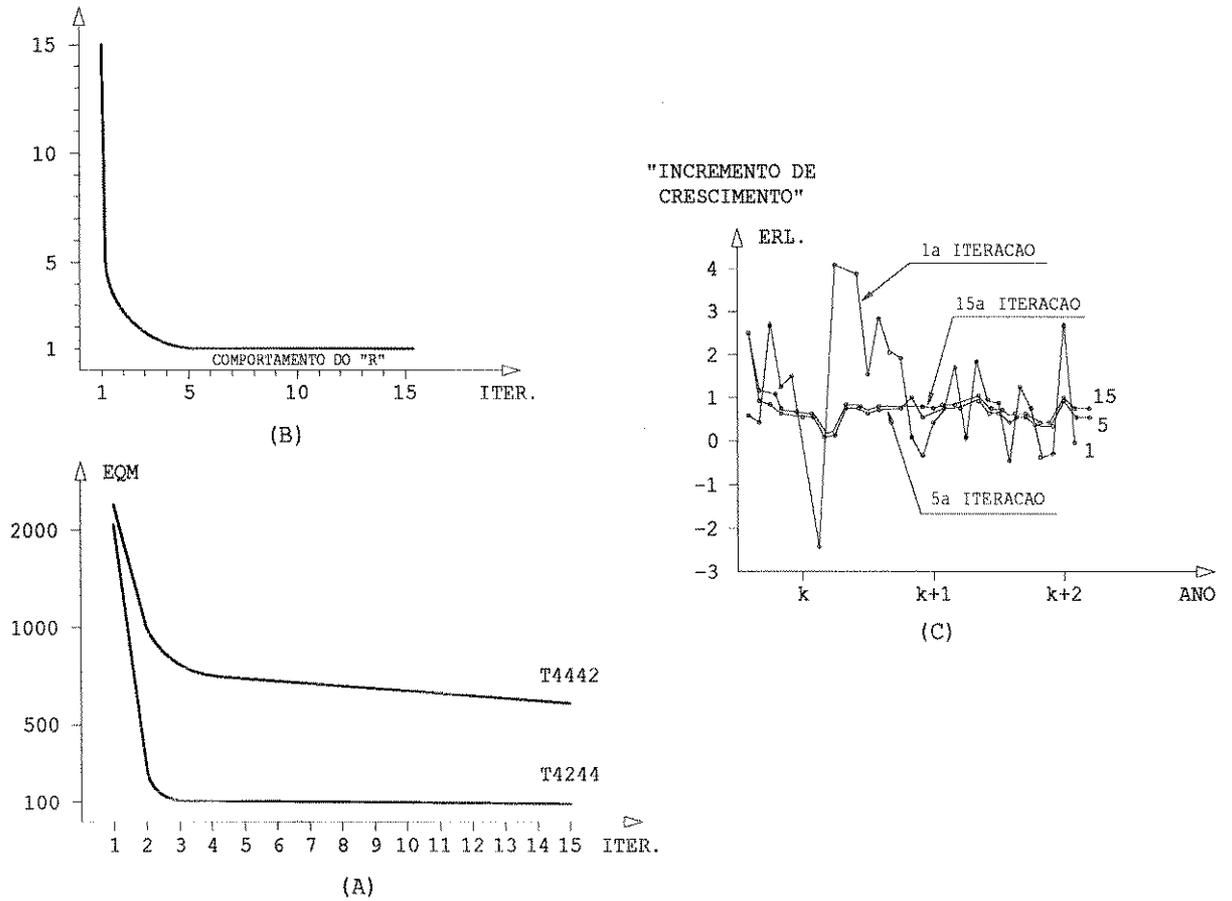


Figura 3.6: Aplicação do Algoritmo *EM*

Tabela 3.1 Comparação dos *REQMs* produzidos pelos algoritmos *EM* e o proposto por Moreland [12], para a Rota T6362.

PROJ. <i>EM</i>	PROJ. Moreland
$E_{20} = 16,48$	$E_{20} = 24,03$
$E_{21} = 17,49$	$E_{21} = 27,56$
$E_{22} = 17,09$	$E_{22} = 21,43$
$E_{23} = 8,08$	$E_{23} = 23,31$
$E_{24} = 8,24$	$E_{24} = 22,70$
$E_{25} = 1,96$	$E_{25} = 38,18$
$E_{26} = 8,18$	$E_{26} = 24,50$
$E_{27} = 11,38$	$E_{27} = 15,51$
$E_{28} = 42,62$	$E_{28} = 14,16$
$E_{29} = 54,75$	$E_{29} = 1,21$
$E_{30} = 58,47$	$E_{30} = 2,97$
<i>média</i> = 22,25	<i>média</i> = 19,60
<i>desv.pad.</i> = 19,98	<i>desv.pad.</i> = 10,68

E_i é o valor do *REQM* (%) da projeção a 12 passos a partir da projeção i .

A Figura 3.7 (a) mostra os dados da Tabela 3.1, e as projeções, bem como os valores observados, são mostrados na Figura 3.7 (b). Pela figura visualiza-se que a projeção de Moreland, embora não tão precisa quanto a projeção do *EM*, também não comete erros tão grandes quando há mudanças bruscas na tendência da série temporal.

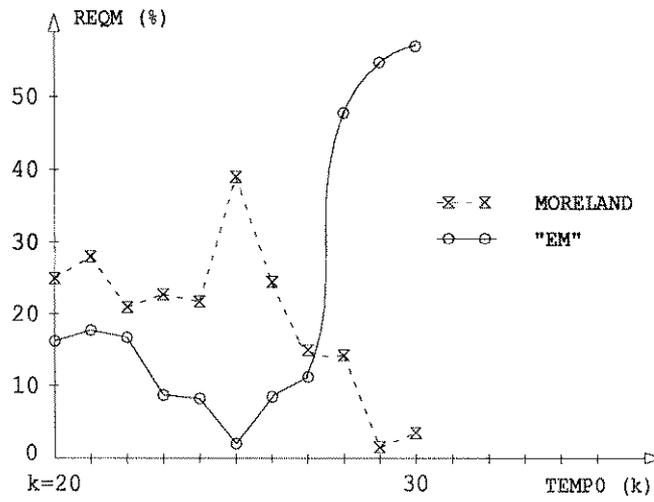
3.8 Otimalidade, Sensibilidade e Robustecimento

Sabe-se que quanto mais o filtro se aproxima do ótimo (por exemplo, com maior número de estados e com os parâmetros estimados com maior precisão), geralmente, mais esforço de computação é necessário e menor é o seu *EQM*. O filtro mais simples (por exemplo, aquele com menos estados e que utiliza ganho fixo) irá requerer um menor esforço computacional, mas, geralmente, apresentará um *EQM* maior. O que deve ser buscado é um compromisso entre essas duas características, dando prioridade a uma ou a outra conforme a aplicação em vista.²⁵

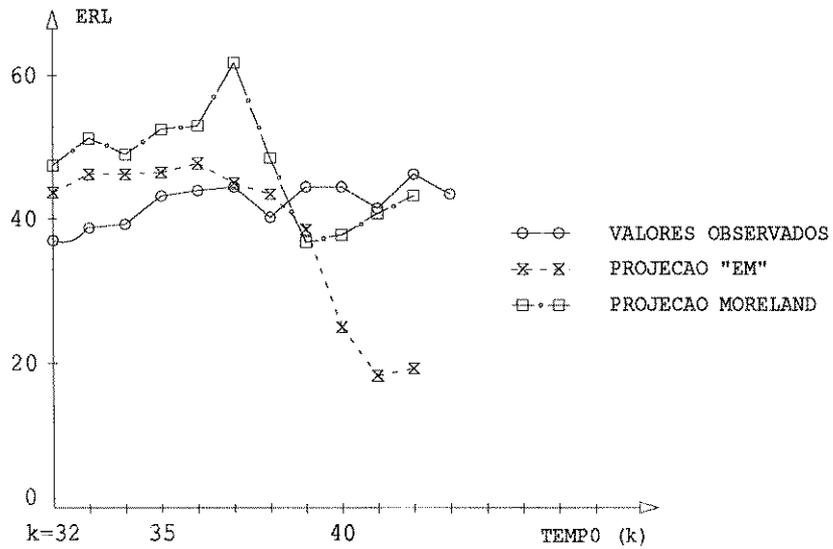
O exemplo da seção anterior mostrou uma situação real em que um filtro mais simples pode obter melhores resultados. O filtro mais complexo e supostamente mais próximo do ótimo, tem, geralmente, os parâmetros estimados com maior precisão e, geralmente, maior número de variáveis de estado (o Apêndice G mostra diversos testes com modelos de diferentes tamanhos de estado). Com relação ao problema do tamanho do vetor de estado, sabe-se que:

- O número de multiplicações necessárias para se computar a matriz de covariância do erro de estimação, em filtros baseados no Filtro de Kalman, varia com a terceira potência da dimensão do estado [24];
- Com a utilização de poucas variáveis de estado, o filtro pode ter um *EQM* maior, mas também com a inclusão de muitas variáveis de estado há a questão de se requerer

²⁵Na prática, não existe o filtro ótimo ideal. O que existem são aproximações melhores ou piores. Por filtro ótimo, neste contexto, entende-se aquele com tamanho de estado compatível com o padrão da série que modela (tendência, sazonalidade, etc.) e com parâmetros otimizados pelo algoritmo *EM*.



(a)



(b)

Figura 3.7: Comparação entre o Algoritmo de Moreland e o Algoritmo *EM* para a Rota T6362

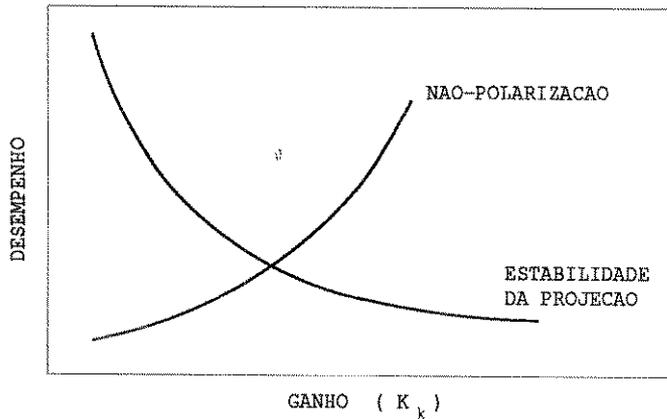


Figura 3.8: Compromissos Empíricos de Desempenho

um maior número de observações para que elas sejam convenientemente estimadas (Apêndice G).

Portanto, um filtro com menos variáveis de estado poderá vir a ter um *EQM* dentro de limites pré-estabelecidos e requerer um menor número de observações para estimar seus parâmetros ²⁶.

Como o valor de todos os parâmetros reflete-se nos valores da seqüência de ganho, pode-se analisar esse problema através da avaliação do desempenho em função do ganho. A Figura 3.8 ilustra genericamente essa questão [23].

Existem dois tipos de alterações na dinâmica do processo — observações aberrantes (*outliers*), que se devem a grandes erros de medição, e eventos determinísticos não previstos —, e que podem ser contornados com valores adequados de ganho. Conforme o ganho aumenta, as projeções do filtro se tornam menos estáveis porque o filtro se apresenta mais sensível às novas observações. Por outro lado, o filtro se torna menos polarizado porque tem condições de processar eventos determinísticos (que no caso do tráfego telefônico são, por exemplo, mudanças de tarifa não-previstas).

Se os parâmetros do filtro estiverem incorretamente estimados (ou se estiverem corretamente estimados, como o faz o algoritmo *EM*, mas variarem no tempo), o resultado será um ganho que não corresponderá ao ganho ótimo. Para essas condições, a equação mais adequada para se calcular a covariância real do erro de estimação passa a ser²⁷ [33, 24, 32]:

$$\bar{P}_k = (I - K_k H) P_k^* (I - K_k H)^T + K_k R K_k^T \quad (3.66)$$

Ao se avaliar um filtro com um ganho qualquer, a covariância do erro de estimação é dada por (3.66) (Apêndice J), mas a covariância do erro de predição, P_k^* , segue a mesma equação dada por (3.36).

A covariância real do erro de estimação pode, desta maneira, fornecer o erro teórico ao se adotar diferentes valores de ganho. Portanto, é possível se ter uma idéia teórica do

²⁶Existe também o problema da precisão da máquina que executa o processamento. Quanto maior o processamento, maior a propagação do erro. Esta questão é fundamental para algoritmos que processam uma quantidade muito grande de observações e não é relevante para dados de tráfego mensais e anuais.

²⁷A equação (3.66) pode ser verificada na dedução das equações do filtro.

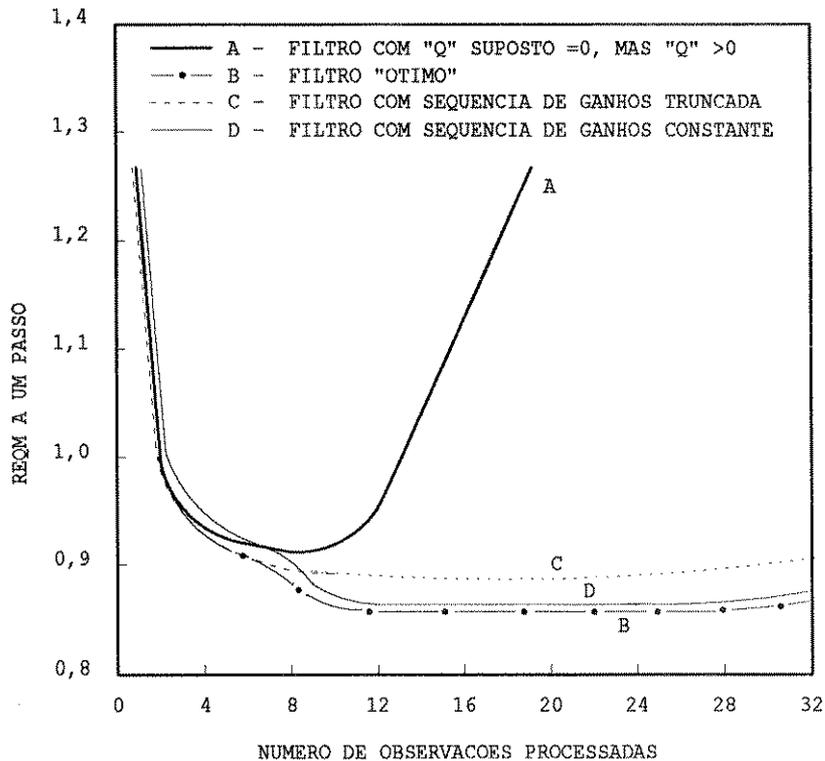


Figura 3.9: Desempenho Teórico do Algoritmo

erro cometido em relação a qualquer parâmetro desconhecido (que se reflete no ganho) em função do número de observações processadas.

A Figura 3.9 de [29, 26] mostra a variação do *EQM* da projeção a um passo em função do número de observações processadas (o *EQM* teórico é a raiz quadrada do elemento \bar{P}_{11} da matriz \bar{P}_k , para o modelo univariável). Pode-se observar que ao se utilizar $Q = 0$ (a exceção é quando o filtro é um mero estimador de parâmetros, ou seja, realiza uma regressão de efeitos fixos, vide o exemplo do efeito de variação diário na Seção 3.3), sem ruído de modelagem, o filtro perde a capacidade de adaptação e apresenta divergência porque a seqüência de ganho tende a zero.

A divergência do filtro pode ser causada por Q próximo de zero ou para R elevado, porque o funcionamento do filtro depende da relação $\frac{Q}{R}$ (Apêndice D e [23]).

Uma forma de se evitar essa divergência é truncar a seqüência de ganho:

$$K_k^t = \begin{cases} K_k, & \text{se } 1 < k \leq k^* \\ K_k^*, & \text{se } k > k^* > 1 \end{cases} \quad (3.67)$$

O valor de k^* para o qual se truncar o ganho é determinado empiricamente e exemplos serão apresentados e discutidos no Capítulo 4.

Uma outra maneira de se contornarem os problemas citados é através da utilização de procedimentos robustos.

Por robustez, entende-se [37]:

“ Se um estimador ou teste, particularmente adequado quando a população tem determinada distribuição, tem propriedades que **resistem** bem quando a verdadeira distri-

buição tem um pequeno afastamento em relação à pressuposta, diz-se um estimador ou teste robusto”. Por erro “pequeno”, entende-se:

- Uma pequena proporção das observações está sujeita a grandes erros;
- Uma apreciável proporção das observações está sujeita a pequenos erros.

O termo **resistem bem** do texto original significa que o estimador ou o teste é pouco sensível a “pequenos” afastamentos da verdadeira distribuição em relação à pressuposta.

Um exemplo do primeiro tipo é a presença de observações duvidosas, anômalas ou aberrantes (*outliers*). Um exemplo do segundo tipo são os erros causados por arredondamento ou tabulagem (quando as observações são em pequena quantidade, o segundo tipo tem pouca importância). O robustecimento para o primeiro tipo corresponde, por exemplo, à utilização de intervalos de confiança, que serão abordados nos Capítulos 4 e 5. O robustecimento para o segundo tipo corresponde, por exemplo, à utilização de variáveis aleatórias para correção dos erros devido à precisão da máquina, como no Capítulo 5.

A utilização do filtro robusto é fundamental quando se dispõe de poucas observações. Esse filtro pode ser robustecido com relação a dois aspectos:

- Pela adoção de um conjunto robusto de parâmetros (estimadores robustos dos parâmetros), em vez de incorrer em erros grosseiros ao se tentar estimá-los com poucas observações;
- Pela adoção de testes com intervalos de confiança e assim evitar que observações duvidosas (*outliers*) possam contaminar os valores estimados.

Portanto, em muitas situações é preferível perder um pouco de precisão, mas ganhar robustez (um resultado individual pode ser pior, mas o resultado global será melhor).

O Capítulo 4 abordará o Algoritmo de Projeção Seqüencial Robusto proposto neste trabalho.

3.9 Aplicações ao Problema de Projeção de Tráfego

Existem três aplicações principais em que as previsões de tráfego de entroncamento²⁸ podem ser utilizadas:

Planejamento de expansão: Requer previsões para dois ou mais anos à frente e trabalha com dados anuais;

Administração de tráfego: Requer previsões de curto prazo, um a dois anos, e pode trabalhar com dados anuais ou com dados mensais;

Gerência de rede: Requer previsão para dados coletados em intervalos de cinco minutos.

É importante ressaltar que segundo estudos da Bell System para os EUA [23], as rotas de tráfego têm um a dez anos de dados relevantes, sendo valores típicos três a cinco anos²⁹. O número limitado de observações inviabiliza algoritmos que requeiram muitas observações para o processo de identificação e estimação dos parâmetros.

²⁸No caso de projeção de tráfego de assinantes, outras situações podem ocorrer.

²⁹Esse período de duração se deve às mudanças na rede, tais como estrutura de encaminhamento, desvio do tráfego para rotas alternativas, etc.

Além disso, o número de rotas de tráfego é muito grande (no estado de São Paulo existem em torno de 10000 rotas). Assim, são requeridas projeções em grande quantidade e a todo instante (ano, mês, cinco minutos) e há a necessidade de padronização e simplificação visando a automatização do sistema. O uso de algoritmos robustos, mais simples, apesar da menor precisão, pode evitar erros grosseiros, sendo portanto um opção a ser considerada.

Para os casos de previsões mais a longo prazo (maiores que dois anos) e portanto para o **Planejamento de expansão**, a aplicação de algoritmos robustos, apoiados por outros algoritmos do tipo extrapolativo com projeção agregada, de ajuste de funções ou com variáveis explicativas, é recomendada. Ao se empregarem os algoritmos robustos, deve-se dar especial atenção à estimação da matriz de transição de estado (Capítulo 4).

Para as previsões de curto prazo (um a dois anos à frente) e portanto para a **Administração de tráfego**, dadas as considerações acima, os algoritmos robustos podem ser empregados, porém, em base mensal, dado que o número de observações disponíveis é maior. No Capítulo 4 há o desenvolvimento de um Algoritmo de Projeção Seqüencial Robusto (APS Robusto).

Para obtenção de maior precisão, às custas de um processamento maior, se houver dados suficientes, também é possível o uso de um algoritmo adaptativo para prazos de um a dois anos, em que os parâmetros são recalculados a cada nova observação. Esse algoritmo propicia um esforço computacional menor que aquele do algoritmo *EM* e tem uma precisão maior que a do APS Robusto. Sabe-se que os parâmetros requerem um certo número de observações para se estabilizarem. Assim, antes da entrada definitiva do algoritmo Adaptativo em operação, deve começar a funcionar o Algoritmo Robusto. Esse algoritmo, como proposto neste trabalho de tese, no Capítulo 5, recebe o nome de Algoritmo de Projeção Seqüencial Adaptativo Modificado, APSAM, porque reúne as características do APS Robusto quando se dispõe de poucas observações e as qualidades do algoritmo Adaptativo para a atualização dos parâmetros.

A característica da aquisição de dados a cada cinco minutos, portanto para a **Gerência de rede**, sugere o uso de um algoritmo de tempo real, o APSAM, conforme descrito na aplicação anterior ³⁰.

O algoritmo *EM* é utilizado como **suporte**, aplicado por amostragem em tempo não real (o APS Robusto, por exemplo, possui os parâmetros estimados por uma amostra da aplicação do algoritmo *EM*). Também para efeito de **identificação** — a aplicação do algoritmo *EM*, junto com o critério *AIC*, por exemplo, permite avaliar os efeitos de tendência, sazonalidade, etc. Sem dúvida, também para efeito de **comparação de precisão e validação** — já que na maioria dos casos, esse algoritmo fornece resultados superiores quando comparado com outros algoritmos de projeção de tráfego [45].

Os capítulos seguintes abordarão os algoritmos propostos: O **Algoritmo de Projeção Seqüencial Robusto** e o **Algoritmo de Projeção Seqüencial Adaptativo Modificado**.

³⁰Testes realizados no Capítulo 5 mostraram que a utilização do APS Robusto na iniciação (ou reiniciação) não mostrou melhorias significativas para aplicações em tempo real, voltadas às projeções de tráfego. Nesses casos, recomenda-se a utilização do APSAM apenas com o algoritmo Adaptativo.

3.10 Sumário e Conclusões

O presente capítulo apresentou o embasamento utilizado para os estudos propostos nos Capítulos 4 e 5. Foram mostrados:

- Os modelos a serem empregados;
- A forma de estimação dos parâmetros e do estado (algoritmo *EM* e Filtro de Kalman);
- Os critérios de verificação do desempenho;
- As questões de otimalidade, sensibilidade e robustecimento;
- O tipo de aplicação em função do prazo (e portanto do número de observações disponíveis).

Algumas conclusões do capítulo:

- Os resultados da aplicação dos diversos modelos da Seção 3.3 (no decorrer do capítulo e no Apêndice G) indicaram que, embora notada a existência de efeitos do tipo sazonalidade, efeito de variação diário, etc., nenhum dos modelos desempenhou de maneira significativamente diferente do modelo com dois estados representando tendência;
- O suavizador utilizado no algoritmo *EM*, empregando as recursões de de Jong, devido a não necessitar inverter a matriz P_k^* , equação (3.36), para o modelo univariável, deve ser o preferido, principalmente quando se utilizam grandes dimensões para o vetor do estado;
- Quando se dispõe de poucas observações é conveniente, ao aplicar o algoritmo *EM*, que se tenha valores iniciais adequados para \bar{P}_0 , conforme proposto neste capítulo;
- Os critérios preferidos para avaliação do desempenho são o *EQM(REQM)* e a variância das projeções, *VP*, porque são os mais comumente empregados e, portanto, mais práticos ao se comparar com outros trabalhos;
- A correta estimação da matriz de transição de estado é fundamental na estabilidade da projeção n passos à frente;
- Um modelo simplificado pode obter melhores resultados que outros modelos mais complexos;
- Além disso, também forneceu as bases para que surgissem as propostas dos novos algoritmos, o Algoritmo de Projeção Seqüencial Robusto (APS Robusto), que será apresentado no Capítulo 4, e o Algoritmo de Projeção Seqüencial Adaptativo Modificado (APSAM), que será focado no Capítulo 5.

O Algoritmo Robusto do Capítulo 4 é fortemente baseado na aplicação do algoritmo *EM*, como suporte para a estimação do seu ganho e de sua matriz de transição de estado. O Capítulo 5 se utiliza também, além dos conceitos gerais deste capítulo, de conceitos específicos, como por exemplo, a estimação da matriz de transição de estado.

Capítulo 4

Algoritmo de Projeção Seqüencial Robusto: Uma Proposta

Alguns algoritmos mais elaborados, como o algoritmo *EM* apresentado no Capítulo 3, utilizado para estimar os parâmetros em um modelo de espaço de estados com o Filtro de Kalman, podem apresentar desempenho inferior ao de algoritmos mais simples, em alguns casos específicos, além de requerer uma substancial carga de processamento em tempo não-real. Devido à carência de dados em muitas rotas de tráfego individuais e à grande quantidade de manipulações de dados necessárias para o planejamento da rede, é conveniente utilizar um algoritmo mais simples, mas robusto, que também pode ser automatizado. Este trabalho propõe um algoritmo simples e robusto para projeções de tráfego: O Algoritmo de Projeção Seqüencial Robusto. Os resultados mostram que esse algoritmo funciona adequadamente mesmo com poucas observações. Em geral, comparativamente ao algoritmo *EM*, o algoritmo proposto apresenta um desempenho ligeiramente inferior, no que diz respeito ao erro quadrático médio entre o valor projetado e o valor observado. Entretanto, requer menos esforço computacional além de ser mais robusto.

4.1 Introdução

A utilização do Filtro de Kalman para estimar o estado em um modelo de espaço de estados, com a utilização do algoritmo *EM* para estimar os parâmetros, foi abordada no Capítulo 3 e os resultados se mostraram adequados. No entanto, constatou-se que tal algoritmo requeria uma substancial carga de processamento em tempo não-real, além de desempenhar pior que algoritmos mais simples em alguns casos específicos. Além do mais, devido ao fato do planejamento da rede envolver manipulação de grande quantidade de dados e haver carência de dados em várias séries temporais relacionadas com rotas de tráfego, é conveniente utilizar-se um algoritmo de projeção mais simples que possa também ser automatizado.

Este trabalho propõe um algoritmo simples e robusto para projeções n passos à frente com o principal objetivo de testar a estabilidade de suas projeções. Além disso, este algoritmo encontra aplicação no planejamento a longo prazo. Comparativamente a um algoritmo similar, como o Algoritmo Robusto com ganho fixo proposto por Moreland [12], o algoritmo sugerido apresenta dois pontos favoráveis:

- Maior estabilidade para a projeção n passos à frente;

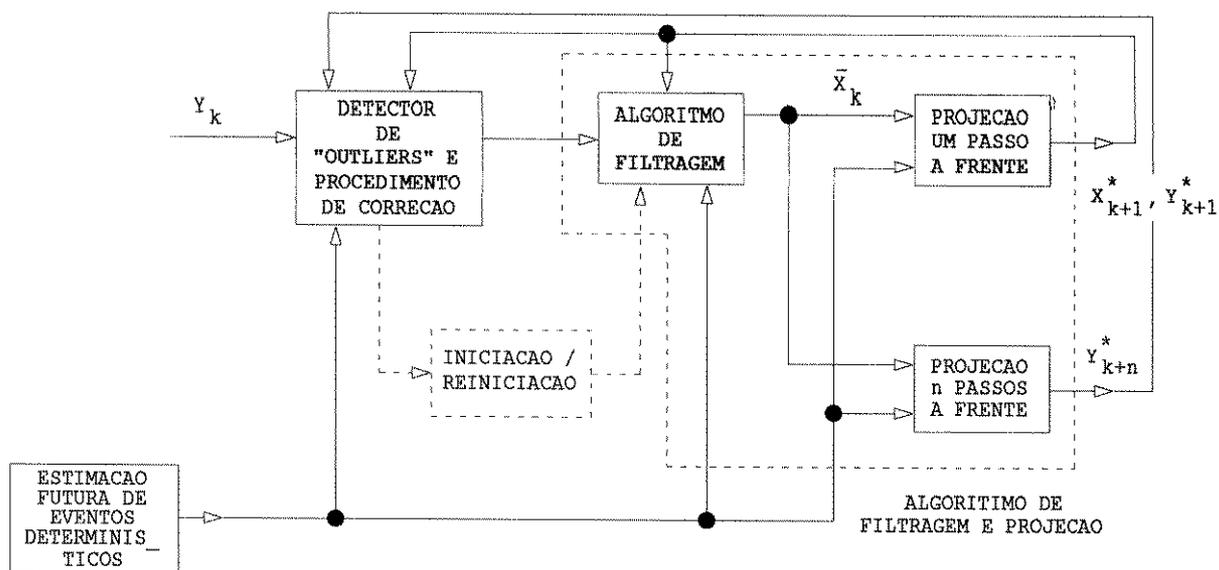


Figura 4.1: Algoritmo de Projeção Sequencial

- Maior robustez (menor sensibilidade) com relação ao valor verdadeiro do estado inicial.

A Figura 4.1 ilustra o funcionamento do Algoritmo de Projeção Sequencial (APS) proposto e é baseada em ilustração de [22, 23]. A seguir o APS Robusto é explicado brevemente ¹:

Algoritmo de Filtragem e Projeção: É o filtro em si. É baseado no modelo de Espaço de Estados utilizando duas variáveis de estado. Como usual, o estado desse modelo é estimado por meio do Filtro de Kalman;

Detector de Outliers: É utilizado para verificar se os dados observados estão dentro de um dado intervalo de confiança especificado;

Estimação de Eventos Determinísticos Futuros: É utilizado para corrigir o valor da carga projetada, dado que algum evento determinístico pode ocorrer (e.g., quando o número de assinantes é aumentado em uma dada rota de tráfego);

Iniciação/Reiniciação: Utilizado para iniciar/reiniciar o filtro.

Existem duas fontes principais de erro para esse tipo de algoritmo [12]:

- A taxa de crescimento agregado da rota, \hat{g} , que, em geral, diverge do crescimento individual da rota, g ;
- As medições da carga de tráfego, que, em geral são efetuadas por diferentes métodos e tem como parâmetro descrevendo o erro de medição a sua covariância R .

¹A explicação das variáveis e parâmetros do filtro é deixada para as seções seguintes.

O algoritmo proposto objetiva ser menos sensível a essas fontes de erro além de prover maior estabilidade para a projeção n passos à frente.

A Seção 4.2 descreve o algoritmo de filtragem, enfocando seus Aspectos Gerais, Iniciação e Seqüência de Ganho. A Seção 4.3 aborda a influência da matriz de transição de estado no desempenho do filtro e propõe uma forma de robustecê-la. A Seção 4.4 descreve o problema da reiniciação e do robustecimento e, finalmente, a Seção 4.5 apresenta as considerações finais e conclusões.

4.2 O Modelo do Algoritmo de Filtragem do APS

Esta seção enfoca três aspectos relacionados ao Algoritmo de Projeção Seqüencial (APS), como se segue:

- Características Gerais;
- Iniciação;
- Seqüência de Ganho e Robustecimento.

4.2.1 Características Gerais

No APS a ser proposto, um modelo univariável, tendo duas variáveis de estado, é utilizado. Estas variáveis de estado representam a carga de tráfego e estão relacionadas ao seu nível, T_k , e ao incremento de crescimento, \dot{T}_k . Além disso, supõe-se que a matriz de transição de estado, Φ , é genérica e seus elementos são estimados pela aplicação do algoritmo *EM* sobre uma amostra de um conjunto de rotas de tráfego com características similares, utilizando o critério da mediana². O modelo proposto é uma variante do modelo fornecido por (3.10) e (3.11)³.

As equações gerais do modelo de Espaço de Estados (2.13) e (2.12), se tornam:

$$X_{k+1} = \Phi X_k + W_k \quad (4.1)$$

e

$$Y_k = H X_k + v_k \quad (4.2)$$

onde

$$X_k = \begin{bmatrix} T_k \\ \dot{T}_k \end{bmatrix}, \quad \Phi = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix}, \quad H = \begin{bmatrix} 1 & 0 \end{bmatrix}, \quad W_k = \begin{bmatrix} w_{11k} \\ w_{21k} \end{bmatrix} \quad (4.3)$$

onde $E[v_i W_j] = 0$ para todo i, j . Em (4.3), $\dot{T}_k = gT_k$, onde g é a taxa de crescimento verdadeira de cada rota. Além do mais, W_k é um vetor Gaussiano de média nula ($E[W_k] =$

²No critério da mediana os valores estimados são colocados em ordem crescente e seu elemento central é escolhido. Quando somente algumas observações são disponíveis ou quando as amostras provêm de uma distribuição de cauda acentuada (*heavy tailed*), o critério da mediana é mais eficaz que o da média aritmética [37]. Os parâmetros do filtro, uma vez obtidos das amostras, são utilizados para toda a população.

³O Apêndice G apresenta vários testes realizados com modelos mais simples e modelos mais sofisticados, com maior número de variáveis de estado. Nenhum deles desempenhou substancialmente melhor que o modelo com dois estados, representando nível e incremento de crescimento, principalmente quando se dispõe de poucas observações, que é o objetivo da atuação dos algoritmos robustos.

0) e v_k representa um processo de ruído branco de média nula. A equação (3.30) da estimação do Filtro de Kalman é dada por:

$$\bar{X}_k = X_k^* + K_k(Y_k - HX_k^*) \quad (4.4)$$

onde se define o ganho do filtro como $K_k \triangleq \begin{bmatrix} \alpha_k \\ \beta_k \end{bmatrix}$.

Portanto, de (4.3) e (4.4) vem:

$$\bar{X}_k = \begin{bmatrix} \bar{T}_k \\ \dot{\bar{T}}_k \end{bmatrix} = \begin{bmatrix} T_k^* + \alpha_k(Y_k - T_k^*) \\ \dot{T}_k^* + \beta_k(Y_k - T_k^*) \end{bmatrix} \quad (4.5)$$

ou, equivalentemente

$$\begin{bmatrix} \bar{T}_k \\ \dot{\bar{T}}_k \end{bmatrix} = \begin{bmatrix} (1 - \alpha_k)T_k^* + \alpha_k Y_k \\ (1 - \beta_k)\dot{T}_k^* + \beta_k(Y_k - T_k^* + \dot{T}_k^*) \end{bmatrix} \quad (4.6)$$

O incremento medido, I_M , é definido como:

$$I_M \triangleq Y_k - T_k^* + \dot{T}_k^* \quad (4.7)$$

A equação de predição é, então:

$$X_k^* = \begin{bmatrix} T_k^* \\ \dot{T}_k^* \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \begin{bmatrix} \bar{T}_{k-1} \\ \dot{\bar{T}}_{k-1} \end{bmatrix}$$

Ou, na sua forma escalar:

$$T_k^* = \phi_{11}\bar{T}_{k-1} + \phi_{12}\dot{\bar{T}}_{k-1} \quad (4.8)$$

e

$$\dot{T}_k^* = \phi_{21}\bar{T}_{k-1} + \phi_{22}\dot{\bar{T}}_{k-1} \quad (4.9)$$

O incremento predito, I_P , é escrito como:

$$I_P = \dot{T}_k^* = \frac{\phi_{22}}{\phi_{12}}T_k^* - \frac{(\phi_{11}\phi_{22} - \phi_{12}\phi_{21})}{\phi_{12}}\bar{T}_{k-1} \quad (4.10)$$

Agora, se a matriz de transição de estado é suposta ser $\Phi = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, I_M e I_P são dados por:

$$I_M = Y_k - \bar{T}_{k-1} \quad (4.11)$$

e o incremento projetado é:

$$I_P = \dot{T}_k^* = T_k^* - \bar{T}_{k-1} \quad (4.12)$$

A Figura 4.2 de [12] esquematiza os incrementos projetado, medido e estimado. Pela figura pode-se notar que o incremento estimado \bar{T}_k é igual ao incremento projetado.

Também pela Figura 4.2, pode-se notar que a projeção dois anos à frente, (instante $k + 2$), corresponde à soma de dois incrementos projetados ao valor estimado no instante k . Para a projeção n passos à frente utiliza-se $\Phi^n = \begin{bmatrix} 1 & n \\ 0 & 1 \end{bmatrix}$.

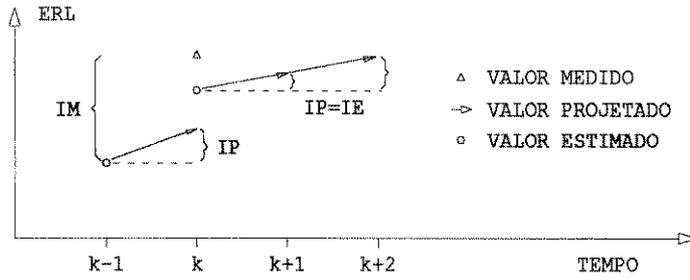


Figura 4.2: Incrementos Projetado, Medido e Estimado

Neste caso, a equação (4.6) produz:

$$\begin{bmatrix} \bar{T}_k \\ \bar{T}_k \end{bmatrix} = \begin{bmatrix} (1 - \alpha_k)T_k^* + \alpha_k Y_k \\ (1 - \beta_k)\bar{T}_k^* + \beta_k(Y_k - \bar{T}_{k-1}) \end{bmatrix} \quad (4.13)$$

É interessante notar que o nível filtrado é dado pela soma ponderada dos valores projetado e medido do nível da carga de tráfego. Da mesma forma, o incremento filtrado é uma soma ponderada dos incrementos predito e medido.

A Figura 4.3 apresenta os erros e conceitos envolvidos no modelo do Filtro de Kalman em uma situação de tráfego real, da Rota Marília 33 para a Central Trânsito, T44. As ilustrações da Figura 4.3 correspondem à uma ampliação do retângulo mostrado na Figura 3.4 dessa mesma rota. A curva verdadeira, não-observável, é suposta entre o valor projetado e o valor filtrado (Apêndice I).

Ao receber a observação Y_k (triângulo), o filtro faz uma interpolação entre o valor da predição T_k^* (flecha) e o valor filtrado \bar{T}_k (bola). Essa interpolação é realizada por meio dos valores da seqüência de ganhos, $\bar{T}_k = T_k^* + r_k \alpha_k = T_k^* + r_k k_{11k}$ (Figura 4.3 (b))⁴. O Filtro de Kalman tem a vantagem de fazer a projeção a partir do valor filtrado \bar{T}_k (bola) e da observação recebida Y_k (triângulo), o que não acarreta acúmulo de erros de observações anteriores à observação Y_k .

Em particular a Figura 4.3 (b) mostra:

- O termo de inovação, r_k , e o valor filtrado, $T_k^* + \alpha_k r_k$;
- O incremento medido, $I_M = Y_k - \bar{T}_{k-1}$ e os incrementos projetado e estimado, \dot{T}_{k+1}^* e \bar{T}_k respectivamente, onde $\dot{T}_{k+1}^* = \bar{T}_k$.

Já a Figura 4.3 (a) mostra:

- $v_k = Y_k - HX_k$: o erro de observação;
- $\hat{T}_{k+1}^* = T_{k+1}^* - T_{k+1}$: o erro de previsão;

⁴A Figura 4.3 (b) traz explicitamente as relações comentadas, porém essa é uma característica de ambas as figuras porque é o próprio funcionamento do filtro.

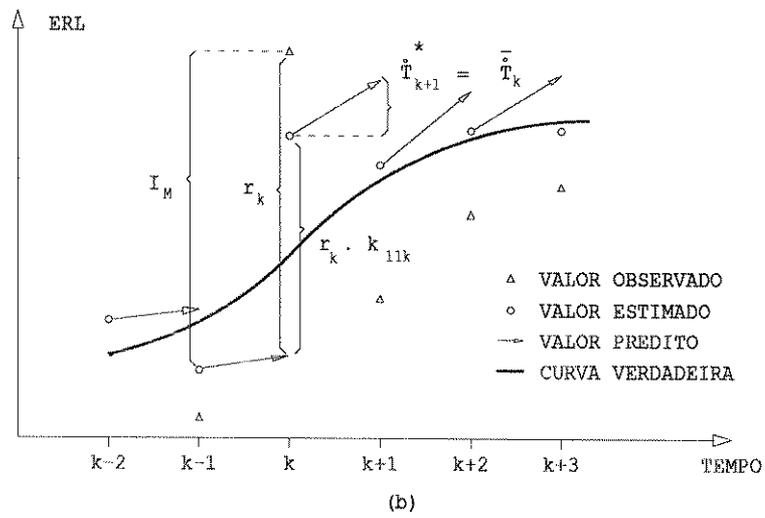
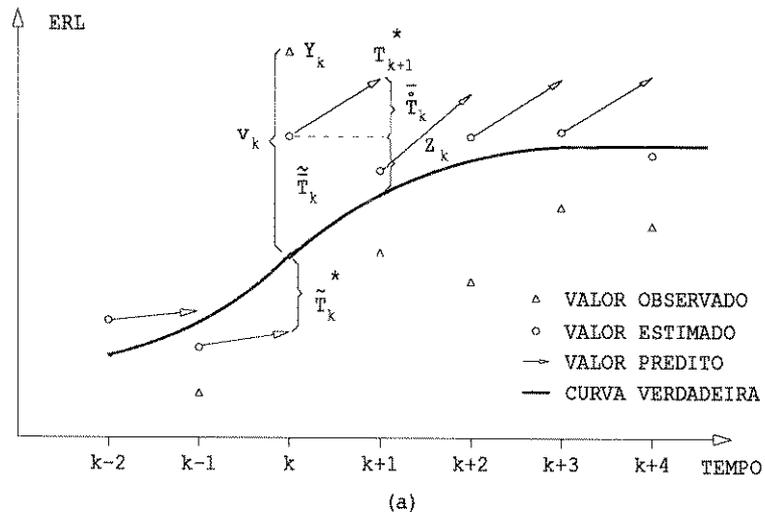


Figura 4.3: Funcionamento do Filtro de Kalman

- $\tilde{T}_k = \bar{T}_k - T_k$: o erro de estimação (ou de filtragem);
- Os incrementos projetado e estimado, \dot{T}_{k+1}^* e \bar{T}_k respectivamente, onde $\dot{T}_{k+1}^* = \bar{T}_k$;
- A variável auxiliar Z_k que será explicada no texto.

O valor da previsão a um passo pode ser escrito como:

$$T_{k+1}^* = \dot{T}_k + \bar{T}_k + Z_k$$

onde

$$Z_k = \tilde{T}_k + I_k + w_{11k}$$

que é uma soma de variáveis aleatórias e Z_k e I_k são apenas variáveis auxiliares para explicação da figura. De fato, nota-se na figura que, em valor absoluto, só o valor de \tilde{T}_k já é maior que Z_k . I_k corresponde a

$$I_k = \dot{T}_k - \bar{T}_k$$

e

$$\tilde{T}_k = \bar{T}_k - T_k$$

é o erro de estimação (filtragem). Ou seja, Z_k reúne várias fontes de erro do algoritmo. Se $\tilde{T}_k = \dot{T}_k \Rightarrow I_k = 0$. Se $T_k = \bar{T}_k \Rightarrow \tilde{T}_k = 0$. E se ambas as condições forem verdadeiras, $T_{k+1}^* = \dot{T}_k + T_k + w_{11k}$, ou seja, $T_{k+1}^* = T_{k+1}$.

A precisão com que o conjunto de parâmetros é estimado se reflete na minimização dos erros citados. Supondo Φ genérico, porque a Figura 4.3 corresponde a $\Phi = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, os parâmetros a serem estimados são:

- A matriz de transição de estado, Φ ;
- A covariância do erro de medição, $R = E[v_k^2]$;
- A covariância do erro de modelagem, $Q = E[W_k W_k^T]$;
- A estimação do estado inicial, \bar{X}_0 ;
- A covariância do erro da estimação inicial, $\bar{P}_0 \triangleq E[(X_0 - \bar{X}_0)(X_0 - \bar{X}_0)^T]$.

4.2.2 Iniciação

Supondo a matriz de transição de estado conhecida⁵, os demais parâmetros a serem estimados são: \bar{X}_0 , \bar{P}_0 , R , e Q .

Devido a que este algoritmo objetiva trabalhar somente com poucas observações, seu desempenho é, em geral, mais dependente da razão $\frac{P_0}{R}$ do que da razão $\frac{Q}{R}$ ([23] e Apêndice D). Neste caso, os parâmetros mais relevantes são, então, \bar{X}_0 , \bar{P}_0 e R . Seguindo Moreland [12], o estado inicial depende da taxa de crescimento agregado, \hat{g} , e a covariância do erro de estimação depende de \hat{g} e R . Portanto, \hat{g} e R são as principais fontes de erro desse algoritmo.

⁵A próxima seção mostra uma proposta para se estimar Φ .

Dado um período de observação composto de M subperíodos. Define-se como crescimento agregado, \hat{g} , do período como:

$$\hat{g} \triangleq (1 + \hat{g}_s)^M - 1 \quad (4.14)$$

onde \hat{g}_s é o crescimento agregado do subperíodo. O crescimento agregado descreve o crescimento de um conjunto de rotas como um todo. Assim para dados anuais e mensais:

$$\hat{g}_{mensal} = \sqrt[12]{1 + \hat{g}_{anual}} - 1$$

A variância, σ_g^2 , devido ao erro entre \hat{g} e g é:

$$\sigma_g^2 = E[(g - \hat{g})^2]$$

estimou-se o desvio-padrão, σ_g , com dados de campo, em um conjunto de 40 rotas de tráfego, obtendo-se valores na faixa de 0,04 - 0,08, com média de 0,05.

Para observações mais a longo prazo (anos, meses, semanas), a covariância R pode ser estimada conforme proposto por Moreland [12]. Para aquisição de dados em tempo real utilizando varredura, esta covariância pode ser estimada como proposto por Nunes [41]⁶.

Consideram-se como valores iniciais do estado o seguinte:

$$\bar{X}_0 = \begin{bmatrix} Y_0 \\ \hat{g}Y_0 \end{bmatrix}$$

onde Y_0 é o valor da primeira observação disponível.

De acordo com Moreland [12], a covariância do erro de estimação inicial é obtida como uma função de \hat{g} , T_0 , e R . Detalhando este resultado tem-se que: Se $\bar{P}_0 = \begin{bmatrix} \bar{P}_{11} & \bar{P}_{12} \\ \bar{P}_{21} & \bar{P}_{22} \end{bmatrix} = E[X_0 - \bar{X}_0]^2$, então $\frac{P_0}{R}$ terá:

$$\bar{P}_{11} = \frac{E[T_0 - Y_0]^2}{R} = 1, \text{ porque } E[T_0 - Y_0]^2 = E[v^2] = R$$

mas

$$\begin{aligned} \bar{P}_{22} &= \frac{E[gT_0 - \hat{g}Y_0]^2}{R} = \frac{E[gT_0 - \hat{g}T_0 - \hat{g}Y_0 + \hat{g}T_0]^2}{R} = \frac{E[(g - \hat{g})T_0 + \hat{g}(T_0 - Y_0)]^2}{R} = \\ &= \frac{T_0^2 E[g - \hat{g}]^2}{R} + \frac{\hat{g}^2 E[T_0 - Y_0]^2}{R} + \frac{2E[(g - \hat{g})T_0 \hat{g}(T_0 - Y_0)]}{R}, \text{ mas } E[g - \hat{g}] = 0, \end{aligned}$$

Portanto:

$$\bar{P}_{22} = \frac{T_0^2 \sigma_g^2}{R} + \hat{g}^2$$

onde $\sigma_g^2 = E[g - \hat{g}]^2$ e $\bar{P}_{12} = \bar{P}_{21}$, então:

$$\begin{aligned} \bar{P}_{12} = \bar{P}_{21} &= \frac{E[(T_0 - Y_0)(gT_0 - \hat{g}Y_0)]}{R} = \frac{E[(T_0 - Y_0)[(g - \hat{g})T_0 + \hat{g}(T_0 - Y_0)]]}{R} = \\ &= \frac{E[(T_0 - Y_0)^2] \hat{g}}{R} + \frac{E[(T_0 - Y_0)(g - \hat{g})T_0]}{R} = \hat{g}, \end{aligned}$$

Finalmente:

$$\frac{\bar{P}_0}{R} = \begin{bmatrix} 1 & \hat{g} \\ \hat{g} & \frac{\sigma_g^2 T_0^2}{R} + \hat{g}^2 \end{bmatrix} \quad (4.15)$$

⁶Testes efetuados com o algoritmo operando em tempo real (Capítulo 5), indicaram que, nos padrões previstos para a variação do nível de tráfego em curto prazo, em geral, não se justifica a utilização do Algoritmo Robusto, mas apenas a utilização do algoritmo Adaptativo.

Portanto de (4.15) nota-se que, para um \hat{g} fixo, o Erro Quadrático Médio, EQM ^{7 8} é uma função do erro de crescimento, σ_g^2 , e do erro de observação, R . Note-se ainda da equação (4.15) que este algoritmo é muito sensível à estimação dos parâmetros σ_g^2 e R . Uma estimação errada de σ_g^2 e R pode ser desastrosa. Daí, no caso em que esses parâmetros não são conhecidos, o uso de um procedimento simples, mas robusto, é mais apropriado.

4.2.3 Seqüência de Ganhos e Robustecimento

Supondo-se conhecido Φ , a estimação de R , \bar{X}_0 , e \bar{P}_0 pode ser feita como explicado anteriormente. Entretanto, nem a estimação da covariância Q nem o ajuste do conjunto de parâmetros se constituem em tarefas simples. Devido ao não conhecimento de todos os parâmetros envolvidos, a utilização de um valor adequado para a seqüência de ganho é de importância fundamental.

Esse valor pode ser obtido por meio de três diferentes procedimentos:

- Filtragem com seqüência de ganho truncada;
- Filtragem com seqüência de ganho constante;
- Filtragem com ganhos fixos.

A principal restrição ao primeiro tipo de procedimento é a dificuldade de se encontrar o ponto de truncamento (Seção 3.8) a partir do qual uma filtragem com ganho fixo é utilizada⁹. A Figura 4.4 mostra o valor do erro quadrático médio de projeção 12 passos à frente na Rota T44, para o truncamento da seqüência de ganhos após a 5^a, a 10^a, e a 15^a observações processadas.

O melhor desempenho, neste caso, ocorreu para o truncamento da seqüência de ganho em $k = 15$. Nos testes realizados em outras rotas, no entanto, o melhor desempenho ocorreu para outros valores de truncamento. Em média, o melhor $REQM$ ocorreu para o truncamento em $k = 10$.

Os outros dois procedimentos, Filtragem com seqüência de ganho constante e Filtragem com ganhos fixos, fazem uso do algoritmo EM com o critério da mediana, como descrito anteriormente, e são mais simples de se aplicar, uma vez determinado o conjunto robusto de parâmetros. Esses procedimentos têm sido testados em várias rotas e o último tem sido utilizado, devido a ser mais simples e diferir apenas nas primeiras observações processadas. A Figura 4.5 ilustra o desempenho desses procedimentos na projeção 12 passos à frente para a Rota T44, utilizando a raiz quadrada do erro quadrático médio, $REQM$. Comparando as Figuras 4.5 e 4.4, para a Rota T44, nota-se que estes dois últimos procedimentos (seqüência de ganho constante e ganhos fixos) têm desempenho superior ao procedimento com seqüência de ganhos truncada, mas esta superioridade também ocorreu

⁷Este valor de \bar{P}_0 também foi proposto para a iniciação do algoritmo EM , apresentado no capítulo anterior.

⁸O ganho ótimo estacionário bem como o EQM ótimo são calculados recursivamente por meio da utilização das equações do filtro de Kalman, equações (3.36), (3.32), e (3.34). Este procedimento corresponde à solução da equação de Ricatti discreta, [12, 17] e Apêndice D. O Apêndice J mostra que a equação de Joseph (3.66) para o cálculo da covariância do erro de estimação, e portanto para avaliação do EQM teórico, vale para qualquer valor do ganho. Entretanto, este é um método teórico supondo o conhecimento de Φ , H , R , e Q , o que não é, usualmente, verdadeiro. Portanto, este capítulo trabalhará com dados reais de campo para se calcular o EQM .

⁹O ponto de truncamento é determinado para um filtro que possui uma matriz Q nula.

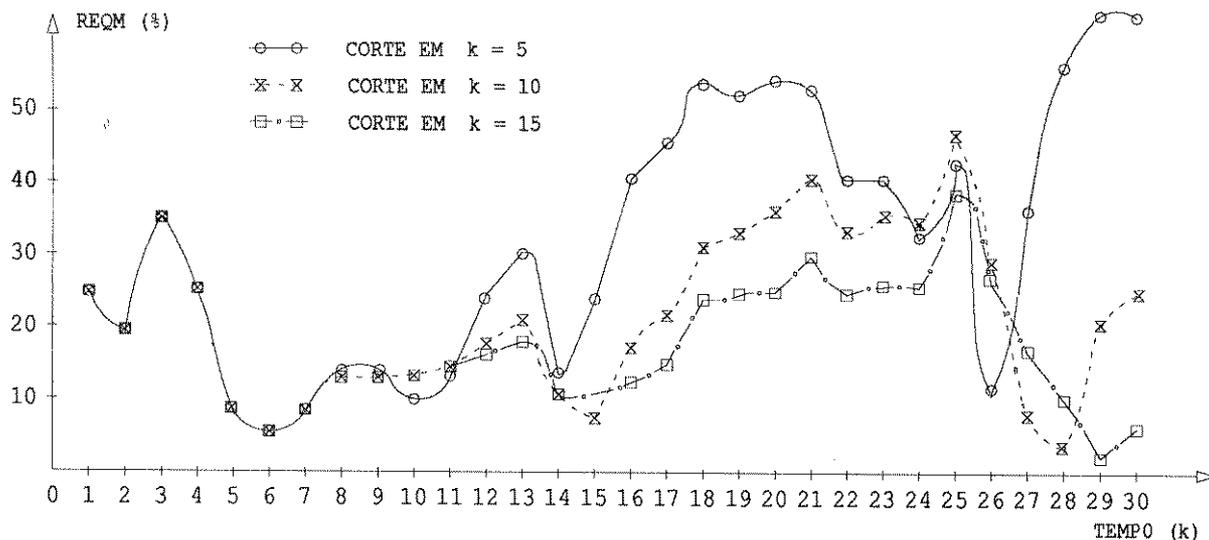


Figura 4.4: Projeção 12 Passos à Frente - Filtro com Sequência de Ganhos Truncada

para as demais rotas testadas, T6362 (Jaboticabal Trânsito para Araraquara Trânsito), T2522 (Guaratinguetá Trânsito para Taubaté Trânsito), T4542 (Lins Trânsito para Bauru Trânsito), T4742 (Avaré Trânsito para Bauru Trânsito) e T8688 (Araçatuba Trânsito para Dracena Trânsito).

O conjunto de rotas utilizado como exemplo produziu um ganho fixo $K_k = \begin{bmatrix} 0,239 \\ 0,015 \end{bmatrix}$.

4.3 Influência da Matriz de Transição de Estado no Desempenho do Filtro

No algoritmo proposto sugere-se ainda que a matriz de transição de estado seja também estimada pelo algoritmo *EM*, utilizando-se o critério da mediana. Este algoritmo foi aplicado à Rota T44 utilizando uma seqüência de ganhos constantes e ganhos fixos. A Figura 4.6 ilustra o desempenho do filtro para a projeção 12 passos à frente utilizando este procedimento.

Comparando-se estes resultados com aqueles da Figura 4.5, nota-se uma considerável melhoria quando o algoritmo proposto é empregado. Com a utilização da matriz robustecida, Φ , o algoritmo de estimação é menos sensível às variações do valor inicial do estado, \bar{X}_0 , e portanto também a \hat{g} . Verificou-se também que esta melhoria se aplica para a projeção um passo à frente como será descrito a seguir.

A Tabela 4.1 mostra os valores médios do *REQM* e seus desvios-padrões (entre parênteses) para 40 rotas de tráfego, utilizando diferentes ganhos ($k_{11} = \alpha_k$ e $k_{21} = \beta_k$) e diferentes valores de \hat{g} ($\hat{g} = 0,015$ e $\hat{g} = 0,030$).

(Araçatuba Trânsito para Dracena Trânsito). Ambos os filtros possuem ganhos fixos, mas um utilizando $\Phi = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ e o outro utilizando Φ robustecido, onde $\Phi =$

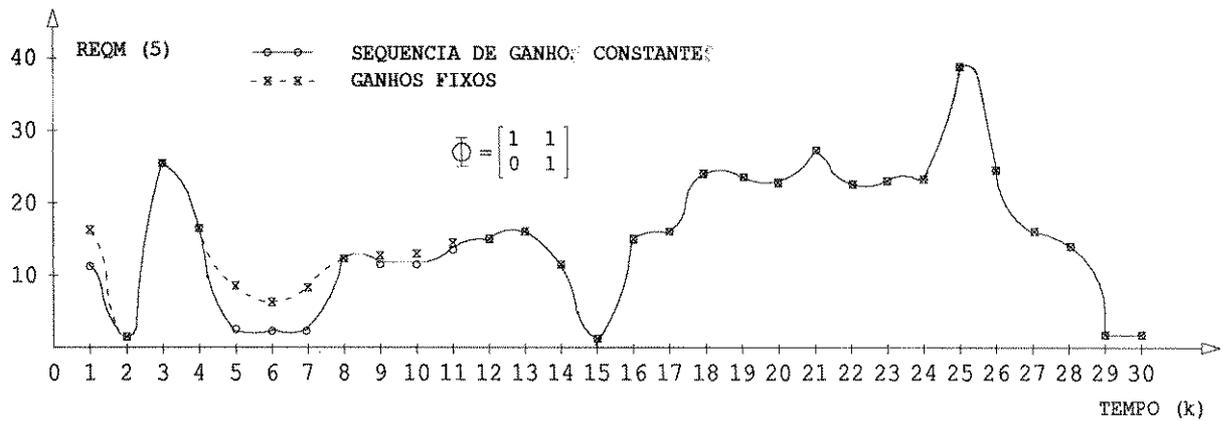


Figura 4.5: Projeção 12 Passos à Frente - Filtro com Sequência de Ganhos Constantes e com Ganhos Fixos

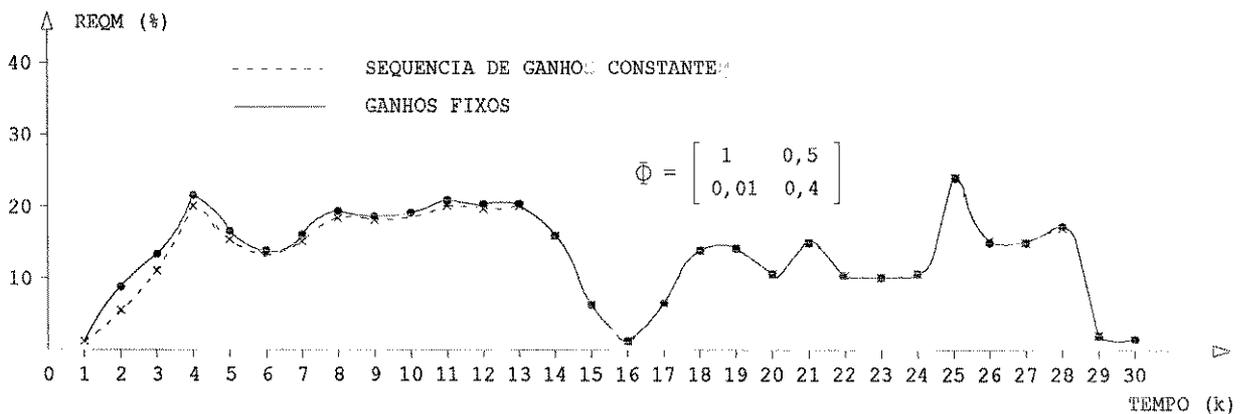


Figura 4.6: Proj. 12 Passos à Frente com Matriz de Transição de Estado Robusta

$\begin{bmatrix} 1 & 0,5 \\ 0,01 & 0,4 \end{bmatrix}$. Nota-se que o desvio-padrão permanece estável para diferentes ganhos quando se utiliza a matriz Φ robustecida. Esta estabilidade também ocorre, em menor escala, para a projeção um passo à frente.

Tabela 4.1 - Aplicação do APS Robusto a 40 Rotas de Tráfego

PARAMETROS	\hat{g}	k_{11} k_{21}	PROJEÇÕES/PASSOS		
			1	6	12
$\Phi = \begin{bmatrix} 1 & 0,5 \\ 0,01 & 0,4 \end{bmatrix}$	0,015	0,82	7,5	11,2	11,3
		0,42	(5,0)	(6,4)	(5,7)
	0,015	0,60	6,9	10,9	11,4
		0,04	(5,1)	(6,2)	(5,5)
0,015	0,50	6,8	10,7	11,3	
	0,00	(5,1)	(6,1)	(5,7)	
0,030	0,60	7,0	10,9	11,4	
	0,04	(5,1)	(6,2)	(5,5)	
$\Phi = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$	0,015	0,82	8,4	18,1	27,9
		0,42	(5,6)	(14,6)	(22,2)
	0,015	0,60	8,2	15,3	18,4
		0,04	(6,2)	(16,4)	(28,0)
	0,015	0,50	6,6	9,6	9,6
		0,00	(4,8)	(5,6)	(6,1)
	0,030	0,60	7,0	10,5	10,9
		0,04	(5,3)	(7,4)	(9,4)

A Figura 4.7 compara o desempenho de dois filtros para cinco rotas de tráfego, T6362 (Jaboticabal Trânsito para Araraquara Trânsito), T2522 (Guaratinguetá Trânsito para Taubaté Trânsito), T4542 (Lins Trânsito para Bauru Trânsito), T4742 (Avaré Trânsito para Bauru Trânsito) e T8688 (Araçatuba Trânsito para Dracena Trânsito). Também aqui ambos os filtros possuem ganhos fixos, e como na tabela anterior, um utilizando $\Phi = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ e o outro utilizando Φ robustecido, onde $\Phi = \begin{bmatrix} 1 & 0,5 \\ 0,01 & 0,4 \end{bmatrix}$.

O desempenho desses dois filtros é também ilustrado na Tabela 4.2, onde é apresentado o desvio-padrão médio do *REQM* das mesmas cinco rotas de tráfego da Figura 4.7.

Tabela 4.2 - Desvios-padrões (%) Observados para a Média de 5 Rotas de Tráfego

\hat{g}	0,015		0,050		0,100	
	$\Phi = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$	$\Phi = \begin{bmatrix} 1 & 0,5 \\ 0,01 & 0,4 \end{bmatrix}$	$\Phi = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$	$\Phi = \begin{bmatrix} 1 & 0,5 \\ 0,01 & 0,4 \end{bmatrix}$	$\Phi = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$	$\Phi = \begin{bmatrix} 1 & 0,5 \\ 0,01 & 0,4 \end{bmatrix}$
seq. de gan. ctes	6,18	3,07	4,58	3,08	6,26	2,48
ganhos fixos	6,40	3,38	7,88	3,36	11,82	3,34

Note que o filtro com Φ robustecido é sensivelmente melhor do que o com $\Phi = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$.

Note também que aqui novamente o desvio-padrão permanece estável para diferentes valores de \hat{g} .

4.4 Reiniciação e Robustecimento

A utilização de um ganho robusto e uma matriz de transição de estado robusta torna o algoritmo de estimação menos sensível à variação nos parâmetros \bar{X}_0 , \bar{P}_0 , R , Q , e Φ .

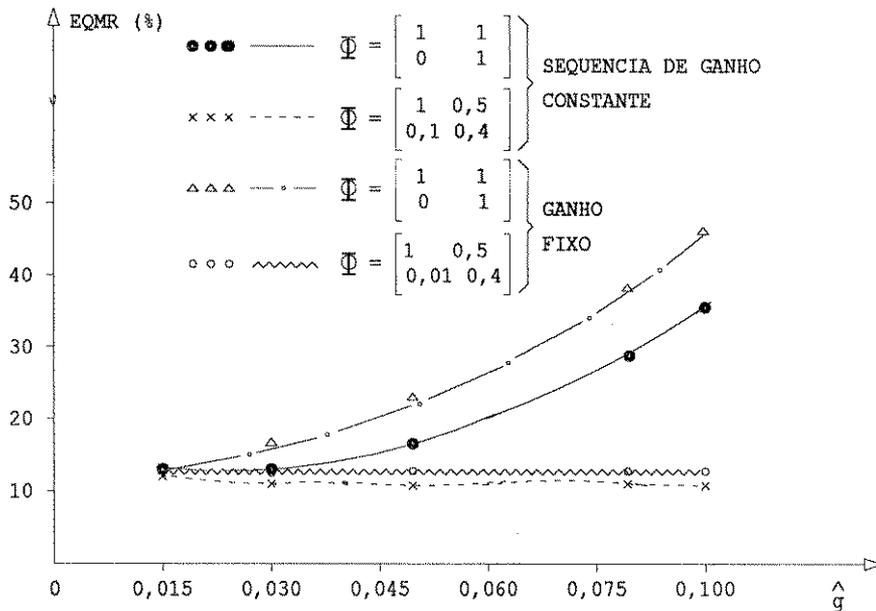


Figura 4.7: Proj. 12 Passos à Frente em Função do Valor do Estado Inicial

A robustez a ser discutida nesta seção se aplica à ocorrência de observações duvidosas (*outliers*).

Além da utilização do ganho e da matriz de transição de estado, como proposto, o algoritmo deve também utilizar o intervalo de confiança para minimizar a ocorrência de *outliers*. Pode-se mostrar que [38, 39] a utilização de um intervalo de confiança na faixa de $1 - 2 REQM$ da projeção um passo à frente assegura limites aceitáveis para a variância do estimador.

4.4.1 Intervalo de Confiança para Detecção de Observações Duvidosas (*Outliers*)

Nesta subseção determina-se o *REQM* para Φ genérico e também para a projeção n passos à frente. Este valor será utilizado na delimitação do intervalo de confiança para se detectar *outliers*¹⁰.

Supondo que:

- O instante de operação inicial é $k = 0$, então, $T_k = T_0 + \Delta$, onde Δ é um valor constante e \bar{T}_0 é Y_0 ;
- $\hat{g} \ll 1$, $E[g - \hat{g}] = 0$, e $\phi_{12} < 1$;

¹⁰Pode-se mostrar que o critério da mediana, como descrito anteriormente, é um caso particular do critério do intervalo de confiança [38].

- Para um Φ genérico:

$$\Phi^n = \begin{bmatrix} \varphi_{11} & \varphi_{12} \\ \varphi_{21} & \varphi_{22} \end{bmatrix}$$

então, $T_{k+n} = \varphi_{11}T_k + \varphi_{12}gT_k$ e $T_{k+n}^* = \varphi_{11}\bar{T}_k + \varphi_{12}\hat{g}\bar{T}_k$ ¹¹.

Portanto, o *EQM* n passos à frente é:

$$E[(T_{k+n}^* - Y_{k+n})^2]$$

produzindo

$$R(1 + \varphi_{11}^2) + \varphi_{12}^2 T_0^2 \sigma_g^2$$

Assim, os limites para o intervalo de confiança são $\pm \sqrt{R(1 + \varphi_{11}^2) + \varphi_{12}^2 T_0^2 \sigma_g^2}$.

4.4.2 Correção da Carga Projetada n Passos à Frente

A fim de se detectar *outliers* e se verificar a precisão da projeção n passos à frente, os valores projetados devem ser corrigidos e comparados com os valores medidos. Neste caso a equação (4.16) se aplica:

$$X_{k+1}^* = \Phi \bar{X}_k + LU_k \quad (4.16)$$

onde L é um ganho e U_k são as entradas determinísticas. Supondo que o melhor valor para \bar{X}_{k+1} é X_{k+1}^* e resolvendo-se a equação (4.16) recursivamente, resulta:

$$X_{k+n}^* = \Phi^n \bar{X}_k + L \sum_{i=0}^{n-1} \Phi^{n-i-1} U_{k+i} \quad (4.17)$$

e

$$Y_{k+n}^* = H X_{k+n}^* \quad (4.18)$$

onde a segunda parcela do lado direito na soma em (4.17) corresponde à soma das entradas provocadas por eventos determinísticos previstos. X_{k+n}^* é a projeção do valor filtrado n passos à frente e Y_{k+n}^* é a saída projetada n passos à frente, ambas a partir do instante k .

4.4.3 Exemplo de Aplicação

A Figura 4.8 ilustra a aplicação do APS Robusto com os mecanismos citados.

A Rota Jaboticabal Trânsito para Araraquara Trânsito, T6362, foi escolhida porque, como mostrado no Capítulo 3, as projeções baseadas no algoritmo *EM*, aplicadas a essa rota, resultaram em grandes erros de projeção. Como se pode observar, o Algoritmo Robusto proposto resulta em uma projeção média entre os valores mais altos e os valores mais baixos, desta forma, embora não se chegue a ter um *REQM* relativamente muito baixo, também não se vai tê-lo relativamente muito elevado. É fácil verificar que um algoritmo com maior poder de rastreamento, que vai acompanhar mais a forma da curva, também apresentará erros relativamente elevados devido à ocorrência de mudança abrupta no valor do nível. Outro comentário sobre a Figura 4.8 é que os valores projetados por meio da utilização

¹¹ T_{k+n} e T_{k+n}^* são os n passos à frente, do instante k , para o nível do estado e o nível do estado filtrado, respectivamente.

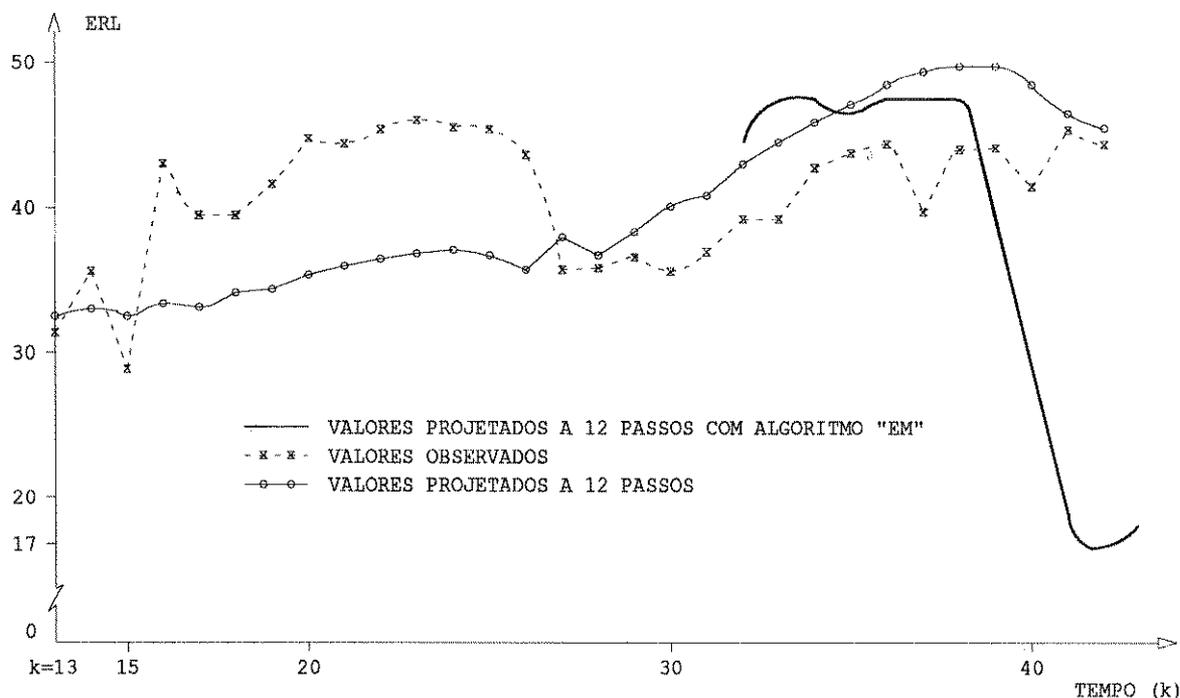


Figura 4.8: Valores Observados e Projetados a 12 Passos para a Rota T6362

do algoritmo *EM* somente começam a aparecer a partir do instante $k = 32$. Isso se deve a que o algoritmo *EM* necessita da utilização das primeiras observações para a estimação dos parâmetros (no caso, 19), e as primeiras projeções se iniciam com $k = 20$, resultando o primeiro valor projetado a 12 passos em $k = 32$. Este aspecto é uma outra desvantagem da utilização do algoritmo *EM*, já que o Algoritmo Robusto começa projetando desde a primeira observação recebida.

Em seguida testou-se esta rota com a introdução de *outliers*. Primeiro colocou-se um *outlier* somando 100 ao 20º valor observado (44,1), tornando-o 144,1. Os resultados mostraram diferença muito pequena (menos de 1%) nas projeções subsequentes. Em seguida colocaram-se dois *outliers* consecutivos e depois três e o algoritmo teve o desempenho degradado em ambos os casos, com erros superiores a 100%. Algumas considerações a respeito:

- O detector funciona adequadamente na presença de um *outlier*;
- Se houver dois ou mais pontos consecutivos fora do intervalo de confiança, todos para cima ou todos para baixo, é importante que realmente seja um evento determinístico não-previsto, caso contrário a projeção é deteriorada porque o filtro com ganhos fixos não possui capacidade de rastreamento (se houver dúvida, é melhor aumentar o número de pontos consecutivos antes de se tomar decisão);
- Quando se tem a previsão do evento determinístico, então T_k^* , e portanto X_k^* , são corrigidos automaticamente antes de se comparar com Y_k ;

- O intervalo a n passos serve somente para se avaliar o erro de previsão.

4.5 Conclusões

Neste Capítulo propõe-se um algoritmo de projeção seqüencial utilizando:

- Um ganho robusto fixo;
- Uma matriz de transição de estado robusta;
- Um intervalo de confiança para minimizar o efeito dos *outliers* e detectar a presença de eventos determinísticos não-previstos.

O ganho e a matriz de transição de estado são estimados pelo critério da mediana após a aplicação do algoritmo *EM* a uma amostra de um conjunto de rotas de tráfego.

O Algoritmo de Projeção Seqüencial Robusto desempenha adequadamente mesmo quando somente poucas observações são disponíveis (e.g., para dados mensais ou anuais). Por outro lado, se há uma grande quantidade de dados para serem feitas previsões, o algoritmo pode ser facilmente automatizado por sua simplicidade.

Foi mostrado que o algoritmo produz resultados estáveis, dentro de algumas condições específicas, mesmo com uma grande faixa de variação de seus parâmetros.

Em geral, comparativamente ao algoritmo *EM*, o algoritmo proposto apresenta um desempenho ligeiramente inferior com relação ao *EQM*. Entretanto, é muito mais simples de ser aplicado e é mais robusto.

O capítulo seguinte apresenta um algoritmo Adaptativo que pode rastrear variações nos parâmetros, evitar erros grosseiros e também ser aplicado em situações de tempo real. Ele deve apresentar um *EQM* similar ao do algoritmo *EM* bem como a robustez do algoritmo proposto neste capítulo.

Capítulo 5

Algoritmo de Projeção Seqüencial Adaptativo Modificado: Uma Proposta

Neste capítulo propõe-se um novo tipo de algoritmo de projeção, o Algoritmo de Projeção Seqüencial Adaptativo Modificado, APSAM, que combina **precisão**, como aquela dada pela utilização do algoritmo *EM* para se estimarem os parâmetros em um modelo de Espaço de Estados, e a **robustez**, como aquela fornecida pelo Algoritmo de Projeção Seqüencial Robusto. O algoritmo foi aplicado a um conjunto de rotas de tráfego, bem como a séries temporais geradas por simulação de Monte Carlo, e os resultados se mostraram adequados. No que diz respeito à Raiz Quadrada dos Erros Quadráticos Médios, este algoritmo superou o desempenho do algoritmo *EM*, além de requerer menos esforço computacional. Este algoritmo é adequado para previsão de processos estocásticos com perturbações não-estacionárias, como a previsão da carga de tráfego em prazo muito curto. Portanto, ele também pode ser utilizado para aplicações em tempo real do tipo Gerência de Rede.

5.1 Introdução

A utilização do Filtro de Kalman para a estimação do estado em modelos de Espaço de Estados foi abordada no Capítulo 3, onde a utilização do algoritmo *EM* para a estimação dos parâmetros do filtro produziu resultados adequados. No Capítulo 4 propôs-se o Algoritmo de Projeção Seqüencial Robusto, um modelo de algoritmo mais simplificado, também baseado no modelo de Espaço de Estados, que, às custas de uma perda de precisão, provê maior robustez.

Para combinar precisão, como aquela dada pelo algoritmo *EM*, e robustez, como aquela fornecida pelo APS Robusto, este capítulo propõe o Algoritmo de Projeção Seqüencial Adaptativo Modificado (APSAM) que também pode ser aplicado a sistemas variantes no tempo. Este último aspecto torna-o adequado para aplicações em tempo real. O desempenho do algoritmo é analisado para a projeção n passos à frente com o principal objetivo de testar a estabilidade das projeções (Capítulo 3). As principais características desse algoritmo são:

- Precisão;
- Robustez;

- Estabilidade da projeção n passos à frente.

O algoritmo *EM* realiza uma estimação a priori dos parâmetros do Filtro de Kalman, por meio da maximização da função de verossimilhança. O APSAM, ao contrário, minimiza, a posteriori, os fatores de causa de erros, i.e., o algoritmo realiza uma correta estimação dos parâmetros.

A Figura 5.1 esquematiza o funcionamento do APSAM, composto pelo APS Adaptativo, a ser descrito neste capítulo, e pelo APS Robusto, descrito no Capítulo 4 (é uma evolução das ilustrações de [22, 23]). Inicialmente, o APS Robusto está em estado operacional e o APS Adaptativo está no estado *stand-by* até que haja o chaveamento de um para o outro. Esse chaveamento é função do tipo de dados amostrados (mensais, anuais, etc., e do tipo de modelo adotado; no caso mensal, em torno de 20 observações se mostraram suficientes para a estimação dos parâmetros do APS Adaptativo, antes dele entrar em atividade). A seguir apresenta-se uma breve descrição do APSAM¹.

Algoritmo de Filtragem e Projeção: É o filtro em si. É baseado no modelo de Espaço de Estados e inclui a estimação dos parâmetros;

Detector de outliers: Implementa um procedimento para verificar se os dados observados estão dentro de um dado intervalo de confiança;

Estimação Futura de Eventos Determinísticos: É usado para corrigir o valor da carga projetada dado que algum evento pode ocorrer (e.g., quando o número de assinantes é aumentado em uma dada rota de tráfego);

Iniciação/Reiniciação: Usado para iniciar/reiniciar o filtro;

APS Robusto: Como descrito no Capítulo 4 e em [46].

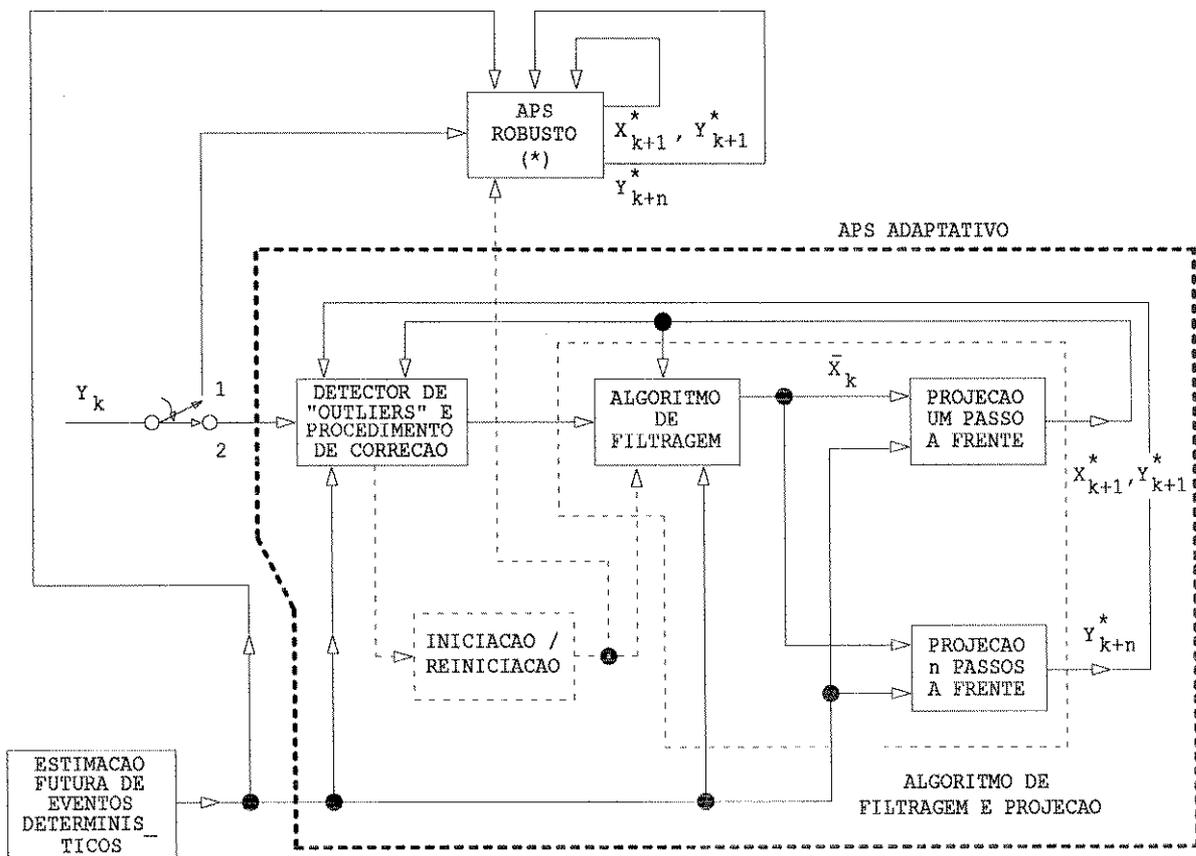
O desempenho do filtro depende do conhecimento de vários parâmetros: Φ , R_k , Q_k , \bar{X}_0 e \bar{P}_0 . No caso do filtro Adaptativo, a questão dos parâmetros iniciais não é tão crítica porque inicialmente estará em operação o APS Robusto (evidentemente que isso é função do número de observações processadas com o APS Robusto como ativo, até o momento do chaveamento e do tipo de modelo utilizado). A estimação de Φ , Q_k e R_k , e sugestões para a escolha dos parâmetros iniciais serão abordadas nas próximas seções. As Seções 5.2 e 5.3 enfocam a proposta de estimação dos parâmetros para o funcionamento do algoritmo de filtragem e projeção. A Seção 5.4 apresenta a questão da iniciação do algoritmo. A Seção 5.5 discute os problemas de reiniciação e robustecimento, a Seção 5.6 aborda a aplicação do algoritmo em tempo real, e, finalmente, a Seção 5.7 apresenta as considerações finais e conclusões.

5.2 Estimação das Covariâncias dos Erros Q_k e R_k

Em [17] Chemouil e Garnier propuseram um algoritmo Adaptativo para projeção de tráfego². Neste trabalho, as covariâncias dos erros foram estimadas recursivamente utilizando as equações adicionais de Sage-Husa [17, 18]. Entretanto, a computação dessas

¹A explicação das variáveis e parâmetros do filtro é deixada para as seções seguintes.

²Em [45] há um gráfico que mostra que este algoritmo, ao receber um *outlier*, demora para estabilizar os valores de sua projeções. Além disso, seu *EQM* é significativamente maior que o obtido pela aplicação do algoritmo *EM*.



(*) DETECTOR DE OBSERVACOES ABERRANTES ("OUTLIERS"), FILTRAGEM E PROJECAO. A ESTIMACAO FUTURA DE EVENTOS DETERMINISTICOS E COMUM TANTO PARA O APS ROBUSTO QUANTO PARA O APS ADAPTATIVO.

Figura 5.1: Algoritmo de Projecção Sequencial Adaptativo Modificado - APSAM = (APS Robusto + APS Adaptativo)

matrizes, utilizando as equações propostas, pode conduzir à instabilidade devido a matrizes das covariâncias dos erros semi-definidas negativas, principalmente ocasionadas por valores iniciais inadequados. Em [18] Knottnerus propôs um conjunto de equações numericamente estável. ³ Além de possuir um valor do *EQM* ligeiramente acima do fornecido pelo algoritmo *EM*, esse procedimento, em geral, possui uma baixa taxa de convergência com relação à estimação de Q_k . Para melhorar a precisão da estimação da covariância do erro de modelagem, a consistência estatística entre as inovações e suas estatísticas, para um sistema univariável, é utilizada (será apresentada em seguida, nesta seção), $r_{k+1}^2 = E[r_{k+1}^2]$, $i = 1, 2, \dots, N$, onde N é o número de observações disponíveis.

Esse método foi primeiramente proposto por Jazwinski [35]; entretanto requeria uma computação *off-line*. Para computação recursiva da matriz Q_k , um filtro adicional, denominado Subfiltro de Kalman, foi utilizado por Rios Neto e Kuga em [40]. Neste trabalho, as seguintes contribuições são introduzidas:

- O desenvolvimento de um método para se estimar a matriz Q_k completa;
- Testes de sensibilidade, relativos ao uso de diferentes valores para a matriz de observação H ;
- Considerações sobre a equação a ser empregada para se estimar R_k , bem como sua taxa de convergência;
- Desenvolvimento do modelo de estimação de Q_k para multivariáveis, detalhado no Apêndice A.

5.2.1 Estimação da Covariância do Erro de Modelagem (Q_k)

Em seguida serão apresentados, i) o modelo de Espaço de Estados para Q_k , denominado Submodelo de Espaço de Estados, e ii) o Subfiltro de Kalman.

Submodelo de Espaço de Estados para os Elementos de Q_k

Para se obter um modelo de Espaço de Estados para os elementos da matriz Q_k , o erro de predição verdadeiro é definido ⁴:

$$r_{k+1}^v \triangleq (Y_{k+1} - HX_{k+1}^*) - (Y_{k+1} - HX_{k+1}) \quad (5.1)$$

A Figura 5.2 ilustra o erro de predição verdadeiro para um filtro com dois estados e $H = \begin{bmatrix} 1 & 0 \end{bmatrix}$.

De (5.1), (2.13) e (3.31), obtém-se:

$$r_{k+1}^v = r_{k+1} - V_{k+1} \quad (5.2)$$

³Testes realizados o algoritmo de Knottnerus nas Rotas Bauru Trânsito para Ourinhos Trânsito, T4342, Ourinhos Trânsito para Bauru Trânsito, T4243, São João da Boa Vista Trânsito para Campinas Trânsito, T9692, e Marília Trânsito para Bauru Trânsito, T442, mostraram uma *REQM* de 8,18% com um desvio-padrão de 3,73%. Para o algoritmo de Sage-Husa, com essas mesmas rotas, os resultados foram 11,55% e 5,66%, respectivamente. O algoritmo *EM* obteve, para essas mesmas rotas, 4,81% e 0,39%, respectivamente.

⁴A utilização do índice $k + 1$ nesta seção é apenas para diferenciar os diferentes instantes de tempo. Por exemplo, Q_k (Q_{k+1} ainda não foi calculada) e r_{k+1} .

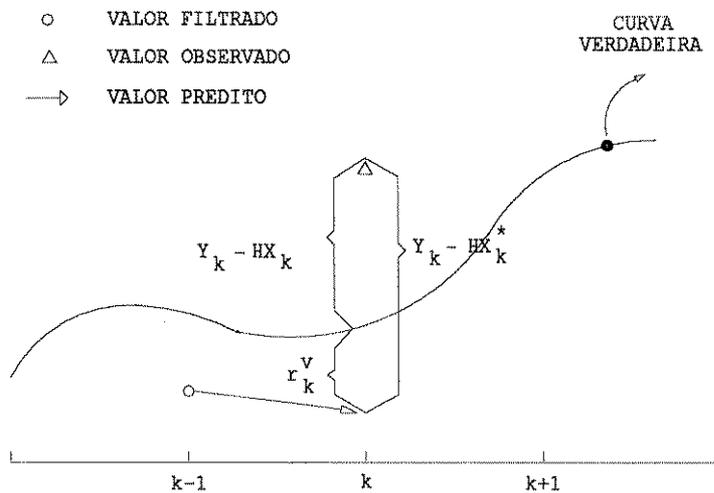


Figura 5.2: Erro de Predição Verdadeiro

Cancelando os valores Y_{k+1} em (5.1), e substituindo X_{k+1}^* e X_{k+1} por suas respectivas equações dadas por (3.35) e (2.12) com $U_k = 0$ ⁵:

$$r_{k+1}^v = H[\Phi(X_k - \bar{X}_k) + \Gamma W_k] \quad (5.3)$$

Portanto

$$E[(r_{k+1}^v)^2] = H(\Phi \bar{P}_k \Phi^T + \Gamma Q_k \Gamma^T) H^T = H \Phi \bar{P}_k \Phi^T H^T + H \Gamma Q_k \Gamma^T H^T \quad (5.4)$$

De (5.2), obtém-se:

$$(r_{k+1}^v)^2 = r_{k+1}^2 - 2r_{k+1}V_{k+1} + V_{k+1}^2 \quad (5.5)$$

Seja q_{k+1} o vetor com os componentes diferentes da matriz Q_k simétrica (como Q_k possui $r \times r$ componentes, q_{k+1} possuirá $\frac{r(r+1)}{2}$ componentes). Segundo Kuga [60], ao se maximizar a probabilidade de ocorrência do resíduo verdadeiro com $q_{k+1} \geq 0$, ou seja, ao se supor que o resíduo verdadeiro das observações tem distribuição Gaussiana, determinam-se as variâncias q_{k+1} do ruído que ajustam o ruído verdadeiro ocorrido, r_{k+1}^v :

$$\text{maximizar } p[r_{k+1}^v], \text{ com } q_{k+1} \geq 0$$

onde $p[r_{k+1}^v]$ representa a densidade de probabilidade do resíduo verdadeiro e o vínculo $q_{k+1} \geq 0$ é coerente com a propriedade de positividade das variâncias. A aplicação desta operação para o resíduo verdadeiro escalar (modelo univariável)⁶ leva à hipótese generalizada de consistência estatística dada por:

$$(r_{k+1}^v)^2 = E[(r_{k+1}^v)^2]$$

⁵Com relação à utilização de U_k , vide Capítulo 4 (eventos determinísticos previstos).

⁶O modelo multivariável é desenvolvido no Apêndice A.

de fato, supõe-se que uma realização é igual à sua esperança matemática, o que, em princípio, não teria significado estatístico nem para um processo ergódico. A fim de introduzir significância estatística, utiliza-se esse resultado, que é parte do valor da pseudo-observação (que será apresentada em seguida). A pseudo-observação é um dado de saída de um novo Filtro de Kalman cujo estado é o vetor q_{k+1} que corresponde aos elementos de Q_k que se querem estimar (também será apresentado a seguir). Portanto, esta hipótese terá validade após algumas observações processadas, dado que o filtro é um estimador seqüencial do estado. Neste processo de filtragem, é como se fosse realizada a média aritmética no tempo, só que esta média tem a vantagem da ponderação pelas covariâncias do filtro. Dada a hipótese da consistência estatística e a definição do resíduo verdadeiro, pode-se notar também que $r_{k+1}^2 = E[r_{k+1}^2]$.

Em seguida, usando a hipótese da consistência estatística e devido a $E[r_{k+1}V_{k+1}] = 0$ e $R_{k+1} = E[V_{k+1}^2]$, obtém-se:

$$E[(r_{k+1}^v)^2] = r_{k+1}^2 + R_{k+1} \quad (5.6)$$

Definindo o erro esperado observado da covariância do erro de predição verdadeiro, denominado pseudo-observação, como:

$$z_{k+1} \triangleq E[(r_{k+1}^v)^2] - E[(r_{k+1}^v)^2 | Q_k = 0] \quad (5.7)$$

De (5.4) e (5.6), tem-se:

$$z_{k+1} = r_{k+1}^2 + R_{k+1} - H\Phi\bar{P}_k\Phi^T H^T \quad (5.8)$$

A equação da pseudo-observação é dada por:

$$z_{k+1} = Mq_{k+1} + \eta_{k+1} \quad (5.9)$$

Da hipótese da consistência estatística, $(r_{k+1}^v)^2 = E[(r_{k+1}^v)^2]$, de (5.4), e de (5.5), segue-se que:

$$Mq_{k+1} = H\Gamma Q_k \Gamma^T H^T \quad (5.10)$$

e

$$\eta_{k+1} = 2r_{k+1}V_{k+1} - V_{k+1}^2 + R_{k+1} \quad (5.11)$$

onde $E[\eta_{k+1}] = 0$.

Essa equação⁷ é chamada de equação da pseudo-observação do submodelo de Espaço de Estados, cujo vetor de estado q_{k+1} estimará os componentes da covariância Q_k . M é um vetor formado por componentes de H e de Γ e η_{k+1} é uma perturbação de média nula.⁸ Q possui $r \times r$ elementos, portanto M possuirá $\frac{r(r+1)}{2}$ elementos ou $r + p$ (onde $p = \binom{r}{2} = \frac{r(r-1)}{2}$), ou seja o número de elementos diferentes da matriz Q_k simétrica. Se forem estimados apenas os elementos da diagonal principal de Q_k , $p = 0$. Como H possui $1 \times r$ elementos e Γ possui $n \times r$ elementos, B possuirá $1 \times r$ elementos. Para a

⁷Existe uma maneira alternativa de se chegar à equação da pseudo-observação, que é forçando-se $E[\eta_{k+1}] = 0$ (ao se definir $E[\eta_{k+1}]$ como em (5.11), isso ocorre automaticamente) e rearranjando os outros termos de 5.8, porque é conveniente que a perturbação possua média nula [40].

⁸Para se obter a matriz da pseudo-observação M , pode-se utilizar a identidade matricial $BQ_k B^T = TR\{B^T BQ_k\}$, onde TR é o traço da matriz e B é o vetor $H\Gamma$.

determinação dos coeficientes da matriz M sugere-se utilizar o produto $B^T B$, $B = H\Gamma$, e os coeficientes são os elementos diferentes de B (B é simétrica).

Por exemplo, $H = \begin{bmatrix} \gamma & 1 - \gamma \end{bmatrix}$, Γ é a matriz identidade e Q_k é uma matriz 2×2 . O produto $B^T B$ fornece:

$$\begin{bmatrix} \gamma^2 & \gamma(1 - \gamma) \\ (1 - \gamma)\gamma & (1 - \gamma)^2 \end{bmatrix}$$

O produto $B^T B Q_k$ fornece:

$$\begin{bmatrix} \gamma^2 Q_{11} + \gamma(1 - \gamma)Q_{21} & \gamma^2 Q_{12} + \gamma(1 - \gamma)Q_{22} \\ (1 - \gamma)\gamma Q_{11} + (1 - \gamma)^2 Q_{21} & (1 - \gamma)\gamma Q_{12} + (1 - \gamma)^2 Q_{22} \end{bmatrix}$$

Considerando que $Q_{12} = Q_{21}$, o traço da matriz resulta:

$$\gamma^2 Q_{11} + (1 - \gamma)^2 Q_{22} + 2\gamma(1 - \gamma)Q_{12}$$

Portanto:

$$M = \begin{bmatrix} \gamma^2 & (1 - \gamma)^2 & 2\gamma(1 - \gamma) \end{bmatrix}$$

Se apenas a diagonal principal de Q_k for estimada, M é:

$$M = \begin{bmatrix} \gamma^2 & (1 - \gamma)^2 \end{bmatrix}$$

Pode-se verificar que o produto Mq_{k+1} , onde $q_{k+1}^T = \begin{bmatrix} Q_{11} & Q_{22} & Q_{12} \end{bmatrix}$, fornece o mesmo resultado que $H\Gamma Q_k \Gamma^T H^T$. O modelo do estado para esse subsistema é:

$$q_{k+1} = \Phi^q q_k + W_k^q \quad (5.12)$$

Onde Φ^q é a matriz de transição de estado, $\Phi^q \in R^{(r+p) \times (r+p)}$; p é o número de elementos na matriz triangular superior de Q_k (ou na inferior, porque Q_k é simétrica), sem a diagonal principal que possui r elementos. Como a dinâmica desse subsistema não é conhecida, a matriz identidade, para essa matriz de transição, é usada. W_k^q é um ruído Gaussiano de média nula com covariância σ_q^2 . Os valores dos elementos dessa matriz estão relacionados à precisão do computador. Por exemplo, se o computador possui d dígitos de precisão, esta covariância é dada por $\sigma_q^2 = 10^{-2d} I$, $I \in R^{(r+p) \times (r+p)}$, onde I é a matriz identidade.

O estado desse submodelo é estimado por um Filtro de Kalman, que é denominado Subfiltro de Kalman.

Estimação do Estado do Subsistema (Subfiltro de Kalman)

Para se estimar a matriz Q_k , i.e., o vetor do estado q_k , aplica-se às equações (5.12) e (5.8), um estimador de estado denominado Subfiltro de Kalman, com as seguintes equações:

- Filtragem

$$K_{k+1}^q = P_{k+1}^{q*} M^T (M P_{k+1}^{q*} M^T + E[\eta_{k+1}^2])^{-1}$$

$$\bar{q}_{k+1} = q_{k+1}^* + K_{k+1}^q (z_{k+1} - M q_{k+1}^*)$$

$$\bar{P}_{k+1}^q = P_{k+1}^{q*} - K_{k+1}^q M P_{k+1}^{q*}$$

• Predição

$$q_{k+1}^* = \Phi^q \bar{q}_k$$

$$P_{k+1}^{q*} = \Phi^q \bar{P}_k^q (\Phi^q)^T + \sigma_q^2$$

onde \bar{P}_{k+1}^q é a covariância do erro de estimação. Esse valor é usualmente iniciado como $\bar{P}_0^q = 10^6 I$, onde I é a matriz identidade. \bar{q}_0 depende do valor inicial de Q_k do Filtro de Kalman. A covariância do erro de pseudo-observação é dada por:

$$E[\eta_{k+1}^2] = 4r_{k+1}^2 R_{k+1} + 2R_{k+1}^2 \quad (5.13)$$

porque $E[V_{k+1}^4] = 3R_{k+1}^2$ e $E[r_{k+1}] = E[V_{k+1}] = 0$.

O vetor q_k é dado por:

$$q_k^T = [Q_{11}, Q_{22} \dots, Q_{rr}, Q_{12}, Q_{13}, \dots, Q_{(r-1)r}] \quad (5.14)$$

A montagem da matriz Q_k , após cada processamento do subfiltro, é:

$$\left. \begin{array}{l} q_1 \rightarrow Q_{11} \\ q_2 \rightarrow Q_{22} \\ \vdots \\ q_r \rightarrow Q_{rr} \end{array} \right\} \begin{array}{l} \text{elementos} \\ \text{da} \\ \text{diagonal} \\ \text{principal} \end{array} \quad \left. \begin{array}{l} q_{r+1} \rightarrow Q_{12} \text{ e } Q_{21} \\ q_{r+2} \rightarrow Q_{13} \text{ e } Q_{31} \\ \vdots \\ q_{r+p} \rightarrow Q_{(r-1)r} \text{ e } Q_{r(r-1)} \end{array} \right\} \begin{array}{l} \text{elementos} \\ \text{fora da} \\ \text{diagonal} \\ \text{principal} \end{array}$$

Após esses valores estimados, a matriz Q_k é testada a fim de se evitar que se torne uma matriz semi-definida negativa ⁹:

$$Q_k = \begin{cases} 0, & \text{se o mínimo } (Q_{ii}, Q_{jj}) < 0 \\ Q_{ij}, & \text{caso contrário} \end{cases}$$

Observações e Considerações

- A perturbação presente na equação da pseudo-observação do subfiltro, η_{k+1} , não é Gaussiana. Em (5.11) vê-se que η_{k+1} resulta do produto de duas Gaussianas V_{k+1} e r_{k+1} , [13, 28], e da subtração de uma distribuição quiquadrado, V_{k+1}^2 , (porque V_{k+1} é Gaussiana por definição), e portanto η_{k+1} não é Gaussiana;
- Se a perturbação é normalmente distribuída, então o Filtro de Kalman produz um estimador não-polarizado de variância mínima. Entretanto, se há *outliers*, este estimador pode ser polarizado (viesado). Nesse caso, um intervalo de confiança será utilizado para melhorar o robustecimento do filtro.

Os pontos acima serão abordados na Seção 5.5.

⁹Existe uma maneira de se evitar o problema de instabilidade numérica, que é ideal para aplicações em tempo real e para computadores com capacidade limitada de operações e poucos dígitos significativos. Nas aplicações em tempo real previstas neste trabalho, para Gerência de Rcd, o número de observações não é muito elevado (no máximo algumas dezenas de observações nos períodos de pico de tráfego porque a coleta é a cada cinco minutos) e a questão de instabilidade numérica é secundária. No entanto, a aplicação desse filtro é recomendada como um trabalho futuro no Capítulo 6 visando outras aplicações em tempo real.

5.2.2 Estimação da Covariância do Erro de Observação (R_k)

A matriz da covariância R_k é estimada utilizando-se uma equação adicional conhecida como Filtro Complementar. Há diferentes procedimentos na literatura (e.g., [18]) para se computar essa matriz recursivamente. Neste trabalho, analisa-se a equação mais apropriada para essa estimação e vários testes são realizados com dados de campo.

Knottnerus [18] e Chemouil [17] utilizam o procedimento de Sage-Husa, estimando a matriz da covariância R_k como:

$$R_k = \frac{k-1}{k} R_{k-1} + \frac{1}{k} (r_k^2 - H P_k^* H^T) \quad (5.15)$$

Entretanto, dependendo dos valores iniciais dos parâmetros, a matriz R_k pode ser semi-definida negativa e o filtro se torna instável. Para evitar essa instabilidade numérica, Fahmeier (vide Knottnerus [18]) utilizou a seguinte equação¹⁰:

$$R_k = \frac{k-1}{k} R_{k-1} + \frac{1}{k} [(Y_k - H \bar{X}_k)(Y_k - H \bar{X}_k)^T + H \bar{P}_k H^T] \quad (5.16)$$

Chow, e na mesma época, Abraham e Ledoiter (vide Knottnerus [18]) mostraram que o estimador de máxima verossimilhança da matriz da covariância R_k é dado por:

$$R_{MV} = \frac{1}{N} \sum_{i=1}^N \frac{R_i r_i^2}{H P_i^* H^T + R_i} \quad (5.17)$$

onde R_i é unitário (procedimento inicial) ou a própria estimação de R . Quando R_i é unitário, R_{MV} estima a covariância R_k após N observações [18]. Este estimador pode ser atualizado recursivamente como¹¹:

$$R_k = \frac{k-1}{k} R_{k-1} + \frac{1}{k} \frac{R_{k-1} r_k^2}{H P_k^* H^T + R_{k-1}} \quad (5.18)$$

No apêndice B, a análise da convergência entre essas duas últimas equações é discutida. Embora a convergência seja mais lenta para a equação de Fahmeier do que para a equação de Chow, a primeira será utilizada neste estágio de testes porque é menos sensível à presença de *outliers* devido a possuir reação mais lenta aos valores observados do que a equação de Chow (Apêndice B).

5.2.3 Testes Realizados e Considerações

Nesta subseção, os seguintes aspectos são analisados:

- O desempenho do filtro utilizando a matriz Q_k estimada como diagonal ou simétrica;
- O desempenho do filtro como uma função da matriz de observação H , dado que ela tem uma influência direta na computação do Subfiltro de Kalman;

¹⁰Essa equação também pode ser obtida a partir de (3.57), substituindo \bar{X}_k^N e \bar{P}_k^N por \bar{X}_k e \bar{P}_k , respectivamente (\bar{X}_k e \bar{P}_k são as melhores informações, em um determinado instante, sobre X_k e P_k).

¹¹É interessante notar que essa equação corresponde a multiplicar o valor de R_i por $\frac{r_i^2}{E[r_i^2]}$, porque o denominador é $E[r_i^2]$, (3.33). Portanto, o valor atualizado de R_k representa, em média, o desvio em relação à hipótese da consistência estatística discutida na seção anterior.

- A influência da escolha da covariância R_k correta.

Os testes da Tabela 5.1 foram desenvolvidos para um sistema com dois estados e o seguinte modelo:

$$H = \begin{bmatrix} \gamma & 1 - \gamma \end{bmatrix} \text{ com } 0 < \gamma \leq 1; \Phi = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}; R_k = [1]; X_k^T = \begin{bmatrix} T_k & \dot{T}_k \end{bmatrix}$$

com Γ sendo a matriz identidade. As variáveis de estado são relacionadas com o nível da carga de tráfego, T_k , e seu incremento de crescimento, \dot{T}_k . Para a estimação de Q_k simétrica, a matriz de pseudo-observação do subfiltro, M , é $[\gamma^2 \quad (1 - \gamma)^2 \quad \gamma(1 - \gamma)]$. Se Q_k é diagonal, $M = [\gamma^2 \quad (1 - \gamma)^2]$ (equação (5.10)).

A Tabela 5.1 apresenta os *REQM* para as projeções a um (*REQM1*) e doze (*REQM12*) passos à frente para as Rotas T4244 (Bauru Trânsito para Marília Trânsito), T4942 (Botucatu Trânsito para Bauru Trânsito) e Tsim (série gerada por meio do método de Monte Carlo, vide Apêndice K).

Tabela 5.1 - *REQM1* e *REQM12* para as Rotas T4244, T4942 e Tsim

Rota	Matriz de Observ.	1	0	0,9	0,1
	Nº Estados	2	3	2	3
T4244	<i>REQM1/REQM12</i>	4,06%/6,29%	4,11%/6,78%	4,06%/5,27%	3,99%/4,34%
Rota	Matriz de Observ.	1	0	0,9	0,1
	Nº Estados	2	3	2	3
T4942	<i>REQM1/REQM12</i>	5,30%/8,51%	5,34%/8,94%	5,26%/9,02%	5,26%/8,25%
Rota	Matriz de Observ.	1	0	0,9	0,1
	Nº Estados	2	3	2	3
Tsim	<i>REQM1/REQM12</i>	2,50%/12,00%	2,50%/12,00%	2,50%/12,00%	2,50%/12,00%

Pelos resultados apresentados na Tabela 5.1 pode-se ver que os *REQM1s* são relativamente independentes do modelo utilizado para a matriz Q_k . Entretanto, se γ é igual a 0,9 e uma matriz Q_k simétrica é estimada, *REQM12* para a Rota T4244 se reduz em quase 50% em relação a quando essa matriz é diagonal e γ igual a um. No entanto, nos dois outros exemplos, Rotas T4942 e Tsim, não há melhoria significativa nem em relação à estimação de Q_k completa ou em relação à utilização de $\gamma = 0,9$.

Os gráficos que serão apresentados a seguir mostram que a utilização de γ diferente de um (mas próximo a um), melhora a precisão das previsões. A Figura 5.3 mostra o *REQM* para a predição doze passos à frente, com as seguintes condições:

- γ igual a 1 e a 0,9;
- Matriz da covariância Q_k diagonal (estimação de dois estados no subfiltro) ou simétrica (estimação de três estados no subfiltro);
- Covariância R_k (modelo escalar, portanto, valor único) variando de 0,125 a 20.

A Figura 5.3 mostra que, quando $\gamma = 1$, o desempenho do Filtro de Kalman depende do método utilizado para se estimar a matriz Q_k . Entretanto, se a covariância do erro de observação aumenta, então o melhor desempenho do filtro é alcançado quando todos os elementos da matriz Q_k são estimados. Se γ é próximo a um, o mínimo *REQM* é sempre obtido quando todos os elementos da matriz Q_k são estimados.

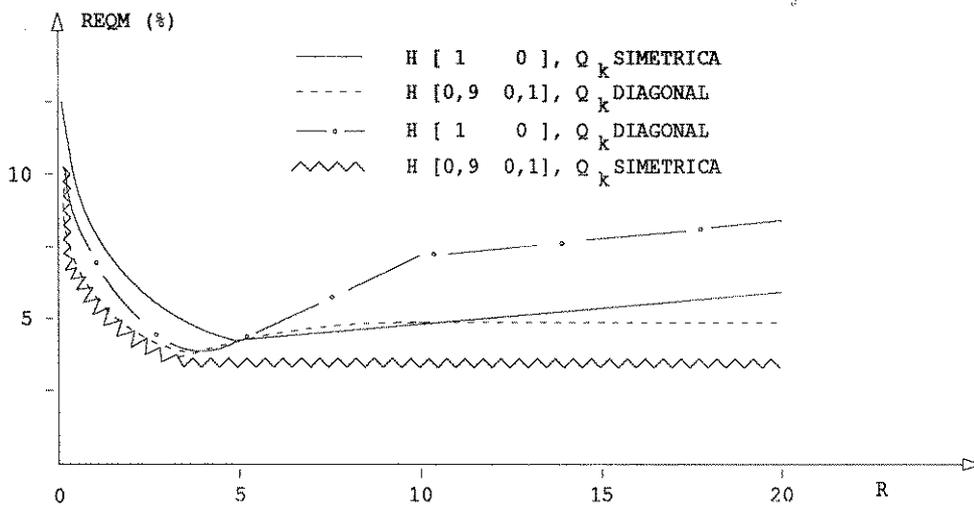


Figura 5.3: REQM em Função da Matriz da Covariância R_k - Rota T4244

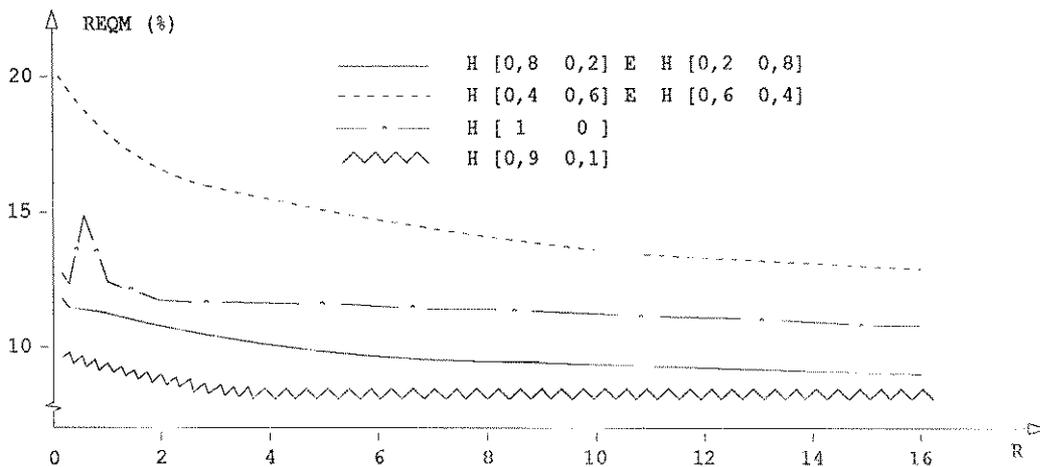


Figura 5.4: REQM em Função da Covariância do Erro de Observação - Rota T4442

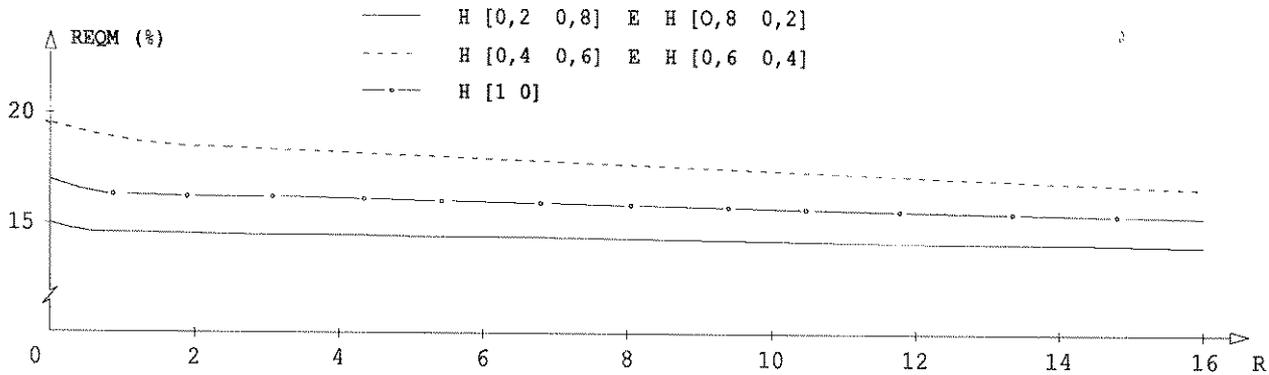


Figura 5.5: *REQM* em Função da Covariância do Erro de Observação - Rota T4942, Botucatu Trânsito para Bauru Trânsito

Na Figura 5.4 R_k varia de 0,125 a 16, γ recebe os valores 1, 0,9, 0,8, 0,6 e 0,4, e uma matriz da covariância Q_k simétrica é utilizada.

Na Figura 5.5, o modelo e as variações dos parâmetros foram idênticos aos da Figura 5.4. Também para a Rota T4249, Bauru Trânsito para Botucatu Trânsito, o resultado foi similar ao apresentado para as outras três curvas.

Como uma conclusão geral, o melhor desempenho do Filtro de Kalman é obtido quando γ é próximo a 0,9, e algumas vezes, quando estima-se a matriz Q_k simétrica.

5.3 Estimação da Matriz de Transição de Estado (Φ_k)

De maneira semelhante à que foi utilizada para a estimação de R_k , a matriz de transição de estado, Φ_k , também é determinada com a inclusão de equações recursivas adicionais ao conjunto de equações do Filtro de Kalman. Nesta seção propõem-se:

- Utilização de um conjunto de equações para se estimar Φ_k recursivamente;
- Efetuar testes com valores reais de algumas rotas de tráfego.

5.3.1 Equações para a Estimação da Matriz de Transição de Estado

Ao invés de se realizar o processo do algoritmo do Capítulo 3 (algoritmo *EM*), ao estimar os parâmetros a partir de um conjunto fixo de N observações, a idéia deste algoritmo é utilizar somente a última observação recebida¹². A atualização da matriz de transição será

¹²Evidentemente que é possível utilizar o algoritmo *EM*, realizando algumas iterações sobre um conjunto de dados recebidos, porém, isso acarretaria uma quantidade adicional de processamento em tempo não real e estaria fora do objetivo deste algoritmo que é também ser utilizado em tempo real.

então (3.55), $\Phi = BA^{-1}$, onde :

$$B = \sum_{k=1}^N [\bar{P}_{k,k-1} + \bar{X}_k(\bar{X}_{k-1})^T]$$

e

$$A = \sum_{k=1}^N [\bar{P}_{k-1} + \bar{X}_{k-1}(\bar{X}_{k-1})^T]$$

e na forma recursiva¹³:

$$B_k = \frac{k-1}{k} B_{k-1} + \frac{1}{k} [\bar{P}_{k,k-1} + \bar{X}_k(\bar{X}_{k-1})^T] \quad (5.19)$$

$$A_k = \frac{k-1}{k} A_{k-1} + \frac{1}{k} [\bar{P}_{k-1} + \bar{X}_{k-1}(\bar{X}_{k-1})^T] \quad (5.20)$$

Portanto, as equações (3.43) e (3.45) do Capítulo 3, com $N = 1$, se tornam:

$$\bar{X}_k^1 = X_k^* Z_{k-1} \quad (5.21)$$

Mas de (3.42), $Z_{k-1} = H^T(\sigma_k^2)^{-1} r_k + 0$, então:

$$X_k^1 + X_k^* + P_k^* H^T(\sigma_k^2)^{-1} r_k = \bar{X}_k \quad (5.22)$$

e

$$\bar{P}_k^1 = P_k^* - P_k^* S_{k-1} P_k^* \quad (5.23)$$

Utilizando-se (3.44), $S_{k-1} = H^T(\sigma_k^2)^{-1} H + 0$, então:

$$\bar{P}_k^1 = P_k^* - P_k^* H^T(\sigma_k^2)^{-1} H P_k^* = (I - K_k H) P_k^* = \bar{P}_k \quad (5.24)$$

Este resultado é intuitivo porque, em um determinado instante, as melhores informações disponíveis sobre X_k e P_k são os valores filtrados.

A equação da covariância de transição de estado também deve ser estimada para $N = 1$. De (3.46), $\bar{P}_{k,k-1}^1 = \bar{P}_{k,k-1} = E[(X_k - \bar{X}_k)(X_{k-1} - \bar{X}_{k-1})^T]$, mas \bar{X}_{k-1} é ortogonal a $X_k - \bar{X}_k$, então:

$$\bar{P}_{k,k-1} = E[(X_k - \bar{X}_k) X_{k-1}^T] = E[(X_k - X_k^* - K_k r_k) X_{k-1}^T] = E[(X_k - \bar{X}_k) X_{k-1}^T] - K_k E[r_k X_{k-1}^T]$$

No apêndice E mostra-se que:

$$E[(X_k - \bar{X}_k) X_{k-1}^T] = L_{k-1} P_{k-1}^* \text{ e } E[r_k X_{k-1}^T] = H L_{k-1} P_{k-1}^*$$

onde $L_{k-1} = \Phi(I - K_{k-1} H)$, portanto:

$$\bar{P}_{k,k-1} = (I - K_k H) L_{k-1} P_{k-1}^* = (I - K_k H) \Phi(I - K_{k-1} H) P_{k-1}^* \quad (5.25)$$

¹³Todas as equações na forma recursiva podem ter os valores $\frac{k-1}{k}$ e $\frac{1}{k}$ substituídos por $1-w$ e w no caso dos valores da matriz serem variantes no tempo, onde $0 < w < 1$ é um parâmetro de suavização.

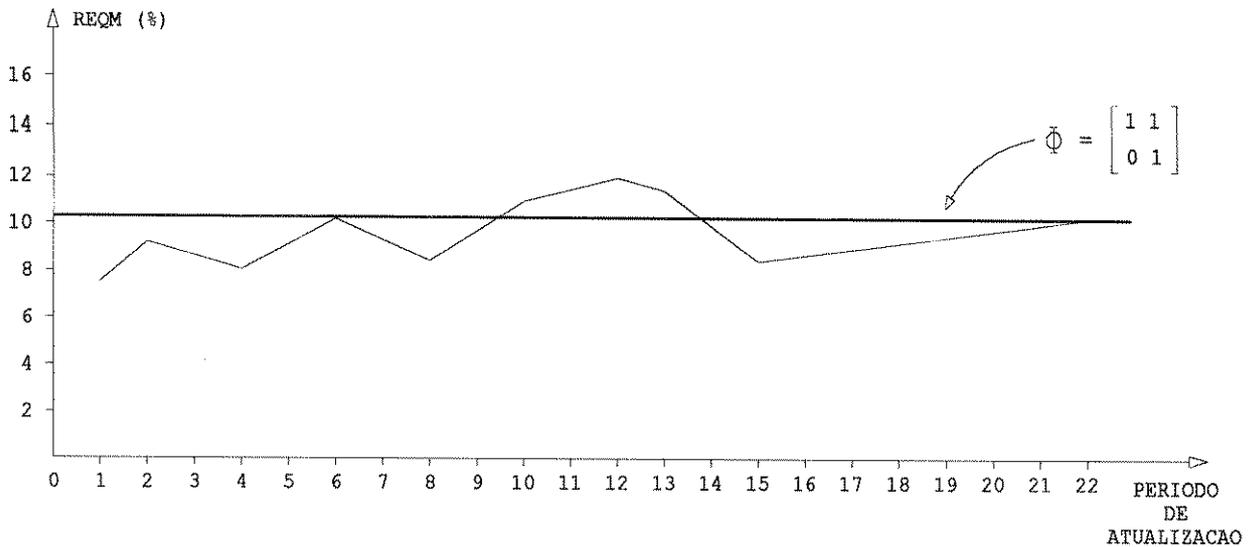


Figura 5.6: Média dos *REQMs* para Quatro Rotas de Tráfego, T4244, T4442, T4249 e T4942, na Projeção Doze Passos à Frente com Atualização de Φ

5.3.2 Testes Realizados e Considerações

Os testes realizados objetivaram a verificação da melhoria no desempenho do filtro ao se fazer a adaptação de Φ_k utilizando o modelo com dois estados. Nestes testes, as melhores condições obtidas nos testes anteriores foram utilizadas, i.e., matriz Q_k simétrica, $\gamma = 0,9$ e estimação de R_k utilizando (5.16). Quatro rotas de tráfego, cada uma com 42 observações mensais disponíveis, foram testadas. Os resultados obtidos após 18 observações processadas foram, então, comparados com aqueles para uma matriz constante, $\Phi = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ ¹⁴.

A Figura 5.6 mostra esses resultados para um período de atualização variando de 1 a 22, i.e., a matriz Φ_k é atualizada a cada observação recebida ou até após 22 observações recebidas.

Da Figura 5.6 pode-se ver que:

- O melhor resultado foi obtido quando Φ_k foi atualizado a cada observação recebida;
- O segundo melhor valor foi obtido para um período de atualização de quatro observações recebidas;
- Embora tenha ocorrido piora do *REQM* para alguns valores de atualização, a atualização de Φ_k deve ainda continuar a ser usada para evitar situações em que esse parâmetro varie com o tempo;

¹⁴A utilização de $\Phi = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ é porque este é o modelo mais simples e mais usual para a matriz de transição de estado representando tendência. No capítulo anterior, estimou-se uma matriz robusta, $\Phi = \begin{bmatrix} 1 & 0,5 \\ 0,01 & 0,4 \end{bmatrix}$. No entanto, essa matriz é específica para um conjunto de rotas considerado e neste capítulo o objetivo é se estimar a matriz Φ genérica e para qualquer tipo de aplicação.

- Em geral, o *REQM* obtido neste caso foi menor que o obtido quando se utiliza o algoritmo *EM* [45]. Em [45], o valor *REQM* para a aplicação do algoritmo *EM* a 40 rotas de tráfego foi 8,4 %. Entretanto, se Φ_k não tivesse sido atualizado, os resultados seriam ligeiramente piores que aqueles obtidos em [45] (na Figura 5.6, nota-se que para a atualização de Φ com frequências 1,2,3 e 4, o *REQM* obtido é inferior ao obtido pela aplicação do algoritmo *EM*). Considerando-se a aplicação do algoritmo *EM* às mesmas rotas da Figura 5.6, T4244, T4442, T4249, e T4942, ainda assim há uma pequena diferença a favor do APSAM (7,2% a 7,34%) [47]. A aplicação do APSAM à Rota T6362, a partir da 10^a observação ¹⁵, forneceu um *REQM* de 21,00% e um desvio padrão de 9,00%. O algoritmo *EM* produziu 22,25% e 19,98%, respectivamente. Outro fato importante é que o algoritmo APSAM requer menos computação que o algoritmo *EM* para ser aplicado.

5.4 Iniciação

Ao receber as primeiras observações, o algoritmo APSAM funciona em paralelo com o algoritmo APS Robusto (descrito no capítulo anterior), mas se ativa o Algoritmo Robusto. Nessa fase de *standby*, cuja duração depende do tipo de dado amostrado (dados mensais, anuais, etc.), o APSAM inicia seus parâmetros que serão utilizados na sua fase ativa. No entanto, enquanto o APS Robusto é ativo, é necessário que o APS Adaptativo desempenhe da melhor maneira possível. Portanto, é conveniente que seus parâmetros iniciais: $(\bar{X}_0, \bar{P}_0, R_0, Q_0$ e $\Phi_0)$, sejam estimados, sendo que a necessidade de melhor ou pior precisão é função do tipo de dados amostrados e do modelo adotado. Para os parâmetros iniciais, considerou-se a seguinte proposta:

- Covariância do Erro de Observação (R_0)

Inicia-se com o valor de $R_0 = 1$ e utiliza-se (5.16), mas com w e $1-w$ substituindo $\frac{k-1}{k}$ e $\frac{1}{k}$, respectivamente, para as 18 primeiras observações quando o APS Robusto está ativo. Depois do período transiente (supondo 18 observações para dados mensais), utiliza-se a equação de Chow (5.18) que tem condições de detectar mudanças na dinâmica do sistema mais rapidamente (Apêndice B). A sugestão para a escolha de w é a adoção de $w^{I_N} \simeq 0$, onde I_N é o comprimento do período em que opera o APS Robusto, que é função do tipo de dados observados.

- Matriz de Transição de Estado (Φ_0)

Inicia-se com Φ_k inicial = $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ ¹⁶. Em seguida, utilizam-se as equações (5.20), (5.19) e (3.55) para atualizar o valor de Φ_k a cada quatro observações recebidas, com w e $1-w$ em lugar de $\frac{k-1}{k}$ e $\frac{1}{k}$, onde o valor de w é determinado com no item anterior. Quando o algoritmo Adaptativo passa a ativo, pode-se continuar com w e $1-w$ ou utilizar $\frac{k-1}{k}$ e $\frac{1}{k}$, dependendo dos parâmetros serem variantes no tempo ou não.

¹⁵Outros resultados podem decorrer da aplicação do algoritmo a partir de outro ponto. É interessante notar que o algoritmo *EM* requer, em testes efetuados com dados mensais, em torno de vinte observações para a estimação adequada de seus parâmetros.

¹⁶Isto é válido para o modelo de crescimento linear dado por (3.10) e (3.11); no caso de outros modelos, como o sazonal por exemplo, as condições de iniciação devem ser coerentes com o Capítulo 3.

Na prática, o que ocorre é o pagamento de um preço pelo fato de se supor que os parâmetros são variantes no tempo, e de fato não o são, e vice-versa. Evidente que se os parâmetros são invariantes no tempo, os resultados serão melhores quando se utiliza $\frac{k-1}{k}$ e $\frac{1}{k}$ (média aritmética recursiva) porque haverá um maior número de observações disponíveis para a correta estimação dos parâmetros. A experiência deve ditar que fator utilizar. Outra alternativa é a adoção de um valor de w correspondendo a uma janela de tamanho suficiente para a correta estimação dos parâmetros. Esse valor só pode ser obtido pela realização de testes de campo.

- Matriz da Covariância do Erro de Modelagem (Q_0)

Como o valor de Q_0 é estimado por meio da utilização do subfiltro, conforme apresentado na Seção 5.2, seu valor inicial verdadeiro não é crítico: quando o filtro Adaptativo se torna ativo, o valor de Q_k deve representar adequadamente o valor verdadeiro. Os elementos de Q_0 tomam valores da mesma ordem de R_0 .

- Valor do estado inicial (\bar{X}_0) Se supõe-se um modelo com duas variáveis de estado, dado por (3.10) e (3.11), \bar{X}_0 é iniciado como no APS Robusto. Se for suposta a variante dois do modelo dado por (3.10) e (3.11), então o estado inicial é:

$$X_0 = \begin{bmatrix} Y_0 \\ \hat{g}Y_0 \\ 1-\gamma \end{bmatrix}$$

Onde Y_0 é o primeiro valor observado e \hat{g} é a taxa de crescimento agregado¹⁷.

- Valor da covariância do erro de estimação inicial (\bar{P}_0)

Se o modelo for dado por (3.10) e (3.11), adota-se a estimação proposta para o algoritmo APS Robusto. Em outros casos, adota-se:

$$\bar{P}_0 = mI$$

Onde m é um número grande para mostrar o desconhecimento em relação ao valor verdadeiro do parâmetro (quando não se conhece a distribuição verdadeira, costuma-se admitir que a variância é grande), por exemplo $m = 10^6$, e I é a matriz identidade de ordem compatível com o tamanho do vetor de estado.

Algumas observações finais:

- No período de iniciação, o filtro Adaptativo não utiliza o mecanismo de reiniciação, porque este já é utilizado pelo filtro Robusto;

¹⁷Outros modelos devem ser iniciados de forma coerente com os exemplos de modelos aditivos do Capítulo 3. Como exemplo, se o modelo é dado por (3.8), com dois estados, então:

$$\bar{X}_0 = \begin{bmatrix} Y_0 \\ Y_0 \end{bmatrix}$$

Porque os dois valores correspondem à mesma variável de estado defasada no tempo.

No caso do modelo dado por (3.16), o vetor \bar{X}_0 deve ser nulo.

Desta forma, para cada modelo do Capítulo 3 há uma maneira mais adequada de iniciação do estado inicial.

- Dependendo do modelo específico a ser adotado, diferentes valores iniciais devem ser utilizados. Em [18], alguns exemplos são apresentados;
- Para modelos aditivos que consideram vários padrões para série temporal do tráfego, tais como, tendência, sazonalidade, etc., a iniciação de cada modelo individual é feita independentemente (e.g., [18, 19, 20, 21]);
- Os parâmetros \bar{X}_0 e \bar{P}_0 também podem ser iniciados por meio dos Mínimos Quadrados Generalizados [34, 26], quando há dados iniciais disponíveis.

5.5 Reiniciação e Robustecimento

Ao receber as primeiras observações, o APS Adaptativo não possui ainda informações suficientes para se estimarem adequadamente seus parâmetros. Da mesma forma, se houver reiniciação, também há um tempo até que os parâmetros possam ser reestimados. Com relação à questão do robustecimento, o filtro deve ser capaz de evitar a ocorrência de observações duvidosas (*outliers*) e também responder a mudanças na tendência devidas à ocorrência de eventos determinísticos não-previstos. Além disso, no algoritmo proposto, a perturbação da equação da pseudo-observação não é Gaussiana (Seção 5.2). Portanto, é fundamental que haja um controle sobre os resíduos para garantir que a aplicação seqüencial do subfiltro produza resultados equivalentes ao do filtro linear Gaussiano.

Esta seção propõe a utilização de uma transformação adequada, T_{R_k} , para que o filtro se torne robusto dentro da faixa de um a dois desvios-padrões para o estimador da covariância do erro de estimação [39, 38]. No caso do APS Robusto (Capítulo 4), supõe-se que os resíduos são constantes (termos de inovação constantes) e calcula-se um intervalo de confiança que deve ser utilizado por todas as observações (que corresponde a um nível de saturação, acima (ou abaixo) do qual adota-se seu valor em substituição ao valor observado). Para o APS Adaptativo, o intervalo de confiança a ser adotado não é constante, mas varia em função do resíduo, sendo diferente para cada observação recebida. A transformação utilizada, T_{R_k} , desempenha o papel de um tipo de normalização, fazendo com que a regra de decisão seja sempre a mesma, independente do valor absoluto do resíduo. No Apêndice A desenvolve-se esta transformação também para o modelo multivariável.

5.5.1 Intervalo de Confiança para Reiniciação e Robustecimento

Ao receber uma observação aberrante (*outlier*) a observação é simplesmente descartada e o valor limite do intervalo de confiança utilizado para proteção é adotado em seu lugar.

Regra utilizada¹⁸

Se dois ou mais dados consecutivos são *outliers*, todos excedendo o limite superior do intervalo de confiança ou todos abaixo de seu limite inferior, o algoritmo é reiniciado pois supõe-se a existência de um evento determinístico não-previsto.

¹⁸No Capítulo 6 será proposta, como trabalho futuro, a utilização de um teste de hipóteses para detectar a presença de eventos determinísticos não-previstos.

Na reiniciação¹⁹, os dados representando um período completo de aquisição (mês, anos, etc.), são processados. Após a reiniciação, as primeiras observações são também processadas pelo APS Robusto. Para aplicações em tempo real é, provavelmente, desejável ter um ajuste automático para o nível de tendência, sem reiniciação. Quando o ruído da perturbação da equação da pseudo-observação apresenta uma distribuição de cauda acentuada (*heavy tailed*)²⁰, a equação da estimação do Filtro de Kalman é mudada, conforme proposto por Martin e Masreliez [39]:

$$\bar{X}_k = X_k^* + P_k^* H^T T_{R_k}^T \Psi(\nu_k) \quad (5.26)$$

onde $T_{R_k}^T$ é proposta como:

$$T_{R_k} = (HP_k^* H^T + R_k)^{-\frac{1}{2}} \quad (5.27)$$

porque com este valor, para observações contidas dentro do intervalo de confiança previsto, o filtro Adaptativo terá um ganho igual ao Filtro de Kalman padrão. A função $\Psi(\nu_k)$ é definida por [39]:

$$\Psi(\nu_k) = \begin{cases} \nu_k, & |\nu_k| \leq NIC \\ NIC \operatorname{sgn}(\nu_k), & |\nu_k| > NIC \end{cases} \quad (5.28)$$

onde $\nu_k = T_{R_k} r_k$, r_k é dado por (3.31), $\operatorname{sgn}(\cdot)$ é a função sinal e NIC é o número de desvios-padrões da distribuição normal padrão.

Como observações finais:

- Quando há ocorrência de eventos determinísticos previstos, a correção da projeção n passos à frente é utilizada como desenvolvido no Capítulo 4 e em [46];
- Com relação ao intervalo de confiança para n passos à frente, [18] e equação (2.65).

5.6 Aplicações em Tempo Real

Além dos testes realizados nas seções anteriores, nesta seção objetiva-se verificar a adequação do algoritmo para situações em que há mudanças na dinâmica do sistema amostrado, tanto por mudança no valor do nível da carga de tráfego, quanto por mudanças no valor dos parâmetros.

¹⁹Na fase de iniciação, o filtro Adaptativo não deve usar o procedimento de reiniciação, dado que ele já está sendo utilizado pelo filtro Robusto, nem deve utilizar a correção do nível estimado em sua plenitude (e.g., tolerar dois desvios-padrões para o erro do termo de inovação em vez de somente um desvio-padrão), a fim de permitir uma convergência mais rápida.

²⁰Quando a perturbação da equação da observação for modelada por uma distribuição normal contaminada, tem-se:

$$V_k \sim (1 - \epsilon)\mathcal{N}(0, R_k) + \epsilon\mathcal{N}(0, jR_k)$$

onde $0 \leq \epsilon \leq 0,2$ para que valham as condições de robustecimento de Huber, [38], e $j > 1$. Devido a ambas as Gaussianas terem média nula, a variância verdadeira de V_k é:

$$\operatorname{VAR}[V_k] = (1 - \epsilon)R_k + \epsilon jR_k$$

A covariância verdadeira do termo de inovação é:

$$\operatorname{VAR}[Y_k - HX_k^*] = \operatorname{VAR}[Y_k - HX_k + HX_k - HX_k^*] = \operatorname{VAR}[V_k] + HP_k^* H^T = (1 - \epsilon)R_k + \epsilon jR_k + HP_k^* H^T$$

Para detecção dessas mudanças, utiliza-se o para dados acumulados [48] e também verifica-se a seqüência de inovações, que deve apresentar um padrão descorrelacionado (ruído branco Gaussiano). A seguir uma breve descrição desses métodos.

5.6.1 Fator de Tendência de Laplace

O Fator de Tendência de Laplace²¹ para dados acumulados é [48]:

$$u_M = \frac{\frac{\sum_{i=1}^M (i-1)Y_i}{Y_T} - \frac{M-1}{2}}{\sqrt{\left(\frac{M^2-1}{12Y_T}\right)}} \quad (5.29)$$

onde M é o valor da janela de tamanho fixo e corresponde às últimas M observações recebidas, e Y_T é o valor acumulado dos Y_i valores de tráfego recebidos. É possível mostrar, [48], que u_M se distribui normalmente com média zero e variância um quando o processo não apresenta nem crescimento nem decrescimento. O significado de u_M é o quanto a média aritmética se desviou da mediana dos tempos observados, porque quando a distribuição é Gaussiana, a média aritmética e a mediana são coincidentes.

Implementação do Fator de Tendência de Laplace

Primeiro, desloca-se um vetor que possui M posições, uma a uma para a frente para que o último valor lido ocupe a primeira posição. A seguir, calcula-se a parcela $D_i = \sum_{i=1}^M (i-1)Y_i$.

Y_T é calculada como:

$$Y_T \leftarrow Y_T + Y_k - Y_{k-M}$$

ou seja, adiciona-se a observação corrente e elimina-se a mais antiga. Se tiverem ocorrido mais que M observações, calcula-se u_M e guarda-se N_V valores.

Vários testes foram efetuados e para aplicação em tempo real, adotou-se um intervalo de confiança para u_M . Para evitar que haja acentuada ocorrência de crescimentos ou decrescimentos, de tal forma que o filtro não disponha de tempo suficiente para adaptar seus parâmetros, utiliza-se o seguinte:

- Quando um número de vezes consecutivas, N_V , u_M esteve acima do limite superior previsto, admite-se que há tendência de crescimento;
- Quando um número de vezes consecutivas, N_V , u_M esteve abaixo do nível inferior previsto, admite-se que há tendência de decrescimento;
- Quando um número de vezes consecutivas, N_V , u_M esteve dentro do intervalo de confiança previsto, admite-se que o nível nem cresceu nem decresceu.

Na prática, utilizou-se um intervalo de confiança de um desvio-padrão (68% de confiança) e 2,33 desvios-padrões (98% de confiança), e $N_V = 2$ e $N_V = 3$.

²¹Os testes realizados indicaram que o Fator de Tendência de Laplace detecta com rapidez a mudança de tendência do padrão de tráfego, bastando algumas observações.

5.6.2 Verificação se a Seqüência de Inovações é Descorrelacionada

Para que o filtro esteja desempenhando a contento, é preciso que a seqüência de inovações seja um ruído branco Gaussiano. A equação para cálculo da autocorrelação amostral é (2.50), e alguns detalhes de implementação são apresentados a seguir.

Implementação da Detecção do Ruído Branco Gaussiano

As seguintes operações, de maneira geral, foram utilizadas para a implementação da Detecção do Ruído Branco Gaussiano. Primeiro desloca-se M posições um vetor que possui $M + 1$ posições, uma a uma para a frente, para que o último valor lido ocupe a primeira posição. Para cálculo da média aritmética no intervalo de tamanho M ²² tomou-se:

$$S_M \leftarrow S_M + r_k - r_{k-M}$$

e

$$\bar{r} \leftarrow \frac{S_M}{M}$$

onde S_M é uma variável auxiliar para fazer o somatório, \bar{r} é a média aritmética dos M valores, M é o tamanho fixo da janela e r_k são os valores das inovações. Em seguida calcula-se o denominador da função de autocorrelação:

$$\nu_0 \leftarrow \nu_0 + (r_k - \bar{r})^2 + (r_{k-M} - \bar{r}_a)^2$$

onde \bar{r}_a é a média aritmética anterior, antes de receber a informação atual. Para se calcular as L autocorrelações, procede-se da seguinte forma:

Para j variando de 1 a L , faça

Para i variando de 1 até $M - j$ faça

$$\nu_j \leftarrow \nu_j + (r_i - \bar{r})(r_{i+j} - \bar{r})$$

Fim

$$\text{Calcule } \rho_j = \frac{\nu_j}{\nu_0}, \nu_j \leftarrow 0$$

Fim.

para cada valor de ν_j calculado, deve-se verificar (deve ser inserido dentro do programa descrito) se está dentro do intervalo de confiança $\pm t_{\alpha} \sqrt{\frac{1}{M}}$, onde t_{α} é o valor da distribuição t de Student com $M - 1$ graus de liberdade. Na prática, adotou-se $M = 30$, $L = 10$ e $t_{\alpha} = 2$. Se $L = 10$, com 90% de certeza pode-se ter no máximo um ponto fora do intervalo de confiança, ou seja, admite-se que não se trata de ruído branco se mais de um ponto estiver fora do intervalo.

²²Em geral, o valor de M para a Tendência de Laplace é menor de que para a Detecção de Ruído Branco (neste exemplo, um é dez (Laplace) e o outro é trinta (Ruído Branco)).

Comparação entre os Testes para Aplicação em Tempo Real

A Tabela 5.2 mostra o valor comparativo, em *up* (unidade padrão de tempo), entre os diversos métodos utilizados para aplicação em tempo real. O objetivo de se utilizar *up* é o fato de que diversos processadores têm velocidades diferentes e é preciso escolher o computador adequado para a aplicação em tempo real desejada, tendo vista o requisito de tempo.

Tabela 5.2 - Comparação entre Métodos para Aplicação em Tempo Real (em *up*)

Nº estados	2	3	4	5	6
Atualiz. de Φ	1	1,67	2,56	3,87	5,50
Tend. de Laplace	2,5	2,5	2,5	2,5	2,5
Detecção R. Branco	47	47	47	47	47

Como pode-se observar, apenas a atualização de Φ (Seção 5.3) é função do número de estados e os demais métodos não. A detecção de Ruído Branco é o teste que demanda maior processamento porque requer um maior número de observações para que sua decisão seja confiável. O teste de Tendência de Laplace, por ser um valor praticamente fixo com o tamanho do estado (também o de Detecção de Ruído Branco o é), vai ficando mais viável (em termos de tempo) conforme aumentam o número de estados. De qualquer forma, para até seis estados, nem o teste de Tendência de Laplace, nem o de Detecção de Ruído Branco, e nem a atualização de Φ (Seção 5.3) apresentaram carga de processamento maior que em torno de 10 % da carga para a realização de todas as outras funções do algoritmo.

5.6.3 Testes Realizados e Considerações

Para efeito de se testar padrões de séries temporais que podem ser encontradas em aplicações em tempo real, utilizou-se séries fora dos padrões da maioria das rotas testadas, visando avaliar o desempenho do algoritmo em diversas situações. Para isso escolheu-se, entre 40 rotas disponíveis, uma rota que apresenta uma grande descontinuidade (um salto abrupto), T6362²³, e uma outra rota que apresenta *REQM* elevado em relação aos demais testes efetuados, Rota Bauru Trânsito para Jaboticabal Trânsito, T4249. Além disso, gerou-se, pelo método de Monte Carlo, uma série temporal, Tsim, que apresenta crescimento, decrescimento e um trecho nem de crescimento nem de decrescimento (Apêndice K). Essa última série temporal procura modelar uma situação de perfil de tráfego comum durante as horas do dia, em que há crescimento, estabilização e decrescimento do valor da carga de tráfego. O modelo de Espaço de Estados é o mesmo (dois estados) utilizado anteriormente e adotaram-se os melhores resultados dos testes precedentes com relação à estimação de R_k ²⁴, Q_k , Φ_k , e $\gamma = 0,9$. Os resultados indicaram que:

- As três séries temporais testadas apresentaram *REQM* (Figuras 5.7, 5.8 e 5.9) maior para utilização do Fator de Tendência de Laplace que pela utilização da adaptação de Φ em tempo real. De fato, o Fator de Tendência de Laplace indica onde houve

²³Todavia para aplicações em tempo real, em geral, não ocorrem saltos abruptos.

²⁴Com relação à estimação de R_k , as previsões mais estáveis ocorrem com a utilização da equação de Fahmeier nas primeiras observações e, posteriormente utilizando a equação de Chow, ou de Abraham e Ledoiter (vide Apêndice B).

mudanças na série temporal ²⁵ (quando o Fator de Tendência de Laplace indica crescimento, utiliza-se a matriz $\Phi = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, quando indica decrescimento, utiliza-se $\Phi = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$, mas quando não há nem crescimento nem decrescimento, utiliza-se $\Phi = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$), mas como essa informação chega atrasada, devido à necessidade de uma janela de tempo (de tamanho M observações; utilizaram-se $M = 10$ e $M = 6$ nos testes) utilizada para a sua tomada de decisão, a introdução de uma matriz de transição de estado mais de acordo com a tendência apontada pelo teste (crescimento, decrescimento, nem crescimento nem decrescimento) não apresenta melhoria nos resultados;

- Além disso, há vários parâmetros que estão sendo adaptados em conjunto pela técnica de relaxação e, melhorar um isoladamente (Φ), não implica em uma melhoria global. Isto pode ser observado na série gerada por Monte Carlo, Tsim, em que todos os parâmetros eram inicialmente conhecidos. O fato do algoritmo ter estimado com bastante precisão alguns dos parâmetros não implicou em *REQM* menor do que em outra situação em que nenhum dos parâmetros foi estimado com precisão, mas seu conjunto alcançou melhor desempenho (isto apenas valeria caso todos os parâmetros fossem estimados com precisão). O melhor resultado para a série Tsim (Figura 5.7 (b)), sem a utilização do Fator de Tendência de Laplace, foi com parâmetros estimados com menos precisão que no teste utilizando Tendência de Laplace. Um outro exemplo é o fato de que o desempenho do filtro depende da relação $\frac{Q}{R}$ e não isoladamente de Q e R ([23] e Apêndice D);
- A verificação se a seqüência de inovações é branca não se mostrou conveniente porque o algoritmo indica a necessidade de reiniciação antes do teste indicar a presença de ruído não-branco. Isto se deve à necessidade que o teste tem de possuir um número significativo de valores para poder decidir ²⁶. Além do mais, conforme a Tabela 5.2, o teste para detecção do Ruído Branco também é o mais demorado;
- Ambas as rotas com dados de campo (T6362 e T4249, Figuras 5.9 e 5.8, respectivamente), tiveram indicação de reiniciação, a primeira na 27^a observação e a segunda na 32^a observação processada. Essa reiniciação ocorre, em ambos os casos, tanto para $NIC = 1$ quanto para $NIC = 2$, 33 desvios-padrões. Como os valores de projeção

²⁵Vários testes foram realizados apenas para verificar a eficácia do Fator de Tendência de Laplace na detecção de mudanças de Tendência. De fato, nas Rotas T6362, T4249, T4942, na série Tsim, e na série representando o número de falhas da Central Telefônica Trópico RA [48], as indicações de mudança de tendência foram bastante precisas. Em particular, para a série Tsim, houve indicação de tendência de crescimento em $k = 12$, de tendência de decrescimento em $k = 60$ e indicação de nem crescimento nem decrescimento em $k = 31$. Como a série Tsim foi gerada por simulação, sabe-se que esses valores são $k = 1$, 26 e 51, respectivamente. Os valores dos parâmetros do Fator de Tendência de Laplace que possibilitaram melhor detecção foram: $M = 10$ (tamanho da janela), $N_V = 3$ (número de vezes consecutivas que um ponto está fora do intervalo de confiança, todas as vezes para cima ou todas as vezes para baixo), e intervalo de confiança de um desvio-padrão.

²⁶Outras rotas também foram testadas para verificar a presença de ruído branco: Bauru Trânsito para Marília Trânsito, T4244, Botucatu Trânsito para Bauru Trânsito, T4942, Marília Trânsito para Bauru Trânsito, T4442 e Avaré Trânsito para Bauru Trânsito, T4742. Todos os testes indicaram a presença de ruído branco na seqüência de inovações.

obtidos antes da reiniciação são superiores àqueles (sabe-se que o algoritmo Adaptativo, em geral, inicia com erros relativamente altos devido ao desconhecimento dos seus parâmetros) que é possível se obter com a aplicação do APS Robusto para um conjunto de rotas com características semelhantes, justifica-se a utilização do APS Robusto após a reiniciação do algoritmo (uma observação interessante é que o período de atividade do APS Robusto coincide com o período (número de observações) que o algoritmo *EM* necessita para estimar os parâmetros do filtro);

- Ao se utilizar $NIC = 1$ em vez de $NIC = 2,33$ desvios-padrões, melhora-se a precisão do filtro na presença de *outliers*, no entanto, quando ocorrem eventos determinísticos não-previstos, o filtro demora mais para se reiniciar;
- O filtro estima os parâmetros com os valores do termo de inovação, r_k , substituídos (robustecidos) por $\frac{+}{-} NIC(HP_k^*H^T + R_k)^{-1}$, quando este valor é ultrapassado (Seção 5.5). Esse procedimento é fundamental para a estimação adequada de R_k , principalmente quando se utiliza a equação de Chow (5.18), e também para a estimação conjunta dos demais parâmetros;
- Ao se estimarem Φ , Q_k e R_k somente quando as observações estão dentro do intervalo de confiança para o termo de inovação permite a estimação mais precisa de Φ ; no entanto, os *REQMs* globais se mostraram piores ao se adotar $NIC = 1$ ou $NIC = 2,33$;
- As Rotas T6362 e T4249 tiveram, após a 18^a observação, um desempenho melhor com Φ adaptado do que com $\Phi = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. Até a 18^a observação, a variância dos erros observados é maior do que a ocorrida com a utilização do APS Robusto, o que justifica sua utilização no início;
- As Figuras 5.8 (a) e 5.9 (b) para as Rotas T4249 e T6362 mostram, respectivamente, os valores de tráfego observados, bem como as projeções a seis passos com Φ fixo, com a utilização do Fator de Tendência de Laplace e com a adaptação de Φ . As Figuras 5.8 e 5.9 mostram os respectivos *REQMs*. Uma conclusão é que a utilização do Fator de Tendência de Laplace, por apresentar *REQM* mais elevado, também provoca mais rapidamente a reiniciação do algoritmo. A reiniciação implica em começar novamente com erros mais altos, porém, sob controle (APS Robusto);
- Na aplicação do algoritmo à Rota T6362, ficou claro o compromisso entre otimalidade e sensibilidade. Devido a esta rota apresentar uma descontinuidade (queda abrupta no nível do tráfego mensal, que não é peculiar aos processos de tempo real) nota-se que um algoritmo mais preciso, que acompanha mais as variações, também cometerá maiores erros. Nesta situação, o algoritmo deve se readaptar rapidamente ao novo nível ou se reiniciar;
- Os valores de média e desvio padrão, na projeção para a Rota T6362 são, respectivamente, 0,19 e 0,18. No entanto, esses valores foram calculados a partir da 4^a observação e tanto incluem os altos valores iniciais, que estariam sendo tratados pelo APS Robusto, quanto os valores finais que determinaram a reiniciação do algoritmo;

- Os mesmos valores, com as mesmas considerações, para a Rota T4249, são, respectivamente, 0,12 e 0,07.

A Figura 5.7 ilustra as projeções seis passos à frente e os *REQMs* para a Rota Tsim, cujo processo de geração dos dados e valores se encontram no Apêndice K.

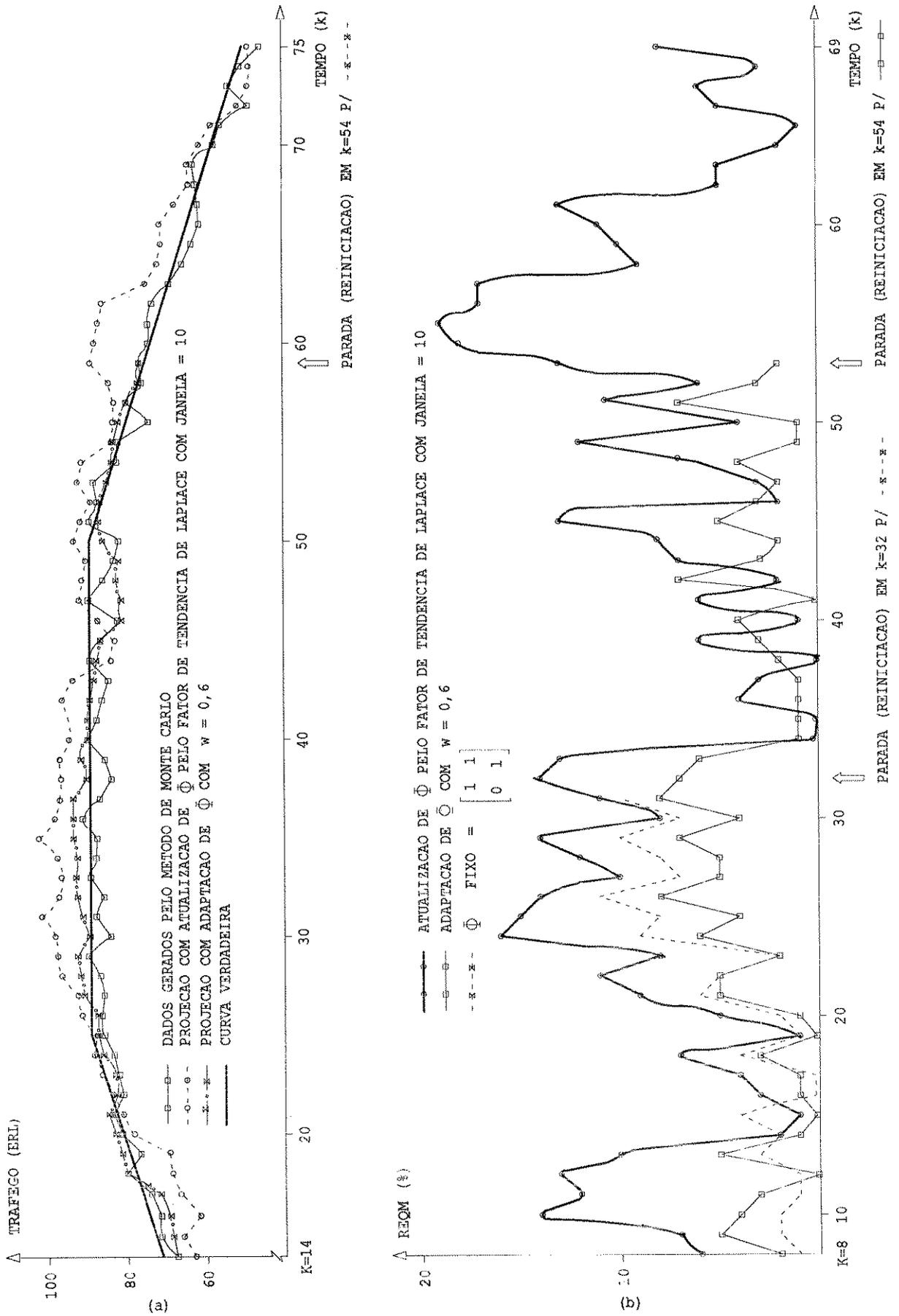


Figura 5.7: Projeções a Seis Passos e *REQM* para a Rota Tsim

A seguir, algumas considerações específicas sobre a Figura 5.7 e a série Tsim:

- As projeções a seis passos mostram o algoritmo com adaptação de Φ com $w = 0,6$, e de outra maneira pela substituição de Φ utilizando a informação fornecida pelo Fator de Tendência de Laplace. O procedimento utilizando o Fator de Tendência de Laplace, substituí o valor de Φ quando há indicação de mudança de tendência, conforme descrito a seguir: a tendência é de decrescimento, utiliza-se $\Phi = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$, quando há crescimento da tendência, utiliza-se $\Phi = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, e quando a tendência nem cresce nem decresce, utiliza-se $\Phi = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. Com Φ adaptado, o *REQM* para a projeção a seis passos é 0,03 e seu desvio-padrão é 0,02. Os valores para a utilização do Fator de Tendência de Laplace são, respectivamente, 0,05 e 0,03;
- A parada ocorrida no processamento do algoritmo é, na verdade, uma indicação da necessidade de reiniciação que se deve ao fato de se ter usado $NIC = 1$ e duas observações consecutivas fora do intervalo de confiança (Seção 5.5). Com $NIC = 2,33$ e três observações consecutivas fora do intervalo de confiança, o algoritmo não acusou necessidade de reiniciação, e embora seu desempenho tivesse se tornado pior após a 54^a observação, ele logo recuperou seu desempenho anterior;
- Com a reiniciação, o algoritmo não conseguiu se adaptar com a mesma velocidade que quando se deixa em funcionamento mais livre ($NIC = 2,33$ em vez de $NIC = 1$), ou seja, não se justifica o emprego do APS Robusto neste caso. Também na iniciação, não se recomenda a utilização do APS Robusto, principalmente porque para padrões de tráfego muito gerais, do tipo das aplicações em gerência de rede com valores coletados a cada cinco minutos, não se possui uma avaliação segura do valor inicial dos parâmetros;
- A Figura 5.7 (a) mostra também os dados gerados pelo método de Monte Carlo e a curva verdadeira, que é a função determinística suposta. O lado direito da figura mostra os *REQMs* para três situações: com $\Phi = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, para Φ adaptado com $w = 0,6$ (ambos com janela de tamanho 6), e para Φ imposto ($\Phi = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, $\begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$, ou $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$) com o auxílio do Fator de Tendência de Laplace com janela de tamanho 10;
- A utilização de Φ fixo foi forçada porque se sabia que no início os dados foram gerados com $\Phi = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. Quando o valor da matriz de transição de estado mudou houve um aborto no processamento, que significa que o algoritmo deve se reiniciar. Como era de se esperar, a utilização do valor de Φ correto (já conhecido) possibilitou melhores resultados nas primeiras observações processadas, mas logo em seguida o filtro com Φ adaptado superou seu desempenho, mostrando que se adaptou à mudança no valor de Φ ;
- A utilização da informação do Fator de Tendência de Laplace não melhorou o desempenho do algoritmo e, com a utilização da janela de tamanho 6 ainda piora com

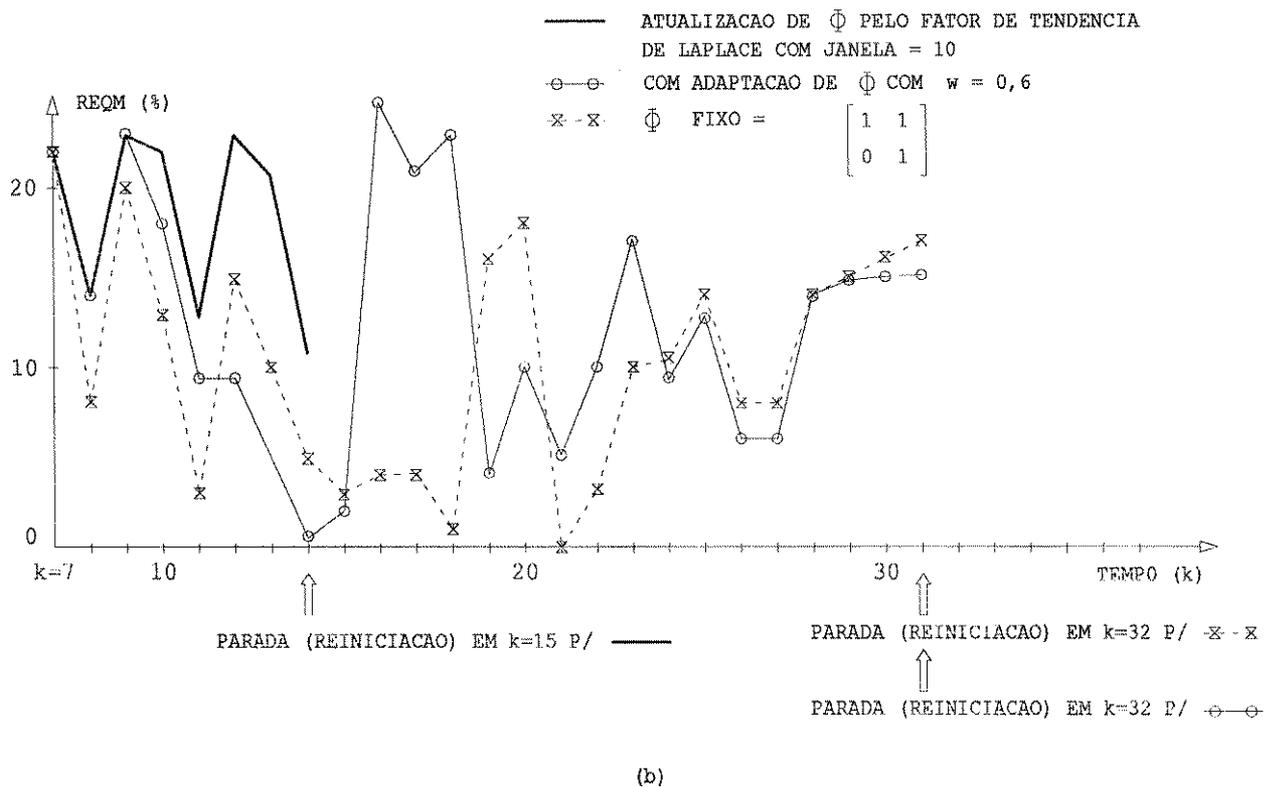
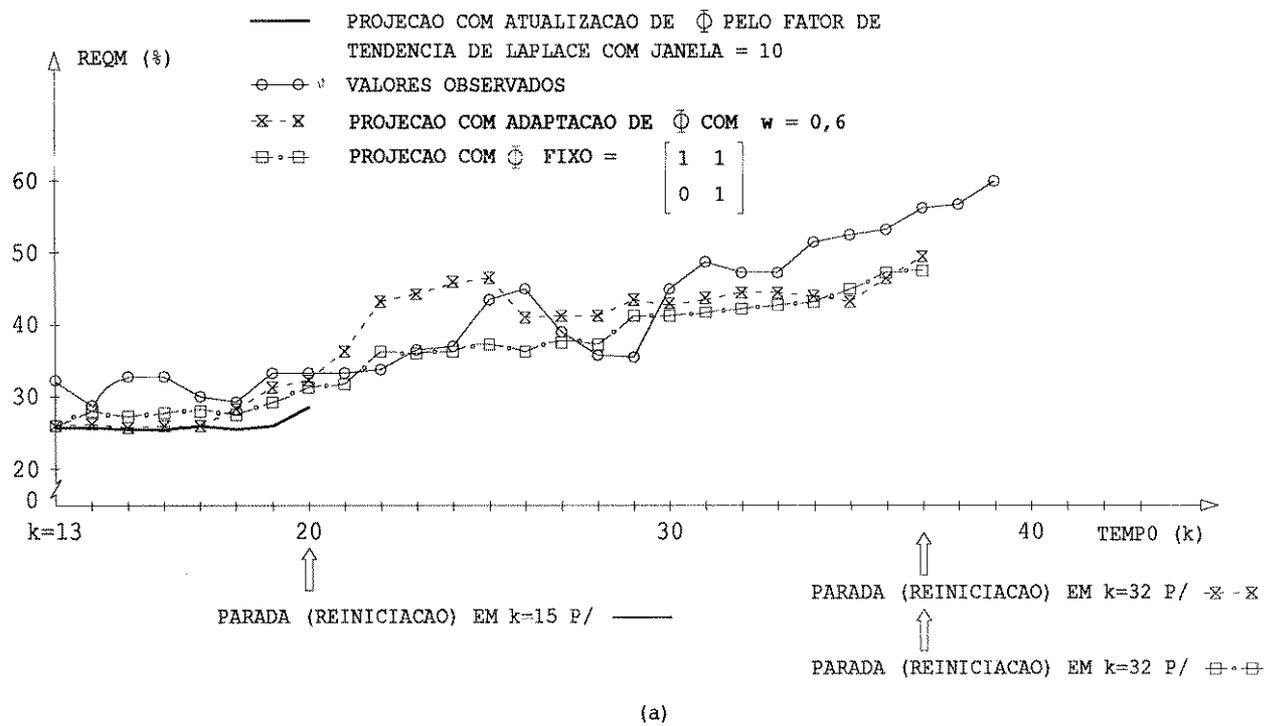


Figura 5.8: Projeções a Seis Passos e *REQM* para a Rota T4249

relação à janela de tamanho 10. Isto se deve ao fato deste tamanho reduzido de janela possibilitar um maior número de detecções de mudança de tendência, não possibilitando a estabilização da estimação do conjunto de parâmetros;

- O fato do algoritmo utilizando Tendência de Laplace não ter reiniciado se deve à utilização de $NIC = 2,33$ e com três *outliers* consecutivos permitidos, enquanto para os outros exemplos que se reiniciaram (marcado aborto na figura), utilizou-se $NIC = 1$ e dois *outliers* consecutivos permitidos;
- Como conhece-se os parâmetros verdadeiros (Q , R , Φ , \bar{X}_0 , e \bar{P}_0) da Rota Tsim, obtidos pelo método de Monte Carlo (Apêndice K), nota-se que o melhor ajuste nem sempre corresponde ao melhor desempenho em termos de EQM . Isso se explica porque o desempenho do filtro não depende de um parâmetro isoladamente, mas de seu conjunto. Como exemplo, sabe-se que o desempenho do filtro depende da relação $\frac{Q}{R}$ (Apêndice D) e não de Q e R isoladamente. Portanto, a estimação inadequada de um parâmetro pode ser compensada pela estimação do outro parâmetro.

5.7 Conclusões

O APSAM foi aplicado a cinco rotas de tráfego, escolhidas aleatoriamente entre 40 rotas disponíveis, a uma rota que apresenta uma grande descontinuidade (diminuição abrupta da carga de tráfego), e a uma série temporal gerada pelo Método de Monte Carlo, e os resultados se mostraram adequados.

No que diz respeito ao $REQM$, o algoritmo supera o desempenho do algoritmo EM , além de requerer menos esforço computacional. Além do mais, é adequado para processos com perturbações não-estacionárias, como a predição da carga de tráfego telefônico em prazo muito curto. Portanto, pode ser utilizado para aplicações em tempo real, e.g., em Gerência de Rede.

Pelos resultados das simulações, o melhor desempenho foi obtido para o parâmetro γ da matriz de observação H próximo a um, quando todos os elementos da matriz da covariância do erro de modelagem são estimados e também atualizando a matriz de transição de estado quando cada nova observação torna-se disponível.

Resultados de simulação mostraram que os $REQMs$ obtidos por meio da atualização da matriz da covariância do erro de observação pela utilização de uma equação recursiva (5.16) são quase os mesmos que aqueles obtidos quando a matriz R é constante e fixa. Isto é devido ao fato que o algoritmo pode estimar convenientemente a relação $\frac{Q}{R}$ ([12, 26] e Apêndice D), mesmo quando a matriz R não é convenientemente estimada. Entretanto, se o sistema tiver perturbações não-estacionárias, com os parâmetros variantes no tempo, a estimação da matriz R melhora a convergência da relação $\frac{Q}{R}$.

A utilização do Fator de Tendência de Laplace e os testes para verificar a presença de Ruído Branco Gaussiano na Sequência de Inovações não resultaram em benefício adicional para a melhoria da capacidade de adaptação do algoritmo, mas ratificaram as seções anteriores mostrando que o APS Adaptativo é uma ferramenta bastante útil para previsões em tempo real.

Devido a este algoritmo ser simples de implementar, ele pode ser útil em Planejamento de Rede onde requer-se uma grande quantidade de manipulação de dados. Além disso,

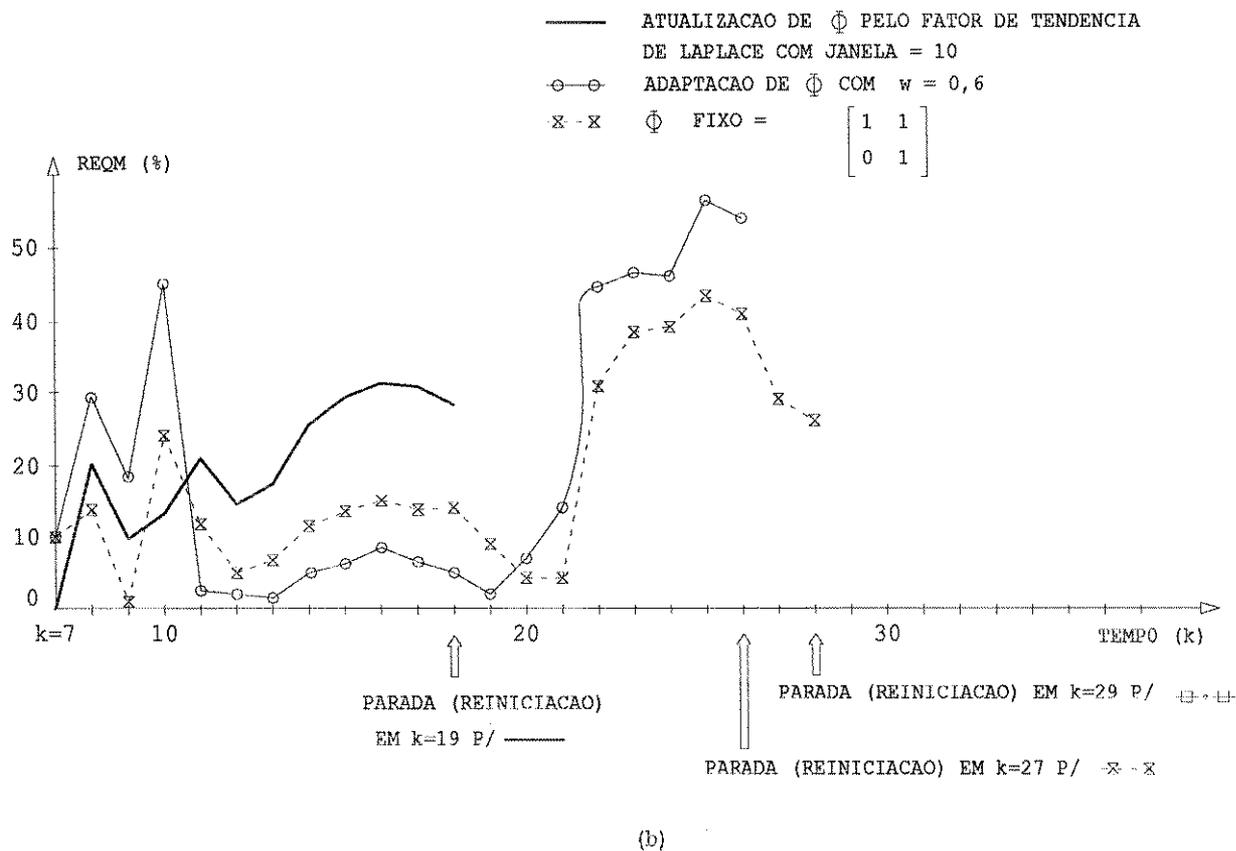
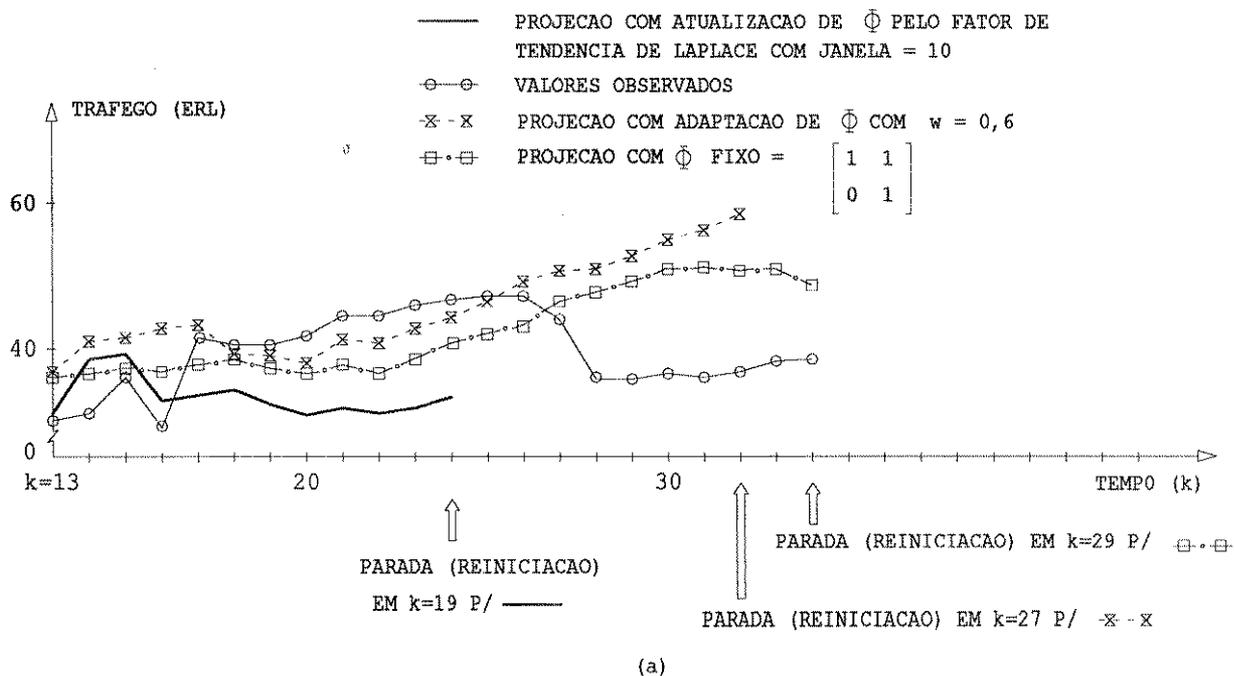


Figura 5.9: Projeção a Seis Passos e *REQM* para a Rota T6362

dado que utiliza mecanismos de robustecimento, sua aplicação pode ser interessante onde há carência de dados ou medições não-confiáveis.

v

Capítulo 6

Conclusões e Trabalhos Futuros

Este capítulo apresenta um sumário das conclusões dos capítulos anteriores, recomenda os métodos de projeção mais adequados, de acordo com a periodicidade dos dados amostrados, e termina com a proposição de trabalhos futuros.

A seguir, algumas considerações gerais sobre a utilização dos modelos de previsão apresentados no Capítulo 2:

- Um modelo complexo tende a apresentar maior precisão. Porém, esse aumento de complexidade acarreta aumento da carga de processamento que adiciona outras fontes de erro ao modelo;
- Quando o propósito do modelo é a relação causa-efeito entre suas variáveis componentes, os modelos que trabalham com a série histórica vinculada à estrutura que gerou os dados são os indicados (exemplo, modelos econométricos ou com variáveis explicativas);
- Quando há poucos dados, os modelos Entrada×Saída não são adequados, devendo-se preferir um modelo que leve em conta a estrutura do sistema ou permita identificação física;
- Quando é difícil de se identificar a estrutura que gerou os dados, os modelos dinâmicos, pelo fato de trabalharem com valores passados das observações e das perturbações, possuem maior eficácia (menor número de parâmetros a serem estimados para alcançar o mesmo objetivo).

Com relação aos modelos de Espaço de Estados, sua utilização se deve aos seguintes aspectos principais:

- Identificação (determinação da ordem do modelo) baseada no sentido físico do problema e não apenas no conjunto de observações disponíveis;
- Versatilidade em se formular intuitivamente modelos aditivos considerando os efeitos de tendência, sazonalidade e componentes irregulares;
- Manutenção da estrutura das matrizes para grande quantidade de rotas de tráfego (apenas estimando os parâmetros quando necessário);

- Simplicidade para a projeção a n passos;
- Existência de modelos de Espaço de Estado multivariáveis que podem trabalhar com observações faltantes;
- Representam um papel fundamental na busca de um arcabouço de metodologia unificada para os modelos de previsão;
- Ao se utilizar o Filtro de Kalman como estimador do estado, tem-se um preditor do estado futuro e um estimador intrinsecamente incorporado ao seu funcionamento;
- Modelos de Espaço de Estados mais simplificados, embora geralmente menos precisos que os mais sofisticados, possuem maior grau de robustecimento;
- Há uma tendência de utilização dos modelos de Espaço de Estados em trabalhos recentes de projeções de rotas de tráfego.

O Capítulo 3 apresentou a base em que se fundamentaram os capítulos seguintes, resumida a seguir:

- Os diversos tipos de modelos de Espaço de Estados aditivos;
- A forma de estimação dos parâmetros (algoritmo *EM*) e do estado (Filtro de Kalman);
- Os critérios para avaliação de desempenho;
- As questões de otimalidade, sensibilidade e robustecimento;
- O tipo de aplicação em função da periodicidade das medições e do número de observações disponíveis.

O Capítulo 3 forneceu também as bases para que surgissem as propostas dos novos algoritmos. O Algoritmo Robusto (Capítulo 4) é fortemente baseado na aplicação do algoritmo *EM* como suporte para a estimação de seu ganho e de sua matriz de transição de estado. O APSAM (Capítulo 5) também se utiliza de conceitos gerais e específicos fornecidos pelo Capítulo 3.

O Capítulo 4 propõe um APS Robusto com:

- Um ganho robusto fixo;
- Uma matriz de transição de estado robusta;
- Um intervalo de confiança para minimizar a ocorrência de *outliers* e detectar a presença de eventos determinísticos não-previstos.

O APS Robusto desempenha adequadamente mesmo quando somente poucas observações são disponíveis. Quando há grande quantidade de dados para serem feitas previsões, o algoritmo pode ser facilmente automatizado por sua simplicidade. Também se mostrou que o Algoritmo Robusto produz resultados estáveis, mesmo com uma grande faixa de variação

dos parâmetros.

O Capítulo 5 propõe o algoritmo Adaptativo Modificado, que pode rastrear variações nos parâmetros, evitar erros grosseiros e também ser utilizado em aplicações de tempo real. Em relação aos valores do erro quadrático médio, *EQM*, o APSAM supera o desempenho do algoritmo *EM*, além de requerer menor esforço computacional. Além do mais, é adequado para sistemas com perturbações não-estacionárias, como a previsão da carga de tráfego telefônico em prazo muito curto. Portanto, pode ser utilizado para aplicações em tempo real, e.g., em Gerência de Rede. Os melhores resultados, com relação ao *EQM*, foram obtidos quando todos os elementos da covariância do erro de modelagem são estimados, e a matriz de transição de estado é estimada a cada nova observação disponível. A estimação da matriz da covariância do erro de observação melhora a convergência do Filtro de Kalman. Além disso, o algoritmo mostrou-se adequado para aplicações em tempo real, podendo ser útil em planejamento de rede onde é requerida uma grande quantidade de dados. Pelo fato de utilizar mecanismos de robustecimento, sua aplicação pode ser interessante onde há carência de dados ou medições não-confiáveis.

O trabalho atingiu os objetivos propostos, tanto no que diz respeito ao ganho em robustez propiciado pelo APS Robusto quanto em relação ao ganho em precisão alcançado pelo APSAM.

Com relação à periodicidade com que os dados são obtidos, e com base nos resultados deste trabalho e experiências anteriores, os seguintes métodos são recomendados:

Planejamento de Expansão, onde as previsões são de longo prazo, com tempos entre dois anos a dez anos: Utilização do APS Robusto e/ou algoritmos de ajuste de funções;

Administração de Tráfego, onde as previsões são de médio prazo, com tempos entre alguns meses a dois anos: Utilização do APS Robusto e/ou do APSAM;

Gerência de Rede, onde as previsões são de curto prazo, e se tem aplicações em tempo real: Utilização do APS Adaptativo Modificado e/ou APS Adaptativo.

Vários outros aspectos podem ainda ser pesquisados e os seguintes trabalhos futuros são sugeridos:

- Busca de aumento de precisão:
 - Aplicar o Filtro de Kalman utilizando linearização estatística ao produto ΦX_k (utilizar tanto no APSAM quanto no algoritmo *EM*).
- Busca de aumento de robustez:
 - Estimação do ganho por meio de procedimento que minimize (projeto de mínima sensibilidade) os erros em relação aos parâmetros especificados pelo projetista, dado o máximo erro com respeito aos parâmetros desconhecidos (estimação min-max) [24].
- Busca de aumento da capacidade de rastreamento:

- Detecção de *outliers* ou eventos determinísticos não-previstos por meio de testes de hipóteses.
- Busca de melhoria da estabilidade numérica (precisão):
 - Utilização de um filtro com fatorização ou filtro UD [60]. Define-se $\bar{P}_k = UDU^T$, onde D é matriz diagonal e U é triangular superior unitária (na sua diagonal). Este filtro mantém as vantagens dos filtros de raiz quadrada em relação às questões de estabilidade numérica, porém, não necessita de operações aritméticas de raiz quadrada que sobrecarregam o processador.
- Aplicação em novos serviços ou em novas tecnologias:
 - Aplicação dos algoritmos em rede de pacotes, em rede de dados e serviços integrados, etc.
- Comparação dos tempos de execução dos diversos algoritmos com base em parâmetros estatísticos (média e variância):

Lista de Símbolos

A Os símbolos apresentados aqui são os que foram consistentemente usados e não apenas eventualmente como variáveis auxiliares ou temporárias.

- Y_k - Valor da saída observada do sistema;
- V_k - Perturbações (ruídos) das observações (v_k para sistemas univariáveis);
- μ_k - Nível constante (ou aproximadamente constante) da série temporal;
- T_k - Ciclotendência da série temporal (ou tendência);
- S_k - Componente sazonal da série;
- T_{E_k} - Tendência estocástica, geralmente representada por um processo auto-regressivo de ordem p ;
- V_{D_k} - Componente de variação diário, representado por uma regressão que converge a valores fixos;
- ν_j - Autocovariâncias de um processo estocástico;
- $\rho_j = \frac{\nu_j}{\nu_0}$ - Autocorrelações de um processo estocástico;
- $Y_{k+n|k}^* = Y_{k+n}^*$ - Projeção n passos à frente a partir do instante k ;
- Y_k^* - Projeção um passo à frente a partir de $k - 1$;
- $\Phi(B)$ - Polinômio em B (operador deslocamento para trás) que representa os coeficientes de um modelo AR (auto-regressivo);
- $\Theta(B)$ - Polinômio em B que representa os coeficientes de um modelo MA (*moving average*);
- $\xi(B)$ - Polinômio em B que representa os coeficientes das entradas de controle;
- X_k - Vetor de variáveis de estado;
- L - Matriz de entradas (ganhos) associada às entradas de controle;
- U_k - Vetor de entradas de controle;
- Φ - Matriz de transição de estado (ou Φ_k , se for variante no tempo);
- Γ - Matriz de entradas associada às perturbações (ruídos) no estado;
- W_k - Vetor de erros de modelagem (ruídos no estado);
- H - Matriz de observação;
- \bar{X}_k - Estimativa do vetor do estado (filtragem);
- X_k^* - Vetor do estado estimado um passo à frente a partir de $k - 1$ (predição);

- K_k - Matriz do ganho do Filtro de Kalman;
- \bar{P}_k - Covariância do erro de estimação (associada a \bar{X}_k);
- P_k^* - Covariância do erro de predição (associada a X_k^*);
- r_k - Termo de inovação (resíduo);
- σ_k^2 - Covariância do termo de inovação;
- $R_k = E[V_k V_k^T]$ - Covariância do erro de observação (apenas R se for invariante no tempo);
- $Q_k = E[W_k W_k^T]$ - Covariância do erro de modelagem (apenas Q se for invariante no tempo);
- \bar{X}_k^N - Estado estimado (suavizado) utilizando N observações;
- \bar{P}_k^N - Covariância do erro de estimação para o estado suavizado;
- $\bar{P}_{k-1,k}^N$ - Covariância de transição de estado (do erro entre o valor suavizado em k e o estado verdadeiro e o valor suavizado em $k - 1$ e o estado verdadeiro);
- g, \hat{g} - Taxas de crescimento real e taxa de crescimento agregado estimada;
- σ_g^2 - Variância do erro entre a taxa de crescimento real da rota e a taxa de crescimento agregado estimada;
- r_v - Erro de predição (resíduo) verdadeiro;
- q_k - Vetor com elementos de Q_k (estado do subfiltro);
- z_k - Vetor da pseudo-observação do subfiltro;
- η_k - Perturbação (ruído) da equação da pseudo-observação do subfiltro;
- M - Matriz da pseudo-observação do subfiltro (corresponde à matriz H do filtro principal);
- Φ^q - Matriz de transição de estado do subfiltro;
- W_k^q - Vetor de perturbações (ruídos) de modelagem do subfiltro;
- $\sigma_q^2 = E[W_k^q (W_k^q)^T]$ - Covariância do erro de modelagem do subfiltro;
- \bar{P}_k^q - Coivariância do erro de estimação do subfiltro;
- P_k^{q*} - Covariância do erro de predição do subfiltro;
- \bar{q}_k - Valor do estado estimado (filtragem) do subfiltro;
- q_k^* - Valor da predição do subfiltro (estado estimado um passo à frente a partir de $k - 1$);
- K_k^q - Matriz do ganho do subfiltro.

Bibliografia

- [1] Eykoff, P., *Parameter and State Estimation*, Wiley, Londres, 1974.
- [2] Shannon, R. E., *Systems Simulation The Art and Science*, Prentice Hall, Englewood Cliffs, New Jersey, 1975.
- [3] Lavelha, A. C., *Notas de Aula do Curso Simulação de Sistemas*, CPqD - Telebrás, maio/1990.
- [4] Amaral, W. C., *Identificação Recursiva de Sistemas Multivariáveis*, Tese de Doutorado - Fac. Engenharia Elétrica - UNICAMP.
- [5] Fasol, K. H.; Jorgl, H. P., *Tutorials on System identification*, 5th IFAC - Symposium on Identification and System Parameter Estimation, Darmstadt, setembro/1979.
- [6] Strejc, V., *On Least Squares and Regression Methods in Identification: A Survey of Progresses and Trends*, Proc. of the 7th IFAC World Congress, Helsinki, 1978.
- [7] Morettin, P. A.; Tolói, C. M. C., *Previsão de Séries Temporais*, Atual, São Paulo, 1987.
- [8] Morettin, P. A.; Tolói, C. M. C., *Séries Temporais*, Atual, São Paulo, 1986.
- [9] Anderson, T. W., *The Statistical Analysis of Time Series*, John Wiley & Sons, Inc, New York, 1971.
- [10] CCITT (atual ITU-TS), Livro azul, Seção 2, rec. E.501, E.506 e E.507, *Forecasting of Traffic*, 1988.
- [11] Haykin, S., *Adaptive Filter Theory*, Prentice Hall, 1986.
- [12] Moreland, J. P., *A Robust Sequential Projection Algorithm for Traffic Load Forecasting*, BSTJ, janeiro/1982.
- [13] Tomé, F. M.; Cunha, J. A., *Traffic Forecasting with a State-Space Model*, 11th International Teletraffic Congress, Kyoto, Japão, 1985.
- [14] Arcirio, R., *Curso sobre Séries Temporais*, Telebrás - Centro de Treinamento, abril/1980.
- [15] Arcirio, R.; Joseli, C., *Modelo de Box-Jenkins para Previsão do Índice Inflacionário*, Telebrás/Universidade de Brasília..

- [16] Galiana, F. D.; Handschin, E.; Fiechter, A. R., *Identification of Stochastic Electric Load Models from Physical Data*, IEEE Transactions on Automatic Control, novembro/1974.
- [17] Chemouil, P.; Garnier, B., *Adaptive Short-term Forecasting Procedure Using Kalman Filter*, 11th International Teletraffic Congress, Kyoto, Japão, 1985.
- [18] Knottnerus, P., *Forecasting: Kalman Filtering and Prediction Intervals*, 12th International Teletraffic Congress, Torino, Itália, 1988.
- [19] Kitagawa, G.; Gersch, W., *Space Modelling of Time Series with Trend and Seasonality*, Journal of the American Statistical Association, volume 79, número 386, junho/1984.
- [20] Shumway, R. H., *Multivariate State-space Estimation and Forecasting Using the EM Algorithm*, 4th International Symposium on Forecasting, Londres, julho/1984.
- [21] Shumway, R. H.; Stoffer, D. S., *An Approach to Time Series Smoothing and Forecasting Using the EM Algorithm*, Journal of the Time Series Analysis, 1982.
- [22] David, A. J.; Pack, C. D., *The Sequential Projection Algorithm: A New and Improved Traffic Forecasting Procedure*, Proceedings 9th International Teletraffic Congress, Torremolinos, Espanha, 1979.
- [23] Pack, C. D.; Whitaker, B. A., *Kalman Filter Models for Network Forecasting*, Bell System Technical Journal, janeiro/1982.
- [24] Gelb, A.; et al, *Applied Optimal Estimation*, MIT Press, Cambridge, Massachusetts, 1974.
- [25] Jazwinski, A. M., *Stochastic Processes and Filtering Theory*, Academic Press, Nova Iorque, 1970.
- [26] Szelag, C. R., *A Short-term Forecasting Algorithm for Trunk Demand Servicing*, Bell System Technical Journal, janeiro/1982.
- [27] Kalman, R. E., *A New Approach to Linear Filtering and Prediction Problems*, Journal of Basic Eng., número 82, março/1960.
- [28] Amaral, W. C., *Notas de Aula do Curso de Controle Estocástico e Adaptativo*, 2^o semestre/1989.
- [29] Ionescu-Graff, A., *A Sequential Projection Algorithm for Special Services Demand*, Bell System Technical Journal, janeiro/1982.
- [30] Shumway, R. H., *Notas de Aula sobre a Dedução do Suavizador Usando as Recursões de De Jong*, IME-USP, 1990.
- [31] Dempster, A. P.; Laird, N. M.; Rubin, D. B., *Maximum Likelihood from Incomplete Data via the EM Algorithm*, Journal of the Royal Statistical Society, 1977.
- [32] Mehra, R. K., *On the Identification of Variances and Adaptive Kalman Filtering*, IEEE Transactions on Automatic Control, abril/1970.

- [33] Heffes, H., *The Effects of Erroneous Models on the Kalman Filter Response*, IEEE Transactions on Automatic Control (short papers), julho/1966.
- [34] Mizala, A.; Bugarin, M. N. S., *Estimação de Parâmetros Variáveis pelo Método de Filtro de Kalman: uma Nota Didática*, Universidade de Brasília, Departamento de Economia, julho/1988.
- [35] Sage, A. P.; Husa, G. W., *Algorithms for Sequential Adaptive Estimation of Prior Statistics*, 8th Symposium on Adaptive Processes, Decision and Control, IEEE, New York, novembro/1969.
- [36] Breiman, L., *Statistics with a View Toward Applications*, Mifflin Company, Boston, 1973.
- [37] Murteira, B. J. F., *Probabilidade e Estatística*, Volume 2, 2^a edição, Mc Graw-Hill de Portugal, 1990.
- [38] Huber, P. J., *Estimation of a Location Parameter*, Annals of Mathematical Statistics, número 35, pág. 73-110, 1964.
- [39] Masreliez, C. J.; Martin, R. D., *Robust Bayesian Estimation for the Linear Model and Robustifying the Kalman Filter*, IEEE Transactions on Automatic Control, Vol AC-22, n^o 3, junho/1977.
- [40] Rios Neto, A.; Kuga H. K., *Estimação Adaptativa do Ruído no Estado para o Filtro de Kalman*, 4^o Congresso Brasileiro de Automática, Campinas, 1982.
- [41] Nunes, G., *Bases de Tempo para Medição de Tráfego*, Revista Telebrás, março/1984.
- [42] Zolfaghari, A., *Traffic Estimation and Forecasting for Developing Telecommunications Networks*, 13th International Teletraffic Congress, Copenhagen, Dinamarca, 1991.
- [43] Cole, A. C., *Progress in Forecasting and Traffic Measurement Techniques as Seen at Recent Teletraffic Congresses*, 12th International Teletraffic Congress, Torino, Itália, 1988.
- [44] Moreland, J. P., *Estimation of Point-to-Point Telephone Traffic*, Bell System Technical Journal, outubro/1978.
- [45] Girolami, A.; Ursini, E. L., *Yearly and Monthly Data Forecasting Using Kalman Filter*, 12th International Teletraffic Congress, Torino, Itália, 1988.
- [46] Ursini, E. L.; Yacoub, M. D.; Amaral, W. C.; Girolami, A., *A Robust Sequential Projection Algorithm*, ITC Regional Seminar, Brasília, Brasil, setembro/1993.
- [47] Ursini, E. L.; Yacoub, M. D.; Amaral, W. C.; Girolami, A., *Adaptive Modified Sequential Projection Algorithm: A proposal*, aprovado para o 14th International Teletraffic Congress, France, junho/1994.
- [48] Martini, M. R. B.; Souza, J. M.; Kanoun, K., *Software Reliability Evaluation of the Tropico-R Switching System*, IEEE Transactions on Reliability, Vol. 39, agosto/1990.

- [49] Fonseca, J. S.; Martins, G. A.; Toledo, G. L., *Estatística Aplicada*, Ed. Atlas, São Paulo, 2ª edição, 1982.
- [50] Kitagawa, G., *A Nonstationary Time Series Model and Its Fitting by a Recursive Filter*, Journal of Time Series Analysis, Vol.2, nº 2, 1981.
- [51] Girolami, A.; Ursini, E. L., *Projeção dos Valores Representativos Mensais e Anuais de Tráfego - Modelo Analítico e Aplicação*, Atas da 2ª Escola de Econometria e Séries Temporais, ENCE, Rio de Janeiro, 1987.
- [52] Girolami, A.; Ursini, E. L., *Procedimento de Projeção de Valores Representativos Anuais e Mensais de Tráfego Usando Filtro de Kalman*, 5º Simpósio Brasileiro de Telecomunicações, Campinas - SP, 1987.
- [53] Girolami, A.; Ursini, E. L., *Application of One Simulating Model of the Telephone System*, 13th International Teletraffic Congress, Copenhagen, Dinamarca, 1991.
- [54] Rapp, Y., *Calculation of Traffic Distribution in Multi-exchange Networks*, Ericsson Technics, 1962.
- [55] Revista Siemens 1989 - International Telecommunication Statistics.
- [56] Papoulis, A., *Probability, Random Variables, and Stochastic Processes*, Mc Graw Hill Kogakusha, Ltd., Tokio, Japão, 1965.
- [57] Solon, L. R.; Tiba, C. R.; Cunha, E., *Programa Totem - Análise do Tráfego Escoado*, EST-TPL-039-1, Telesp, novembro/1978.
- [58] Girolami, A.; Galante, M.; Shigeno, E. Y., *SIDAT Results - Data and Traffic Analysis Integrated System*, ITC Regional Seminar, Brasília, Brasil, setembro/1993.
- [59] Arraes, A. L. A., *Bayesian Forecasting in Dynamic Environments*, ITC Regional Seminar, Brasília, Brasil, setembro/1993.
- [60] Kuga, H. K., *Determinação de Órbitas de Satélites Artificiais Terrestres através de Técnicas de Estimção Combinadas a Técnicas de Suavização de Estado*, Tese de Doutorado em Ciência Espacial / Mecânica Orbital, INPE, São José dos Campos, maio/1989.
- [61] Pratt, C. W., *Contributions to Telecommunications Networks and Services*, 13th International Teletraffic Congress, The ITC - Past, Present and Future - Contributions to the Closing Session, Copenhagen, Dinamarca, 1991.
- [62] Maddala, G. S., *Econometrics*, McGraw-Hill International Editions, 1977.
- [63] Padulo, L.; Arbib, M. A., *System Theory - a Unified State-Space Approach to Continuous and Discrete Systems*, W. B. Saunders Company, Library of Congress, 1974.
- [64] Abe, T.; Saito, H., *Bayesian Forecasting with Multiple State Space Model*, 11th International Teletraffic Congress, Kyoto, Japão, setembro/1985.

Apêndice A

Extensão do Algoritmo Adaptativo para Multivariáveis

Neste apêndice serão mostrados os passos necessários para tornar possível a utilização do algoritmo para observações múltiplas — Y_k vetorial — a partir de analogia com o algoritmo escalar apresentado no Capítulo 5. Para tornar o algoritmo adequado para trabalhar com multivariáveis basta adaptar a estimação de Q_k e o intervalo de confiança para detecção de *outliers* para multivariáveis. Inicialmente será abordada a estimação de Q_k , ou seja, a estimação do modelo de Espaço de Estados para os elementos de Q_k , denominado de Submodelo de Espaço de Estados, e o correspondente filtro para se estimar seu estado, denominado Subfiltro de Kalman.

A.1 Submodelo de Espaço de Estados para os Elementos de Q_k

A fim de se obter um modelo de Espaço de Estados para os elementos da matriz Q_k , o erro de predição verdadeiro é definido ¹:

$$r_{k+1}^v \triangleq (Y_{k+1} - HX_{k+1}^*) - (Y_{k+1} - HX_{k+1}) \quad (\text{A.1})$$

De (A.1), (2.13), e (3.31), obtém-se:

$$r_{k+1}^v = r_{k+1} - V_{k+1} \quad (\text{A.2})$$

Cancelando os valores Y_{k+1} em (A.1), em seguida substituindo X_{k+1}^* e X_{k+1} por suas respectivas equações, dadas por (3.35) e (2.12) com $U_k = 0$ (o significado de U_k é discutido no Capítulo 4 ao considerar eventos determinísticos previstos), obtém-se:

$$r_{k+1}^v = H[\Phi(X_k - \bar{X}_k) + \Gamma W_k] \quad (\text{A.3})$$

Portanto:

$$E[r_{k+1}^v (r_{k+1}^v)^T] = H(\Phi \bar{P}_k \Phi^T + \Gamma Q_k \Gamma^T) H^T = H \Phi \bar{P}_k \Phi^T H^T + H \Gamma Q_k \Gamma^T H^T \quad (\text{A.4})$$

De (A.2):

$$r_{k+1}^v (r_{k+1}^v)^T = r_{k+1} r_{k+1}^T - r_{k+1} V_{k+1}^T - V_{k+1} r_{k+1}^T + V_{k+1} V_{k+1}^T \quad (\text{A.5})$$

Utilizando a hipótese de consistência estatística (Capítulo 5), e devido a $E[r_{k+1} V_{k+1}] = 0$ e $R_{k+1} = E[V_{k+1} V_{k+1}^T]$, segue-se que:

$$E[r_{k+1}^v (r_{k+1}^v)^T] = r_{k+1} r_{k+1}^T + R_{k+1} \quad (\text{A.6})$$

¹A utilização do índice $k+1$ nesta seção e na seção que mostra o exemplo de obtenção das matrizes é apenas para mostrar a diferença de instantes de tempo entre Q_k (porque Q_{k+1} está sendo calculada) e r_{k+1} , por exemplo.

Definindo o erro observado esperado da covariância do erro de predição verdadeiro, denominado pseudo-observação, como:

$$z_{k+1} \triangleq E[r_{k+1}^v(r_{k+1}^v)^T] - E[r_{k+1}^v(r_{k+1}^v)^T | Q_k = 0] \quad (\text{A.7})$$

Utilizando (A.6) e (A.4), então:

$$z_{k+1} = r_{k+1} r_{k+1}^T + R_{k+1} - H \Phi \bar{P}_k \Phi^T H^T \quad (\text{A.8})$$

A equação da pseudo-observação é dada por:

$$z_{k+1} = M q_{k+1} + \eta_{k+1} \quad (\text{A.9})$$

Da hipótese de consistência estatística (Capítulo 5), $r_{k+1}^v(r_{k+1}^v)^T = E[r_{k+1}^v(r_{k+1}^v)^T]$, de (A.4), e de (A.5), obtém-se:

$$M q_{k+1} = H \Gamma Q_k \Gamma^T H^T \quad (\text{A.10})$$

e 2 3

$$\eta_{k+1} = r_{k+1} V_{k+1}^T + V_{k+1} r_{k+1}^T - V_{k+1} V_{k+1}^T + R_{k+1} \quad (\text{A.11})$$

Essa equação é chamada de equação da pseudo-observação do submodelo de Espaço de Estados, cujo vetor de estado q_{k+1} estima os componentes da covariância Q_k , M é um vetor formado por componentes de H e de Γ e η_{k+1} é uma perturbação de média nula.

Para obtenção da matriz M , pode-se dividi-la em linhas e colunas. O número de linhas de M é função de m , o tamanho do vetor de observação Y_k , $\frac{m(m+1)}{2}$. O número de colunas de M é função do tamanho da matriz Q_k ; como Q_k tem tamanho $r \times r$, M terá $\frac{r(r+1)}{2} = r + p$, onde $p = \binom{r}{2} = \frac{r(r-1)}{2}$ colunas porque Q_k é simétrica (ou r , se Q_k for diagonal). O procedimento para o cálculo dos coeficientes da matriz M é o mesmo que para o modelo univariável, só que agora ele é feito por linha. Pode-se utilizar a identidade matricial $b_i Q_k (b_j)^T = TR\{(b_i)^T b_j Q_k\}$ $i, j = 1, \dots, m$, onde $H \Gamma = B$ (como H tem dimensão $m \times n$ e Γ tem dimensão $n \times r$, então B tem dimensão $m \times r$) e b_i é a i -ésima linha de B . Esta forma de abordagem facilita o cálculo dos elementos da matriz M . Cada linha de M é formada pelo produto de duas linhas de B (todas as combinações possíveis sem repetição de elementos, que fornece $\frac{m(m+1)}{2}$ combinações). Sejam b_i e b_j duas linhas quaisquer de B . O produto $(b_i)^T b_j$ resulta:

$$\begin{bmatrix} b_{i1} b_{j1} & b_{i1} b_{j2} & \cdots & b_{i1} b_{jr} \\ b_{i2} b_{j1} & b_{i2} b_{j2} & \cdots & b_{i2} b_{jr} \\ \vdots & \vdots & \vdots & \vdots \\ b_{ir} b_{j1} & \cdots & \cdots & b_{ir} b_{jr} \end{bmatrix}$$

onde b_{kl} representa o elemento l da linha k . Essa matriz resultante possui todos os elementos de uma linha de M . Os coeficientes dos elementos Q_{ij} e Q_{ji} (que são visualizados ao se multiplicar por Q_k e fazer o traço) devem ser somados porque Q_k é simétrica. A ordem dos elementos na linha é coerente com a ordem dos elementos para a obtenção de z'_{k+1} e η'_{k+1} que serão apresentadas a seguir. A ordem das linhas da matriz M também segue essa coerência, iniciando-se pelos produtos de linhas iguais. Ao final do apêndice há uma seção com um exemplo para um modelo bivariável.

Dado que o filtro trabalha na forma vetorial, é preciso identificar cada elemento de z_{k+1} (A.8), que possui $m \times m$ elementos, com o vetor z'_{k+1} que possui $\frac{m(m+1)}{2}$ elementos:

$$(z'_{k+1})^T = [z_{11} \quad z_{22} \quad \cdots \quad z_{mm} \quad z_{12} \quad \cdots \quad z_{1m} \quad z_{23} \quad \cdots \quad z_{2m} \quad \cdots \quad z_{(m-1)m}] \quad (\text{A.12})$$

O modelo do estado para este subsistema é:

$$q_{k+1} = \Phi^q q_k + W_k^q \quad (\text{A.13})$$

²Na verdade, ainda será preciso colocar tanto z_{k+1} quanto η_{k+1} em formas vetoriais, denominadas z'_{k+1} e η'_{k+1} , a serem apresentadas posteriormente.

³Existe uma maneira alternativa de se chegar à equação da pseudo-observação e que é forçando $E[\eta_{k+1}] = 0$ e rearranjando os outros termos.

onde Φ^q é a matriz de transição de estado, $\Phi^q \in \mathbb{R}^{(r+p) \times (r+p)}$; p é o número de elementos na matriz triangular superior de Q_k (ou na inferior, porque Q_k é simétrica), sem a diagonal principal que possui r elementos. Devido ao desconhecimento da dinâmica para este subsistema, adota-se a matriz identidade para esta matriz de transição. W_k^q é um vetor de ruído branco de média nula com covariância σ_q^2 . Os valores dos elementos da matriz são relacionados à precisão do computador. Por exemplo, se o computador tem d dígitos de precisão, esta covariância é dada por $\sigma_q^2 = 10^{-2d}I$, $I \in \mathbb{R}^{(r+p) \times (r+p)}$, onde I é a matriz identidade.

Um Filtro de Kalman, denominado subfiltro, vai processar seqüencialmente a pseudo-observação z'_{k+1} .

A.2 Estimação do Estado do Subsistema (Subfiltro de Kalman)

Para se estimar o estado, o Filtro de Kalman precisa conhecer as covariâncias do erro de pseudo-observação e do erro de modelagem. A covariância do erro de pseudo-observação é estimada como descrito a seguir. Como Y_k é vetorial, η_{k+1} toma a forma de uma matriz, expandindo (A.11):

$$\eta_{k+1} = \begin{bmatrix} 2r_1V_1 - V_1^2 + R_{11} & r_1V_2 + r_1V_2 - V_1V_2 + R_{12} & \cdots & r_1V_m + V_1r_m - V_1V_m + R_{1m} \\ \vdots & 2r_2V_2 - V_2^2 + R_{22} & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ r_mV_1 + V_mr_1 - V_mV_1 + R_{m1} & \cdots & \cdots & 2r_mV_m - V_m^2 + R_{mm} \end{bmatrix} \quad (\text{A.14})$$

onde:

$$r_{k+1} = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_m \end{bmatrix}, \quad V_{k+1} = \begin{bmatrix} V_1 \\ V_2 \\ \vdots \\ V_m \end{bmatrix}, \quad R_{k+1} = \begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1m} \\ \vdots & R_{22} & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ R_{m1} & \cdots & \cdots & R_{mm} \end{bmatrix}$$

como os elementos $R_{ij} = R_{ji}$ ($i \neq j$), a matriz η_{k+1} é simétrica. Para o modelo de Espaço de Estados interessa um η_{k+1} vetorial, portanto define-se η'_{k+1} :

$$(\eta'_{k+1})^T \triangleq [\eta_{11} \quad \eta_{22} \quad \cdots \quad \eta_{mm} \quad \eta_{12} \quad \cdots \quad \eta_{1m} \quad \eta_{23} \quad \cdots \quad \eta_{2m} \quad \cdots \quad \eta_{(m-1)m}] \quad (\text{A.15})$$

onde η'_{k+1} possui $\frac{m(m+1)}{2}$ elementos. Portanto:

$$\eta'_{k+1}(\eta'_{k+1})^T = \begin{bmatrix} \eta_{11}^2 & \eta_{11}\eta_{22} & \cdots & \eta_{11}\eta_{(m-1)m} \\ \eta_{22}\eta_{11} & \eta_{22}^2 & \cdots & \eta_{22}\eta_{(m-1)m} \\ \vdots & \vdots & \vdots & \vdots \\ \eta_{1m}\eta_{11} & \eta_{1m}\eta_{22} & \cdots & \eta_{1m}\eta_{(m-1)m} \\ \vdots & \vdots & \vdots & \vdots \\ \eta_{(m-1)m}\eta_{11} & \cdots & \cdots & \eta_{(m-1)m}^2 \end{bmatrix} \quad (\text{A.16})$$

$\eta'_{k+1}(\eta'_{k+1})^T$ é uma matriz simétrica de tamanho $\frac{m(m+1)}{2} \times \frac{m(m+1)}{2}$, portanto o número de elementos a serem calculados é:

$$\left\{ \frac{m(m+1)}{2} \left[\frac{m(m+1)}{2} + 1 \right] \right\} = \frac{(m^4 + 2m^3 + 3m^2 + 2m)}{8}$$

É fácil ver que para $m = 1$ (modelo univariável), basta se calcular um elemento. O número de elementos cresce muito com m e pode se tornar inviável para m alto.

Sabe-se que o valor esperado de uma matriz é o valor esperado de cada um de seus elementos. Na

forma geral, para o valor esperado dos elementos da matriz da equação (A.16), tem-se ⁴:

$$E[\eta_{ij}\eta_{kl}] = r_j r_k R_{il} + r_j r_l R_{ik} + r_i r_k R_{jl} + r_i r_l R_{jk} - R_{ij} R_{kl} + E[V_i V_j V_k V_l] \quad (\text{A.20})$$

onde $R_{ij} = E[V_i V_j]$, $E[V_i V_j V_k] = 0$, quaisquer que sejam i, j e k , $E[V_i^2 V_j^2] = 2R_{ij}^2 + R_{ii} R_{jj}$, $E[V_i^3 V_j] = 3R_{ii} R_{ij}$, $E[V_i^4] = 3R_{ii}^2$ e $E[r_i^2] = r_i^2$.

Finalmente, pode-se escrever:

$$E[\eta_{ij}\eta_{kl}] = r_j r_k R_{il} + r_j r_l R_{ik} + r_i r_k R_{jl} + r_i r_l R_{jk} + R_{ik} R_{jl} + R_{jk} R_{il} \quad (\text{A.21})$$

As equações do subfiltro são então:

- Filtragem

$$K_{k+1}^q = P_{k+1}^{q*} M^T (M P_{k+1}^{q*} M^T + E[\eta'_{k+1}(\eta'_{k+1})^T])^{-1} \quad (\text{A.22})$$

$$\bar{q}_{k+1} = q_{k+1}^* + K_{k+1}^q (z'_{k+1} - M q_{k+1}^*) \quad (\text{A.23})$$

$$\bar{P}_{k+1} = P_{k+1}^{q*} - K_{k+1}^q M P_{k+1}^{q*} \quad (\text{A.24})$$

- Predição

$$q_{k+1}^* = \Phi^q \bar{q}_k \quad (\text{A.25})$$

$$P_{k+1}^{q*} = \Phi^q \bar{P}_k^q (\Phi^q)^T + \sigma_q^2 \quad (\text{A.26})$$

onde \bar{P}_{k+1}^q é a covariância do erro de estimação. Seu valor é normalmente iniciado como $\bar{P}_0^q = 10^6 I$. \bar{q}_0 depende do valor inicial de Q_k do filtro principal. As dimensões das matrizes desse subfiltro devem ser compatíveis com a matriz M , de dimensão $\frac{m(m+1)}{2} \times (r+p)$. Se $p = 0$, Q_k é a matriz diagonal. Com relação à notação e ao significado, as equações do subfiltro são as mesmas do Filtro de Kalman, apresentadas no Capítulo 3.

O vetor q_k , $q_k \in R^{(r+p) \times 1}$, é dado por:

$$q_k^T = [Q_{11} \quad Q_{22} \quad \cdots \quad Q_{rr} \quad Q_{12} \quad Q_{13} \quad \cdots \quad Q_{(r-1)r}] \quad (\text{A.27})$$

A montagem da matriz Q_k , após cada processamento do subfiltro, é:

$$\left. \begin{array}{l} q_1 \rightarrow Q_{11} \\ q_2 \rightarrow Q_{22} \\ \vdots \\ q_r \rightarrow Q_{rr} \end{array} \right\} \begin{array}{l} \text{elementos} \\ \text{da} \\ \text{diagonal} \\ \text{principal} \end{array} \quad \left. \begin{array}{l} q_{r+1} \rightarrow Q_{12} \text{ e } Q_{21} \\ q_{r+2} \rightarrow Q_{13} \text{ e } Q_{31} \\ \vdots \\ q_{r+p} \rightarrow Q_{(r-1)r} \text{ e } Q_{r(r-1)} \end{array} \right\} \begin{array}{l} \text{elementos} \\ \text{fora da} \\ \text{diagonal} \\ \text{principal} \end{array}$$

Após esses valores estimados, a matriz Q_k é testada para evitar que se torne semi-definida negativa:

$$Q_k = \begin{cases} 0, & \text{se o mínimo } (Q_{ii}, Q_{jj}) < 0 \\ Q_{ij}, & \text{caso contrário} \end{cases}$$

Após essa verificação, os Q_{ij} do filtro principal que forem zerados devem zerar também os correspondentes q_i do subfiltro, para que a coerência seja mantida.

⁴Para o cálculo dos momentos de $E[\eta'_{k+1}(\eta'_{k+1})^T]$, utiliza-se [56]:

$$\frac{\partial^u \varphi(0, \dots, 0)}{\partial w_1^{\theta_1} \dots \partial w_n^{\theta_n}} = (J)^u E[V_1^{\theta_1} \dots V_n^{\theta_n}] \quad (\text{A.17})$$

onde $u = \theta_1 + \dots + \theta_n$, $J = \sqrt{-1}$, e w_i , $i = 1, \dots, n$ são variáveis auxiliares da Transformada de Fourier. Para n variáveis aleatórias conjuntamente distribuídas, tem-se:

$$\varphi(w_1, \dots, w_n) = E[e^{J(w_1 V_1 + \dots + w_n V_n)}] \quad (\text{A.18})$$

para n variáveis aleatórias Gaussianas conjuntamente distribuídas, tem-se:

$$\varphi(w_1, \dots, w_n) = e^{-\frac{1}{2} w R_{k+1} w^T} \quad (\text{A.19})$$

onde $w^T = \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix}$

A.3 Robustecimento

A perturbação presente na equação da pseudo-observação do subfiltro, η_k^i , não é Gaussiana, conforme comentado no Capítulo 5. Além disso, sabe-se que o Filtro de Kalman produz um estimador não-polarizado e de variância mínima quando a perturbação é normalmente distribuída. Entretanto, se há *outliers*, este estimador pode ser polarizado. Nesse caso, um intervalo de confiança é utilizado para melhorar a robustez do filtro. Esses problemas implicam no robustecimento do filtro principal e na validade do subfiltro. Portanto, é preciso que haja controle sobre o termo de inovação, r_k .

Quando o ruído da perturbação da equação da observação apresenta uma distribuição de cauda acentuada (*heavy tailed*), a equação da estimação do Filtro de Kalman é mudada, conforme proposta de Martin e Masreliez [39]:

$$\bar{X}_k = X_k^* + P_k^* H^T T_{R_k}^T \Psi(\nu_k) \quad (\text{A.28})$$

onde T_{R_k} é dada por:

$$T_{R_k} = (HP_k^* H^T + R_k)^{-\frac{1}{2}} \quad (\text{A.29})$$

O valor de T_{R_k} deve ser calculado como:

$$T_{R_k}^T T_{R_k} = (HP_k^* H^T + R_k)^{-1} = (\sigma_k^2)^{-1}$$

devido a σ_k^2 ser uma matriz simétrica, fixa-se um dos elementos de T_{R_k} e calcula-se os demais. Por exemplo, no caso bivariável, supondo:

$$T_{R_k}^T = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

Portanto:

$$T_{R_k}^T T_{R_k} = \begin{bmatrix} a^2 + b^2 & ac + bd \\ ca + bd & c^2 + d^2 \end{bmatrix}$$

Fixando-se o valor de a , calcula-se b (porque σ_k^2 é conhecida), e tendo-se a e b , calcula-se c e d . A função ν_k é definida por [39]:

$$\Psi(\nu_k) = \begin{cases} \nu_k^i, & |\nu_k^i| \leq NIC \\ NIC \operatorname{sgn}(\nu_k^i), & |\nu_k^i| > NIC \end{cases} \quad (\text{A.30})$$

onde $\nu_k = T_{R_k} r_k$, r_k é dado pela equação (3.31), $\operatorname{sgn}(\cdot)$ é a função sinal, NIC é o número de desvios-padrões da distribuição normal padrão e o índice i de ν_k indica a ordem da variável no vetor de variáveis e a comparação deve ser feita para cada uma delas. Por exemplo, no caso bivariável, com T_{R_k} calculado como anteriormente:

$$T_{R_k} = \begin{bmatrix} a & c \\ b & d \end{bmatrix}$$

então:

$$\nu_k^1 = ar_1 + cr_2 \text{ e } \nu_k^2 = br_1 + dr_2$$

onde $r_k = \begin{bmatrix} r_1 \\ r_2 \end{bmatrix}$.

A utilização de um a dois desvios-padrões provê robustez para o estimador da covariância do erro de estimação [38].

A.4 Exemplo de Obtenção das Matrizes para um Modelo Bivariável

Dado um sistema com duas variáveis de estado ($n = 2$), duas séries temporais correlacionadas ($m = 2$ observações), matriz da covariância do erro de modelagem de tamanho 2×2 ($r = 2$), e matriz Γ igual à matriz identidade ($n \times r$). A equação da observação é:

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} x_{11} \\ x_{12} \end{bmatrix} + \begin{bmatrix} V_1 \\ V_2 \end{bmatrix}$$

o termo de inovação é:

$$r_{k+1} = \begin{bmatrix} r_1 \\ r_2 \end{bmatrix} = Y_{k+1} - H X_{k+1}^*$$

- Cálculo da matriz M

$$m_1 = h_1^T h_1 = \begin{bmatrix} h_{11} \\ h_{12} \end{bmatrix} [h_{11} \quad h_{12}] = \begin{bmatrix} h_{11}^2 & h_{11}h_{12} \\ h_{12}h_{11} & h_{12}^2 \end{bmatrix}$$

onde a_i é a i -ésima linha da matriz A , e a_{ij} é o elemento ij da matriz A . Multiplicando pela matriz $Q_k = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix}$, obtém-se:

$$\begin{bmatrix} Q_{11}h_{11}^2 + Q_{21}h_{11}h_{12} & Q_{12}h_{11}^2 + Q_{22}h_{11}h_{12} \\ Q_{11}h_{12}h_{11} + Q_{21}h_{12}^2 & Q_{12}h_{12}h_{11} + Q_{22}h_{12}^2 \end{bmatrix}$$

cujo traço é:

$$Q_{11}h_{11}^2 + Q_{21}h_{11}h_{12} + Q_{12}h_{12}h_{11} + Q_{22}h_{12}^2$$

Como $Q_{12} = Q_{21}$, o resultado é:

$$Q_{11}h_{11}^2 + h_{12}^2 Q_{22} + 2h_{11}h_{12}Q_{12}$$

este mesmo resultado poderia ter sido obtido por HQ_kH^T . Para a obtenção desse resultado ao se fazer o produto m_1Q_k , deve-se ter:

$$m_1 = [h_{11}^2 \quad h_{12}^2 \quad 2h_{11}h_{12}], \text{ e } q_k^T = [Q_{11} \quad Q_{22} \quad Q_{12}]$$

Procedendo-se da mesma maneira para as demais linhas, obtém-se:

$$h_2^T h_2 = \begin{bmatrix} h_{21} \\ h_{22} \end{bmatrix} [h_{21} \quad h_{22}] = \begin{bmatrix} h_{21}^2 & h_{21}h_{22} \\ h_{22}h_{21} & h_{22}^2 \end{bmatrix}$$

portanto:

$$m_2 = [h_{21}^2 \quad h_{22}^2 \quad 2h_{21}h_{22}]$$

m_3 é decorrente do produto:

$$h_1^T h_2$$

portanto:

$$m_3 = [h_{11}h_{21} \quad h_{12}h_{22} \quad h_{11}h_{22} + h_{12}h_{21}]$$

Desta forma, todas as combinações diferentes estão esgotadas e a matriz M é:

$$M = \begin{bmatrix} h_{11}^2 & h_{12}^2 & 2h_{11}h_{12} \\ h_{21}^2 & h_{22}^2 & 2h_{21}h_{22} \\ h_{11}h_{21} & h_{12}h_{22} & h_{11}h_{22} + h_{12}h_{21} \end{bmatrix}$$

Quando Γ for diferente da matriz identidade, basta fazer $H\Gamma = B$ e trabalhar da mesma maneira com as linhas de B .

- Cálculo de z_{k+1} e z'_{k+1}

Da equação (A.8), tem-se:

$$z_{k+1} = \begin{bmatrix} r_1^2 & r_1r_2 \\ r_2r_1 & r_2^2 \end{bmatrix} + \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix} - \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \begin{bmatrix} \bar{P}_{11} & \bar{P}_{12} \\ \bar{P}_{21} & \bar{P}_{22} \end{bmatrix} \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix}$$

Utilizando a mesma regra de formação, z'_{k+1} , com $\frac{m(m+1)}{2}$ elementos, é:

$$z'_{k+1} = [z_{11} \quad z_{22} \quad z_{12}]^T = \begin{bmatrix} r_1^2 + R_{11} - \dots \\ r_2^2 + R_{22} - \dots \\ r_1r_2 + R_{12} - \dots \end{bmatrix}$$

- Cálculo de η_{k+1} , η'_{k+1} , e $E[\eta'_{k+1}(\eta'_{k+1})^T]$

Da equação (A.11), tem-se:

$$\eta_{k+1} = \begin{bmatrix} r_1 \\ r_2 \end{bmatrix} [V_1 \quad V_2] + \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} [r_1 \quad r_2] - \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} [V_1 \quad V_2] + \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix}$$

portanto:

$$\eta_{k+1} = \begin{bmatrix} 2r_1V_1 - V_1^2 + R_{11} & r_1V_2 + V_1r_2 - V_1V_2 + R_{12} \\ r_2V_1 + V_2r_1 - V_2V_1 + R_{21} & 2r_2V_2 - V_2^2 + R_{22} \end{bmatrix}$$

notar que devido à simetria de R_{k+1} , $R_{12} = R_{21}$ e portanto $\eta_{12} = \eta_{21}$. Define-se o vetor η'_{k+1} , de dimensão $\frac{m(m+1)}{2}$:

$$(\eta'_{k+1})^T = [\eta_{11} \quad \eta_{22} \quad \eta_{12}]$$

portanto:

$$\eta'_{k+1}(\eta'_{k+1})^T = \begin{bmatrix} \eta_{11}^2 & \eta_{11}\eta_{22} & \eta_{11}\eta_{12} \\ \eta_{22}\eta_{11} & \eta_{22}^2 & \eta_{22}\eta_{12} \\ \eta_{12}\eta_{11} & \eta_{12}\eta_{22} & \eta_{12}^2 \end{bmatrix}$$

O valor esperado $E[\eta'_{k+1}(\eta'_{k+1})^T]$ é o valor esperado de cada elemento de $\eta'_{k+1}(\eta'_{k+1})^T$. De (A.21) resulta:

$$\begin{aligned} E[\eta_{11}^2] &= 4r_1^2R_{11} + 2R_{11}^2 \\ E[\eta_{11}\eta_{22}] &= 4r_1r_2R_{12} + 2R_{12}^2 \\ E[\eta_{11}\eta_{12}] &= 2r_1^2R_{12} + 2r_1r_2R_{11} + 2R_{11}R_{12} \\ E[\eta_{22}^2] &= 4r_2^2R_{22} + 2R_{22}^2 \\ E[\eta_{22}\eta_{12}] &= 2r_1r_2R_{22} + 2r_2^2R_{21} + 2R_{21}R_{22} \\ E[\eta_{12}^2] &= 2r_1r_2R_{12} + r_2^2R_{11} + r_1^2R_{22} + R_{11}R_{22} + R_{12}^2 \end{aligned}$$

Observações Gerais

- Os modelos aditivos apresentados no Capítulo 3 são válidos para cada valor observado do vetor Y_k . No exemplo com dois estados, todas as considerações feitas para a série univariável valem para Y_1 e para Y_2 , e portanto, cada valor medido tem a sua própria matriz de transição de estado. A expressão do relacionamento entre as variáveis fica evidenciada na matriz da covariância do erro de observação, que neste caso é de ordem 2×2 . O processo de estimação dos parâmetros é o mesmo, e compatível tanto para o modelo univariável como para o multivariável;
- Conforme comentários do Capítulo 2, há exemplos de modelo de Espaço de Estados que são tratados como modelos multivariáveis apenas porque se dispõe de duas ou mais séries temporais referidas à mesma variável observada e as séries apresentam dados faltantes [21]. Nesse caso, o relacionamento entre as variáveis também é feito pela matriz da covariância do erro de observação, mas a matriz de transição de estado é a mesma. Os valores da matriz de observação H (zeros ou uns) é que definem que séries têm elementos observados que devem ser considerados em cada instante.

Apêndice B

Comparação entre as Equações de Fahmeier e de Chow (ou de Abraham e Ledoiter)

A equação de Fahmeier é (Capítulo 5):

$$\hat{R} = \frac{1}{N} \sum_{i=1}^N (Y - H\bar{X})(Y - H\bar{X})^T + H\bar{P}H^T \quad (\text{B.1})$$

onde a colocação do instante de tempo k foi omitida por simplicidade. A equação de Chow para multi-variáveis é:

$$R_{MV} = \frac{1}{N} \sum_{i=1}^N (Y - HX^*)(Y - HX^*)^T R(HP^*H^T + R)^{-1} \quad (\text{B.2})$$

Para se chegar à equação (B.2), será feito o desenvolvimento de (B.1). Primeiro, substituindo \bar{X} por $X^* + K(Y - HX^*)$ e \bar{P} por $(I - KH)P^*$:

$$\begin{aligned} & \{Y - H[X^* + K(Y - HX^*)]\}\{Y - H[X^* + K(Y - HX^*)]\}^T + H(I - KH)P^*H^T = \\ & = [Y - HX^* - HK(Y - HX^*)][Y - HX^* - HK(Y - HX^*)]^T + H(I - KH)P^*H^T = \\ & = (I - HK)(Y - HX^*)[(I - HK)(Y - HX^*)]^T + H(I - KH)P^*H^T = \\ & = (I - HK)(Y - HX^*)(Y - HX^*)^T(I - HK)^T + H(I - KH)P^*H^T = \\ & = [I - HP^*H^T(HP^*H^T + R)^{-1}](Y - HX^*)(Y - HX^*)^T[I - HP^*H^T(HP^*H^T + R)^{-1}]^T \\ & \quad + H(I - KH)P^*H^T \end{aligned} \quad (\text{B.3})$$

desenvolvendo o valor $[I - HP^*H^T(HP^*H^T + R)^{-1}]$, ou seja, multiplicando à esquerda e dividindo por $HP^*H^T + R$, resulta:

$$\frac{HP^*H^T + R - HP^*H^T}{HP^*H^T + R} = \frac{R}{HP^*H^T + R} \quad (\text{B.4})$$

Nota-se que a transposta de (B.4) é a própria (B.4). Desenvolvendo o termo mais à direita de (B.3), obtém-se:

$$\begin{aligned} H(I - KH)P^*H^T &= HP^*H^T + HKHP^*H^T = (I - HK)HP^*H^T = \\ & [I - HP^*H^T(HP^*H^T + R)^{-1}]HP^*H^T \end{aligned} \quad (\text{B.5})$$

multiplicando e dividindo a equação (B.5) por $HP^*H^T + R$, obtém-se:

$$\frac{(HP^*H^T + R - HP^*H^T)}{HP^*H^T + R} HP^*H^T \quad (\text{B.6})$$

substituindo (B.6) e (B.4) em (B.3) e escrevendo recursivamente na forma temporal, resulta:

$$\hat{R} = \frac{k-1}{k} R_{k-1} + \frac{1}{k} \left\{ \frac{R_{k-1}}{HP_k^* H^T + R_{k-1}} \left[(Y_k - HX_k^*)(Y_k - HX_k^*)^T \frac{R_{k-1}}{HP_k^* H^T + R_{k-1}} + HP_k^* H^T \right] \right\} \quad (\text{B.7})$$

A equação (B.2) escrita na forma recursiva temporal é:

$$R_{MV} = \frac{k-1}{k} R_{k-1} + \frac{1}{k} \left[(Y_k - HX_k^*)(Y_k - HX_k^*)^T \frac{R_{k-1}}{HP_k^* H^T + R_{k-1}} \right] \quad (\text{B.8})$$

Considerações

- A equação (B.8) responde mais rapidamente aos valores observados, e portanto deve convergir mais rapidamente ao verdadeiro valor R ;
- A equação (B.7) responde mais lentamente aos valores observados, e embora mais lenta que (B.8), é menos sensível à presença de *outliers*;
- Quando r_k não está sob controle, i.e., quando não se utiliza um intervalo de confiança para delimitar a faixa de variação do termo de inovação, (B.8) deve ser escolhida;
- Na condição de estado estacionário pode-se notar que, quando $R_{MV} \rightarrow R$, então $\hat{R} \rightarrow R$.

Apêndice C

Estabilidade e Estacionariedade do Modelo de Espaço de Estados

Este apêndice pretende discutir alguns conceitos ligados à estabilidade entendida em um contexto amplo, abrangendo também as questões de observabilidade, controlabilidade e “suavizabilidade” (*smoothability*) e também aspectos ligados à estacionariedade das séries modeladas. Este apêndice é um conjunto de pontos pertinentes ao trabalho de tese, encontrados em trabalhos de vários autores.

- Estabilidade

- Observabilidade [24]

Seja a equação de observação de um sistema dada por (2.13):

$$Y_k = HX_k + V_k$$

Dada uma seqüência de observações Y_0, Y_1, \dots, Y_{n-1} para um sistema de ordem n (n variáveis de estado), a condição de observabilidade define a habilidade em se determinar X_0, X_1, \dots, X_{n-1} , a partir das medições. Supondo $X_{k+1} = \Phi X_k$, que é a equação (2.12) com U_k e W_k iguais a zero, e sem a perturbação V_k , escreve-se:

$$Y_0 = HX_0$$

$$Y_1 = HX_1 = H\Phi X_0$$

\vdots

$$Y_{n-1} = HX_{n-1} = H\Phi^{n-1}X_0$$

Portanto,

$$\begin{bmatrix} Y_0 \\ Y_1 \\ \vdots \\ Y_{n-1} \end{bmatrix} = \begin{bmatrix} H \\ H\Phi \\ \vdots \\ H\Phi^{n-1} \end{bmatrix} X_0 = \Xi^T X_0$$

Para a determinação única de X_0 , Ξ^T deve possuir uma inversa, i.e., ser não-singular. Assim a condição de observabilidade para um sistema de ordem n é que:

$$\Xi = \begin{bmatrix} H^T & \vdots & \Phi^T H^T & \vdots & \dots & (\Phi^T)^{n-1} H^T \end{bmatrix}$$

seja de posto n . A condição de observabilidade pode evitar a inclusão de estados adicionais no modelo que não representariam qualquer efeito em termos de desempenho.

- Controlabilidade [24]

É a capacidade que o sistema de ordem n tem de alcançar qualquer estado a partir de um

determinado estado após n estágios. De maneira análoga à observabilidade, trabalhando com a equação do sistema (2.12), $X_{k+1} = \Phi X_k + LU_k$, com W_k igual a zero:

$$\begin{aligned} X_1 &= \Phi X_0 + LU_0 \\ X_2 &= \Phi X_1 + LU_1 = \Phi^2 X_0 + \Phi LU_0 + LU_1 \\ &\vdots \\ X_n &= \Phi X_{n-1} + LU_{n-1} = \Phi^n X_0 + \Phi^{n-1} LU_0 + \dots + \Phi LU_{n-2} + LU_{n-1} \end{aligned}$$

ou seja, $\Phi^{n-1} LU_0$ é a entrada selecionada a partir de X_0 para que o sistema atinja qualquer estado após n estágios. Portanto:

$$\Theta = \begin{bmatrix} L & \Phi L & \dots & \Phi^{n-1} L \end{bmatrix}$$

deve possuir posto n .

– *Smoothability* (“Suavizabilidade”) [24]

O sistema será “suavizável” se puder ser “controlado” pela entrada de ruído, ou seja, dada a equação do sistema (2.12), $X_{k+1} = \Phi X_k + \Gamma W_k$, com U_k igual a zero, a matriz:

$$\Theta = \begin{bmatrix} \Gamma & \Phi \Gamma & \dots & \Phi^{n-1} \Gamma \end{bmatrix}$$

deve ter posto n . Um estado constante, por exemplo, não é “suavizável”.

As condições de observabilidade, controlabilidade e “suavizabilidade” se verificam nos modelos propostos no Capítulo 3. Mesmo os modelos que não tem Φ definida (Φ é estimada) satisfizeram essas condições.

• Covariâncias não negativas-definidas [10]

A forma quadrática:

$$F_Q = \mathbf{x}^T A \mathbf{x}$$

onde \mathbf{x} é um vetor qualquer e A uma matriz qualquer. A é definida positiva se $F_Q > 0$ para todo $\mathbf{x} \neq 0$ e é definida negativa se $F_Q < 0$ para todo $\mathbf{x} \neq 0$ (semi-definida negativa se $F_Q \leq 0$ para todo $\mathbf{x} \neq 0$ e semi-definida positiva se $F_Q \geq 0$ para todo $\mathbf{x} \neq 0$).

As matrizes de covariâncias não devem ser negativas-definidas (semi-definidas negativas), caso contrário ocasionam a instabilidade do filtro. Essa condição pode ocorrer quando se utilizam equações adicionais para se estimarem as covariâncias do filtro em tempo real e se deve à escolha inadequada de valores iniciais (vide Capítulo 5, que aborda maneiras de evitar esse problema na aplicação enfocada).

• Estacionariedade

As séries temporais tratadas pelo modelo de Espaço de Estados possuem diversos aspectos de não-estacionariedade, por exemplo, tendência, sazonalidade, etc. O modelo da equação do sistema (2.12), $X_{k+1} = \Phi X_k$, com U_k e W_k nulos, é um modelo auto-regressivo de ordem um, *AR1*. Para que esse modelo represente um processo estacionário, na forma escalar, $\Phi < 1$.

Avaliando os modelos do Capítulo 3 (para dimensão n do estado), pode-se verificar que correspondem a ter elementos ϕ_{ij} da matriz Φ unitários (quando for o caso, e os demais elementos nulos) modelando tendência, sazonalidade, etc. Isso corresponde a ter zeros sobre o círculo unitário (vide Capítulo 2, porque a equação do estado (2.12) é um modelo *AR1*) e, segundo Kitagawa e Gersch [19], perfeitamente admissível para a modelagem desses padrões não-estacionários. Além disso, quando a matriz Φ é estimada, pode ocorrer que seus elementos não estejam nem mesmo sobre o círculo unitário, mas correspondam a um padrão não-estacionário qualquer. Segundo Shumway e Stoffer [21], isto é admissível para a matriz de transição de estado, que pode, portanto, modelar qualquer processo não-estacionário. Os testes realizados confirmaram a estabilidade com relação aos valores estimados da matriz de transição de estado.

Outro tipo de não-estacionariedade ocorre com os valores das variâncias das perturbações, que podem mudar com o tempo, [18], (conhecido como heterocedasticidade). Este problema é contornado com a estimação dos parâmetros em tempo real.

Apêndice D

Avaliação da Equação do Ganho do Filtro de Kalman e Equação de Ricatti Discreta

Este apêndice objetiva fazer considerações a respeito da seqüência de ganhos do Filtro de Kalman Discreto, bem como apresentar a dedução da equação de Ricatti discreta para estado estacionário (esta equação permite que o ganho de estado estacionário seja pré-calculado e armazenado).

D.1 Avaliação do Ganho do Filtro de Kalman

A equação do ganho do Filtro de Kalman é (3.32):

$$K_k = P_k^* H^T (H P_k^* H^T + R)^{-1}$$

Multiplicando-se numerador e denominador por $H P_k^* H^T + R$, obtém-se:

$$K_k = \frac{P_k^* H^T}{H P_k^* H^T + R}$$

Multiplicando-se numerador e denominador por $(P_k^* H^T)^{-1}$, obtém-se:

$$K_k = \frac{I}{H + R(P_k^* H^T)^{-1}}$$

onde I é a matriz identidade. Multiplicando-se e dividindo-se $R(P_k^* H^T)^{-1}$ por $P_k^* H^T$, obtém-se:

$$K_k = \frac{I}{H + \frac{R}{P_k^* H^T}} \quad (D.1)$$

Como geralmente H é fixado, o ganho do filtro depende da relação $\frac{R}{P_k^*}$. Algumas considerações a respeito do ganho:

- P_k^* depende diretamente de \bar{P}_k e Q , da equação (3.36), $P_{k+1}^* = \Phi \bar{P}_k \Phi^T + Q$ (evidentemente que \bar{P}_k também depende de H e Φ , geralmente fixos, R e Q);
- No estado estacionário (H , Φ , R e Q fixos), \bar{P}_k estaciona e o ganho vai depender da relação $\frac{Q}{R}$;
- Devido ao desconhecimento do valor do estado inicial, o que geralmente ocorre, \bar{P}_0 é um valor alto (por exemplo, com os elementos da diagonal principal da ordem de 10^6) e portanto \bar{P}_0 tem um peso relativo bem maior que Q nas primeiras observações processadas até que a seqüência de ganhos estacione. Neste período transiente, o ganho depende mais da relação $\frac{\bar{P}_0}{R}$;
- Avaliando-se (D.1), nota-se que o ganho tende a zero quando $Q = 0$ ou quando R tende a infinito (são as causas de divergência do filtro). Geralmente tem-se um controle maior sobre R , inclusive mais facilidades para sua estimação, portanto a questão da divergência vai ocorrer para $Q = 0$ (a exceção é a regressão de efeitos fixos, em que o filtro estaciona com $Q = 0$, Seção 3.3);

D.2 Dedução da Equação de Ricatti Discreta

Dispensando o índice k (estado estacionário), tem-se (3.34):

$$\bar{P} = (I - KH)P^*$$

Mas, de (3.36):

$$P^* = \Phi \bar{P} \Phi^T + Q$$

portanto:

$$P^* = \Phi(I - KH)P^* \Phi^T + Q \quad (D.2)$$

Mas, de (3.32):

$$K = P^* H^T (HP^* H^T + R)^{-1}$$

portanto (D.2) se torna:

$$P^* = \Phi P^* \Phi^T + Q - \Phi P^* H^T (HP^* H^T + R)^{-1} H P^* \Phi^T \quad (D.3)$$

que é a equação de Ricatti discreta e fornece o valor do ganho de estado estacionário, supondo conhecidos Φ , H , R , e Q .

A equação de Ricatti discreta pode ser calculada recursivamente por meio da aplicação sequencial de (Capítulo 4):

$$P_{k+1}^* = \Phi \bar{P}_k \Phi^T + Q \quad (D.4)$$

$$K_k = P_{k+1}^* H^T (HP_{k+1}^* H^T + R)^{-1} \quad (D.5)$$

$$\bar{P}_{k+1} = (I - K_k H) P_{k+1}^* \quad (D.6)$$

No caso escalar e com estado unidimensional, (D.2) pode ser escrita como:

$$P^* = \frac{Q}{1 - \Phi^2(1 - KH)}$$

Por questão de estabilidade resulta:

$$K > \frac{1}{H} \left(\frac{\Phi^2 - 1}{\Phi^2} \right)$$

Apêndice E

Dedução das Equações do Suavizador de Kalman (Kalman smoother)

No Capítulo 3 foram mostradas as equações do Suavizador de Kalman e neste apêndice mostra-se a forma como elas podem ser obtidas [30]. Este apêndice corresponde a uma organização das notas de aula do professor Shumway.

Antes de começar com as equações do suavizador, deve-se desenvolver algumas funções auxiliares. De (2.12), com $U_k = 0$ e (3.35), tem-se:

$$X_{k+1} - X_{k+1}^* = \Phi X_k + W_{k+1} - \Phi \bar{X}_k = \Phi(X_k - \bar{X}_k) + W_{k+1}$$

e utilizando (3.32) :

$$\begin{aligned} &= \Phi(X_k - X_k^* - K_k r_k) + W_{k+1} = \\ &= \Phi(X_k - X_k^*) - \Phi K_k [H(X_k - X_k^*) + V_k] + W_{k+1} = \\ &= \Phi(I - K_k H)(X_k - X_k^*) - \Phi K_k V_k + W_{k+1} = \\ &= L_k(X_k - X_k^*) + F(V_k, W_{k+1}) \end{aligned}$$

onde $L_k = \Phi(I - K_k H)$, equação (3.41) e $F(V_k, W_{k+1}) = -\Phi K_k V_k + W_{k+1}$ de (2.13) e (3.31), tem-se:

$$r_{k+1} = H X_{k+1} + V_{k+1} - H X_{k+1}^* = H(X_{k+1} - X_{k+1}^*) + V_{k+1}$$

mas, usando a equação anterior:

$$\begin{aligned} r_{k+1} &= H[L_k(X_k - X_k^*) - \Phi K_k V_k + W_{k+1}] + V_{k+1} = \\ &= H L_k(X_k - X_k^*) - H \Phi K_k V_k + H W_{k+1} + V_{k+1} = \\ &= H L_k(X_k - X_k^*) + F(V_k, W_{k+1}, V_{k+1}) \end{aligned} \tag{E.1}$$

onde $F(V_k, W_{k+1}, V_{k+1}) = -H \Phi K_k V_k + H W_{k+1} + V_{k+1}$. Notar que $E[F(V_k, W_{k+1})] = E[F(V_k, W_{k+1}, V_{k+1})] = 0$.

Antes de começar com as equações do suavizador, um conceito que será empregado na maioria das deduções é o de ortogonalidade. Como exemplo de ortogonalidade, supõe-se que se deseja a projeção de X_k a partir do instante anterior, ou seja, X_k^* . A melhor projeção X_k^* será aquela que fornece o menor erro entre $X_k - X_k^*$ e X_k^* . Geometricamente, ao se representar X_k como um vetor, X_k^* também como um vetor representando sua projeção, o menor erro ocorre quando $X_k - X_k^*$ for ortogonal a X_k^* . Como exemplo de aplicação, sabe-se que a covariância do erro de predição é dada por:

$$P_k^* = E[(X_k - X_k^*)(X_k - X_k^*)^T]$$

dadas as considerações acima, $E[(X_k - X_k^*)(X_k^*)^T] = 0$, portanto:

$$P_k^* = E[(X_k - X_k^*)X_k^{*T}]$$

E.1 Dedução da Equação do Estado Suavizado

Pode-se partir agora para o desenvolvimento das equações do suavizador propriamente dito. Sabe-se que $\bar{X}_k^N = E[X_k | Y_1, Y_2, \dots, Y_N]$. Ao projetar X_k sobre $Y_1, \dots, Y_k, r_{k+1}, \dots, r_N$ (porque, ao vir de trás para a frente, ou de N até 1, ao chegar em k será dependente de todas as observações até k do caminho para a frente e de todos os termos de inovação que ficaram para trás), pode-se escrever:

$$\bar{X}_k^N = b_0 + \sum_{i=1}^k b_i Y_i + \sum_{i=k+1}^N b_i r_i \quad (\text{E.2})$$

Sabe-se também que:

$$E[(X_k - b_0 - \sum_{i=1}^k b_i Y_i - \sum_{i=k+1}^N b_i r_i)(Y_u^T)] = 0, \quad u = 1, \dots, k \quad (\text{E.3})$$

pela ortogonalidade com Y_1, \dots, Y_k . Então, a partir de (E.3), pode-se escrever:

$$E[X_k] = b_0 + \sum_{i=1}^k b_i E[Y_i]$$

ou seja:

$$\bar{X}_k = b_0 + \sum_{i=1}^k b_i Y_i \quad (\text{E.4})$$

Projetando (E.2) em r_u , $u = k+1, \dots, N$, obtém-se:

$$E[(X_k - \bar{X}_k - \sum_{i=k+1}^N b_i r_i) r_u^T], \quad u = k+1, \dots, N$$

ou seja:

$$E[X_k r_u^T] = b_u E[r_u r_u^T]$$

lembrando que $E[r_u r_u^T] = \sigma_u^2$ (3.33), pode-se escrever:

$$b_u \sigma_u^2 = E[X_k r_u^T] \rightarrow b_u = E[X_k r_u^T] [\sigma_u^2]^{-1}$$

então pode-se escrever (E.2) como:

$$\bar{X}_k^N = \bar{X}_k + \sum_{i=k+1}^N E[X_k r_i^T] [\sigma_i^2]^{-1} r_i$$

mas $E[X_k r_k^T] = P_k H^T$, que é facilmente verificado pela ortogonalidade entre X_k^* e $X_k - X_k^*$. Da dedução (E.1), pode-se escrever:

$$E[X_k r_{k+1}^T] = E[X_k ((X_k - X_k^*) L_k^T H^T + F(V_k, W_{k+1}, V_{k+1}))] = P_k^* L_k^T H^T \quad (\text{E.5})$$

já para r_{k+2} pode-se escrever:

$$\begin{aligned} r_{k+2} &= H L_{k+1} (X_{k+1} - X_{k+1}^*) + F(V_{k+1}, W_{k+2}, V_{k+2}) = \\ &= H L_{k+1} L_k (X_k - X_{k+1}^*) + F(V_k, W_{k+1}, V_{k+1}, W_{k+2}, V_{k+2}) \end{aligned}$$

portanto:

$$\begin{aligned} E[X_k r_{k+2}^T] &= P_k^* L_k^T L_{k+1}^T H^T \\ &\vdots \\ E[X_k r_N^T] &= P_k^* L_k^T L_{k+1}^T \dots L_{N-1}^T H^T \end{aligned}$$

portanto:

$$\begin{aligned}\overline{X}_k^N &= \overline{X}_k + P_k^* L_k^T H^T (\sigma_{k+1}^2)^{-1} r_{k+1} + P_k^* L_k^T L_{k+1}^T H^T (\sigma_{k+2}^2)^{-1} r_{k+2} + \\ &\quad \dots + P_k^* L_k^T L_{k+1}^T \dots L_{N-1}^T H^T (\sigma_N^2)^{-1} r_N\end{aligned}$$

chamando:

$$Z_{k-1} = H^T (\sigma_k^2)^{-1} r_k + L_k^T Z_k, \quad Z_N = 0 \quad (\text{E.6})$$

esta é a equação (3.42) do texto principal. Portanto:

$$\begin{aligned}Z_{N-1} &= H^T (\sigma_2)^{-1} r_N \\ Z_{N-2} &= H^T (\sigma_{N-1}^2)^{-1} r_{N-1} + L_{N-1}^T H^T (\sigma_N^2)^{-1} r_N = H^T (\sigma_{N-1}^2)^{-1} r_{N-1} + L_{N-1}^T Z_{N-1} \\ Z_{N-3} &= H^T (\sigma_{N-2}^2)^{-1} r_{N-2} + L_{N-2}^T Z_{N-2}.\end{aligned}$$

Lembrando que o cálculo está sendo feito de trás para a frente (se $N-3 = k$, $N-2 = k+1$, $N-1 = k+2$ e $N = k+3$), pode-se verificar que:

$$X_k^N = \overline{X}_k + P_k^* L_k^T Z_k \quad (\text{E.7})$$

o estado inicial é

$$X_0^N = \overline{X}_0 + \overline{P}_0 \Phi^T Z_0 \quad (\text{E.8})$$

porque no instante $k = 0$ não se possui os valores de P_0^* e L_0 , que são substituídos por \overline{P}_0 e Φ respectivamente.

E.2 Dedução da Equação da Covariância do Erro de Estimação do Suavizador

Para se calcular a covariância \overline{P}_k^N , tem-se:

$$\overline{P}_k^N = E[(X_k - \overline{X}_k^N)(X_k - \overline{X}_k^N)^T | Y_1, Y_2, \dots, Y_N] = E[(X_k - \overline{X}_k^N)X_k^T]$$

porque \overline{X}_k^N é ortogonal a $X_k - \overline{X}_k^N$, portanto:

$$\overline{P}_k^N = E[((X_k - \overline{X}_k^N) - P_k^* L_k^T Z_k)X_k^T] = \overline{P}_k - P_k^* L_k^T E[Z_k X_k^T]$$

desenvolvendo $E[Z_k X_k^T]$, somando e subtraindo \overline{X}_k^N , obtém-se:

$$E[Z_k X_k^T] = E[Z_k((X_k - \overline{X}_k^N) + \overline{X}_k^N)^T]$$

porém como Z_k depende de $r_{k+1}, r_{k+2}, \dots, r_N \rightarrow Z_k$ é ortogonal a $X_k - \overline{X}_k^N$, logo:

$$E[Z_k X_k^T] = E[Z_k(\overline{X}_k^N)^T] = E[Z_k(\overline{X}_k + P_k^* L_k^T Z_k)^T]$$

como \overline{X}_k depende de Y_1, \dots, Y_k , também é ortogonal a Z_k , então:

$$E[Z_k X_k^T] = E[Z_k Z_k^T] L_k P_k^*, \quad (P_k^*)^T = P_k^*$$

chamando $E[Z_k Z_k^T] = S_k$:

$$\begin{aligned}S_k &= E[(H^T (\sigma_{k+1}^2)^{-1} r_{k+1} + L_{k+1}^T Z_{k+1})(r_{k+1}^T (\sigma_{k+1}^2)^{-1} H + Z_{k+1} L_{k+1})] = \\ &= H^T (\sigma_{k+1}^2)^{-1} H + L_{k+1} E[Z_{k+1} Z_{k+1}^T] L_{k+1}\end{aligned}$$

portanto, como o suavizador vem de trás para frente:

$$S_{k-1} = H^T (\sigma_k^2)^{-1} H + L_k^T S_k L_k \quad (\text{E.9})$$

esta é a equação (3.44) do texto principal. Da mesma forma que $Z_N = 0$, também $S_N = 0$. Finalmente:

$$\overline{P}_k^N = \overline{P}_k - P_k^* L_k^T S_k L_k P_k^* \quad (\text{E.10})$$

e como valor inicial:

$$\overline{P}_0^N = \overline{P}_0 - \overline{P}_0 \Phi^T S_0 \Phi \overline{P}_0 \quad (\text{E.11})$$

E.3 Simplificações de Devine

As equações (E.7) e (E.10) podem ser simplificadas [30]:

$$\bar{X}_k^N = \bar{X}_k + P_k^* L_k^T Z_k$$

mas:

$$L_k^T Z_k = Z_{k-1} - H^T (\sigma_k^2)^{-1} r_k$$

então:

$$\bar{X}_k^N = \bar{X}_k + P_k^* Z_{k-1} - P_k^* H^T (\sigma_k^2)^{-1} r_k$$

porém sabe-se que:

$$\bar{X}_k = X_k^* + P_k^* H^T (\sigma_k^2)^{-1} r_k = X_k^* - K_k r_k$$

portanto:

$$\bar{X}_k^N = X_k^* + P_k^* Z_{k-1} \quad (\text{E.12})$$

esta é a equação (3.43) do texto principal. Para simplificar a covariância, sabe-se que:

$$L_k^T S_k L_k = S_{k-1} - H^T (\sigma_k^2)^{-1} H$$

portanto:

$$\begin{aligned} \bar{P}_k^N &= \bar{P}_k - P_k^* L_k^T L_k P_k^* = \\ \bar{P}_k - P_k^* S_{k-1} P_k^* + P_k^* H^T (\sigma_k^2)^{-1} H P_k^* &= \bar{P}_k - P_k^* S_{k-1} P_k^* + K_k H P_k^* \end{aligned}$$

porém:

$$\bar{P}_k = (I - K_k H) P_k^*$$

então:

$$\bar{P}_k^N = (I - K_k H) P_k^* - P_k^* S_{k-1} P_k^* + K_k H P_k^*$$

finalmente:

$$\bar{P}_k^N = P_k^* - P_k^* S_{k-1} P_k^* \quad (\text{E.13})$$

esta é a equação (3.45) do texto principal.

E.4 Dedução da Covariância de Transição de Estado do Suavizador

Esta covariância é fundamental para possibilitar a estimação da matriz de transição de estado. $\bar{P}_{k,k-1}^N = E[(X_k - \bar{X}_k^N)(X_{k-1} - \bar{X}_{k-1}^N)^T | Y_1, \dots, Y_N]$, mas \bar{X}_{k-1}^N é ortogonal a $X_k - \bar{X}_k^N$, portanto $P_{k,k-1}^N = E[(X_k - \bar{X}_k^N)X_{k-1}^T]$. Desenvolvendo:

$$\bar{P}_{k,k-1}^N = E[(X_k - \bar{X}_k - P_k^* L_k^T Z_k)X_{k-1}^T] = E[(X_k - \bar{X}_k)X_{k-1}^T] - P_k^* L_k^T E[Z_k X_{k-1}^T] \quad (\text{E.14})$$

desenvolvendo o primeiro termo do lado direito da equação (E.14):

$$E[(X_k - \bar{X}_k)X_{k-1}^T] = E[((X_k - X_k^*) - K_k r_k)X_{k-1}^T] = \bar{P}_{k,k-1}^{k-1} - K_k E[r_k X_{k-1}^T]$$

de acordo com (E.5):

$$E[r_{k+1}X_k^T] = H L_k P_k^*, \text{ portanto } E[r_k X_{k-1}^T] = H L_{k-1} P_{k-1}^*$$

$$\bar{P}_{k,k-1}^{k-1} = E[(X_k - X_k^*)X_{k-1}^T]$$

mas:

$$X_k - X_k^* = L_{k-1}(X_{k-1} - X_{k-1}^*) - \Phi K_{k-1} V_{k-1} + W_k$$

portanto:

$$\overline{P}_{k,k-1}^{k-1} = L_{k-1} E[(X_{k-1} - X_{k-1}^*) X_{k-1}^T] = L_{k-1} P_{k-1}^*$$

finalmente:

$$E[(X_k - \overline{X}_k) X_{k-1}^*] = L_{k-1} P_{k-1}^* - K_k H L_{k-1} P_{k-1}^* = (I - K_k H) P_{k-1}^*$$

desenvolvendo agora o segundo membro do lado direito de (E.14):

$$E[Z_k X_{k-1}^T] = E[Z_k (X_{k-1} - \overline{X}_{k-1}^N + \overline{X}_{k-1}^N)^T]$$

mas Z_k depende de r_{k+1}, \dots, r_N , sendo portanto ortogonal a $X_{k-1} - \overline{X}_{k-1}^N$, então:

$$E[Z_k X_{k-1}^T] = E[Z_k (\overline{X}_{k-1}^N)^T] = E[Z_k (\overline{X}_{k-1} + Z_{k-1}^T L_{k-1} P_{k-1}^*)^T]$$

como \overline{X}_{k-1} depende de Y_1, \dots, Y_{k-1} , é também ortogonal a Z_k . Portanto:

$$E[Z_k (X_{k-1}^N)^T] = E[Z_k Z_{k-1}^T] L_{k-1} P_{k-1}^* = E[Z_k (r_k^T (\sigma_k^2)^{-1} H^T + Z_{k-1}^T L_k) L_{k-1} P_{k-1}^*]$$

como Z_k é função de r_{k+1}, \dots, r_{k+N} , é ortogonal a r_k , portanto:

$$E[Z_k (X_{k-1}^N)^T] = E[Z_k Z_k^T] L_k L_{k-1} P_{k-1}^*$$

finalmente o segundo membro do lado direito de (E.14) se torna:

$$P_k^* L_k^T S_k L_k L_{k-1} P_{k-1}^*$$

e portanto:

$$\overline{P}_{k,k-1}^N = (I - K_k H) L_{k-1} P_{k-1}^* - P_k^* L_k^T S_k L_k L_{k-1} P_{k-1}^*$$

sabe-se que:

$$L_k^T S_k L_k = S_{k-1} - H^T (\sigma_k^2)^{-1} H \quad \text{e} \quad K_k = P_k^* H^T (\sigma_k^2)^{-1}$$

portanto:

$$\overline{P}_{k,k-1}^N = L_{k-1} P_{k-1}^* - P_k^* S_{k-1} L_{k-1} P_{k-1}^* = (I - P_k^* S_{k-1}) L_{k-1} P_{k-1}^* \quad (\text{E.15})$$

esta é a equação (3.46) do texto principal. E seu valor inicial:

$$\overline{P}_{1,0}^N = (I - P_1^* S_0) \Phi \overline{P}_0 \quad (\text{E.16})$$

Apêndice F

Desenvolvimento das Relações entre Critérios Empregados na Avaliação de Desempenho dos Algoritmos de Previsão

Desenvolvendo as equações (3.61), (3.62) e (3.64):

$$EQM = E[(Y_{k+n}^* - Y_{k+n})(Y_{k+n}^* - Y_{k+n})^T] =$$

$$VP = E[Y_{k+n}^*(Y_{k+n}^*)^T] - E[Y_{k+n}^*]E[(Y_{k+n}^*)^T]$$

$$EMP = E[Y_{k+n}^*] - E[Y_{k+n}]$$

chamando $Y_{k+n}^* = a$ e $Y_{k+n} = b$, e sabendo-se que $E[b] = b$ e $E[b^T] = b^T$, pode-se escrever:

$$EQM = E[aa^T] - E[ab^T] - E[ba^T] + E[bb^T] = E[aa^T] - E[a]b^T - bE[a^T] + E[bb^T]$$

porque $E[Y_{k+n}] = E[HX_{k+n} + V_{k+n}] = HX_{k+n} = \text{constante}$. Desenvolvendo $E[bb^T]$:

$$E[bb^T] = E[(HX_{k+n} + V_{k+n})(HX_{k+n} + V_{k+n})^T] = HX_{k+n}(HX_{k+n})^T + E[V_{k+n}V_{k+n}^T] = bb^T + R_{k+n}$$

Da mesma forma:

$$EMP = E[a] - b$$

e

$$VP = E[aa^T] - E[a]E[a^T]$$

Também pode-se escrever:

$$EMP(EMP)^T = E[a]E[a^T] - E[a]b^T - bE[a^T] + bb^T$$

Portanto:

$$EQM = EMP(EMP)^T + VP + R_{k+n}$$

Apêndice G

Aplicação do Algoritmo EM aos Modelos da Seção 3.3

Realizaram-se diversos testes para fins de verificação da adequação dos modelos a várias séries temporais. Além dos testes realizados para efeito de convergência dos parâmetros e também dos testes para ressaltar a importância da matriz de transição de estado na projeção a n passos e apresentados no Capítulo 3, este apêndice vai apresentar outros testes específicos, visando à comparação entre os vários modelos apresentados na Seção 3.3.

Para efeito de se ter uma base para discutir os resultados obtidos, o Critério de Informação de Akaike (*Akaike Information Criterion, AIC*) será utilizado. O *AIC* (Capítulo 3) é definido como:

$$AIC = -2 \log L + 2(n^{\circ} \text{ de parâmetros}) \quad (\text{G.1})$$

Onde o fator $\log L$ é o logaritmo da função de verossimilhança (3.59) e o n° de parâmetros é o número de variáveis de estado mais o número de elementos desconhecidos do conjunto de parâmetros, $(R_k, \Phi_k, \text{ e } Q_k)$, necessários ao funcionamento do filtro.

O modelo a ser escolhido é o que possui o mínimo *AIC*. Uma vez definido o modelo, a quantidade correspondente ao número de parâmetros é um valor fixo e a parcela do fator $-2 \log L$ é quem vai definir a superioridade de um modelo em relação a outro.

O que se esperaria, ao se aumentar o número de parâmetros é que automaticamente houvesse uma redução no fator $-2 \log L$, porque deveria ocorrer um aumento de precisão. Porém, isso nem sempre ocorre. Quando não há a diminuição do fator $-2 \log L$, a inclusão de novos parâmetros é rejeitada e trabalha-se com o modelo mais simples (claro que o *AIC* também indica o modelo mais simples porque suas duas parcelas crescem). Quando a inclusão de novos parâmetros provoca a diminuição do fator $-2 \log L$, o critério *AIC* mostra qual o melhor modelo. No entanto, existem outros fatores importantes que devem ser considerados. Entre eles podem-se destacar:

- O **número de observações** disponíveis para a estimação dos parâmetros. Como a parcela relativa ao número de parâmetros é fixa, o *AIC* varia com o número de observações. Assim, um modelo que possui *AIC* maior, mas um fator $-2 \log L$ menor, acaba se tornando viável com o aumento do número de observações disponíveis;
- A **precisão e estabilidade** na projeção n passos à frente (Seção 3.6). Este fator é bastante explorado nos Capítulos 3, 4 e 5 do texto principal.

Com exceção da série temporal para testes de Tendência, Sazonalidade e Efeito de Variação Diário¹, todos os testes foram realizados com valores reais de medição em rotas de tráfego no Estado de São Paulo.

Estes testes, a título de ilustração, foram realizados com a utilização da calculadora HP28S, que possui recursos de *software* altamente desenvolvidos com orientação a objetos e rotinas prontas para trabalhar com matrizes. No entanto, como a área permitida ao usuário é pequena (32 kbytes), não é possível

¹Nesse caso foi utilizada a série do Índice de Produto Industrial (IPI), de Morettin, porque essa série possui os padrões de Tendência e Sazonalidade [7]. Além disso, o problema de projeção da carga de tráfego guarda similaridade com os problemas de projeção de séries econômicas.

testar modelos com muitas variáveis de estado e muitas observações ao mesmo tempo. Estes modelos, já testados, vão ser passados para microcomputadores pela Seção de Estudos de Tráfego da Telesp. Testes com a utilização de programa na linguagem BASIC e utilizando microcomputador do tipo PC haviam sido realizados em 40 rotas de tráfego [45] para o algoritmo *EM* utilizando o suavizador com Jota. Os testes na HP28S foram efetuados para ambos os suavizadores e os resultados puderam ser comparados com aqueles dos microcomputadores tipo PC.

Séries Testadas

1) Rota Marília 33 para a Central Trânsito, T44

- Número e tipo de observações disponíveis: 43 observações mensais;
- Número de observações utilizadas nos testes para estimação dos parâmetros: 24 observações mensais;
- a) Modelo: Dado por (3.10) e (3.11) da Seção 3.3;
 - Valores iniciais: \bar{X}_0 , \bar{P}_0 e R_0 como apresentado no Capítulo 4 (APS Robusto), Q_0 com elementos possuindo valores da ordem de grandeza de R_0 e Φ fixa;
 - Número de iterações: 15;
 - Suavizador utilizado: De Jong;
 - Valores finais:

$$AIC = 2(n^0) \text{ de parâmetros} - 2 \log L = 2(2 + 1 + 3) + 45,66 = 57,66$$

- Evolução do fator $\log L$:

$$\begin{Bmatrix} -26,29 & -23,13 & -22,83 & -22,78 & -22,77 \\ -22,78 & -22,78 & -22,79 & -22,79 & -22,80 \\ -22,80 & -22,81 & -22,81 & -22,82 & -22,83 \end{Bmatrix}$$

- b) Modelo: Dado por (3.10) e (3.11) da Seção 3.3;
 - Valores iniciais: \bar{X}_0 , \bar{P}_0 e R_0 como apresentado no Capítulo 4 (APS Robusto), Q_0 com elementos possuindo valores da ordem de grandeza de R_0 e Φ fixa;
 - Número de iterações: 15;
 - Suavizador utilizado: com J ;
 - Valores finais:

$$AIC = 2(n^0) \text{ de parâmetros} - 2 \log L = 2(2 + 1 + 3) + 46,88 = 58,88$$

- Evolução do fator $\log L$:

$$\begin{Bmatrix} -24,22 & -23,36 & -23,38 & -23,39 & -23,40 \\ -23,40 & -23,41 & -23,41 & -23,41 & -23,42 \\ -23,42 & -23,43 & -23,43 & -23,43 & -23,44 \end{Bmatrix}$$

- c) Modelo: Dado (3.10) e (3.11) da Seção 3.3, variante 1 (com Φ também estimado);
 - Valores iniciais: São os mesmos que os do modelo anterior;
 - Número de iterações: 15;
 - Suavizador utilizado: De Jong;
 - Valores finais:

$$AIC = 2(2 + 1 + 3 + 4) + 44,56 = 64,56, \quad \Phi = \begin{bmatrix} 1 & 0,93 \\ 0,01 & 0,62 \end{bmatrix}$$

– Evolução do fator $\log L$:

$$\begin{Bmatrix} -26,21 & -22,68 & -22,39 & -22,33 & -22,32 \\ -22,31 & -22,30 & -22,30 & -22,29 & -22,29 \\ -22,29 & -22,29 & -22,28 & -22,28 & -22,28 \end{Bmatrix}$$

• d) Modelo: Dado por (3.8), ou seja, um modelo de tendência polinomial com dois estados;

– Valores Iniciais: R_0 , Q_0 e \bar{P}_0 como nos itens anteriores. Φ é fixo e dado por (3.8). Devido a \bar{X}_0 representar dois valores da mesma variável de estado (defasados no tempo), adota-se:

$$\bar{X}_0 = \begin{bmatrix} Y_0 \\ Y_0 \end{bmatrix}$$

– Número de iterações: 15;

– Suavizador utilizado: De Jong;

– Valores finais:

$$AIC = 2(2 + 1 + 1) + 45,88 = 53,88$$

– Evolução do fator $\log L$:

$$\begin{Bmatrix} -25,32 & -23,29 & -22,99 & -22,93 & -22,92 \\ -22,92 & -22,91 & -22,91 & -22,92 & -22,92 \\ -22,92 & -22,93 & -22,93 & -22,93 & -22,94 \end{Bmatrix}$$

• e) Modelo aditivo conforme a Seção 3.3, com uma parte de tendência, composta do modelo dado por (3.10) e (3.11) e uma parte de um modelo $AR2$, dado por (3.27), com as matrizes:

$$H = [1 \ 0 \ | \ 1 \ 0], \Phi = \begin{bmatrix} 1 & 1 & | & 0 & 0 \\ 0 & 1 & | & 0 & 0 \\ 0 & 0 & | & \phi_{33} & \phi_{34} \\ 0 & 0 & | & 1 & 0 \end{bmatrix}, \Gamma = \begin{bmatrix} 1 & 0 & | & 0 \\ 0 & 1 & | & 0 \\ 0 & 0 & | & 1 \\ 0 & 0 & | & 0 \end{bmatrix}$$

onde ϕ_{33} e ϕ_{34} são os coeficientes do modelo $AR2$;

– Valores iniciais: A parte de tendência é iniciada da mesma forma que para os modelos dados por (3.10) e (3.11), incluindo o mesmo R_0 . Os coeficientes do $AR2$, ϕ_{33} e ϕ_{34} são iniciados por 0,5 e 0,4 respectivamente. Repete-se o valor de Q_{22} para Q_{33} (os demais valores da coluna 3 ou da linha 3 de Q_0 são nulos). Repete-se o valor de \bar{P}_{22} para \bar{P}_{33} e \bar{P}_{44} (os demais valores das linhas 3 e 4 e das colunas 3 e 4 de \bar{P}_0 são nulos). Coloca-se zero no terceiro e quarto valores do vetor \bar{X}_0 .

– Número de iterações: 6;

– Suavizador utilizado: De Jong;

– Valores finais:

$$AIC = 2(4 + 1 + 4 + 2) + 45,46 = 67,46, \quad \phi_{33} = 0,48, \quad \phi_{34} = 0,41$$

– Evolução do fator $\log L$:

$$\{ -26,60 \quad -23,12 \quad -22,81 \quad -22,75 \quad -22,74 \quad -22,73 \}$$

• Número e tipo de observações disponíveis: 10 observações quadrimestrais;

• Número de observações utilizadas nos testes para estimação dos parâmetros: 8 observações quadrimestrais;

• a) Modelo: Dado por (3.10) e (3.11) da Seção 3.3;

- Valores iniciais: Utilizados os mesmos critérios que para o modelo para observações mensais;
- Número de iterações: 5;
- Suavizador utilizado: De Jong;
- Valores Finais:

$$AIC = 2(2 + 1 + 3) + 17,90 = 29,90$$

- Evolução do fator $\log L$:

$$\{ -16,00 \quad -9,43 \quad -9,02 \quad -8,97 \quad -8,95 \}$$

- b) Modelo: Dado por (3.10) e (3.11) da Seção 3.3, variante 1 (com Φ também estimado);

- Valores iniciais: Utilizados os mesmos critérios que para o modelo com observações mensais;
- Número de iterações: 15;
- Suavizador utilizado: De Jong;
- Valores Finais:

$$AIC = 2(2 + 1 + 3 + 4) + 17,22 = 37,22, \quad \Phi_{11} = \begin{bmatrix} 1 & 0,97 \\ 0,01 & 0,70 \end{bmatrix}$$

- Evolução do fator $\log L$:

$$\left\{ \begin{array}{ccccc} -16,00 & -9,18 & -8,80 & -8,72 & -8,70 \\ -8,68 & -8,67 & -8,65 & -8,65 & -8,64 \\ -8,63 & -8,62 & -8,62 & -8,61 & -8,61 \end{array} \right\}$$

- c) Modelo: Dado por (3.7), ou seja, um modelo de tendência polinomial com um estado;

- Valores iniciais: Como o modelo é todo escalar, utilizaram-se os elementos da coluna um e linha um das matrizes dos parâmetros iniciais dos dois modelos anteriores;
- Número de iterações: 5;
- Suavizador utilizado: De Jong;
- Valores finais:

$$AIC = 2(1 + 1 + 1) + 19,58 = 25,58$$

- Evolução do fator $\log L$:

$$\{ -34,40 \quad -10,42 \quad -9,84 \quad -9,79 \quad -9,79 \}$$

2) Rota Botucatu Trânsito para Bauru Trânsito, T4942

- Número e tipo de observações disponíveis: 43 observações mensais;
- Número de observações utilizadas nos testes para estimação dos parâmetros: 24;
- a) Modelo: Dado por (3.10) e (3.11) da Seção 3.3, variante 1 (também com Φ estimado);

- Valores iniciais:

$$\bar{X}_0 = \begin{bmatrix} 22,6 \\ 2,26 \end{bmatrix}, \quad \bar{P}_0 = \begin{bmatrix} 3,94 & 0,06 \\ 0,06 & 9,06 \end{bmatrix}, \quad \Phi = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad Q_0 = \begin{bmatrix} 0,09 & -0,04 \\ -0,04 & 0,05 \end{bmatrix}, \quad R_0 = [1]$$

- Número de iterações: 6;
- Suavizador utilizado: De Jong;

– Valores finais:

$$AIC = 2(2 + 1 + 3 + 4) + 41,62 = 61,62, \quad \Phi = \begin{bmatrix} 0,99 & 1,23 \\ 0,01 & 0,48 \end{bmatrix}$$

– Evolução do fator $\log L$

$$\{ -22,11 \quad -21,13 \quad -21,00 \quad -20,91 \quad -20,85 \quad -20,81 \}$$

• b) Modelo: Dado por (3.10) e (3.11) da Seção 3.3, variante 1 (também com Φ estimado);

– Valores iniciais: os mesmos que o modelo anterior;

– Número de iterações: 6;

– Suavizador utilizado: com J ;

– Valores finais:

$$AIC = 2(2 + 1 + 3 + 4) + 52,80 = 72,80, \quad \Phi = \begin{bmatrix} 0,99 & 1,28 \\ 0,01 & 0,69 \end{bmatrix}$$

– Evolução do fator $\log L$

$$\{ -79,36 \quad -29,93 \quad -27,44 \quad -26,73 \quad -26,47 \quad -26,40 \}$$

• c) Modelo aditivo conforme a Seção 3.3, com uma parte de tendência, composta do modelo dado por (3.10) e (3.11) e uma parte de um modelo $AR2$, dado por (3.27), com as matrizes:

$$H = [1 \quad 0 \quad | \quad 1 \quad 0], \quad \Phi = \begin{bmatrix} 1 & 1 & | & 0 & 0 \\ 0 & 1 & | & 0 & 0 \\ \hline 0 & 0 & | & \phi_{33} & \phi_{34} \\ 0 & 0 & | & 1 & 0 \end{bmatrix}, \quad \Gamma = \begin{bmatrix} 1 & 0 & | & 0 \\ 0 & 1 & | & 0 \\ \hline 0 & 0 & | & 1 \\ 0 & 0 & | & 0 \end{bmatrix}$$

onde ϕ_{33} e ϕ_{34} são os coeficientes do modelo $AR2$;

– Valores iniciais: A parte de tendência é iniciada da mesma forma que para os modelos dados por (3.10) e (3.11), incluindo o mesmo R_0 . Os coeficientes do $AR2$, ϕ_{33} e ϕ_{34} são iniciados por 0,5 e 0,4 respectivamente. Repete-se o valor de Q_{22} para Q_{33} (os demais valores da coluna 3 ou da linha 3 de Q_0 são nulos). Repete-se o valor de \bar{P}_{22} para \bar{P}_{33} e \bar{P}_{44} (os demais valores das linhas 3 e 4 e das colunas 3 e 4 de \bar{P}_0 são nulos). Coloca-se zero no terceiro e quarto elementos do vetor \bar{X}_0 ;

– Número de iterações: 6;

– Suavizador utilizado: De Jong;

– Valores finais:

$$AIC = 2(4 + 1 + 4 + 2) + 51,08 = 73,08, \quad \phi_{33} = 0,57, \quad \phi_{34} = 0,38$$

– Evolução do fator $\log L$

$$\{ -56,63 \quad -25,99 \quad -25,45 \quad -25,51 \quad -25,55 \quad -25,54 \}$$

3) Rota Bauru Trânsito para Botucatu Trânsito, T4249

• Número e tipo de observações disponíveis: 43 observações mensais;

• Número de observações utilizadas nos testes para estimação dos parâmetros: 24;

• a) Modelo: Dado por (3.10) e (3.11) da Seção 3.3;

– Valores iniciais:

$$\bar{X}_0 = \begin{bmatrix} 22,7 \\ 2,27 \end{bmatrix}, \bar{P}_0 = \begin{bmatrix} 1,15 & 0,02 \\ 0,02 & 1,85 \end{bmatrix}, R_0 = [1], \Phi = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, Q_0 = \begin{bmatrix} 0,43 & -0,41 \\ -0,41 & 0,53 \end{bmatrix}$$

– Número de iterações: 5;

– Suavizador utilizado; De Jong;

– Valores finais:

$$AIC = 2(2 + 1 + 3) + 35,50 = 47,50$$

– Evolução do fator $\log L$

$$\{ -20,03 \quad -18,42 \quad -18,10 \quad -17,89 \quad -17,75 \}$$

• b) Modelo: Dado por (3.10) e (3.11), variante 1, da Seção 3.3;

– Valores iniciais: Os mesmos que o modelo anterior:

– Número de iterações: 5;

– Suavizador utilizado: De Jong;

– Valores finais:

$$AIC = 2(2 + 1 + 3 + 4) + 33,18 = 53,18, \quad \Phi = \begin{bmatrix} 0,56 & -0,20 \\ -0,20 & 0,18 \end{bmatrix}$$

– Evolução do fator $\log L$

$$\{ -20,03 \quad -16,78 \quad -16,60 \quad -16,59 \quad -16,59 \}$$

4) Rota Bauru Trânsito para Marília Trânsito, T4244

• Número e tipo de observações disponíveis: 43 observações mensais;

• Número e tipo de observações utilizadas nos testes para estimação dos parâmetros: 24;

• a) Modelo: Dado por (3.10) e (3.11), variante 1, da Seção 3.3;

– Valores iniciais:

$$\bar{X}_0 = \begin{bmatrix} 31,8 \\ 3,18 \end{bmatrix}, \bar{P}_0 = \begin{bmatrix} 4,53 & 0,07 \\ 0,07 & 3,67 \end{bmatrix}, R_0 = [1], \Phi = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, Q_0 = \begin{bmatrix} 0,26 & -0,15 \\ -0,15 & 0,28 \end{bmatrix}$$

– Número de iterações: 6;

– Suavizador utilizado: De Jong;

– Valores finais:

$$AIC = 2(2 + 1 + 3 + 4) + 45,90 = 65,90, \quad \Phi = \begin{bmatrix} 0,99 & 1,34 \\ 0,01 & 0,48 \end{bmatrix}$$

– Evolução do fator $\log L$

$$\{ -39,68 \quad -30,15 \quad -28,66 \quad -28,21 \quad -28,04 \quad -27,95 \}$$

5) Rota Marília Trânsito para Bauru Trânsito, T4442

• Número e tipo de observações disponíveis: 43 observações mensais;

• Número e tipo de observações utilizadas nos testes para estimação dos parâmetros: 24;

- a) Modelo: Dado por (3.10) e (3.11), variante 1, da Seção 3.3;

– Valores iniciais:

$$\bar{X}_0 = \begin{bmatrix} 80,10 \\ 8,01 \end{bmatrix}, \bar{P}_0 = \begin{bmatrix} 10,82 & 0,16 \\ 0,16 & 22,93 \end{bmatrix}, R_0 = [1], \Phi = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, Q_0 = \begin{bmatrix} 0,30 & -0,06 \\ -0,06 & 0,10 \end{bmatrix}$$

– Número de iterações: 5;

– Suavizador utilizado: De Jong;

– Valores finais:

$$AIC = 2(2 + 1 + 3 + 4) + 56,88 = 76,88, \Phi = \begin{bmatrix} 0,99 & 1,33 \\ 0,01 & 0,51 \end{bmatrix}$$

– Evolução do fator $\log L$

$$\{ -125,59 \quad -32,24 \quad -29,45 \quad -28,62 \quad -28,44 \}$$

6) Série Temporal sobre o Índice do Produto Industrial

- Número e tipo de observações disponíveis: 28 observações quadrimestrais;
- Número de observações utilizadas nos testes para estimação dos parâmetros: 7;
- a) Modelo: Dado por (3.10) e (3.11) da Seção 3.3, variante 1 (também com Φ estimado);

– Valores iniciais:

$$\bar{X}_0 = \begin{bmatrix} 7871 \\ 78,71 \end{bmatrix}, \bar{P}_0 = \begin{bmatrix} 10 & 0 \\ 0 & 1 \end{bmatrix}, Q_0 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, R_0 = [1], \Phi = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

– Número de iterações: 15;

– Suavizador utilizado: De Jong;

– Valores finais:

$$AIC = 2(2 + 1 + 3 + 4) + 44,18 = 64,18, \Phi = \begin{bmatrix} 0,99 & 0,81 \\ -0,01 & 0,81 \end{bmatrix}$$

– Evolução do fator $\log L$

$$\left\{ \begin{array}{cccccc} -457467,5 & -24,33 & -22,78 & -22,46 & -22,35 \\ -22,31 & -22,28 & -22,25 & -22,23 & -22,20 \\ -22,18 & -22,16 & -22,13 & -22,11 & -22,09 \end{array} \right\}$$

- b) Modelo: Dado por (3.10) e (3.11) da Seção 3.3, variante 1 (com Φ também estimado);

– Valores iniciais: Os mesmos do modelo anterior, variando apenas os valores da covariância do erro de modelagem:

$$Q_0 = \begin{bmatrix} 200 & 200 \\ 200 & 200 \end{bmatrix}$$

– Número de iterações: 15;

– Suavizador utilizado: De Jong;

– Valores finais:

$$AIC = 2(2 + 1 + 3 + 4) + 46,78 = 66,78, \Phi = \begin{bmatrix} 0,98 & 0,34 \\ -0,01 & 0,34 \end{bmatrix}$$

– Evolução do fator $\log L$

$$\left\{ \begin{array}{ccccc} -23,28 & -23,39 & -23,39 & -23,39 & -23,39 \\ -23,39 & -23,39 & -23,39 & -23,39 & -23,39 \\ -23,39 & -23,39 & -23,39 & -23,39 & -23,39 \end{array} \right\}$$

- c) Modelo aditivo conforme a Seção 3.3, com uma parte de tendência, composta do modelo dado por (3.10) e (3.11), variante 1 e uma parte de um modelo sazonal, dado por (saz1), com as matrizes:

$$H = [1 \ 0 \ | \ 1 \ 0], \Phi = \left[\begin{array}{cc|cc} \phi_{11} & \phi_{12} & 0 & 0 \\ \phi_{21} & \phi_{22} & 0 & 0 \\ \hline 0 & 0 & -1 & -1 \\ 1 & 0 & 0 & 0 \end{array} \right], \Gamma = \left[\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \hline 0 & 0 & 1 \\ 0 & 0 & 0 \end{array} \right]$$

– Valores iniciais:

$$R_0 = [1], Q_0 = \left[\begin{array}{cc|c} 1 & 1 & 0 \\ 1 & 1 & 0 \\ \hline 0 & 0 & 1 \end{array} \right], \bar{P}_0 = \left[\begin{array}{cccc} 10 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{array} \right], \bar{X}_0 = \left[\begin{array}{c} 7871 \\ 78,71 \\ 0 \\ 0 \end{array} \right]$$

– Número de iterações: 15;

– Suavizador utilizado: De Jong;

– Valores finais:

$$AIC = 2(4 + 1 + 4 + 4) + 40,20 = 66,20, \Phi = \left[\begin{array}{cc|cc} 1,01 & 0,74 & 0 & 0 \\ 0,01 & 0,74 & 0 & 0 \\ \hline 0 & 0 & -1 & -1 \\ 0 & 0 & 1 & 0 \end{array} \right]$$

– Evolução do fator $\log L$

$$\left\{ \begin{array}{ccccc} -81162,86 & -24,61 & -21,56 & -20,83 & -20,56 \\ -20,43 & -20,36 & -20,31 & -20,27 & -20,24 \\ -20,21 & -20,18 & -20,15 & -20,13 & -20,10 \end{array} \right\}$$

- d) Modelo: o mesmo que o modelo anterior;

– Valores iniciais: os mesmos que o modelo anterior, apenas variando a covariância do erro de modelagem:

$$Q_0 = \left[\begin{array}{cc|c} 200 & 200 & 0 \\ 200 & 200 & 0 \\ \hline 0 & 0 & 200 \end{array} \right]$$

– Número de iterações: 15;

– Suavizador utilizado: De Jong;

– Valores finais:

$$AIC = 2(4 + 1 + 4 + 4) + 42,50 = 68,50, \Phi = \left[\begin{array}{cc|cc} 1,05 & -0,58 & 0 & 0 \\ 0,05 & -0,58 & 0 & 0 \\ \hline 0 & 0 & -1 & -1 \\ 0 & 0 & 1 & 0 \end{array} \right]$$

– Evolução do fator $\log L$

$$\left\{ \begin{array}{ccccc} -24,91 & -22,71 & -21,65 & -20,99 & -20,67 \\ -20,63 & -20,77 & -20,96 & -21,10 & -21,19 \\ -21,23 & -21,25 & -21,26 & -21,25 & -21,25 \end{array} \right\}$$

- e) Modelo aditivo conforme a Seção 3.3, com uma parte de tendência, composta do modelo dado por (3.10) e (3.11), variante 1, uma parte sazonal dada por (3.14) e uma parte formada por um modelo do efeito de variação diário, dado por (3.25) e (3.26), com as matrizes:

$$H = \left[\begin{array}{cc|ccc} 1 & 0 & 1 & 0 & d_{1k} & d_{2k} & d_{3k} \end{array} \right], \Phi = \left[\begin{array}{cc|cc|ccc} 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & -1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array} \right], \Gamma = \left[\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \hline 0 & 0 & 1 \\ \hline 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{array} \right]$$

– Valores iniciais:

$$R_0 = [1], Q_0 = \left[\begin{array}{cc|c} 1 & 1 & 0 \\ 1 & 1 & 0 \\ \hline 0 & 0 & 1 \end{array} \right], \bar{X}_0 = \begin{bmatrix} 7871 \\ 78,71 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \bar{P}_0 = \begin{bmatrix} 10 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Neste caso, os valores de H (elementos d_{1k} , d_{2k} e d_{3k} , que variam a cada observação) são:

$$\left\{ \begin{array}{l} \text{observ. 1 } H = [1 \ 0 \ 1 \ 0 \ 21 \ 15 \ 16] \\ \text{observ. 2 } H = [1 \ 0 \ 1 \ 0 \ 21 \ 18 \ 15] \\ \text{observ. 3 } H = [1 \ 0 \ 1 \ 0 \ 17 \ 17 \ 16] \\ \text{observ. 4 } H = [1 \ 0 \ 1 \ 0 \ 21 \ 16 \ 15] \\ \text{observ. 5 } H = [1 \ 0 \ 1 \ 0 \ 21 \ 18 \ 16] \\ \text{observ. 6 } H = [1 \ 0 \ 1 \ 0 \ 19 \ 13 \ 18] \\ \text{observ. 7 } H = [1 \ 0 \ 1 \ 0 \ 22 \ 17 \ 17] \end{array} \right\}$$

Os valores d_{1k}^* correspondem ao número de domingos e sábados no quadrimestre, os valores d_{2k}^* correspondem ao número de segundas e sextas feiras no quadrimestre e os valores d_{3k}^* correspondem ao número de terças e quintas feiras no quadrimestre. Finalmente, os valores d_{4k} são obtidos subtraindo-se o número de quartas feiras (d_{4k}^*), Seção 3.3. Os feriados são somados aos sábados e domingos;

- Número de iterações: 15;
- Suavizador utilizado: De Jong;
- Valores finais:

$$AIC = 2(7 + 1 + 4 + 4) + 39,28 = 71,28, \Phi = \left[\begin{array}{cc|cc|ccc} 1,03 & 0,07 & 0 & 0 & 0 & 0 & 0 \\ 0,03 & 0,07 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & -1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array} \right]$$

– Evolução do fator $\log L$

$$\left\{ \begin{array}{ccccc} -54612,99 & -20,05 & -19,73 & -19,72 & -19,70 \\ -19,69 & -19,68 & -19,67 & -19,67 & -19,66 \\ -19,66 & -19,65 & -19,65 & -19,64 & -19,64 \end{array} \right\}$$

- f) Modelo: idêntico ao anterior;

- Valores iniciais: os mesmos que o anterior, com exceção da covariância do erro de modelagem:

$$Q_0 = \left[\begin{array}{cc|c} 200 & 200 & 0 \\ 200 & 200 & 0 \\ \hline 0 & 0 & 200 \end{array} \right]$$

- Número de iterações: 15;
- Suavizador utilizado: De Jong;
- Valores finais:

$$AIC = 2(7 + 1 + 4 + 4) + 42,12 = 74,12, \quad \Phi = \left[\begin{array}{cc|cc|ccc} -1,05 & -0,63 & 0 & 0 & 0 & 0 & 0 \\ 0,05 & -0,58 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & -1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array} \right]$$

- Evolução do fator L

$$\left\{ \begin{array}{ccccc} -24,91 & -22,71 & -21,65 & -20,99 & -20,67 \\ -20,62 & -20,75 & -20,91 & -21,04 & -21,11 \\ -21,13 & -21,14 & -21,12 & -21,10 & -21,06 \end{array} \right\}$$

Apêndice H

Exemplo de Obtenção das Matrizes do Modelo de Decomposição Aditiva da Série Temporal

Uma série temporal pode ser escrita genericamente como:

$$Y_k = \mu_k + T_k + S_k + V_k \quad (\text{H.1})$$

onde:

- μ_k é o valor de um nível quase constante, que pode se mover suavemente no tempo;
- T_k é o componente de tendência da série no instante k ;
- S_k é o componente sazonal no instante k ;
- V_k é uma variável aleatória Gaussiana, de média zero e covariância R_k .

Nas séries temporais do tráfego telefônico, ainda é possível decompor a média μ_k como fazendo parte da tendência T_k , da sazonalidade S_k ou ainda tendo flutuações semanais V_{D_k} . A tendência também pode ser decomposta em um tipo de tendência determinística T_k e uma parte chamada tendência estocástica, T_{E_k} que, geralmente, corresponde a um processo auto-regressivo.

Desta forma, no modelo formado por decomposição aditiva da série temporal, a observação escalar pode ser escrita como:

$$Y_k = T_k + S_k + V_{D_k} + T_{E_k} + V_k \quad (\text{H.2})$$

onde:

- Y_k é o valor da observação no instante k ;
- T_k é o valor tendência da série no instante k ;
- S_k é o valor do componente sazonal no instante k ;
- V_{D_k} é o valor do componente de variação diário, representado por uma regressão que converge a valores fixos;
- T_{E_k} é o valor da tendência estocástica, representada por um processo auto-regressivo de ordem p ;
- V_k é uma variável aleatória gaussiana, de média zero e covariância R_k .

As equações gerais do modelo de Espaço de Estados linear, supondo as entradas de controle U_k nulas, são:

$$Y_k = H_k X_k + V_k \quad (\text{H.3})$$

e

$$X_{k+1} = \Phi X_k + \Gamma W_k \quad (\text{H.4})$$

Se o estado X_k é formado por:

$$X_k = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix}$$

com

$$X_1 = \begin{bmatrix} T_k \\ \vdots \end{bmatrix} \quad X_2 = \begin{bmatrix} S_k \\ \vdots \end{bmatrix} \quad X_3 = \begin{bmatrix} \gamma_{1k} \\ \vdots \\ \gamma_{6k} \end{bmatrix} \quad X_4 = \begin{bmatrix} T_{E_k} \\ \vdots \end{bmatrix}$$

- X_1 é um vetor de tamanho $n_1 \times 1$;
- X_2 é um vetor de tamanho $n_2 \times 1$;
- X_3 é um vetor de tamanho 6×1 ;
- X_4 é um vetor de tamanho $n_4 \times 1$;

O tamanho dos vetores X_1 e X_4 é variável dependendo do tipo de modelo adotado (ver Capítulo 3), o vetor X_2 varia de acordo com o tipo de modelo e com o período aquisitivo dos dados (meses, trimestres, anos, etc.) e o vetor X_3 possui o tamanho fixo $n_3 = 6$, que corresponde à variação semanal.

O vetor de observação H_k é formado por:

$$H_k = [H_1 \quad H_2 \quad H_{3k} \quad H_4]$$

- H_1 tem dimensão $1 \times n_1$;
- H_2 tem dimensão $1 \times n_2$;
- H_{3k} tem dimensão 1×6 e é variante no tempo;
- H_4 tem dimensão $1 \times n_4$.

As matrizes H_j devem ser tais que:

$$H_k X_k = T_k + S_k + V_{D_k} + T_{E_k} \tag{H.5}$$

Geralmente H_k terá apenas o primeiro elemento não-nulo (igual à unidade); uma exceção são os coeficientes sazonais apresentados no Capítulo 3.

Outra exceção é a formação de V_{D_k} . Se H_{3k} possui os valores:

$$[d_{1k} \quad d_{2k} \quad d_{3k} \quad d_{4k} \quad d_{5k} \quad d_{6k}]$$

e X_3 possui os valores:

$$[\gamma_{1k} \quad \gamma_{2k} \quad \gamma_{3k} \quad \gamma_{4k} \quad \gamma_{5k} \quad \gamma_{6k}]^T$$

de tal forma que:

$$H_{3k} X_3 = \sum_{i=1}^6 d_{ik} \gamma_{ik} = V_{D_k}$$

Para que a forma dos efeitos aditivos separados seja mantida, a equação (H.4) deve ser escrita como:

$$X_k = \begin{bmatrix} T_k \\ \vdots \\ S_k \\ \vdots \\ \gamma_{1k} \\ \vdots \\ \gamma_{6k} \\ T_{E_k} \\ \vdots \end{bmatrix} = \begin{bmatrix} \Phi_1 & 0 & 0 & 0 \\ 0 & \Phi_2 & 0 & 0 \\ 0 & 0 & \Phi_3 & 0 \\ 0 & 0 & 0 & \Phi_4 \end{bmatrix} \begin{bmatrix} T_{k-1} \\ \vdots \\ S_{k-1} \\ \vdots \\ \gamma_{1(k-1)} \\ \vdots \\ \gamma_{6(k-1)} \\ T_{E_{k-1}} \\ \vdots \end{bmatrix} + \begin{bmatrix} \Gamma_1 & 0 & 0 & 0 \\ 0 & \Gamma_2 & 0 & 0 \\ 0 & 0 & \Gamma_3 & 0 \\ 0 & 0 & 0 & \Gamma_4 \end{bmatrix} \begin{bmatrix} W_{1k} \\ W_{2k} \\ W_{3k} \\ W_{4k} \end{bmatrix} \quad (\text{H.6})$$

onde

$$W_k = \begin{bmatrix} W_{1k} \\ W_{2k} \\ W_{3k} \\ W_{4k} \end{bmatrix}$$

É fácil ver que esta forma corresponderá a ter 4 modelos de Espaço de Estados independentes:

$$\begin{bmatrix} T_k \\ \vdots \end{bmatrix} = \Phi_1 \begin{bmatrix} T_{k-1} \\ \vdots \end{bmatrix} + \Gamma_1 W_{1k} \quad (\text{H.7})$$

$$\begin{bmatrix} S_k \\ \vdots \end{bmatrix} = \Phi_2 \begin{bmatrix} S_{k-1} \\ \vdots \end{bmatrix} + \Gamma_2 W_{2k} \quad (\text{H.8})$$

$$\begin{bmatrix} \gamma_{1k} \\ \vdots \\ \gamma_{6k} \end{bmatrix} = \Phi_3 \begin{bmatrix} \gamma_{1(k-1)} \\ \vdots \\ \gamma_{6(k-1)} \end{bmatrix} + \Gamma_3 W_{3k} \quad (\text{H.9})$$

Este caso possui a particularidade de ter $\Phi_3 = I_6$ (matriz identidade de dimensão 6×6) e $\Gamma_3 = 0$ (matriz nula), devido ao terceiro componente ser uma regressão de efeitos fixos, de tal forma que $\gamma_{ik} = \gamma_{i(k-1)}$.

$$\begin{bmatrix} T_{E_k} \\ \vdots \end{bmatrix} = \Phi_4 \begin{bmatrix} T_{E_{k-1}} \\ \vdots \end{bmatrix} + \Gamma_4 W_{4k} \quad (\text{H.10})$$

Geralmente W_{2k} e W_{4k} são modelados por um ruído branco escalar, já W_{1k} pode ser um vetor de componentes de ruído branco.

Como ilustração, se o modelo para dados mensais possuir tendência (crescimento linear), sazonalidade, tendência estocástica (AR de ordem 2) e efeito de variação diário, então:

$$X_1 = \begin{bmatrix} T_k \\ T_k \end{bmatrix}, \Phi_1 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, W_{1k} = \begin{bmatrix} w_{11k} \\ w_{12k} \end{bmatrix}, E[W_{1k}W_{1k}^T] = Q_1, \Gamma_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, H_1^T = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$X_2 = \begin{bmatrix} T_{E_k} \\ T_{E_{k-1}} \end{bmatrix}, \Phi_2 = \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix}, W_{2k} = w_{2k}, E[W_{2k}^2] = q_2, \Gamma_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, H_2^T = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$X_4 = \begin{bmatrix} S_k \\ S_{k-1} \\ \vdots \\ S_{k-10} \end{bmatrix}, \Phi_4 = \begin{bmatrix} -1 & \dots & -1 \\ 1 & \dots & 0 \\ & \ddots & \vdots \\ & & 1 & 0 \end{bmatrix}, W_{4k} = w_{4k}, E[W_{4k}^2] = q_4, \Gamma_4 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, H_4^T = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

X_3 já foi descrito e Γ_4 será um vetor nulo de dimensão 6×1 .

O modelo de Espaço de Estados por completo (H, Φ, Γ) , corresponde à ilustração apresentada no Capítulo 3.

Apêndice I

Considerações sobre a Posição do Estado Verdadeiro (Não-observável) no Funcionamento do Filtro de Kalman

Alguns aspectos observados, relativos ao funcionamento do Filtro de Kalman, são:

- **1)** O ganho do filtro é calculado de forma a minimizar $\bar{P}_k = E[(X_k - \bar{X}_k)(X_k - \bar{X}_k)^T]$. Portanto, ao ser realizada no instante k , a filtragem \bar{X}_k é a melhor informação disponível sobre o estado verdadeiro X_k (supondo que o filtro está funcionando adequadamente);
- **2)** De (3.30):

$$\bar{X}_k = X_k^* + K_k(Y_k - HX_k^*)$$

Portanto:

$$\bar{X}_k = X_k^* \pm \Delta$$

onde Δ é um valor a ser adicionado ou subtraído de X_k^* .

$$\begin{aligned} \text{se } Y_k > HX_k^* &\Rightarrow \bar{X}_k = X_k^* + \Delta \\ \text{se } Y_k < HX_k^* &\Rightarrow \bar{X}_k = X_k^* - \Delta \end{aligned}$$

O ganho K_k de (3.32) é $\frac{P_k^* H^T}{HP_k^* H^T + R_k}$, o que implica na norma de K_k menor que um porque, geralmente, a norma do denominador é maior que a norma do numerador.

Para o modelo com dois estados e $H = [1 \ 0]$, o valor do elemento k_{11} de K_k , para cerca de 50 rotas analisadas, resultou < 1 (com elementos de K_k positivos). Portanto, o elemento \bar{x}_k de \bar{X}_k será uma interpolação entre x_k^* e Y_k ($\bar{X}_k = \begin{bmatrix} \bar{x}_k \\ \bar{x}_k \end{bmatrix}$ e $X_k^* = \begin{bmatrix} x_k^* \\ \dot{x}_k^* \end{bmatrix}$). Desta forma, neste caso, o valor do estado verdadeiro está entre o valor da predição a um passo e o valor observado.

- **3)** Para o caso univariável e estado unidimensional, tem-se:

$$K_k = \frac{P_k^* H}{P_k^* H^2 + R_k}$$

Para $K_k < 1$ é preciso que $P_k^* < \frac{R_k}{H(1-H)}$. Para $H = 1$, K_k é com certeza < 1 . Se $H = 0,9$, é preciso que $P_k^* < 11,11R_k$, o que tem ocorrido com os valores da prática.

Apêndice J

Diferentes Formas da Equação da Covariância do Erro de Estimação (\bar{P}_k)

Existem duas formas possíveis para a covariância do erro de estimação, (3.34) e (3.66), e portanto, para uma avaliação teórica do *EQM*. Partindo de (3.66), tem-se:

$$(I - K_k H)P_k^*(I - K_k H)^T + K_k R K_k^T$$

Mas esse termo pode ser escrito como:

$$(I - K_k H)P_k^* - (I - K_k H)P_k^* H^T K_k^T + K_k R K_k^T \quad (\text{J.1})$$

Os dois termos da direita de (J.1) podem ser escritos como:

$$-P_k^* H^T K_k^T + K_k H P_k^* H^T K_k^T + K_k R K_k^T = -P_k^* H^T K_k^T + K_k (H P_k^* H^T + R) K_k^T$$

Utilizando $K_k = P_k^* H^T (H P_k^* H^T + R)^{-1}$, obtém-se:

$$-P_k^* H^T K_k^T + P_k^* H^T (H P_k^* H^T + R)^{-1} (H P_k^* H^T + R) K_k^T = 0$$

Portanto, resta o primeiro termo de (J.1), que é (3.34).

A forma da equação (3.34), embora mais simples, é computacionalmente inferior à (3.66). A equação (3.66) evita a propagação de erros de arredondamento ou de truncamento no cálculo do ganho e é conhecida como algoritmo de Joseph [24].

Apêndice K

Geração de Séries Temporais pelo Método de Monte Carlo

K.1 Introdução

A seguir serão abordados alguns conceitos e definições que serão empregados para a geração dos dados pelo método de Monte Carlo.

K.1.1 Geração da Distribuição Gaussiana

Para gerar a distribuição Gaussiana de média nula utilizou-se a distribuição uniforme, $s_i \sim \mathcal{U}[0, 1]$, $E[s_i] = \frac{1}{2}$, e $VAR[s_i] = \frac{1}{12}$. Pelo Teorema do Limite Central, a soma infinita de variáveis aleatórias de mesma distribuição, qualquer que seja essa distribuição, converge para a distribuição Gaussiana. Seja:

$$X = s_1 + s_2 + \dots + s_n - \frac{1}{2}n$$

X é Gaussiana de média:

$$E[X] = n\frac{1}{2} - \frac{1}{2}n = 0$$

e variância:

$$VAR[X] = n\frac{1}{12}$$

na prática, a soma de doze amostras já é suficiente para garantir que a distribuição é Gaussiana. Note-se que quando se desejar $VAR[X] = K$, utiliza-se $12K$ amostras da distribuição $\mathcal{U}[0, 1]$.

K.1.2 Distribuição Normal Contaminada

Se:

$$V_k \sim (1 - \epsilon)\mathcal{N}(0, R_k) + \epsilon\mathcal{N}(0, jR_k)$$

onde $0 \leq \epsilon \leq 1$, $2 \leq j > 1$; diz-se que V_k é uma distribuição normal contaminada, de média nula e variância $(1 - \epsilon)R_k + \epsilon jR_k$.

K.1.3 Estimação da Correlação Cruzada entre Duas Variáveis Aleatórias Gaussianas

Seja $W_k = [w_{1k} \ w_{2k}]^T$, um vetor de variáveis aleatórias normais, geradas como acima. A matriz da covariância Q é gerada como:

$$\hat{Q} = \frac{1}{m} \sum_{i=1}^m W_k W_k^T$$

onde m é um valor suficientemente grande para permitir \hat{Q} próximo a Q .

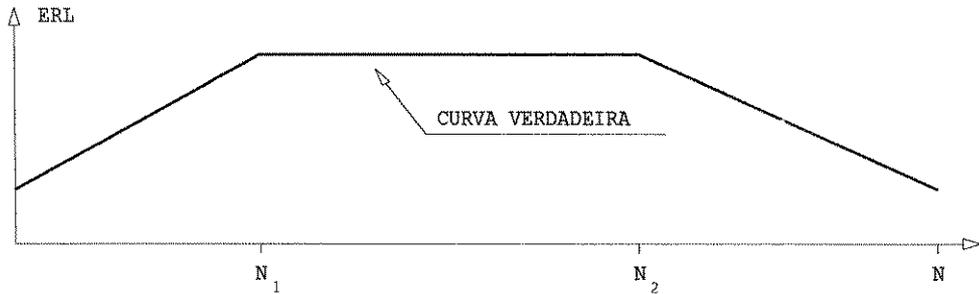


Figura K.1: Curva Verdadeira (não-Observável)

K.1.4 Estimação de \bar{X}_0 e \bar{P}_0

Como se conhece o valor da curva verdadeira, X_V , porque é uma função determinística imposta, estima-se \bar{X}_0 como:

$$Y_0 = X_{V_0} + V_k$$

onde V_k é a distribuição normal contaminada. \bar{X}_0 é estimado como:

$$\hat{\bar{X}}_0 = \frac{1}{m} \sum_{i=1}^m Y_0$$

onde m é um valor suficientemente grande para garantir $\hat{\bar{X}}_0$ próximo de \bar{X}_0 . Note-se que os valores de Y_0 são diferentes devido a diferentes valores de V_k .

Por sua vez, \bar{P}_0 é dado por:

$$\bar{P}_0 = E[(X_0 - \bar{X}_0)(X_0 - \bar{X}_0)^T]$$

e então sua estimação é:

$$\hat{\bar{P}}_0 = \frac{1}{m} \sum_{i=1}^m [(X_{V_0} - \hat{\bar{X}}_0)(X_{V_0} - \hat{\bar{X}}_0)^T]$$

K.2 Geração das Observações

Supondo que os dados sejam como na Figura K.1, procede-se da seguinte maneira:

- Geram-se os vetores Gaussianos V_k e W_k ;
- Calcula-se $X_{V_k} = \Phi_j X_{V_{k-1}}$;
- Calcula-se $X_k = X_{V_k} + W_k$;
- Calcula-se $Y_k = H X_{V_k} + V_k$ ($H = [1 \ 0]$).

onde X_V corresponde à curva verdadeira, teoricamente não acessível, X_k corresponde à equação do sistema e Y_k são os dados observados (equação da observação do Filtro de Kalman). Para $k \leq N_1$,

$$\Phi_1 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \text{ para } N_1 < k \leq N_2, \Phi_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \text{ e para } k > N_2, \Phi_3 = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}.$$

K.3 Exemplo de Aplicação

Supondo $\epsilon = 0, 1$; $R_k = 2$, $j = 20$, então:

$$V_k = (1 - \epsilon)\mathcal{N}(0, 2) + \epsilon\mathcal{N}(0, 20)$$

$$X_{V_0} = \begin{bmatrix} 50 \\ 1,5 \end{bmatrix}$$

$$W_k = \begin{bmatrix} w_{11k} \\ w_{21k} \end{bmatrix}$$

onde $w_{11k} \sim \mathcal{N}(0, 1)$, $w_{21k} \sim \mathcal{N}(0, 2)$, $N_1 = 25$, $N_2 = 50$ e $N = 75$, os seguintes dados foram gerados como na Tabela A.1.

Tabela A.1 - Dados Gerados pelo Método de Monte Carlo

k	X_{V_k}								
1	51,96	16	72,67	31	87,36	46	82,57	61	73,77
2	54,94	17	74,28	32	85,87	47	89,65	62	72,84
3	56,24	18	79,20	33	88,32	48	85,01	63	68,71
4	59,92	19	76,64	34	87,46	49	83,34	64	65,22
5	52,45	20	80,36	35	87,79	50	81,99	65	63,53
6	60,55	21	82,88	36	90,03	51	88,52	66	61,06
7	59,16	22	80,93	37	87,02	52	86,45	67	61,16
8	56,49	23	82,09	38	83,77	53	86,55	68	60,67
9	63,07	24	82,98	39	85,06	54	81,40	69	62,13
10	63,56	25	86,56	40	88,91	55	80,57	70	57,32
11	68,77	26	86,63	41	87,20	56	74,34	71	55,51
12	70,62	27	85,94	42	86,72	57	79,06	72	49,07
13	69,66	28	86,81	43	84,89	58	75,11	73	53,02
14	67,37	29	89,75	44	88,80	59	75,39	74	50,93
15	71,55	30	84,07	45	85,78	60	73,91	75	46,02

Os seguintes parâmetros também foram estimados pelo método anteriormente descrito:

$$\hat{X}_0 = \begin{bmatrix} 50 \\ 1,5 \end{bmatrix}, \hat{P}_0 = \begin{bmatrix} 1,11 & 0,03 \\ 0,03 & 10^{-3} \end{bmatrix}, \hat{Q} = \begin{bmatrix} 1,39 & 0,19 \\ 0,19 & 0,77 \end{bmatrix}$$

Esta série temporal gerada foi utilizada nos testes do Capítulo 5.

Apêndice L

Lista de Símbolos

Os símbolos apresentados aqui são os que foram consistentemente usados e não apenas eventualmente como variáveis auxiliares ou temporárias.

- Y_k - Valor da saída observada do sistema;
- V_k - Perturbações (ruídos) das observações (v_k para sistemas univariáveis);
- μ_k - Nível constante (ou aproximadamente constante) da série temporal;
- T_k - Ciclotendência da série temporal (ou tendência);
- S_k - Componente sazonal da série;
- T_{E_k} - Tendência estocástica, geralmente representada por um processo auto-regressivo de ordem p ;
- V_{D_k} - Componente de variação diário, representado por uma regressão que converge a valores fixos;
- ν_j - Autocovariâncias de um processo estocástico;
- $\rho_j = \frac{\nu_j}{\nu_0}$ - Autocorrelações de um processo estocástico;
- $Y_{k+n|k}^* = Y_{k+n}^*$ - Projeção n passos à frente a partir do instante k ;
- Y_k^* - Projeção um passo à frente a partir de $k - 1$;
- $\Phi(B)$ - Polinômio em B (operador deslocamento para trás) que representa os coeficientes de um modelo AR (auto-regressivo);
- $\Theta(B)$ - Polinômio em B que representa os coeficientes de um modelo MA (*moving average*);
- $\xi(B)$ - Polinômio em B que representa os coeficientes das entradas de controle;
- X_k - Vetor de variáveis de estado;
- L - Matriz de entradas (ganhos) associada às entradas de controle;
- U_k - Vetor de entradas de controle;
- Φ - Matriz de transição de estado (ou Φ_k , se for variante no tempo);
- Γ - Matriz de entradas associada às perturbações (ruídos) no estado;
- W_k - Vetor de erros de modelagem (ruídos no estado);
- H - Matriz de observação;
- \bar{X}_k - Estimativa do vetor do estado (filtragem);
- X_k^* - Vetor do estado estimado um passo à frente a partir de $k - 1$ (predição);
- K_k - Matriz do ganho do Filtro de Kalman;
- \bar{P}_k - Covariância do erro de estimação (associada a \bar{X}_k);

- P_k^* - Covariância do erro de predição (associada a X_k^*);
- r_k - Termo de inovação (resíduo);
- σ_k^2 - Covariância do termo de inovação;
- $R_k = E[V_k V_k^T]$ - Covariância do erro de observação (apenas R se for invariante no tempo);
- $Q_k = E[W_k W_k^T]$ - Covariância do erro de modelagem (apenas Q se for invariante no tempo);
- \bar{X}_k^N - Estado estimado (suavizado) utilizando N observações;
- \bar{P}_k^N - Covariância do erro de estimação para o estado suavizado;
- $\bar{P}_{k-1,k}^N$ - Covariância de transição de estado (do erro entre o valor suavizado em k e o estado verdadeiro e o valor suavizado em $k - 1$ e o estado verdadeiro);
- g, \hat{g} - Taxas de crescimento real e taxa de crescimento agregado estimada;
- σ_g^2 - Variância do erro entre a taxa de crescimento real da rota e a taxa de crescimento agregado estimada;
- r_v - Erro de predição (resíduo) verdadeiro;
- q_k - Vetor com elementos de Q_k (estado do subfiltro);
- z_k - Vetor da pseudo-observação do subfiltro;
- η_k - Perturbação (ruído) da equação da pseudo-observação do subfiltro;
- M - Matriz da pseudo-observação do subfiltro (corresponde à matriz H do filtro principal);
- Φ^q - Matriz de transição de estado do subfiltro;
- W_k^q - Vetor de perturbações (ruídos) de modelagem do subfiltro;
- $\sigma_q^2 = E[W_k^q (W_k^q)^T]$ - Covariância do erro de modelagem do subfiltro;
- \bar{P}_k^q - Coivariância do erro de estimação do subfiltro;
- P_k^{q*} - Covariância do erro de predição do subfiltro;
- \bar{q}_k - Valor do estado estimado (filtragem) do subfiltro;
- q_k^* - Valor da predição do subfiltro (estado estimado um passo à frente a partir de $k - 1$);
- K_k^q - Matriz do ganho do subfiltro.