



COPPE/UFRJ

MODELOS NÃO-LINEARES E LINEARES PARA PREVISÃO DO FLUXO DE
RESERVAS BANCÁRIAS DE UM BANCO COMERCIAL

Marco Antonio Freire Allemão

Tese de Doutorado apresentada ao Programa de Pós-graduação em Engenharia Civil, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Doutor em Engenharia Civil.

Orientador(es): Nelson Francisco Favilla
Ebecken
Luiz Pereira Calôba

Rio de Janeiro
Fevereiro de 2009

Livros Grátis

<http://www.livrosgratis.com.br>

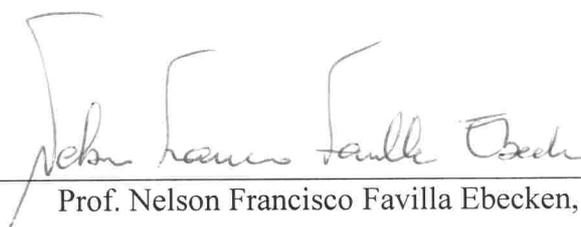
Milhares de livros grátis para download.

MODELOS NÃO-LINEARES E LINEARES PARA PREVISÃO DO FLUXO DE
RESERVAS BANCÁRIAS DE UM BANCO COMERCIAL

Marco Antonio Freire Allemão

TESE SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ
COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA (COPPE) DA
UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS
REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR EM
CIÊNCIAS EM ENGENHARIA CIVIL.

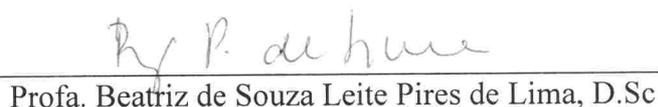
Aprovada por:



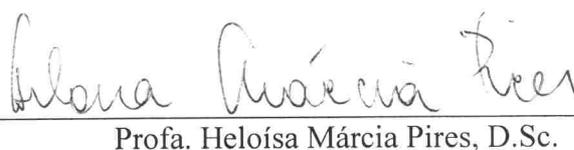
Prof. Nelson Francisco Favilla Ebecken, D.Sc.



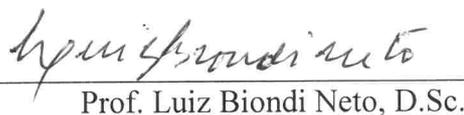
Prof. Luiz Pereira Calôba., Dr.



Profa. Beatriz de Souza Leite Pires de Lima, D.Sc



Profa. Heloisa Márcia Pires, D.Sc.



Prof. Luiz Biondi Neto, D.Sc.

RIO DE JANEIRO, RJ - BRASIL

FEVEREIRO DE 2009

Allemão, Marco Antonio Freire

Modelos Não-Lineares e Lineares para Previsão do Fluxo de Reservas Bancárias de um Banco Comercial/ Marco Antonio Freire Allemão. – Rio de Janeiro: UFRJ/COPPE, 2009.

IX, 192 p.: il.; 29,7 cm.

Orientador: Nelson Francisco Favilla Ebecken

Luiz Pereira Calôba

Tese (doutorado) – UFRJ/ COPPE/ Programa de Engenharia Civil, 2009.

Referencias Bibliográficas: p. 186-192.

1. Séries Temporais. 2. Modelos Não-Lineares e Lineares. 3. Previsão do Caixa Mínimo 4. Reservas Bancárias. I. Ebecken, Nelson Francisco Favilla *et al.* II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia Civil. III. Título.

AGRADECIMENTOS

Muitos foram os que contribuíram de forma decisiva para a conclusão deste trabalho. Agradeço, sinceramente, a todos.

Ao Banco do Brasil que através do programa de doutorado me deu a oportunidade de desenvolver este trabalho.

Aos amigos Gilberto Gabriel e Marcelo Duarte e Imenes, da DIFIN, que nunca deixaram de vibrar com os resultados e incentivar novas pesquisas.

A Cristiane Freire pelo seu companheirismo, auxílio e estímulo.

Aos professores Nelson Ebecken e Luiz Calôba, pelo profissionalismo e ensinamentos transmitidos.

Pelo carinho e compreensão durante meus estudos, dedico um especial agradecimento aos meus filhos Carla e Felipe.

As minhas queridas filhas Carolina e Raquel que se viram, de repente, sem pai para acompanhá-las durante minhas intermináveis horas de estudos.

A Marília, minha mãe, pelo carinho e apoio.

Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Doutor em Ciências (D.Sc.)

MODELOS NÃO-LINEARES E LINEARES PARA PREVISÃO DO FLUXO DE
RESERVAS BANCÁRIAS DE UM BANCO COMERCIAL

Marco Antonio Freire Allemão

Fevereiro/2009

Orientadores: Nelson Francisco Favilla Ebecken

Luiz Pereira Calôba

Programa: Engenharia Civil

As instituições bancárias realizam diariamente milhares de transações comerciais entre clientes e com outros agentes financeiros bancários e não bancários. Os montantes envolvidos nestas operações são muito elevados, fazendo com que o planejamento referente às necessidades de recursos seja uma importante etapa para a melhoria da rentabilidade dos valores administrados pelo Banco. O controle do fluxo das reservas bancárias garante o equilíbrio entre os pagamentos e recebimentos e possibilita o direcionamento de maior volume de recursos para as operações de crédito, através de redução dos valores destinados a assegurar a liquidez, que corresponde à capacidade de uma empresa em saldar seus compromissos no prazo certo e na quantia exata. A realização do planejamento exige estudos e modelos de projeções das variáveis financeiras envolvidas, a fim de calcular os valores movimentados o mais próximo do realizado. Diversas heurísticas vêm sendo pesquisadas e aplicadas pelas empresas buscando-se o aumento da rentabilidade dos recursos disponíveis no caixa. Esta tese avalia as séries financeiras envolvidas na movimentação de um banco comercial brasileiro e procura desenvolver um modelo neural que prediga o montante de recursos a serem transacionados em $d + 2$ (onde d representa o dia vigente) visando auxiliar os executivos no planejamento financeiro.

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Doctor of Science (D.Sc.)

NON-LINEAR AND LINEAR MODELS IN THE FORECAST OF BANKING
RESERVE FLOW OF A COMMERCIAL BANK

Marco Antonio Freire Allemão

February/2009

Advisors: Nelson Francisco Favilla Ebecken

Luiz Pereira Calôba

Department: Civil Engineering

Daily, banking institutions carry out thousands of commercial transactions between customers and other banking or non banking financial agents. The sums involved in these operations are very high, which makes resource planning an important stage for the improvement of the value profitability managed by the bank. Bank reserve flow control guarantees the balance between withdrawals and incomes. In addition, it enables the direction of a bigger resource volume to credit operations through the reduction of values directed to assure liquidity, which corresponds to the capacity of a company to liquidate its debts in the due time and in the exact amount. The accomplishment of this planning demands studies and financial variable projection models in order to calculate all the involved movement as closer as possible of the real one. Several heuristics have been researched and applied by companies in the search of increasing profitability of available resources. This thesis not only analyses the financial series involved in the general movement of a Brazilian commercial bank, but also seeks to develop a neural model that forecasts the resource amount to be dealt in $d+2$ (where d represents the current day) aiming at assisting financial planning executives.

ÍNDICE

1 INTRODUÇÃO	1
1.1 ADMINISTRAÇÃO FINANCEIRA	1
1.2 FLUXO DE CAIXA	1
1.2.1 <i>Gestão do Fluxo de Caixa</i>	3
1.3 GESTÃO DA RESERVA BANCÁRIA.....	5
1.3.1 <i>Previsão</i>	8
1.4 O PROBLEMA	8
1.5 OBJETIVOS	9
1.6 ORGANIZAÇÃO	11
2 BASE DE DADOS	13
2.1 RESERVAS BANCÁRIAS	13
2.1.1 <i>Variáveis do Fluxo de Reservas Bancárias</i>	14
2.1.2 <i>Classificação das Variáveis do Fluxo de Reservas Bancárias</i>	17
2.1.3 <i>Gestão do Fluxo de Reservas Bancárias</i>	19
2.2 TRATAMENTO DA DIMENSIONALIDADE.....	19
2.2.1 <i>Redução da Dimensionalidade - 1ª Fase</i>	19
2.2.2 <i>Redução da Dimensionalidade - 2ª Fase</i>	20
2.3 TRATAMENTO DOS DADOS	21
2.3.1 <i>Valores Inexistentes</i>	28
2.3.2 <i>Normalização dos Dados</i>	29
2.4 DECOMPOSIÇÃO DA SÉRIE.....	30
2.4.1 <i>Extração da Tendência</i>	31
2.4.2 <i>Extração da Sazonalidade</i>	32
2.4.3 <i>Extração dos Ciclos</i>	33
2.4.4 <i>Retirada de Semanas Incompletas</i>	36
2.4.5 <i>Irregularidade</i>	39
2.5 OBSERVAÇÕES	41
2.6 HIPÓTESES	43
3 ESTUDO DA ALEATORIEDADE DA SÉRIE	45
3.1 INTRODUÇÃO.....	45
3.2 AMOSTRAGEM ALEATÓRIA	45
3.3 MÉTODO DA AUTOCORRELAÇÃO	46
3.4 MÉTODO DO QUI-QUADRADO	47
3.5 MÉTODO DE KOLMOGOROV-SMIRNOV	48
3.6 MÉTODO DE CORRIDAS	50

3.6.1	<i>Corrida</i>	50
3.6.2	<i>Análise das Repetições</i>	51
3.6.3	<i>Teste de repetições de dois tipos de observações</i>	52
3.6.4	<i>Teste de repetições acima e abaixo da mediana</i>	53
3.6.5	<i>Teste de repetições acima e abaixo em relação ao anterior</i>	53
3.7	RESULTADOS	54
3.7.1	<i>Teste da autocorrelação</i>	55
3.7.2	<i>Teste do qui-quadrado</i>	66
3.7.3	<i>Teste de Kolmogorov-Smirnov</i>	67
3.7.4	<i>Teste de repetições de 2 tipos de observações</i>	69
3.7.5	<i>Teste de repetições acima e abaixo da mediana</i>	69
3.7.6	<i>Teste de repetições acima e abaixo em relação ao anterior</i>	70
3.8	CONCLUSÕES	70
4	MODELOS NEURAI	73
4.1	INTRODUÇÃO	73
4.2	PESQUISAS COM MODELOS NEURAI	74
4.3	MODELOS NEURAI APLICADOS	79
4.4	TRATAMENTO DOS DIAS DE FERIADOS	88
4.5	TRATAMENTO DOS CICLOS SENOIDAIS	92
4.6	AVALIAÇÃO DA MÉDIA DOS DIAS ÚTEIS	95
4.7	ANÁLISE GRÁFICA DO ERRO	99
4.8	CONCLUSÃO	102
5	OUTRAS ABORDAGENS NEURAI	105
5.1	INTRODUÇÃO	105
5.2	TRATAMENTO DAS DEMAIS VARIÁVEIS INTEGRANTES DO FLUXO DE CAIXA	106
5.3	VARIÁVEIS EXTERNAS	119
5.4	AGREGAÇÃO DE VARIÁVEIS	121
5.5	CONCLUSÃO	130
6	ESTRATÉGIAS NÃO NEURAI E AVALIAÇÕES	132
6.1	INTRODUÇÃO	132
6.2	MODELO ARIMA - AUTOREGRESSIVE INTEGRATED MOVING AVERAGE	133
6.2.1	<i>Aplicação do Modelo ARIMA</i>	135
6.3	MODELO DIFUSO	137
6.3.1	<i>Aplicação do Modelo Difuso</i>	143
6.4	AVALIAÇÃO DOS MODELOS	145
6.4.1	<i>Extração de Componentes Determinísticas</i>	146
6.4.2	<i>Modelo Neural</i>	148

6.4.3 Modelo ARIMA.....	148
6.4.4 Modelo Difuso.....	149
6.5 CONCLUSÕES	150
7 CONCLUSÃO	151
7.1 OBJETIVOS	151
7.2 PRINCIPAIS DIFICULDADES ENCONTRADAS.....	153
7.3 TRABALHOS FUTUROS.....	155
8 ANEXOS	157
1. VARIÁVEIS DO FLUXO DE CAIXA.....	157
2. VARIÁVEIS DO FLUXO DE CAIXA COM REGRAS DE AGREGAÇÃO	161
3. VARIÁVEIS APÓS REDUÇÃO DA DIMENSIONALIDADE – 1ª ETAPA	164
4. VARIÁVEIS APÓS REDUÇÃO DA DIMENSIONALIDADE – 2ª ETAPA	166
5. CARACTERÍSTICAS DOS ARQUIVOS DE DADOS E TOPOLOGIAS DE REDES.....	168
6. QUADROS DE FREQUÊNCIAS DAS VARIÁVEIS	177
7. CÁLCULOS DE CORRELAÇÕES	181
9 REFERÊNCIAS	186

1 INTRODUÇÃO

1.1 ADMINISTRAÇÃO FINANCEIRA

Na administração financeira de uma instituição, qualquer que seja o grau de complexidade de suas operações ou a abrangência de sua atuação, são três os conceitos básicos que determinam o sucesso ou o fracasso de uma gestão financeira (SILVEIRA, 2008):

- Planejamento financeiro;
- Acompanhamento das entradas e desembolsos de recursos; e
- Transparência nas operações.

O planejamento financeiro é o processo de administrar os recursos visando a sua alocação em iniciativas estratégicas, ou seja, é o processo no qual se avaliam os investimentos, rendas, patrimônio, despesas e dívidas para subsidiar as decisões que refletirão na continuidade operacional da empresa.

O acompanhamento dos eventos financeiros é realizado mediante o tratamento dispensado ao fluxo de caixa, que permite saber de imediato se haverá disponibilidade para honrar os desembolsos que irão ocorrer no futuro. Permite, também, programar os remanejamentos de recursos necessários para fazer face a compromissos e prever entradas de valores antecipadas, entre outras informações necessárias à boa gestão da empresa.

A transparência é assegurada através de sua escrituração contábil apoiada em documentação legal, visando o atendimento da legislação em vigor bem como proporcionar legitimidade aos atos dos administradores.

1.2 FLUXO DE CAIXA

As instituições bancárias de todo o mundo realizam, diariamente, milhares de transações comerciais entre clientes e com outros agentes financeiros bancários e não bancários, a exemplo do Banco Nacional de Desenvolvimento Econômico e Social –

BNDES e Banco Central do Brasil – BACEN. Os montantes envolvidos nestas operações são muito elevados, fazendo com que o planejamento referente às necessidades de recursos seja uma importante etapa com vistas à melhoria da rentabilidade dos valores administrados pelo banco.

Entretanto, trata-se de um exercício extremamente complexo. Além das variáveis inerentes ao negócio bancário, existem as variáveis externas que, de uma maneira ou de outra, afetam direta ou indiretamente as reservas financeiras da instituição, como por exemplo o PIB – Produto Interno Bruto e os índices de inflação. Pode-se citar, também, as variáveis que não são mensuráveis, mas que influenciam a movimentação financeira de uma instituição bancária, tais como dia da semana, o dia do mês, o mês do ano, o dia anterior ou imediatamente após um feriado e outras mais. Até o clima tem influência na movimentação dos recursos de uma instituição financeira.

De forma resumida pode-se definir fluxo de caixa como a movimentação entre a origem de todo o dinheiro que entra na empresa, ou seja, no caixa; e a aplicação de todo o dinheiro que sai em um período de tempo.

Zdanowicz (ZDANOWICZ, 1989), conceitua fluxo de caixa como:

“Denomina-se fluxo de caixa de uma empresa ao conjunto de ingressos e desembolsos de numerário ao longo de um período determinado. O fluxo de caixa consiste na representação dinâmica da situação financeira de uma empresa, considerando todas as fontes de recursos e todas as aplicações em itens do ativo.

De forma mais sintética pode-se conceituar: é o instrumento de programação financeira que compreende as estimativas de entradas e saídas de caixa em certo período de tempo projetado.”

A movimentação do fluxo de caixa irá indicar o que ocorreu em um determinado período em termos de saída e entrada de dinheiro na empresa bem como o resultado desse fluxo. Além disso, propicia ao gestor a elaboração de um melhor planejamento financeiro pois no universo econômico em que se inserem as empresas não é aconselhável manter excessos no caixa, mas apenas o estritamente necessário para fazer face aos seus compromissos, ou seja, auxilia nas análises do montante correto de empréstimos para cobrir a falta (insuficiência) de fundos, bem como quando e quanto aplicar no mercado financeiro o excesso de dinheiro existente, evitando a corrosão

inflacionária e proporcionando maior rendimento à empresa.

O fluxo de caixa é considerado um importante instrumento de análise e avaliação de uma empresa, pois proporciona ao gestor uma visão futura dos recursos financeiros disponíveis. Dessa forma, uma eficiente projeção do fluxo de caixa permitirá a avaliação da capacidade da empresa em gerar recursos para suprir o aumento das necessidades de capital de giro oriundas do nível de atividades, remunerar os acionistas da empresa, efetuar pagamento de impostos e realizar compromissos com fornecedores dentre outras obrigações.

Além de permitir analisar a forma como uma empresa desenvolve sua política de captação e aplicação de recursos, o acompanhamento entre o fluxo projetado e o efetivamente realizado permite identificar as variações ocorridas e as causas dessas variações. Estas variações são subsídios para aperfeiçoamento de novas projeções de fluxos de caixa.

Portanto, o gerenciamento do fluxo de caixa é uma das maiores preocupações das empresas modernas, particularmente dos bancos. É, também, uma das atribuições que mais compromete o tempo dos executivos financeiros, pelo fato de haver grande quantidade de variáveis das quais depende o dimensionamento dos valores disponíveis convenientes à organização.

1.2.1 Gestão do Fluxo de Caixa

O fluxo de caixa está diretamente ligado à gestão financeira dos recursos das empresas, buscando o equilíbrio entre os pagamentos e recebimentos, constituindo-se, desta forma, em um instrumento gerencial fundamental ao processo de tomada de decisões financeiras (QUEJI, 2002).

A execução do fluxo de caixa movimenta, diariamente, bilhões de reais e o objetivo desta atividade é, primordialmente, gerenciar as disponibilidades de recursos de forma a assegurar a liquidez¹. Neste sentido, procura-se identificar os melhores investimentos para o saldo de caixa operacional disponível ou as melhores ações para a

¹ Liquidez significa dispor de fundos para atender, de imediato, a todos os compromissos que estiverem vencendo, inclusive depósitos à vista e os registrados em contas de compensação.

cobertura das necessidades existentes.

O controle do nível de caixa se baseia na premissa da necessidade em garantir reservas mínimas de liquidez e da oportunidade de direcionamento de maior volume de recursos para as operações de crédito (uma das atividades-fim dos bancos).

Além das variáveis externas já mencionadas existem aquelas inerentes ao negócio bancário, como por exemplo os tipos de investimentos ou empréstimos, prazos, taxas de rendimento, regras de resgate e pagamento, risco de crédito, impostos etc. Assim, o planejamento eficiente do fluxo de caixa melhora a performance geral da empresa assegurando liquidez e reduzindo seus custos financeiros.

Desde o trabalho pioneiro publicado por Baumol (*apud* SOUSA, BARROS, 2000), em meados do século passado, o desenvolvimento de modelos para o gerenciamento dos saldos de caixa das empresas tem atraído o interesse de respeitável contingente de pesquisadores das áreas de finanças e economia. Até o presente momento pode-se contabilizar dezenas de tentativas de se conferir um tratamento matemático objetivando o equilíbrio da conta caixa, de forma a minimizar os custos associados à sua gestão ou otimizar o valor da empresa.

Os modelos mais conhecidos foram desenvolvidos a partir do tratamento do equilíbrio de caixa como um estoque a mais a ser gerenciado pela empresa. Tendo esta analogia como ponto de partida, é possível reunir condições para a otimização entre os custos referentes ao excesso de caixa e os relacionados à sua falta.

Mesmo tendo em comum a visão do caixa como um estoque, as proposições dos diversos autores que se dedicaram ao tema diferem entre si em aspectos substantivos. Provavelmente, o fator de distinção mais importante entre os vários modelos é o enfoque que cada um dá à questão da incerteza dos fluxos de dinheiro esperados pela empresa. Nas primeiras tentativas de equacionar o problema da gestão do caixa, os pesquisadores tenderam a construir modelos determinísticos, em tudo assemelhados às técnicas de administração de estoques físicos. Num segundo momento procurou-se acrescentar a variável risco às formulações, sofisticando-se seu tratamento.

De uma forma ou de outra, todos os modelos têm sua aplicabilidade limitada aos casos em que as condições reais da empresa e de seu ambiente atendem, ao menos de forma aproximada, aos pressupostos implícitos ou explícitos nas equações determinísticas dos modelos matemáticos desenvolvidos (SOUSA; BARROS, 2000).

Vários modelos já foram desenvolvidos tentando refletir o comportamento dos fluxos de fundos de investimentos e o objetivo principal é o de servir de ferramenta de gestão, tentando minimizar perdas de rentabilidade derivadas de deficiente previsão do caixa. Entretanto, é necessário reconhecer que todos estes modelos possuem uma série de restrições que limitam sua utilidade prática (VILLALBA; SOUSA, 2001). Dessa forma, para se obter a eficiência necessária na gestão dos recursos é fundamental o acompanhamento constante através de instrumentos de controle, tais como *softwares* e relatórios, avaliando os resultados e, principalmente, procedendo a ajustes e correções necessários. A melhoria do desempenho financeiro é oriunda de uma análise contínua e ajustes permanentes e a adequada gestão dos recursos financeiros tende a diminuir a necessidade de capital de giro, reduzindo as despesas financeiras e trazendo, em consequência, maiores lucros.

1.3 GESTÃO DA RESERVA BANCÁRIA

Apesar de se intitular como Fluxo de Caixa toda a movimentação financeira de uma empresa, seja ela financeira, industrial, comercial ou de serviços, o termo, por si só, nos remete a uma concepção distorcida daquilo que efetivamente se realiza nas instituições financeiras. Quando se trata de bancos é mais correto afirmar que o acompanhamento e planejamento dos recursos segue um foco diferente daquele existente nas demais empresas comerciais, industriais e de serviços. Aqui é mais apropriado denominarmos Fluxo de Caixa como sendo a administração das entradas e saídas de recursos das reservas bancárias – RB e consequentes impactos na liquidez.

Liquidez corresponde à capacidade de uma empresa em saldar seus compromissos no prazo certo e na quantia exata. É a disponibilidade imediata do dinheiro.

Ainda conceituando liquidez, o Ministério da Fazenda o define como (FAZENDA, 2005):

“o grau de agilidade na conversão de um investimento em dinheiro, sem perda significativa de valor. Um investimento tem maior liquidez quanto mais fácil for a conversão em dinheiro e quanto menor for a perda de valor envolvida nesta transação”.

Operacionalmente, liquidez, para o Banco sob estudo, é definida como:

$$\text{Liquidez} = \text{saldo de RB} + \text{carteira de títulos} - \text{operações compromissadas de clientes} \quad \text{Eq. (1)}$$

Onde:

- Saldo de reservas bancárias: corresponde aos recursos que sobram na conta de reservas bancárias após o cumprimento do compulsório sobre recursos à vista menos recursos que faltam para o cumprimento do compulsório sobre recursos à vista;
- Carteira de Títulos: corresponde aos valores referentes aos títulos públicos federais de propriedade da instituição; e
- Clientes (operações compromissadas): corresponde aos títulos públicos federais vendidos a clientes com compromisso de recompra.

Uma liquidez é eficaz quando os meios de pagamento suprem tempestivamente as necessidades de pagamentos, ou seja, quando tais meios se convertem rapidamente em dinheiro a tempo de cobrir as obrigações que a empresa possui. Em suma, a liquidez envolve a contínua conversão de ativos ao longo do tempo a fim de satisfazer as obrigações nos respectivos prazos (DALBELLO, 1999).

A temporalidade da liquidez corresponde ao prazo necessário para que um ativo se torne apto a realizar um pagamento. Além disso, os giros de meios de pagamentos devem acompanhar os giros das necessidades de pagamentos, ou seja, é necessário que em tempo hábil os estoques de mercadorias ou de produtos sejam vendidos e as contas a receber sejam recebidas, para que exista dinheiro suficiente para cobrir tudo o que a empresa precisa pagar, pois só assim haverá eficácia em sua liquidez (DALBELLO, 1999).

Entretanto, nem sempre esta equação acontece de forma harmoniosa. Há situações em que a empresa possui ativos para fazer face aos seus compromissos mas a sua liquidez, no momento, não é suficiente para cobrir suas necessidades. Nestes casos será inevitável recorrer a fontes externas para equalizar as receitas e despesas. Portanto, é muito importante um efetivo planejamento do fluxo de caixa, até porque será através deste mecanismo que se terá a indicação antecipada das necessidades de dinheiro para o atendimento dos compromissos da empresa.

Outro importante papel desempenhado no planejamento do fluxo de caixa se

refere à possibilidade de se evitar a programação de desembolsos vultosos para períodos em que os ingressos orçados serão baixos.

Dessa forma, a administração dos ativos só será eficiente se houver um bom planejamento da movimentação financeira. Por sua vez, um bom planejamento dependerá, fortemente, de um modelo de previsão das entradas e saídas de recursos.

O gerenciamento da liquidez é considerado o item mais importante na área financeira de uma instituição bancária e em função da sua relevância para a saúde da empresa foi criado, pelo Comitê de Supervisão de Bancos da Basiléia, em 2000, um documento onde são listados os princípios referentes às boas práticas no gerenciamento da liquidez de organizações bancárias (BASILÉIA, 2000).

Os princípios se aplicam a todos os bancos, em particular aos grandes bancos onde uma crise pode afetar toda a saúde financeira do sistema bancário. Em passado recente, nos anos de 1997 e 1998, o Brasil atravessou uma fase onde houve necessidade de forte intervenção do Banco Central a fim de evitar uma crise de proporções agudas. Segundo Loyola (LOYOLA, 2006), ex-presidente do BACEN, a crise brasileira foi decorrente de 4 fatores:

1. Entrada de capitais de curto prazo incentivada pela existência de mecanismos explícitos ou implícitos de garantia bancária.
2. Saída abrupta e massiva de capitais em razão do aumento do risco de perda de capital.
3. Perda de passivos bancários acarretando crise de liquidez que evoluiu para crise de solvência.
4. Depreciação cambial com piora da qualidade da carteira de crédito dos bancos.

Ainda segundo o ex-presidente do BACEN, uma boa macroeconomia não significa que um país está livre de crises bancárias. Entretanto, crises importantes podem ser evitadas com planejamento e controle da liquidez de cada banco.

Verifica-se, portanto, a importância do tema não só para as instituições financeiras como para o país. Além disso, para se realizar os passos acima é necessário, antes de mais nada, a criação de modelos matemáticos responsáveis por previsões robustas das principais variáveis componentes da rubrica reserva bancária a fim de se prever com antecedência as movimentações que poderão ocorrer e realizar o

planejamento necessário para minimizar os seus efeitos.

1.3.1 Previsão

Uma vez que as condições econômicas e de negócios variam ao longo do tempo, os responsáveis pela condução da empresa precisam encontrar maneiras de se manterem a par dos efeitos que essas mudanças terão em suas operações. Uma técnica que pode ser empregada, como ajuda no planejamento de necessidades operacionais futuras é a previsão, definida por Faria (1967) como:

“... a verificação antecipada dos resultados futuros das prováveis ocorrências, dentro de um lapso de tempo que permita tirar o melhor partido da situação; em última análise, consiste em evitar a casualidade através do estudo da causalidade”.

Existem diversas técnicas para se fazer previsão e todas elas se baseiam no pressuposto de que os fatores que influenciaram padrões da atividade no passado e no presente continuarão a fazê-lo no futuro (LEVINE *et al.*, 2000). Dessa forma, o estudo das séries temporais – série temporal é qualquer conjunto de observações ordenadas no tempo (MORETTIN; TOLOI, 1987) (STEVENSON, 2001) – pode trazer informações importantes sobre o futuro, pois normalmente existe correlação entre as variáveis em diversos instantes. É claro que algum grau de incerteza virá agregado, pois o futuro nunca refletirá exatamente as ocorrências passadas mas, mesmo assim, a prática da previsão tem sido usada como auxílio ao planejamento e à tomada de decisões.

1.4 O PROBLEMA

Um bom modelo de fluxo de caixa para um banco necessita de projeções e análises eficazes das variáveis financeiras. Entretanto, apesar de ser um problema que afeta todas as instituições bancárias do mundo, ainda assim não se encontra vasto material disponível sobre o assunto. Acredita-se que em razão do segmento a que se destina – bancário, onde a confidencialidade é um dos requisitos básicos, tais trabalhos devam estar restritos aos seus usuários. Além do mais, qualquer solução para este problema será, necessariamente, personalizada, pois cada banco possui características próprias que dificultam o uso de um modelo único.

Portanto, os esforços despendidos para se desenvolver um modelo de fluxo

de reservas bancárias tem por objetivo a formulação de modelos matemáticos que possam descrever o comportamento das séries sob estudo e, assim, permitir um conjunto de ações pró-ativas visando a melhoria da rentabilidade das aplicações da empresa.

Entretanto, a dinâmica que acompanha o sistema bancário faz com que hajam alterações frequentes não só na forma de registro dos recursos que entram e saem da instituições como mudanças contábeis e conceituais dos diversos produtos oferecidos. Com isto, valores que em um momento eram contabilizados em uma determinada rubrica passam, em um instante seguinte, a serem registrados em novas rubricas ou agregados em uma outra pré-existente, dificultando o acompanhamento da série histórica e inviabilizando o uso dos modelos existentes.

Em abril de 2002, por exemplo, mudanças decorrentes da reestruturação do Sistema de Pagamentos Brasileiro – SPB trouxeram substanciais alterações nos meios de pagamentos, com significativas alterações nas formas de registrar os valores existentes e que ingressam na instituição. Com isto, parte dos recursos que eram direcionados a uma rubrica passaram a ser registrados em outra rubrica já existente ou em uma nova conta. Ou seja, estes remanejamentos foram responsáveis por inviabilizar o uso das séries históricas anteriores a esta data.

Outro problema existente relaciona-se com as diretrizes estratégicas da instituição. Dependendo do momento e da conjuntura econômica o banco canaliza seus esforços para determinados produtos, aumentando a sua atratividade. Com isto, valores históricos que eram aplicados em determinado segmento passam a ser direcionados para outro. Estas mudanças nem sempre se caracterizam por ciclos ou sazonalidades bem definidos sendo, geralmente, acontecimentos sem nenhuma correlação com o passado.

1.5 OBJETIVOS

Na busca de uma representação matemática que reflita o comportamento das variáveis financeiras dos bancos, diversas heurísticas vêm sendo pesquisadas e aplicadas. Dentre todas, os modelos estatísticos clássicos são, talvez, os mais difundidos e utilizados no tratamento de séries temporais financeiras. Técnicas de amortecimento exponencial, redes bayesianas e aqueles pertencentes à família dos modelos ARIMA (*AutoRegressive Integrated Moving Average*) são as mais difundidas. Entretanto, segundo Morettin (2008), estes modelos podem não ser apropriados para

descreverem tais séries financeiras por apresentarem características particulares que séries de outras áreas podem não ter, tais como a volatilidade. Além disso, por se tratarem de modelos lineares empregados em séries cuja variância não é constante para todos os intervalos de tempo, os resultados nem sempre são adequados aos propósitos.

Nos últimos anos, uma série de fatores tem permitido um avanço considerável da utilização de técnicas matemáticas e computacionais para a criação de modelos a partir de bases de dados. Muitas destas técnicas, apesar de conceitualmente apresentadas há décadas, como as redes neurais artificiais (RNA), não podiam ser utilizadas de maneira efetiva, tendo em vista não haver capacidade computacional para tanto. Atualmente, porém, mesmo os modelos financeiros que exigem forte desempenho computacional como os não-lineares podem ser explorados com as RNA de maneira mais fácil (ZANETI JR; ALMEIDA, 2004).

As redes neurais artificiais (RNA) têm se mostrado muito eficazes na previsão de situações baseadas em séries históricas. As RNA podem ser tratadas como aproximadores de funções matemáticas complexas e, nestes casos, alcançam, geralmente, um desempenho superior aos modelos estatísticos convencionais (ZHANG; PATUWO; HU, 1988). Aliado a isto, as RNA ganham atenção pela sua habilidade de aprendizado, capacidade de generalização e de lidar com dados não lineares. Estas qualidades as tornam capazes de identificar e assimilar as características mais marcantes das séries, tais como sazonalidade, periodicidade, tendência, entre outras, tornando-as eficientes na representação dos dados (ABELÉM, 1994).

O principal objetivo deste trabalho é estudar as séries financeiras integrantes do fluxo de caixa do banco, visando conceber um modelo computacional que forneça previsões para as séries mais importantes em um horizonte de $d + 2$, onde d corresponde ao dia corrente, para servir de apoio no planejamento e gerenciamento das reservas financeiras e liquidez da instituição.

O trabalho de pesquisa envolveu estudos no universo das RNA, avaliações e análises de séries temporais e aplicações de modelos computacionais na previsão de séries financeiras.

A etapa de pré-processamento dos dados também será foco deste trabalho, pois é de fundamental importância para o sucesso da pesquisa. O estudo dos dados, seu comportamento, a influência de feriados nos dias que os antecedem e os precedem, a

ausência de valores na série, a relação entre dias de semanas equivalentes assim como dias de meses ou meses do ano serão alguns dos tópicos analisados e avaliados.

Para esta pesquisa foi utilizado, de início, um arquivo com 764 registros, referente ao período de 01/05/2002 a 31/04/2005 (36 meses) de todas as séries históricas envolvidas no fluxo de caixa (157 variáveis). Posteriormente, agregou-se registros até 31/04/2007, perfazendo um total de 59 meses. Com a inserção destes dados optou-se pelo descarte dos valores até 31/03/2003, em função dos reflexos oriundos das mudanças decorrentes da reestruturação do SPB. Dessa forma, ao término da pesquisa existiam 36 meses de dados para treinamento e validação dos modelos, relativos ao períodos de 01/04/2003 a a 31/03/2006, e 12 meses para teste, referentes ao período de 01/04/2006 a 31/03/2007.

1.6 ORGANIZAÇÃO

Esta tese trata no capítulo 2 sobre a base de dados. Assim, no seu transcorrer, são definidas e conceituadas as variáveis integrantes do fluxo de reservas bancárias para, em seguida, apresentar a metodologia dos procedimentos adotados na redução da dimensionalidade. Posteriormente tem-se as abordagens relativas ao tratamento dos valores inexistentes, normalização dos dados e decomposição da série. Encerrando o capítulo faz-se uma análise dos resultados e formula-se as hipóteses.

O capítulo 3 se dedica ao estudo da aleatoriedade da série. Nele, encontram-se estudos e análises das diversas técnicas utilizadas. Os resultados dos testes e as conclusões, encerram o capítulo.

A abordagem neural para previsão está inserida no capítulo 4. Nele, além da revisão bibliográfica, encontram-se vários procedimentos adotados e modelos implementados, visando a melhoria da performance de acerto dos modelos desenvolvidos.

O capítulo 5 trata dos modelos neurais alternativos que foram desenvolvidos para resolução do problema. A utilização de variáveis externas assim como variáveis auxiliares integrantes do fluxo de reserva bancária foram avaliadas e utilizadas nos modelos. A agregação de variáveis também fez parte dos estudos. No fim do capítulo tem-se uma análise a respeito dos resultados obtidos não só com arquiteturas neurais distintas mas também com as diversas opções de variáveis utilizadas. Estudos sobre o

pré-tratamento dos dados também é tratado neste capítulo.

A utilização de outras abordagens clássicas e não-clássicas são mostradas no capítulo 6. Modelos econométricos e difusos foram testados e avaliados neste capítulo.

As conclusões do trabalho, assim como as propostas para continuidade e melhoria desta tese encontram-se no capítulo 7.

2 BASE DE DADOS

2.1 RESERVAS BANCÁRIAS

A função básica do sistema de pagamentos é transferir recursos, bem como processar e liquidar pagamentos para pessoas, empresas, governo, Banco Central e instituições financeiras. O cliente bancário utiliza-se do sistema de pagamentos toda vez que emite cheques, faz compras com cartão de débito e de crédito ou, ainda, quando envia um documento de crédito (DOC) ou faz uma transferência eletrônica – TED.

Os pagamentos que transitam pelo sistema STR (Sistema de Transferência de Reservas) observam os princípios de irrevogabilidade e incondicionalidade. O STR acata exclusivamente ordens de crédito, e a informação pertinente ao pagamento somente é transmitida ao creditado quando da validação do lançamento na conta reservas bancárias. Essa validação só ocorre se o saldo disponível na conta da instituição que comandar a ordem tiver valor suficiente.

A reestruturação do SPB alterou de maneira significativa o tratamento da conta reservas bancárias, já que o seu saldo agora é monitorado pelo BACEN em tempo real, não permitindo saldo negativo na conta em nenhum momento do dia. O monitoramento *on-line* da conta reservas bancárias, impactando a conta no momento em que ocorrem as transações, requer dos bancos um esforço maior para gerir o fluxo de pagamentos e recebimentos (fluxo de caixa).

Os bancos comerciais estão subordinados a políticas e diretrizes do Banco Central do Brasil e, como tal, devem seguir estritamente os preceitos definidos por aquela Autarquia. Assim, todas as variáveis descritas neste trabalho são, antes de mais nada, fruto de normas definidas pelo BACEN. Cabe às instituições apropriarem corretamente os valores transacionados nas respectivas rubricas e provisionar os percentuais definidos em lei.

2.1.1 Variáveis do Fluxo de Reservas Bancárias

A gestão do fluxo de reservas bancárias envolve a projeção e acompanhamento da movimentação dos lançamentos a débito e a crédito na rubrica reservas bancárias, existente junto ao BACEN.

As variáveis acompanhadas pelo banco e que sensibilizam a conta reservas bancárias estão relacionadas a seguir, onde se faz uma breve descrição conceitual e comentam-se as suas principais características.

Numerário

Destinada a registrar operações relacionadas com movimentação de numerário com o Banco Central do Brasil ou com o custodiante, atendendo às necessidades das agências bancárias no tocante ao meio circulante, além de registrar as tarifas de prestação de serviços.

Compensação

Destinada a registrar os resultados dos valores transitados no sistema de compensação noturna e diurna, incluindo o pré-depósito, seu complemento e/ou devolução.

Após a reestruturação do SPB a maioria dos lançamentos de valores acima de R\$ 5.000,00 (cinco mil reais) e que transitavam nesta rubrica foram migradas para a CIP ou STR.

Arrecadação via Secretaria do Tesouro Nacional – STN e Sistema de Transferência de Reservas – STR

Destinada a registrar as transferências dos valores referentes a tributos federais, estaduais, municipais e a contribuição para o INSS.

O banco diariamente recebe tributos, taxas e contribuições devidos pelos seus contribuintes e transfere e/ou recebe recursos oriundos dessas arrecadações de/para outras instituições e o Tesouro Nacional.

Ordens Bancárias

Destinada a receber os valores referentes às emissões de ordens bancárias

que todos os órgãos do governo federal remetem ao banco. É através deste documento, também, que o banco envia para o Tesouro Nacional os recursos que são depositados por terceiros para encaminhamento a qualquer órgão do Governo Federal.

Câmbio

Destinada a registrar os valores referentes às operações realizadas no mercado interbancário.

Compulsório

Destinada a registrar as previsões e movimentações das contas de exigibilidades compulsórias em suas diversas modalidades, incluindo os recursos compulsórios de terceiros, cujos valores são lançados na conta reservas bancárias.

Sistema Especial de Liquidação e Custódia – SELIC

Destinada a registrar operações com títulos públicos permitidos pelo SELIC. É através das operações no SELIC que o banco vai diariamente a mercado para nivelar sua posição (tomando ou doando recursos).

Central de Custódia e Liquidação Financeira de Títulos – CETIP

Destinada a registrar as operações com títulos custodiados nesta câmara. A CETIP atua como câmara de compensação e de liquidação das operações com títulos públicos e privados registrados em seus sistemas.

A CETIP é uma empresa criada pelas instituições financeiras para realizar a custódia eletrônica e a negociação de uma ampla gama de produtos, incluindo títulos privados, públicos federais, estaduais, municipais.

Tesouro Nacional – Outras

Destinada a registrar os recursos do próprio Tesouro Nacional e de seus diversos órgãos, principalmente a Secretaria da Receita Federal, Instituto Nacional de Seguridade Social – INSS e Fundo de Amparo ao Trabalhador – FAT.

Tecnologia Bancária – TECBAN

Destinada a registrar as operações realizadas no ambiente TECBAN cuja liquidação ocorrerá pela compensação das transações.

Lançamentos BACEN – SLB

Destinada a registrar o recebimento das cobranças geradas pelo Banco Central, tais como multas, tarifas e custos, folha de pagamentos do BACEN e FGC-Fundo Garantidor de Crédito. Além disso, é através deste sistema que se transfere, diariamente, o saldo da conta onde são lançados os valores de responsabilidade do BACEN bem como as operações de câmbio realizadas com o BACEN.

Redesconto

Destinada a registrar as operações de redesconto do banco. Os valores oriundos do redesconto servem para fazer face a desequilíbrios momentâneos provocados pelo descasamento do fluxo de caixa da instituição financeira. Assim, não há necessidade de se projetar seus valores, pois é o somatório das demais variáveis que vai impor ou não a utilização desta linha de crédito.

Bolsa de Mercadorias e de Futuros – Derivativos

Destinada a registrar as operações cursadas nessa câmara de compensação, liquidação e custódia. A BM&F é contraparte em todos os negócios realizados na Bolsa, garantindo a liquidação de todas as operações, tanto dos negócios realizados no pregão como no sistema eletrônico. A liquidação financeira é realizada pelo valor líquido na conta reservas bancárias do banco liquidante indicado por cada membro da compensação.

Companhia Brasileira de Liquidação e Custódia – CBLC

Destinada a registrar os serviços de compensação e liquidação física e financeira de operações realizadas nos mercados à vista e a prazo da BOVESPA e de outros mercados, bem como a operacionalização dos sistemas de custódia de títulos e valores mobiliários em geral.

Bolsa Mercantil e de Futuros – Ativos

Destinada a registrar as operações de mercado secundário realizadas nos sistemas da CETIP e liquidadas na modalidade diferida pelo valor líquido, tanto com títulos públicos e privados custodiados na CETIP, quanto com títulos públicos custodiados no SELIC.

Sistema de Transferência de Reservas – Instituições Financeiras

Destinada a registrar as transferências de fundos entre instituições financeiras referentes às operações de sua responsabilidade e de terceiros.

Sistema de Transferência de Reservas – Clientes

Destinada a registrar as movimentações realizadas por seus clientes no decorrer do dia. Essas movimentações podem ser divididas em duas categorias:

- Eventos de interesse do cliente, como a transferência da conta corrente de um cliente para a conta corrente de outro cliente ou do próprio em outra instituição; e
- Eventos de transferência para devolução de ordens recebidas indevidamente ou não identificadas.

Câmara Interbancária de Pagamentos – CIP

Esta variável é destinada ao registro, compensação e liquidação eletrônica das transferências interbancárias de fundos.

2.1.2 Classificação das Variáveis do Fluxo de Reservas Bancárias

Parkinson e Kallberg (1993) estabelecem uma distinção entre a programação e a previsão de caixa. Programação de caixa é um termo empregado para representar as projeções dos fluxos de caixa de curto prazo, onde muito dos fluxos são conhecidos ou podem ser estimados facilmente. Já previsão de caixa está associada a um período de tempo mais longo. Os autores sugerem a seguinte classificação:

- **Fluxos Certos:** são fluxos de caixa sem necessidade de serem estimados porque são conhecidos antecipadamente, como os pagamentos de

impostos que são, frequentemente, estabelecidos antes da data de pagamento.

- **Fluxos Previsíveis:** são fluxos que não são certos mas podem ser estimados com razoável previsão, baseando-se em outra variável, como os recibos de pagamentos de clientes regulares, que podem ser estimados dentro de períodos limitados.
- **Fluxos Não Previsíveis:** são os fluxos de caixa que podem permanecer incertos sem condições de uma previsão com sucesso, como depósitos de clientes ou desembolsos não planejados.

O quadro abaixo apresenta os tipos das variáveis e a sua classificação, considerando o fluxo em $d + 1$.

Quadro 1: Variáveis globais de uma instituição financeira e sua classificação.

01. Numerário	FP
02. Compensação	FNP
03. Arrecadação via STN / STR	FP
04. Ordens Bancárias	FP
05. Câmbio	FC
06. Compulsório	FC
07. SELIC	FC
08. CETIP	FNP
09. Tesouro Nacional - Outras	FP
10. TECBAN	FP
11. Lançamentos BACEN – SLB	FNP
12. Redesconto	FC
13. BMF - Derivativos	FC
14. CBLC	FC
15. BMF - Ativos	FC
16. STR – IF	FNP
17. STR – Clientes	FNP
18. CIP	FNP

Todas as variáveis classificadas como de Fluxo Previsível – FP e Fluxo Não Previsível – FNP são projetadas pela instituição a fim de se conhecer a movimentação futura e poder tomar as medidas estratégicas cabíveis.

Dentre as variáveis envolvidas a 17. STR – Clientes e a 18. CIP são as que possuem tratamento mais complexo pois, além de serem de valores vultosos, correspondendo a aproximadamente 45% do movimento total, são as variáveis que apresentam maiores oscilações na sua movimentação, pois representam o

comportamento do público com o banco. Além disso, estas variáveis respondem por mais de 70% dos erros de previsão realizados pelo banco.

2.1.3 Gestão do Fluxo de Reservas Bancárias

Para efeito de gestão do fluxo, o banco subdividiu as 18 variáveis globais descritas anteriormente em 157 subitens. Estes itens e subitens, dadas as suas características, são extremamente dinâmicos, alterando-se suas posições periodicamente, ou seja, agregando-se ou separando-se das variáveis globais dependendo de fatores internos (para melhoria do gerenciamento ou operacionalização da variável) ou de fatores externos (determinação do BACEN). Para efeito de padronização, doravante será referenciado como variável ou item os valores globais; e como subitem os valores componentes de uma variável. Por exemplo: a variável ou item Numerário é composta pelos subitens Depósito, Saque e Custodiante – crédito de tarifa.

No Anexo 1 é apresentada a relação de todas as variáveis e subitens utilizados pelo banco no gerenciamento do seu fluxo de caixa assim como a correspondente classificação do tipo de fluxo.

2.2 TRATAMENTO DA DIMENSIONALIDADE

Em virtude da existência de grande quantidade de variáveis na movimentação do fluxo de caixa, para efeito de modelagem dos dados algumas decisões podem ser adotadas, visando a redução do universo de parâmetros envolvidos.

2.2.1 Redução da Dimensionalidade - 1ª Fase

A primeira etapa de redução da dimensionalidade consistiu na criação de uma metodologia com o objetivo de suprimir, condensar ou simplesmente excluir da base itens e subitens não necessários ao estudo. Para tanto, foram adotadas as seguintes regras:

- Regra 01 – Itens ou subitens criados por determinação do BACEN e/ou para facilitar a operacionalização do Fluxo de Caixa, mas não movimentados, foram excluídos da base.
- Regra 02 – Itens ou subitens considerados de Fluxo Certo – FC foram excluídos da base, pois os valores envolvidos são conhecidos com

antecedência, não havendo necessidade de se fazer previsões. Entretanto, para não distorcer os totais da movimentação do Fluxo de Caixa, um único somatório dos subitens foi mantido na base, denominado Agregação – FC. No caso em que todo o item (variável) é formado por subitens de natureza FC, somente o item permaneceu, sendo os demais campos excluídos.

- Regra 03 – Significativa parte das variáveis componentes do Fluxo de Caixa são consideradas como sendo de Fluxo Previsível, ou seja, apesar de não se poder afirmar quais os valores exatos destes subitens ainda assim, baseado em informações internas e/ou externas, os analistas conseguem obter índices de acertos nas projeções próximos de 100%. Dessa forma, a permanência destes subitens na base servirão tão-somente para se analisar as correlações e/ou influências com os demais valores não havendo necessidade de se criar modelos para sua projeção. Baseado no exposto acima, o item que possuir mais de um subitem enquadrado como Fluxo Previsível terá seus valores somados em um outro campo denominado Agregação – FP com os dados originais excluídos.

No Anexo 2 tem-se quadro ilustrativo da aplicação das regras no conjunto de variáveis.

Após esta primeira etapa de tratamento, o universo de variáveis descrito no Anexo 1, passou a ser representado pelas variáveis descritas no Anexo 3.

2.2.2 Redução da Dimensionalidade - 2ª Fase

Apesar de o conjunto inicial de variáveis haver sido reduzido em aproximadamente 60%, passando de 157 para 66, ainda assim o universo resultante encontra-se com uma dimensionalidade elevada.

Verificou-se que a forma de o banco trabalhar permite que se adotem algumas soluções visando uma maior redução do escopo e, conseqüentemente, facilidade na manipulação dos dados. Como exemplo, tem-se o caso da variável 01 – Numerário, onde os subitens integrantes desta variável são classificados como de Fluxo Certo e Fluxo Previsível. Assim, uma agregação destes itens não afetará a análise, uma

vez que a unificação de Fluxos Certos e Fluxos Previsíveis, para efeito do Banco, não prejudica a análise pois as projeções das variáveis enquadradas como Previsíveis são consideradas muito boas.

Outra decisão assumida se refere às variáveis que possuem desmembrados os valores a débito e a crédito, como, por exemplo, a rubrica Tesouro Nacional – Outras e Sistemas de Transferências de Reservas – Clientes. Nestes casos, optou-se em unificá-los em uma única variável desconsiderando os valores de débito e crédito mas, tão-somente, a soma aritmética de ambos.

Por fim, os campos PIS/PASEP e Contribuição Sindical individualmente são considerados como de Fluxo Não Previsível. Entretanto, quando agregados com os demais campos possuem características de Fluxo Certo, ou seja, não se tem a previsão individual destas colunas no sistema de informações do banco, mas o valor total é apurado pelos analistas com precisão permitindo que se agrupe em um único item.

Concluídos os estudos e tratamento das variáveis, as variáveis descritas no Anexo 3 passaram a ser representadas e classificadas conforme o Anexo 4.

2.3 TRATAMENTO DOS DADOS

Os estudos realizados para um melhor conhecimento dos dados, tais como: o cálculo das médias, desvios padrões, correlações, frequências, análises gráficas etc, estão apresentados a seguir. A fim de facilitar a manipulação dos dados os valores foram inicialmente divididos por 1.000.000.

Quadro 2: Valores estatísticos das variáveis.

	Valores em R\$ milhões					
	Média	Mediana	Mínimos	Máximos	Variância	Desvio
01. Numerário	-166,91	-140,45	-1.376,39	600,74	58.658,38	242,19
02. Compensação	3,58	-7,96	-903,21	2.021,50	50.133,03	223,90
03. Arrecadação STN STR	-556,02	-278,04	-4.082,46	5,47	481.513,66	693,91
04. Ordens Bancárias	442,72	237,94	-929,24	5.152,37	496.809,76	704,85
05. Câmbio	-10,81	0,03	-1.016,81	777,72	40.934,66	202,32
06. Compulsório	-3,90	5,13	-2.932,50	3.005,58	30.457,31	174,52
07. SELIC	8,37	67,24	-6.597,00	6.378,11	2.418.562,74	1.555,17
08. CETIP	-20,69	-2,09	-1.411,88	1.497,00	18.382,96	135,58
09. TN - Outras	379,26	148,17	-976,74	6.670,11	618.420,64	786,40
10. TECBAN	0,71	0,53	0,00	2,57	0,19	0,44
11. Lançamentos BACEN	-3,80	0,84	-1.336,94	343,28	4.658,76	68,26
14. CBLC	4,00	0,55	-1.386,78	1.159,91	6.835,02	82,67
16. STR – IF	-7,24	-13,29	-1.230,55	922,00	12.362,40	111,19
17. STR - Clientes	157,38	61,58	-2.014,56	4.256,37	305.690,18	552,89
18. CIP	-222,16	-114,52	-1.334,75	1.145,66	92.817,19	304,66

No quadro 2 observa-se uma significativa variância na maioria das variáveis, demonstrando grande dispersão destes dados.

Nas análises realizadas, foi detectada a presença de valores extremamente altos/baixos o que pode vir a dificultar a modelagem dos dados. Estes registros não são ruídos e sim uma característica da base de dados. Todos estes pontos foram analisados individualmente e correspondem, efetivamente, ao movimento da variável.

Em uma primeira análise, a melhor forma de se estudar os dados ainda é a visual (CHATFIELD, 2004). Avaliam-se, através de gráficos, a qualidade e características dos dados, tais como a existência de tendências, ciclos e sua distribuição no espaço, dentre outras. Assim, todas as variáveis classificadas como de fluxo não previsível tiveram seus valores apresentados em gráficos e, em seguida, mapeadas suas frequências com vistas a detecção e tratamento dos valores discrepantes, acaso existentes.

Variável Compensação

No gráfico da variável Compensação nota-se a existência de diversos pontos acima da faixa de R\$ 600 milhões. A grande maioria dos dados oscila próximo a zero e não se observam tendências, ciclos ou sazonalidade aparentes. O primeiro ano da série apresentou maior dispersão dos dados em relação aos demais anos. Tal fato decorre da reestruturação do SPB que gerou muitas alterações.

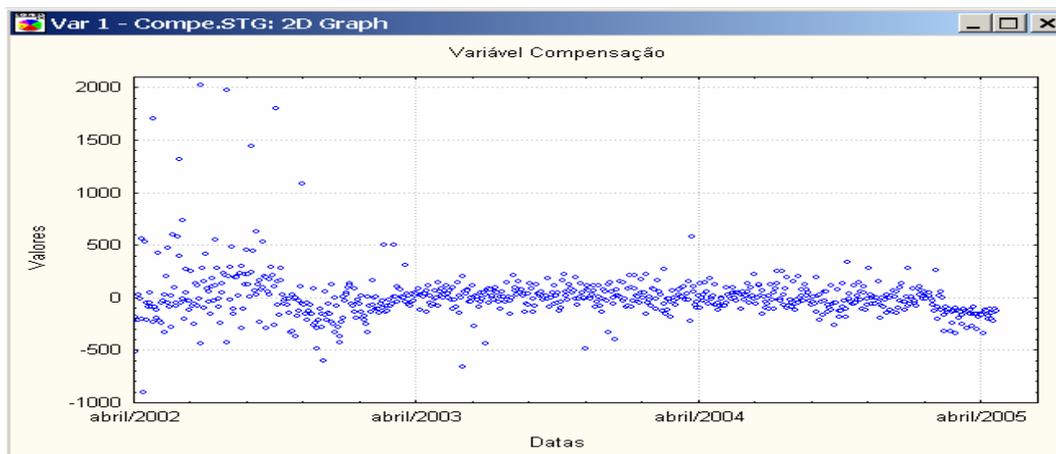


Gráfico 1: Variável Compensação.

Os intervalos de frequência da variável Compensação podem ser vistos no quadro abaixo.

Quadro 3: Análise de frequência da variável Compensação.

Compensação			
R\$ Milhões	Qtd	Percentual	Percentual acumulado
$- 1.000 < X \leq - 500$	4	0,5236	0,5236
$- 500 < X \leq 0$	406	53,1414	53,6649
$0 < X \leq 500$	338	44,2408	97,9058
$500 < X \leq 1.000$	9	1,1780	99,0838
$1.000 < X \leq 1.500$	3	0,3927	99,4764
$1.500 < X \leq 2.000$	3	0,3927	99,8691
$2.000 < X \leq 2.500$	1	0,1309	100,0000

Verifica-se que os valores situados no intervalo $(-1.000, -500]$ se resumem a apenas 4 registros, correspondendo a 0,52% da base e os existentes no intervalo $(500, 2.500]$ representam 2,09%. Ou seja, 97,39% dos registros estão representados no intervalo $(-500, 500]$.

Apesar de discrepantes em relação aos demais registros, estes valores não são ruídos e sim característica da base de dados. Assim, a retirada destes pontos distorceria a série em função do vazio que ficaria no dia. Além disso, há de se considerar que o valor referente à data em que ocorreu o extremo também deve ser previsto, ou seja, com a retirada deste ponto dificulta-se a previsão para aquele dia.

Várias técnicas podem ser aplicadas ao problema, tais como o uso de interpolação, nivelamento para o intervalo mais significativo ou adoção do teto de média ± 2 desvios padrões, dentre outras. Optou-se nesta tese em se trazer os dados para os intervalos mais representativos, ou seja, os valores abaixo de $-R\$ 500$ e acima de $R\$ 500$ foram substituídos por estes patamares. Este processo tenta fazer a rede dedicar os esforços de aproximação na região onde se concentra a maioria dos dados. Em compensação, os valores extremos são abandonados e sua ocorrência deve ser prevista com alguma técnica complementar.

Variável CETIP

Menos dispersa que a variável Compensação, a variável CETIP apresenta a grande maioria do seu movimento bem próximo de zero. Também, neste caso, não são observadas tendências, ciclos ou sazonalidades nos dados.

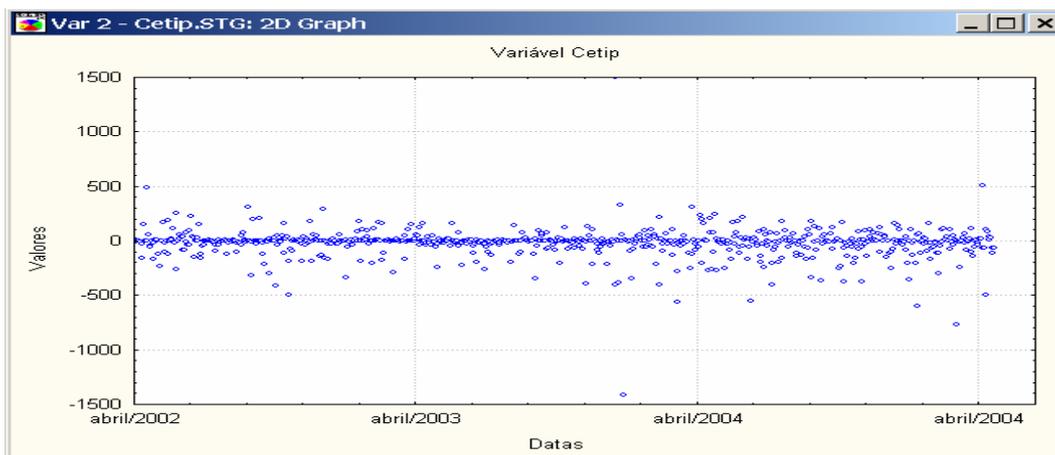


Gráfico 2: Variável CETIP.

A variável CETIP apresentou os intervalos descritos adiante e, após análise, concluímos que tratamento similar à variável Compensação deve ser adotado. Ou seja, os valores abaixo de -R\$ 500 e acima de R\$ 500,00 serão substituídos por estes patamares.

Quadro 4: Análise de frequência da variável CETIP.

CETIP			
R\$ Milhões	Qtd	Percentual	Percentual acumulado
$- 1.500 < X \leq - 1.000$	1	0,1309	0,1309
$- 1.000 < X \leq - 500$	4	0,5236	0,6545
$- 500 < X \leq 0$	416	54,4503	55,1047
$0 < X \leq 500$	341	44,6335	99,7382
$500 < X \leq 1.000$	1	0,1309	99,8691
$1.000 < X \leq 1.500$	1	0,1309	100,0000

Variável Lançamentos BACEN – SLB

A variável SLB também se mostra bem comportada, com característica quase linear. O seu desvio padrão é baixo e sua média próxima de zero.

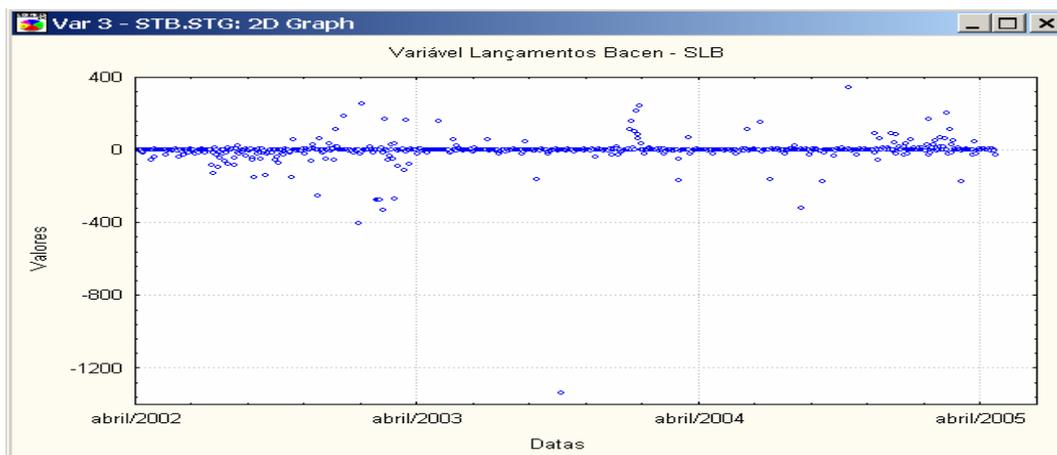


Gráfico 3: Variável Lançamentos BACEN – SLB.

Tratamento idêntico ao das variáveis anteriores será adotado aqui, trazendo os valores situados nos intervalos inferiores a – R\$ 400 e superiores a R\$ 200 para estes patamares.

Quadro 5: Análise de frequência da variável Lançamentos do BACEN.

BACEN - SLB			
R\$ Milhões	Qtd	Percentual	Percentual acumulado
- 1.400 < X <= - 1.200	1	0,1309	0,1309
- 1.200 < X <= -1.000	0	0,0000	0,1309
- 1.000 < X <= - 800	0	0,0000	0,1309
- 800 < X <= - 600	0	0,0000	0,1309
- 600 < X <= - 400	1	0,1309	0,2618
- 400 < X <= - 200	7	0,9162	1,1780
- 200 < X <= 0	289	37,8272	39,0052
0 < X <= 200	461	60,3403	99,3456
200 < X <= 400	5	0,6545	100,0000

Variável STR – IF

Esta variável possui alguma dispersão no início e fim da série. Os dados foram checados, mas os analistas do banco não souberam explicar o fenômeno. Retirando esta característica, o restante da série está bem comportado.

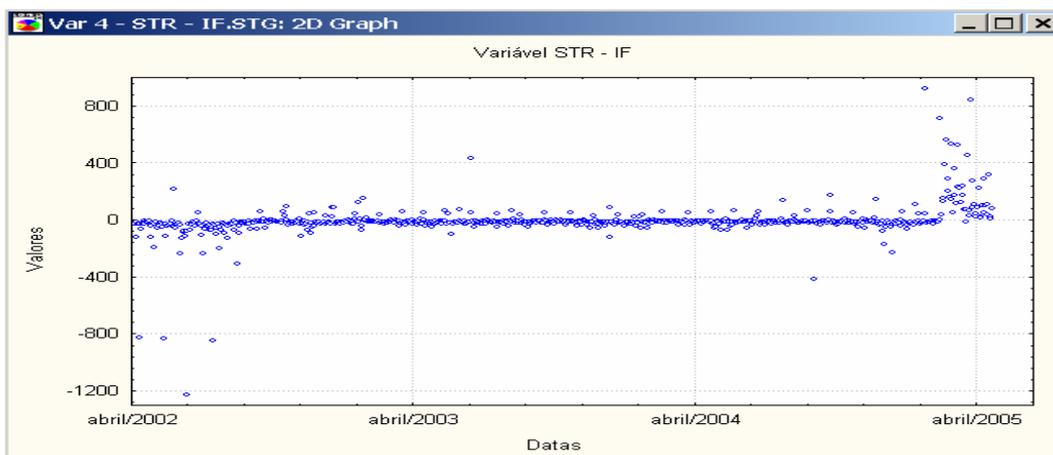


Gráfico 4: Variável STR – IF.

Nesta variável serão agregados os valores existentes nos intervalos inferiores a – R\$ 500 e superior a R\$ 500.

Quadro 6: Análise de frequência da variável STR – IF.

STR - IF			
R\$ Milhões	Qtd	Percentual	Percentual acumulado
$- 1.500 < X \leq - 1.000$	1	0,1309	0,1309
$- 1.000 < X \leq - 500$	3	0,3927	0,5236
$- 500 < X \leq 0$	635	83,1152	83,6387
$0 < X \leq 500$	119	15,5759	99,2147
$500 < X \leq 1.000$	6	0,7853	100,0000

Variável STR – Clientes

Esta variável é considerada pelo banco, juntamente com a CIP, como as de mais difíceis soluções pois todos os modelos de previsão testados não se mostraram satisfatórios. Ela possui uma elevada variância e uma leve tendência positiva no movimento.

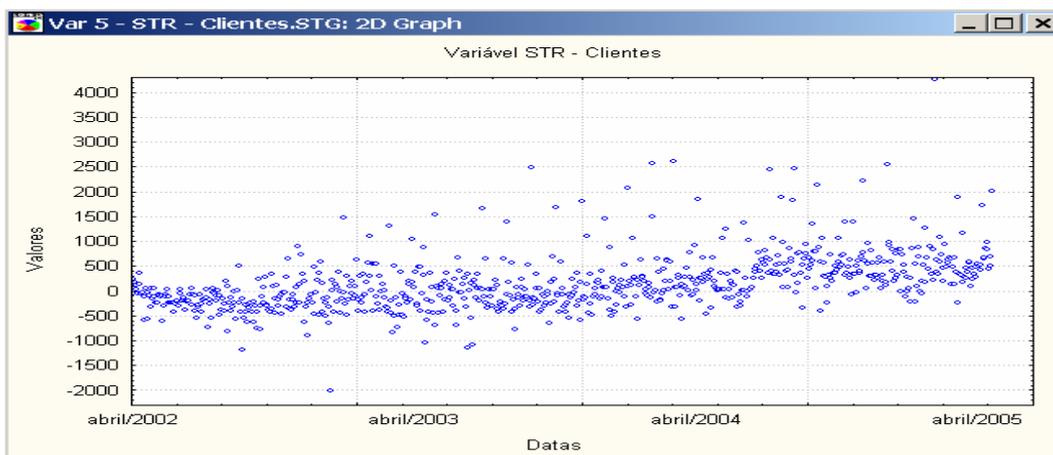


Gráfico 5: Variável STR – Clientes.

Valores inferiores a – R\$ 1.000 e superiores a R\$ 3.000 serão agregados nestes intervalos.

Quadro 7: Análise de frequência da variável STR – Clientes.

STR -Clientes			
R\$ Milhões	Qtd	Percentual	Percentual acumulado
$- 3.000 < X \leq - 2.000$	1	0,1309	0,1309
$- 2.000 < X \leq -1.000$	4	0,5236	0,6545
$- 1.000 < X \leq 0$	325	42,5393	43,1937
$0 < X \leq 1.000$	391	51,1780	94,3717
$1.000 < X \leq 2.000$	32	4,1885	98,5602
$2.000 < X \leq 3.000$	10	1,3089	99,8691
$3.000 < X \leq 4.000$	0	0,0000	99,8691
$4.000 < X \leq 5.000$	1	0,1309	100,0000

Variável CIP

A linearidade existente no primeiro ano da série corresponde à inexistência de movimento. A Câmara Interbancária de Pagamentos só começou a funcionar em dezembro de 2002. A partir daí, vê-se uma grande dispersão dos dados bem como alguma sazonalidade. Esta variável em conjunto com a anterior – STR-Clientes, são as mais importantes para o Banco, não só em função dos vultosos valores movimentados como em relação à dificuldade em realizar previsões corretas.

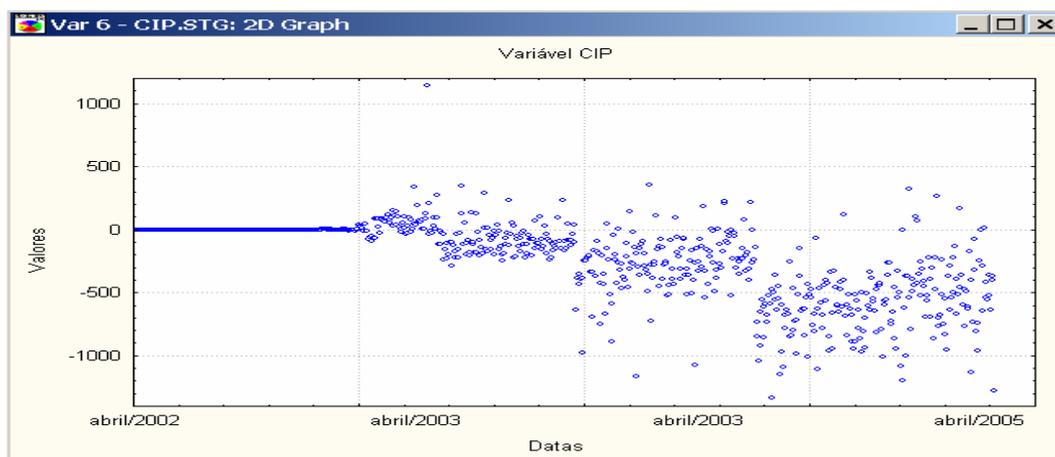


Gráfico 6: Variável CIP.

Nesta variável optou-se por agregar aos intervalos mais próximos somente os valores superiores a R\$ 500.

Quadro 8: Análise de frequência da variável CIP.

CIP			
R\$ Milhões	Qtd	Percentual	Percentual acumulado
- 1.500 < X <= - 1.000	14	1,8325	1,8325
- 1.000 < X <= - 500	137	17,9319	19,7644
- 500 < X <= 0	503	65,8377	85,6021
0 < X <= 500	109	14,2670	99,8691
500 < X <= 1.000	0	0,0000	99,8691
1.000 < X <= 1.500	1	0,1309	100,0000

2.3.1 Valores Inexistentes

Apesar dos problemas decorrentes da falta de dados atingir a maioria das bases existentes, no caso desta pesquisa os casos de inexistência de dados ocorreram somente nos dias referentes a feriados nacionais, em que não há movimento bancário. Os demais dias não apresentam este tipo de problema, pois os dados do movimento do fluxo de caixa são tratados com extrema cautela e contabilizados em rubricas apropriadas, evitando-se, assim, a ocorrência de erros ou mesmo ausência de valores.

Nos casos de feriados, o movimento dos dias que os antecedem e os sucedem são totalmente atípicos. Feriados nas terças e quintas-feiras fazem com que os movimentos das quartas-feiras sejam totalmente divergentes daqueles característicos deste dia. Além disso, as segundas e sextas-feiras que ficam entre os feriados também mudam de comportamento, pois grande parte da população tem por hábito “enforçar”

este dia, diminuindo, conseqüentemente, o movimento bancário.

Conforme explanado no parágrafo anterior, a existência de feriados durante o ano descaracteriza a semana. Dessa forma, há necessidade de se avaliar os efeitos decorrentes desta situação, ou seja, se a ausência de valores em determinadas semanas distorce o modelo ou se os seus efeitos são compensados nos dias subsequentes e não interferem na análise. Com base nessa premissa, optou-se, primeiramente, pela realização de análises considerando a série com a ausência dos dias relativos aos feriados e, posteriormente, foram desenvolvidos estudos com a inserção de registros para os dias faltantes.

Para completar os dias faltantes e, assim, minimizar os efeitos decorrentes da existência de feriados, optou-se pela criação de registros para os dias dos feriados com o valor da média aritmética da série. Outras opções, tais como o uso do valor da véspera poderia ter sido utilizado mas não foi tentado aqui por acreditar-se que não modificaria significativamente os resultados.

2.3.2 Normalização dos Dados

De todos os tipos de transformações, a normalização dos dados é a técnica mais utilizada pois, em geral, os dados são muito sensíveis à diferença de escala ou de magnitude (HAIR; ANDERSON *et al.*, 1998). As normalizações geralmente são realizadas de tal forma que os valores fiquem situados na faixa entre 0 e 1 ou -1 e 1. Entretanto, os dados podem ser normalizados para qualquer intervalo definido. Há normalizações lineares e não lineares. O uso de normalização dos dados facilita a comparação entre as variáveis, pois todas estarão na mesma escala. Dessa forma, os efeitos da escala são eliminados da base de dados.

Utilizou-se, neste trabalho, a chamada normalização estatística (modificada) descrita abaixo:

$$v' = \frac{v - \bar{v}}{2\sigma_v} \quad \text{Eq. (2)}$$

Onde:

\bar{v} = média do conjunto

σ_v = desvio-padrão

Após a transformação dos dados, a série passou a ter média 0 e desvio-padrão de 0,5 com a potência da série, calculada pela equação abaixo, igual a $Pot = 0.249681$.

$$Pot = \frac{1}{N} \sum_1^N (x^2) \quad \text{Eq. (3)}$$

Onde:

N = quantidade de registros da série

A utilização da potência como métrica para verificação da redução de informação da série se baseia no fato de que ao se retirar componentes da série o seu valor deverá ser reduzido pois, caso contrário, se estará incluindo ruído.

2.4 DECOMPOSIÇÃO DA SÉRIE

O estudo das séries temporais pode trazer informações importantes sobre o futuro, pois normalmente existe correlação em diversos instantes. É claro que algum grau de incerteza virá agregado, pois o futuro nunca refletirá exatamente as ocorrências passadas mas, mesmo assim, os resultados têm se mostrado eficazes nas análises de previsão.

A análise de uma série temporal passa pela sua decomposição em outras séries mais simples que são funções determinísticas do tempo. O modelo clássico de decomposição considera as séries compostas dos quatro padrões básicos descritos abaixo, sendo que não necessariamente todos os componentes descritos estão presentes em todas as séries. Podem-se encontrar séries com apenas 1, 2 ou 3 deles.

1. Tendência
2. Variações sazonais
3. Variações cíclicas (senoidais)
4. Variações irregulares e não-lineares

Nesta tese, o objetivo da decomposição é retirar as componentes determinísticas existentes e tratar o resíduo (variações irregulares) através da utilização de redes neurais artificiais, supondo-se a existência de formas complexas não lineares na série residual.

2.4.1 Extração da Tendência

Através da análise gráfica da variável STR–Clientes (Gráfico 3), verifica-se a existência de tendência positiva nos dados. Para a retirada da tendência podem ser utilizadas equações logarítmicas, polinomiais ou outras quaisquer. Entretanto, na prática, deve-se utilizar equações simples, haja vista que os valores da tendência também deverão ser previstos o que, em casos de equações complexas, pode representar bem o passado mas não o futuro. Por este motivo, geralmente é utilizada uma equação linear, conforme a representada a seguir:

$$tend(t) = a_0 t + a_1, \quad t = 1, 2, \dots, N \quad \text{Eq. (4)}$$

Onde:

a_0 = cota da reta em $x = 0$

a_1 = coeficiente angular

Os coeficientes utilizados na série STR-Clientes após a inclusão dos dias relativos aos feriados são dados por:

$$tend(t) = 0,000850t - 0,333430, \quad t = 1, 2, \dots, N \quad \text{Eq. (5)}$$

A potência, calculada pela Eq. 2, após a retirada da tendência, passa a ser $Pot = 0.212717$.

A seguir, o gráfico da série normalizado e com representação da reta de tendência.

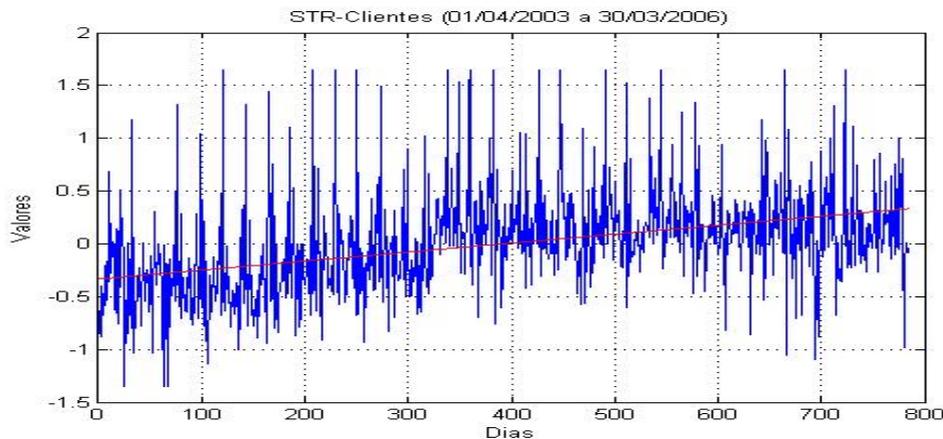


Gráfico 7: Var. STR-Clientes (01/04/2003 a 30/03/2006). Dados com registros nos feriados.

Após a retirada da tendência, a série se apresenta conforme gráfico a seguir.

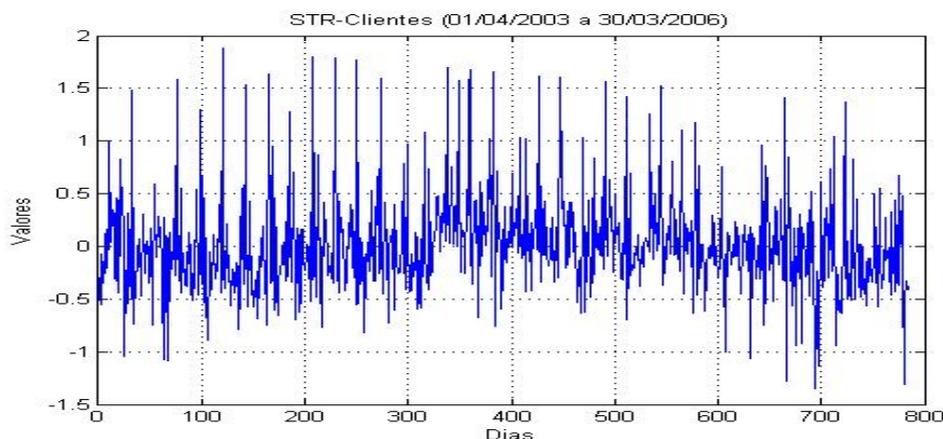


Gráfico 8: STR-Clientes (01/04/2003 a 30/03/2006) depois de retirada da tendência.

2.4.2 Extração da Sazonalidade

É comum em séries temporais a existência de padrões que se repetem a intervalos regulares de tempo. A esta característica dá-se o nome de sazonalidade da série, quando estes períodos são inferiores há um ano, e ciclo para o caso do período ser superior a um ano.

Os períodos da sazonalidade de uma série podem ser determinados pelo conhecimento do comportamento da série. Séries sazonais ocorrem com grande frequência no dia-a-dia como, por exemplo, safra agrícola, consumo de energia, temperatura média mensal, e exportações dentre outras. O estudo das suas características é importante para gerar previsões consistentes.

No caso do movimento financeiro dos clientes, é senso comum imaginar que ele segue o padrão de depósitos no início da semana, caminhando-se para o equilíbrio no meio da semana e saques nos últimos dois dias da semana. Com alguma variação, o comportamento acima tende a se repetir em todas as semanas. Obviamente, outros fatores são responsáveis por modificações na regularidade da série, tais como período do mês em que ocorre o fenômeno, ou seja, se início de mês quando a circulação de dinheiro é maior em função de folhas de pagamentos ou fim de mês quando há uma significativa redução no movimento de numerário; existência ou não de feriado no período; mês do ano em que se insere a semana já que há uma menor atividade econômica nos primeiros meses do ano, refletindo na movimentação do fluxo de caixa da instituição financeira, etc.

Entretanto, apesar de todas estas variações, ainda assim existe uma regularidade semanal que deve ser extraída. Por esta razão, promoveu-se a exclusão da sazonalidade semanal e, para tanto, optou-se pela subtração do valor médio do dia da semana para cada dia, ou seja, para os valores relativos às segundas-feiras foram subtraídos o valor médio de todas as segundas-feiras e o mesmo foi feito para os demais dias da semana. Após esta extração, a potência da série ficou reduzida a $Pot = 0.207513$ e o gráfico da série passou a ter a seguinte característica.

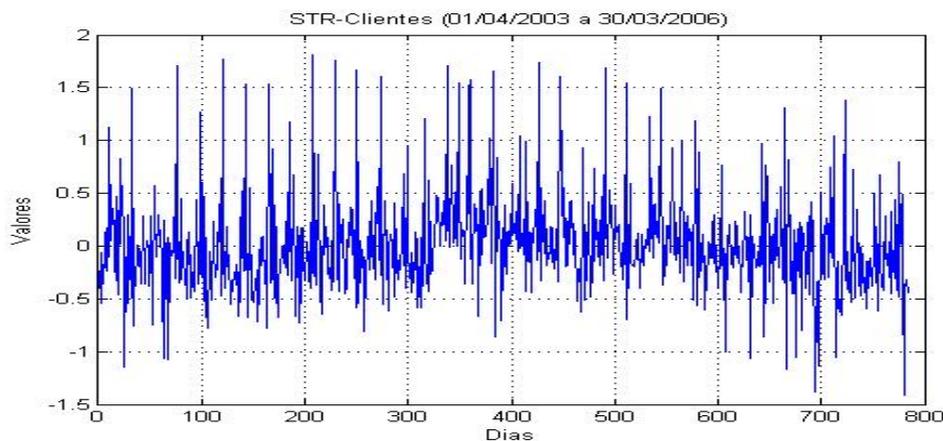


Gráfico 9: STR-Clientes (01/04/2003 a 30/03/2006) sem sazonalidade semanal.

2.4.3 Extração dos Ciclos

Um caso particular de sazonalidade muito frequente na natureza é o ciclo senoidal, em que a sazonalidade tem a forma de uma senóide com período P .

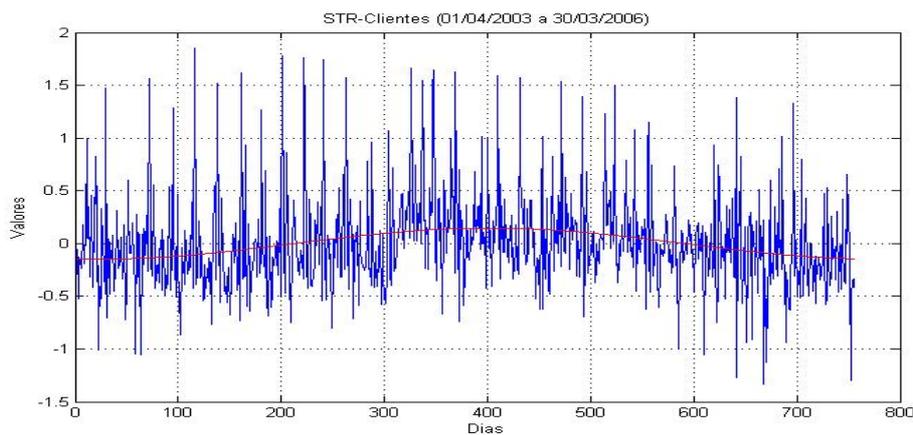


Gráfico 10: STR-Clientes (01/04/2003 a 30/03/2006) com ciclo senoidal.

Vê-se no gráfico acima que uma leve senóide está introduzida na série e, portanto, deve ser retirada por ser determinística. O processo a ser utilizado é a

decomposição através da transformada de Fourier da série, conforme equação abaixo:

$$ciclo(t) = a \cos(2\pi ft) + b \sin(2\pi ft) \quad \text{Eq. (6)}$$

Onde:

a, b = amplitude

f = frequência

A série, transformada para o domínio da frequência, pode ser observada abaixo e indica como as diversas frequências contribuem para a formação da série. Nota-se, claramente, a existência de duas componentes mais significativas na série. A primeira, com frequência $f = 0,001323$, relaciona-se com o período ($P = 1 / f$) $P = 755,86$, caracterizado pelo tamanho total da série e a segunda, equivalente a $0,047862$, corresponde ao período $P = 20,91$ dias. Este último período, de aproximadamente 21 dias úteis, corresponde ao intervalo de tempo equivalente há um mês e é razoável supor a existência de padrões de comportamento repetitivos a cada mês.

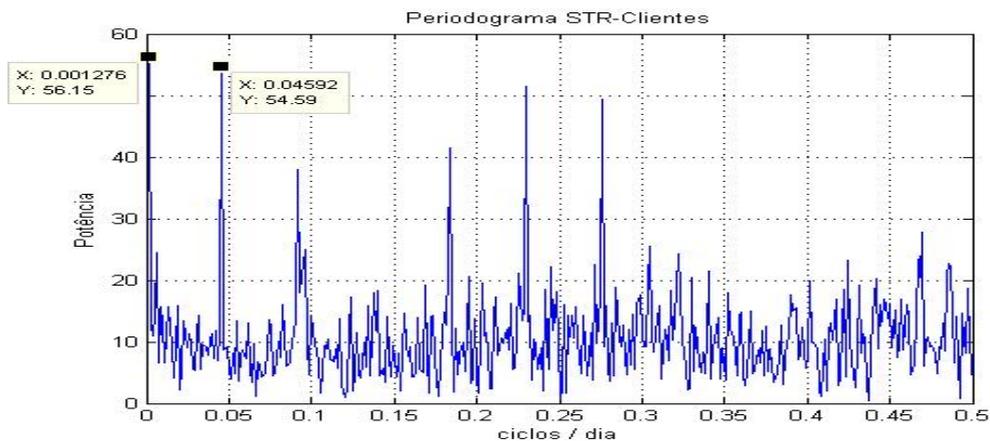


Gráfico 11: STR-Clientes (01/04/2003 a 30/03/2006) com ciclo senoidal.

A partir dos dados acima, para o cálculo dos ciclos C_{s1} e C_{s2} os valores de a e b correspondem a:

$$a_1 = -0.143682 \quad e \quad b_1 = -0.027846$$

$$a_2 = -0.139096 \quad e \quad b_2 = 0.023839$$

Após retirada das duas principais componentes, o periodograma da série passa a apresentar a característica abaixo, com a potência da série valendo $Pot = 0,190715$.

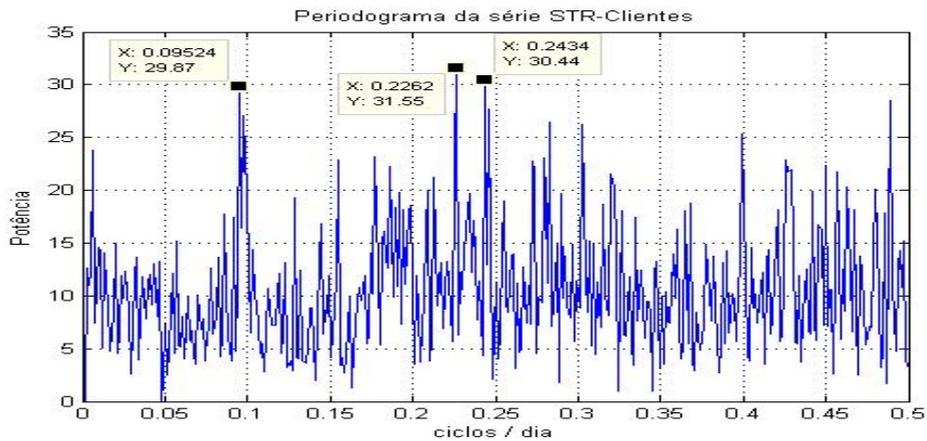


Gráfico 12: Periodograma da série STR-Clientes sem as 2 principais componentes.

Outro processo a ser utilizado na busca da sazonalidade é a autocorrelação, definida no item 3.3. O seu cálculo indica, nos casos em que o coeficiente é alto, que os valores da série tendem a formar seqüências equivalentes de tempos em tempos, produzindo uma ciclicidade, ou seja, a série é caracterizada por seqüências repetitivas em intervalos pré-definidos. Assim, uma amostra onde o coeficiente de autocorrelação é nulo indica que os dados são independentes e identicamente distribuídos.

A autocorrelação da nova série correspondente ao gráfico da figura 12, apresentada abaixo na figura 13, mostrou a existência de significativa correlação com atrasos de 4 e 10 dias. Ao se avaliar o periodograma verifica-se que a componente de frequência $f = 0,09524$ possui um período $P = 10,50$ dias relativo a duas semanas de dias úteis e as outras duas componentes, com frequências $f = 0,2262$ e $f = 0,2434$, são relacionadas a períodos de 4,42 e 4,11 dias, respectivamente.

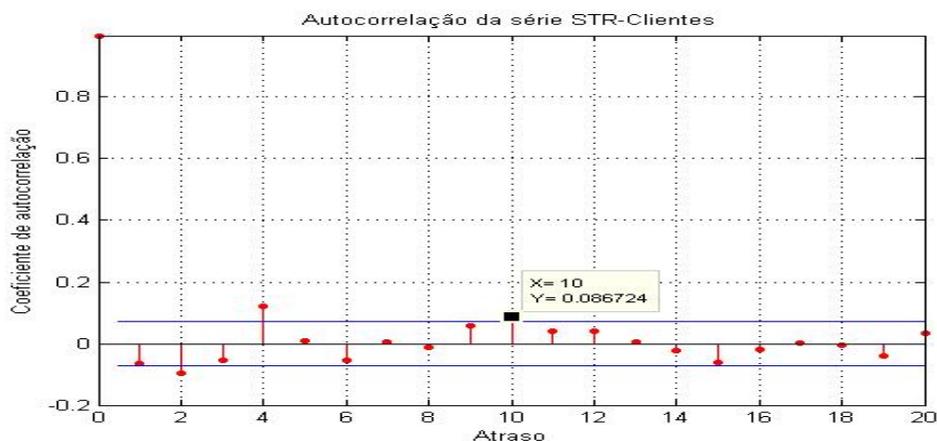


Gráfico 13: Autocorrelograma da série STR-Clientes sem as 2 principais componentes.

Apesar de possuir correlação evidente em atrasos de 4 dias, ainda assim não se justifica com a fenomenologia, pois é sabido que atrasos de 4 dias na série financeira não apresentam correlação justificável. As características das segundas-feiras são diferentes das de sextas-feiras (4 dias de atraso), assim como as das terças-feiras divergem das de segundas-feiras. Por este motivo, optou-se em retirar somente a componente referente à frequência de aproximadamente 10 dias.

Após a retirada do ciclo C_{S_3} , correspondente à um período de 10,50 dias cujas amplitudes valem $a_3 = 0.077932$ e $b_3 = 0.013031$, tem-se o novo autocorrelograma, onde se observa uma pequena redução no índice de correlação.

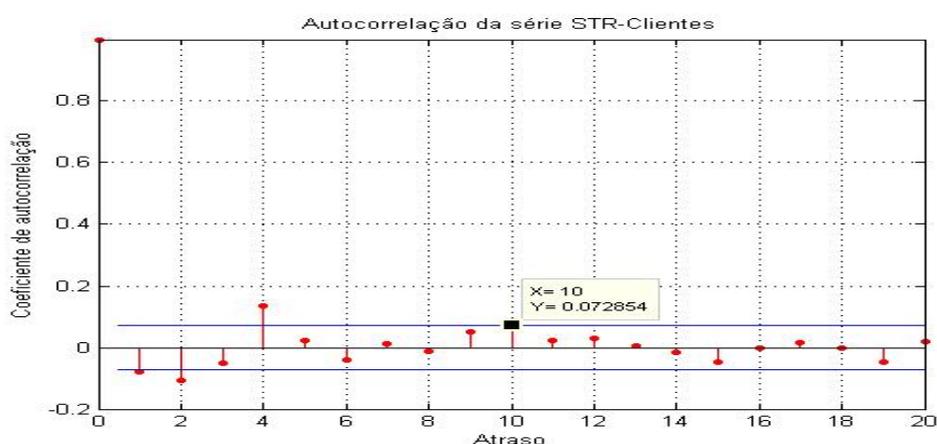


Gráfico 14: Autocorrelação da série STR-Clientes após extração da 3ª componente.

A potência da série foi reduzida novamente, alcançando o valor corresponde a $Pot = 0.187593$.

2.4.4 Retirada de Semanas Incompletas

A atipicidade de semanas com feriados pode ser um dos fatores responsáveis pela ausência de padrões na série, ou seja, a inexistência de correlações fortes entre o dia da semana e o mesmo dia na semana seguinte que seria caracterizada por índices de correlação alto em atrasos de 5 dias.

Os gráficos de autocorrelação apresentados anteriormente mostram uma correlação com atraso de 4 dias, o que não está de acordo com a fenomenologia da série. Isto leva a crer que a ausência de dias relativos a feriados pode estar influenciando a série fazendo com que uma quinta-feira, por exemplo, tenha um comportamento de sexta-feira, quando da existência de feriado nesse dia dificultando a criação de um

modelo em função desta característica.

Baseado na hipótese do parágrafo anterior, optou-se por eliminar da base de dados às semanas que tivessem feriados. Além disso, a fim de ser o mais rigoroso possível no tratamento deste problema, definiu-se que as semanas cujo início e fim não estivessem inseridos dentro de um mesmo mês também seriam excluídas. Tal atitude decorre da premissa de que as características dos dias da semana em final de mês são diferentes daquelas de início de mês, ou seja, uma quarta-feira que ocorra no dia 1º do mês apresentará características de uma sexta-feira pois haverá, neste dia, um aumento da movimentação financeira em função das folhas de pagamento.

O gráfico abaixo apresenta os dados do arquivo que passou pela normalização dos dados, retirada da tendência, da sazonalidade semanal e da primeira componente de Fourier, referente ao período de 756 dias. Após estas transformações, o arquivo teve extraídas as semanas incompletas e as semanas com início e fim em meses distintos. A série neste ponto ficou com 575 registros e sua potência corresponde a $Pot = 0.219079$

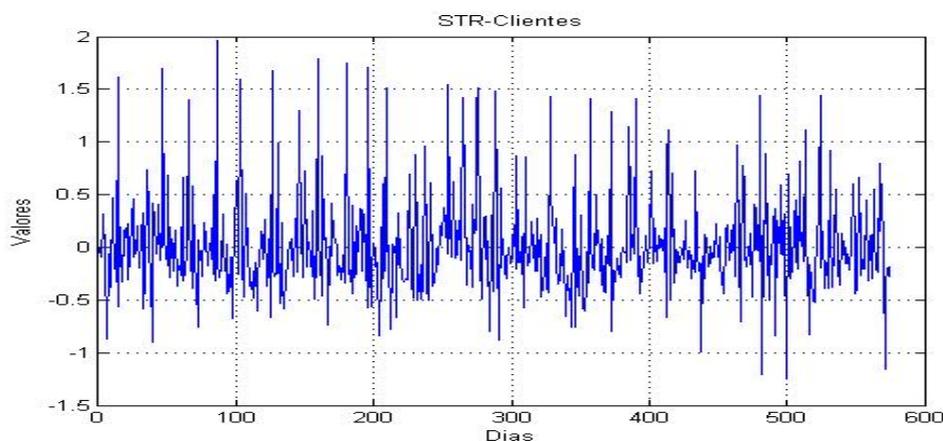


Gráfico 15: Representação da série STR-Clientes.

A fim de analisar a série após as alterações realizadas, foram extraídos gráficos da autocorrelação e o periodograma, conforme apresentados a seguir.

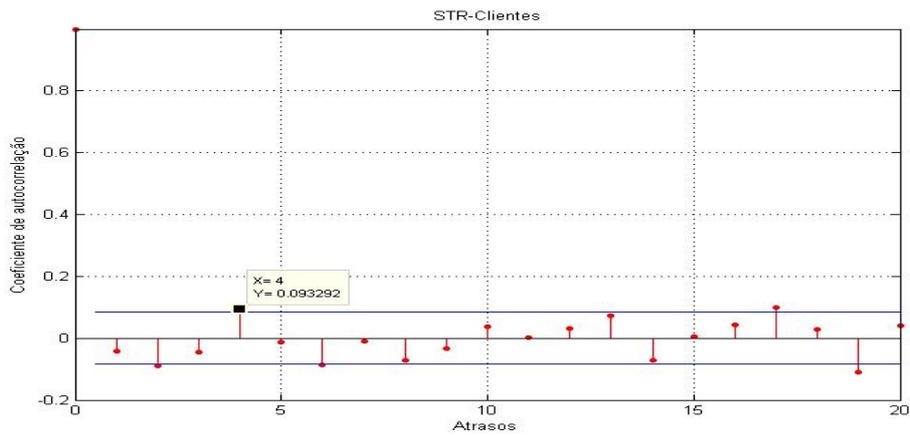


Gráfico 16: Autocorrelação da série STR-Clientes.

Apesar das mudanças efetuadas, a correlação com atraso de 4 dias permanece, derrubando a hipótese de os feriados e o fato de a semana iniciar e terminar em meses diferentes serem responsáveis por este fenômeno.

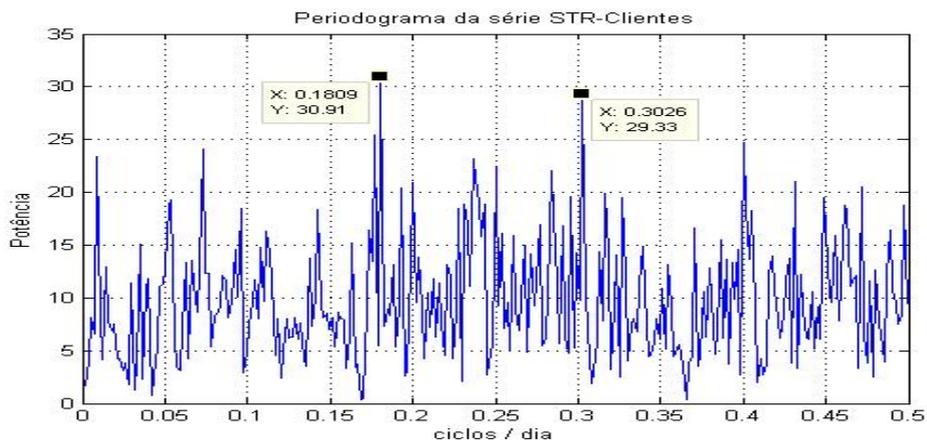


Gráfico 17: Periodograma da série STR-Clientes.

No periodograma acima, as duas maiores componentes, representadas pelos períodos de 5,53 e 3,30 dias, são as mais significativas. Estes ciclos serão retirados da série apesar deles não terem relação conhecida com o fluxo de caixa. Os gráficos da série após a extração encontram-se a seguir, bem como o autocorrelograma. A potência passou a valer $Pot = 0.208098$.

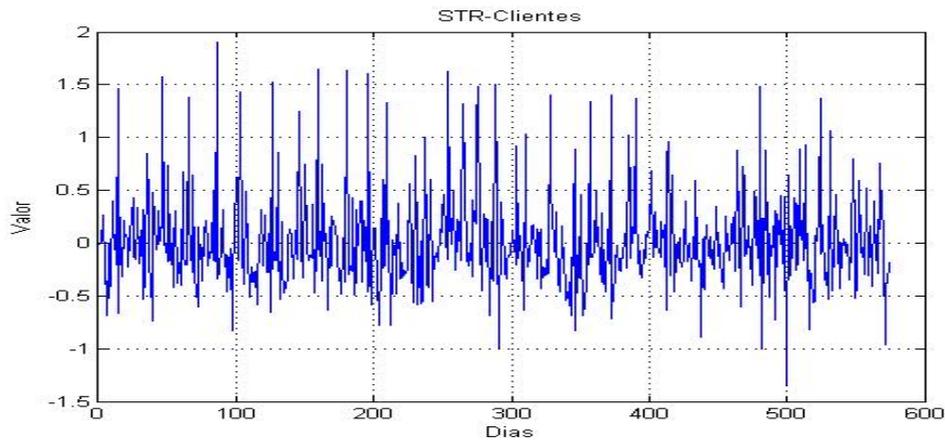


Gráfico 18: Representação da série STR-Clientes.

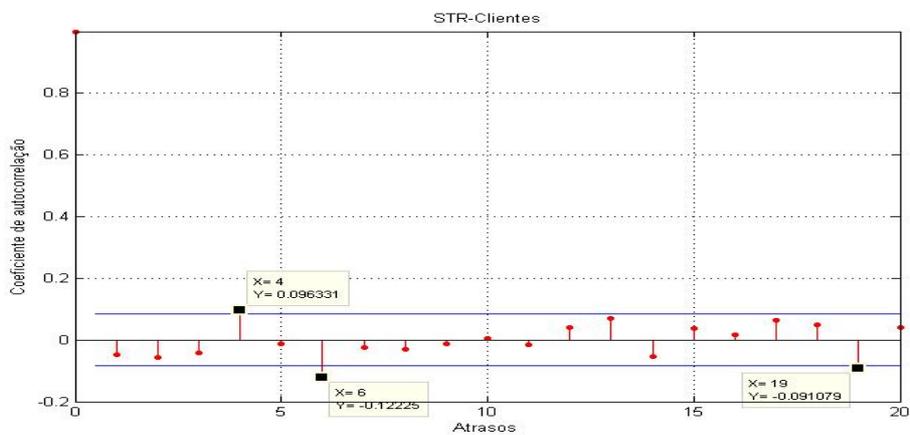


Gráfico 19: Autocorrelograma da série STR-Clientes.

Não só a correlação com atraso de 5 dias não foi evidenciada com a retirada dos registros como houve um pequeno aumento nos valores dos índices de autocorrelação para atraso de 4 dias, passando de 0,093292 para 0,096331. Isto demonstra que o padrão de comportamento semanal verificado no fluxo de caixa não é representado na variável STR-Clientes.

2.4.5 Irregularidade

As séries temporais são formadas pela composição de diversas outras séries. Assim, se torna mais prático analisar uma série a partir da decomposição em outras mais simples. A diferença entre a recomposição destas séries simples e a real corresponde a uma série complexa não linear, chamada pela estatística clássica como irregularidade.

Pelo fato de ser, geralmente, residual, a irregularidade tende a ser de

pequena amplitude e, frequentemente, seguindo o padrão de uma distribuição normal, isto é, aquela na qual os pequenos desvios ocorrem com grande frequência e os grandes com frequência pequena (SPIEGEL, 1961). Não é o que se observa. A amplitude da componente irregular na série STR-Clientes é maior que a das demais componentes. Este fato induz a se acreditar que a forte presença de irregularidade na série dificultará a construção de um modelo de previsão eficaz.

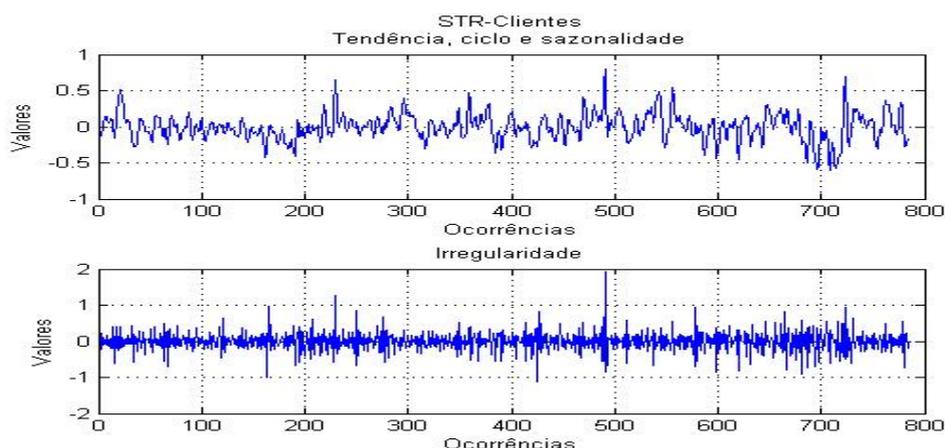


Gráfico 20: Componentes da série STR-Clientes.

Para estudar estatisticamente a série utilizou-se, como auxílio, o *software* forecast.pro. Trata-se de um aplicativo comercial muito utilizado na realização de previsões de séries temporais. Dentre as diversas opções disponíveis, a ferramenta executa, automaticamente, a análise dos dados e informa o percentual representativo de cada componente na base de dados. Assim, a série STR-Clientes foi inserida no aplicativo e os resultados mostraram que após as transformações realizadas pouco se pode extrair dos dados, em matéria de componentes determinísticas haja vista a componente irregularidade, neste ponto, alcançar valores superiores a 80%, conforme pode ser visto no quadro abaixo.

Quadro 9: Decomposição clássica das séries relativas às agências 02, 06 e 13.

	Tendência/Ciclo	Sazonalidade	Irregularidade
STR-CIP	16,58%	0,01%	83,41%

Com relação aos percentuais de tendência e ciclo apresentados pelo aplicativo a sua redução teria que ser feita através da retirada de tendências não-lineares e vários outros ciclos não muito significativos. Para o primeiro caso, retirada de

tendências não-lineares, a sua supressão dificulta a criação de qualquer modelo de previsão pois o seu comportamento terá que ser projetado baseado naquele encontrado nos dados o que nem sempre ocorre. Já para o caso dos ciclos, várias componentes foram excluídas quando da aplicação da transformada de Fourier. Entretanto, verificou-se que as reduções de potência para cada ciclo excluído era muito pequena, desestimulando o seu uso.

Considerando a dificuldade de se mapear a irregularidade com o uso de algoritmos lineares avaliou-se a utilização de algoritmos não lineares, como as redes neurais artificiais.

2.5 OBSERVAÇÕES

É senso comum imaginar que a movimentação financeira do público segue alguns padrões bem definidos como, por exemplo, o de depósitos no início da semana, caminhando-se para o equilíbrio no meio da semana e saques nos últimos dois dias da semana. Das rubricas estudadas, notadamente a STR-Clientes, cuja performance é uma das mais importantes, o padrão esperado não foi verificado.

O gráfico abaixo apresenta os valores médios de cada dia da semana e verifica-se que apesar de a média apresentar movimento receptor para toda a semana, o início se caracteriza com um comportamento bem distinto dos demais. As quartas e quintas-feiras mantêm um padrão de movimentação estabilizado mas, mesmo assim, de recebimento. Na sexta tem-se a entrada de recursos maior que nos dois dias anteriores, o que contradiz a expectativa de fortes saques.

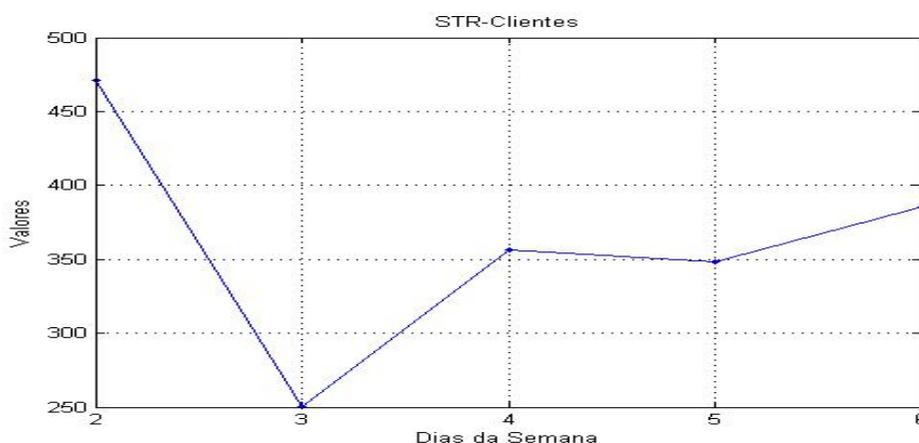


Gráfico 21: Média dos dias da semana da série STR-Clientes.

Sob a ótica do mês acredita-se que os primeiros 5 dias úteis são considerados de movimento pagador e os demais com tendências ao equilíbrio. Não foi observado. O movimento médio é de recebimento em todos os dias do mês.

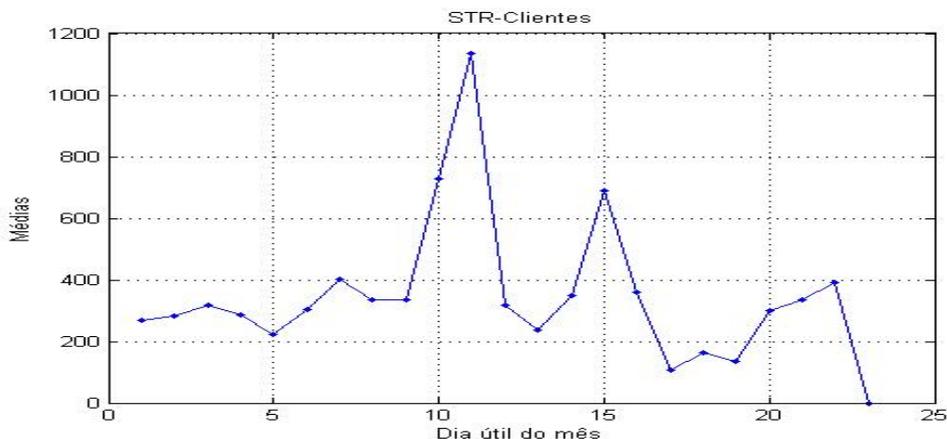


Gráfico 22: Média dos dias úteis do mês da série STR-Clientes.

Também sob o aspecto do mês o comportamento da variável não se mostrou totalmente dentro dos padrões definidos. Algumas das distorções observadas podem ser em decorrência dos feriados existentes que altera as características dos dias antecedentes e subsequentes.

Todas as variáveis integrantes do fluxo de reservas bancárias foram avaliadas. Entretanto, baseado nas análises promovidas alguns conceitos amplamente aceitos não foram confirmados, tais como:

- Contrariamente ao que se espera não se verificou a característica pagadora nos primeiros cinco dias úteis de cada mês;
- Não se observou relação entre as variáveis de forma que pudessem auxiliar na previsão da variável principal;
- Nem mesmo variáveis com similaridades no comportamento, tais como STR-Clientes e CIP apresentaram grau de correlação significativo;
- O fato da existência de vultosos pagamentos de tributos em dias determinados não auxiliou no mapeamento da série pois a característica destes pagamentos serem em dias pré-definidos faz com que eles ocorram em dias da semana distintos, ou seja, é sabido que vários pagamentos são realizados no dia 15 de cada mês. Entretanto, quando o

dia 15 ocorre numa segunda-feira o movimento do dia tem uma característica diferente para o caso de ocorrer numa sexta-feira;

- A ocorrência de valores vultosos não seguiu nenhum padrão. Não foram encontradas situações da existência de grandes valores no mesmo dia e em meses seguidos. No período analisado (01/04/2003 a 31/03/2006) os valores apresentavam uma característica em um mês e no mês seguinte esta característica não se repetia, conforme pode-se observar a movimentação do primeiro dia útil da variável STR-Clientes a seguir.



Gráfico 23: Movimento do primeiro dia útil dos meses entre 01/04/2003 a 31/03/2006.

2.6 HIPÓTESES

Verificou-se, em todas as séries estudadas, um grande número de oscilações no comportamento dos dados, ou seja, tem-se em um dia “d” valores positivos e, logo em seguida, em “d + 1”, valores negativos, ou vice-versa. Este fenômeno evidencia a existência e relevância de componentes de alta frequência, dificultando o mapeamento das séries. Além da alta frequência verificada também observou-se a existência de significativa dispersão dos dados bem como de valores extremamente elevados que, em alguns casos, foram tratados como *outliers* mudando o valor para intervalo mais apropriado.

Com base nas análises iniciais é de se supor que se tratam de séries não lineares onde a aplicação de modelos estatísticos convencionais não deve apresentar bons resultados. O uso, portanto, de técnicas de inteligência computacional deve ser investigado com vistas a obtenção de melhores resultados. Entretanto, mesmo neste caso, há a possibilidade de não se conseguir um bom desempenho em razão do

excessivo número de *turnpoints*, de valores extremos e da grande dispersão dos valores.

Apesar não poder se afirmar, inequivocamente, a superioridade das redes neurais em comparação com os métodos estatísticos, os trabalhos de predição de comportamento futuro utilizando-se dessa técnica vêm obtendo melhores resultados que os métodos estatísticos (FERNANDES; NAVAUUX; PORTUGAL, 1996) (BACHA; MEYER, 1992). Esta é, portanto, a principal motivação para se adotar as técnicas de inteligência computacional na série.

3 ESTUDO DA ALEATORIEDADE DA SÉRIE

3.1 INTRODUÇÃO

Esta pesquisa iniciou-se com registros referentes a 36 meses (22/04/2002 a 31/04/2005) sendo, posteriormente completada para 31/04/2006 e, novamente, para 31/04/2007. Com a inserção destes dados optou-se pela exclusão dos registros até 31/03/2003.

Dos dados utilizados (48 meses) foi separada da base os valores referentes ao período 01/04/2006 a 31/03/2007, para servirem como conjunto de teste e validação.

Cabe comentar, também, que em função das análises promovidas as variáveis STR-Clientes e CIP foram unificadas em uma só. Isto aconteceu em virtude de características operacionais do banco, ou seja, quando da reestruturação do SPB estas 2 variáveis tinham funções similares mas com tratamento distinto. Posteriormente, elas começaram a ser tratadas de forma idêntica e, em vista disto, suas previsões passaram a ser feitas unificadas.

3.2 AMOSTRAGEM ALEATÓRIA

No caso de populações discretas uma amostra aleatória é aquela em que cada item da população tem a mesma chance de ser incluído na amostra. Já nos casos de populações contínuas uma amostra aleatória é aquela em que a probabilidade de incluir na amostra qualquer intervalo de valores é igual à percentagem da população que está naquele intervalo (STEVENSON, 2001).

Os números aleatórios podem ser classificados em 3 grupos:

- Aleatórios verdadeiros: são aqueles que exibem um comportamento verdadeiramente aleatório, ou seja, não apresentam qualquer correlação entre eles. São selecionados por meio não determinístico e não envolvem o cálculo de funções ou o uso de algoritmos. Geralmente são obtidos por intermédio de eventos físicos ou naturais;

- Pseudo-aleatórios: são obtidos por meio de algoritmos, e procuram apresentar um ciclo de repetição tão alto quanto possível, de modo a simular uma distribuição verdadeiramente aleatória. Possuem um aspecto aleatório mas, contudo, apresentam um padrão específico e repetitivo;
- Quase-aleatórios: conhecidos como sequências de baixa discrepância, são obtidos por procedimentos matemáticos. Discrepância é um conceito na teoria dos números que mede o desvio da uniformidade de um conjunto de pontos em uma dimensão d .

Segundo Wasserman (NETER; WASSERMAN; WHITMORE, 1992), em uma população infinita, um conjunto de variáveis X_1, X_2, \dots, X_n gerado por um processo constitui uma amostra aleatória simples se as variáveis são independentes e têm a mesma probabilidade de distribuição. Por variáveis independentes entende-se que as sequências devem ser não correlacionadas, ou seja, qualquer subconjunto de números aleatórios não deve estar correlacionado com nenhum outro subconjunto da mesma população.

A fim de avaliar se um conjunto de números pode ser considerado como aleatório, várias técnicas foram definidas. Dentre os diversos métodos existentes na literatura, optou-se pelo uso de 4 técnicas para avaliação do grau de aleatoriedade das séries sob estudo, quais sejam:

- Teste de autocorrelação;
- Teste do qui-quadrado;
- Teste de Kolmogorov-Smirnov; e
- Teste de corrida

A título de estudo comparativo, inseriu-se na análise uma série com 756 registros gerados através da função *randn*, do aplicativo MatLab. Tal função tem a característica de gerar números pseudo-aleatórios distribuídos segundo uma curva normal, com média 0 e desvio-padrão igual a 1.

3.3 MÉTODO DA AUTOCORRELAÇÃO

A função de autocorrelação corresponde à correlação interna entre elementos de uma série de observações ordenadas no tempo ou no espaço. A força de

uma relação entre duas variáveis em uma população é geralmente medida pelo coeficiente de correlação, cujos valores abrangem desde -1 para a correlação negativa perfeita até $+1$ para a correlação positiva perfeita e é dado por:

$$r = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\sigma_x} \right) \left(\frac{y_i - \bar{y}}{\sigma_y} \right) \quad \text{Eq. (7)}$$

onde n = número de elementos da amostra;

x_i, y_i = elementos no ponto i ;

\bar{x}, \bar{y} = média aritmética dos elementos x e y , respectivamente; e

σ_x, σ_y = desvio padrão dos elementos x e y , respectivamente.

Entretanto, é possível correlacionar os valores de uma variável x , em certas épocas, com os correspondentes x em ocasiões anteriores. Isso é frequentemente denominado autocorrelação e a equação descritiva desta função encontra-se a seguir.

$$r = \frac{\sum_{i=1}^{n-1} (x_i - \bar{x})(x_{i+j} - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad \text{Eq. (8)}$$

onde n = número de elementos da amostra;

x_i = elementos no ponto i ;

\bar{x} = média aritmética dos elementos x ;

j = intervalo de tempo.

O cálculo da autocorrelação indica, nos casos em que o coeficiente é alto, que os valores da série tendem a formar sequências equivalentes de tempos em tempos, produzindo uma ciclicidade, ou seja, a série é caracterizada por sequências repetitivas em intervalos pré-definidos. Assim, uma amostra onde o coeficiente de autocorrelação é nulo indica que os dados são independentes e identicamente distribuídos.

3.4 MÉTODO DO QUI-QUADRADO

Uma forma de verificar se uma distribuição teórica se ajusta ao fenômeno investigado ou não é através de testes de aderência. Através de funções empíricas e certas tolerâncias definidas pelo usuário, estes testes comparam a distribuição teórica

com a aproximada para cada intervalo do diagrama de frequências relativas e, no final, dizem, de acordo com nível de confiança pré-estabelecido, se a distribuição teórica pode representar o fenômeno (LIMA; SAGRILO, 2003).

Os resultados obtidos por meio de amostras nem sempre concordam exatamente com os teóricos esperados, de acordo com as regras de probabilidade. Por exemplo: suponha que, em uma determinada amostra, observou-se que um conjunto de eventos possíveis E_1, E_2, \dots, E_n ocorreram com as frequências x_1, x_2, \dots, x_n denominadas frequências observadas e, de acordo com as regras de probabilidade, esperava-se que elas ocorressem com as frequências e_1, e_2, \dots, e_n , denominadas frequências esperadas ou teóricas. Dessa forma deseja-se, na maioria das vezes, saber se as frequências observadas diferem, de modo significativo, das esperadas.

Uma medida de discrepância existente entre as frequências observadas e esperadas é proporcionada pela estatística X^2 (qui-quadrado) e foi proposta, em 1900, por Karl Pearson, sendo considerado um dos melhores testes estatísticos. Este método compara as frequências observadas x_1, x_2, \dots, x_n , de k valores (ou em k intervalos) de uma variável, com as correspondentes frequências e_1, e_2, \dots, e_n de uma distribuição teórica esperada e é expresso, matematicamente, por:

$$X^2 = \sum_{i=1}^k \frac{(x_i - e_i)^2}{e_i} \quad \text{Eq. (9)}$$

onde k = número de intervalos em que foi subdivida a amostra;

x_i = frequências observadas;

e_i = frequências esperadas.

Quando $X^2 = 0$, as frequências teóricas e observadas concordam exatamente, enquanto que, quando $X^2 > 0$ isso não se dá. Quanto maior for o valor de X^2 , maior será a discrepância entre as frequências observadas e esperadas, ou seja, o teste do qui-quadrado pode ser usado para determinar quanto aproximadamente às distribuições teóricas se ajustam às distribuições empíricas, isto é, as obtidas por meio de dados amostrais (SPIEGEL, 1961).

3.5 MÉTODO DE KOLMOGOROV-SMIRNOV

O procedimento básico do teste de Kolmogorov-Smirnov (K-S) é a

comparação entre a frequência acumulada experimental e a função de distribuição teórica assumida. Se a discrepância é grande em relação ao que seria normalmente esperada para uma dada amostra, o modelo teórico é rejeitado.

Para uma amostra de tamanho n , os dados devem ser organizados em ordem crescente, e a partir deles monta-se a função de frequência acumulada da seguinte forma:

$$S_n(x) = \begin{cases} 0 & x < x_1 \\ \frac{k}{n} & x_k \leq x < x_{k+1} \\ 1 & x \geq x_n \end{cases} \quad \text{Eq. (10)}$$

onde x_1, x_2, \dots, x_n são os valores dos dados amostrais ordenados, e n é o tamanho da amostra. A figura abaixo mostra os valores plotados de $S_n(x)$ e a função de distribuição teórica proposta $F(x)$.

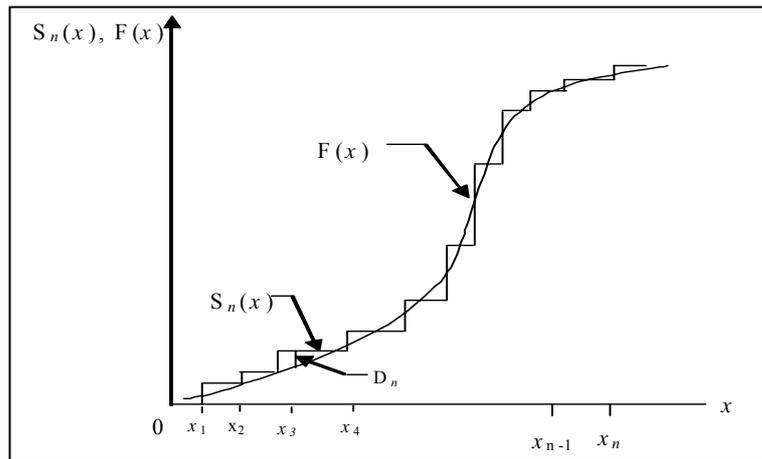


Gráfico 24: Representação da frequência experimental acumulada $S(x)$ e função teórica proposta $F(x)$

No teste K-S a diferença máxima entre $S_n(x)$ e $F(x)$ para todos os valores de X é a medida de discrepância procurada. Este valor máximo é:

$$D_n = \max | F(x) - S_n(x) | \quad \text{Eq. (11)}$$

Para um nível de significância α , o teste K-S compara a diferença máxima observada D_n com o valor crítico D_n^α , que é definido como:

$$P(D_n \leq D_n^\alpha) = 1 - \alpha \quad \text{Eq. (12)}$$

Os valores críticos de D_n^α para vários níveis de significância são tabelados para diferentes valores de n e podem ser encontrados em um bom livro de estatística (NETER; WASSERMAN; WHITMORE, 1992), (AFIF; AZEU, 1979). Se o valor de D_n observado for menor que o valor crítico, então a distribuição proposta é aceitável naquele nível de significância específico; caso contrário, a distribuição assumida deve ser rejeitada (EBECKEN, 2001).

3.6 MÉTODO DE CORRIDAS

A corrida se refere a uma sucessão de eventos similares separados por eventos diferentes, ou seja, uma sequência ininterrupta de observações que apresentam uma característica semelhante. As corridas focalizam a atenção na sequência temporal (ordenada no tempo), como o registro diário da temperatura de uma localidade, ou espacial (ordenada por posição), a exemplo da chegada de homens e mulheres em uma fila. Independentemente da natureza efetiva dos dados, as observações podem ser transformadas numa série de símbolos em que se possa contar o número de repetições (STEVENSON, 2001).

3.6.1 Corrida

Corrida é uma sucessão não interrompida de dígitos da sequência original. Uma corrida de comprimento k consiste em exatamente k dígitos, limitada antes e depois por outro segmento composto por valores opostos. Os testes de corrida verificam se a ocorrência de sequências - crescentes e decrescentes, acima e abaixo etc - apresentam número e comprimento aleatórios

Por exemplo, dada a sequência 0,123 0,357 0,789 0,425 0,224 0,416 0,950 as corridas são contabilizadas da seguinte forma: existe uma corrida crescente de comprimento 3, uma corrida decrescente de comprimento 2 e uma corrida crescente de (pelo menos) 2, dependendo dos próximos valores.

Quadro 10: Sequência hipotética

0,123	0,357	0,789	0,425	0,224	0,416	0,950
Corrida creciente de comprimento 3			Corrida decreciente de comprimento 2		Corrida creciente de comprimento 2	

3.6.2 Análise das Repetições

Para podermos dizer que uma série de observações apresenta as características de um processo aleatório, é preciso primeiro ter uma idéia daquilo que um processo aleatório pode produzir. Isto servirá então de padrão de julgamento de uma situação efetiva. A distribuição amostral proporciona uma definição de aleatoriedade indicando o número esperado (médio) de repetições e revelando qual o grau de variabilidade razoável (para uma série aleatória) entre os números de repetições esperado e observado.

Quando a série em estudo envolve um grande número de observações (> 30) a distribuição amostral do número de repetições é bem aproximada pela distribuição normal.

A hipótese nula num teste de repetições é que a diferença entre o número esperado de repetições e o número efetivamente observado é o resultado de variação aleatória na amostragem. A hipótese alternativa é que a diferença entre o número observado e o número esperado de repetições é demasiado grande para ser devida só ao acaso. Testa-se a diferença comparando-a com a variabilidade esperada das repetições, supondo que haja somente aleatoriedade nos dados. Isto é, divide-se a diferença entre os números observados e esperados de repetições pelo desvio padrão da distribuição adequada. Assim:

$$z = \frac{\text{número observado de repetições} - \text{número esperado de repetições}}{\text{desvio padrão das repetições}} \quad \text{Eq. (13)}$$

ou

$$z = \frac{r - R}{\sigma_R} \quad \text{Eq. (14)}$$

Pequenos valores de z (digamos, menor do que 1,96 para $\alpha = 0,05$, onde α corresponde ao índice de significância) sugerem que a sequência é aleatória, enquanto grandes valores sugerem o contrário (STEVENSON, 2001).

Uma repetição é uma sequência ininterrupta de observações que possuem uma característica comum e focalizam a atenção na ordem de ocorrência das observações. As repetições podem constituir uma sequência temporal (ordenada no tempo) ou uma sequência espacial (ordenada por posição). Em ambos os casos usa-se o teste de repetições para detectar padrões não aleatórios em dados quando se conhece a ordem das observações. Por exemplo, a existência de uma sequência do tipo ABCD ABCD ABCD ABCD ABCD ABCD sugere um ciclo, já uma sequência do tipo AAAAAAAAAA BBBBBB sugere uma tendência de A para B nos dados.

Os testes de repetições se propõem a identificar a existência de algum padrão nos dados a fim de poder classificá-los como aleatórios ou não. Entretanto, pode-se ocorrer de na aplicação dos testes um deles, por exemplo o teste acima e abaixo da mediana, indicar uma não aleatoriedade de variabilidade, enquanto o outro indicar o contrário. Neste caso, a série deve ser considerada como não aleatória.

3.6.3 Teste de repetições de dois tipos de observações

Quando os dados em estudo são dicotomizados, ou seja, pertencem a duas categorias (verdadeiro-falso, positivo-negativo, etc), a definição de aleatoriedade do conjunto deve ser feita após a contagem do número de repetições dos dados e comparação com o número esperado de repetições.

A distribuição amostral para o número de repetições com dois tipos de observação é aproximadamente normal para grandes amostras (> 20) e o número esperado de repetições, para esta situação, é dado por:

$$R = \frac{2n_1n_2}{N} + 1 \quad \text{Eq. (15)}$$

onde R = número esperado de repetições;

n_1 = número de observações de um tipo;

n_2 = número de observações de outro tipo; e

N = número total de observações.

O desvio padrão da distribuição amostral do número de repetições é dado por:

$$\sigma_R = \sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}} \quad \text{Eq. (16)}$$

Que, após alguns procedimentos, pode ser reduzido à equação abaixo:

$$\sigma_R = \sqrt{\frac{(R-1)(R-2)}{N-1}} \quad \text{Eq. (17)}$$

3.6.4 Teste de repetições acima e abaixo da mediana

Quando os dados são numéricos, a detecção da aleatoriedade de uma série pode ser feita em relação à mediana amostral. As medições são caracterizadas como maiores (M) ou menores (m) em relação à mediana e ignoram-se os valores coincidentes. Contam-se, então, os números de repetições M e m.

A conversão de cada valor para M ou m tornará o problema com dois tipos de observações apenas (dicotomizado) podendo-se, então, utilizar as fórmulas anteriores. Entretanto, com o número de M's igualará o de m's (por definição, metade das observações está acima da mediana e metade abaixo), as fórmulas anteriores ficam simplificadas (STEVENSON, 2001):

$$R = \frac{N}{2} + 1 \quad \text{Eq. (18)}$$

$$\sigma_R \approx \sqrt{\frac{(N-1)}{4}} \quad \text{Eq. (19)}$$

onde N corresponde ao número total de observações.

3.6.5 Teste de repetições acima e abaixo em relação ao anterior

Por fim, tem-se o teste de repetições para cima e para baixo como medida de aleatoriedade. Mede-se cada valor da sequência em relação ao valor anterior e cada valor numérico é substituído por U (para cima) e D (para baixo), em relação ao valor anterior. Cabe lembrar que ao primeiro valor não se atribui símbolo pois não possui antecedente.

Para grandes amostras (≥ 40) a distribuição amostral do número de

repetições é normal. Entretanto, as fórmulas da média e do desvio padrão da distribuição amostral não são as mesmas. A razão é que os testes anteriores envolveram elementos com probabilidade constante, como, por exemplo no caso de se ter um valor acima ou abaixo da mediana, cuja probabilidade de se obter um valor de um ou outro lado da mediana é 0,5 e não depende da observação anterior. Não é o caso das repetições para cima e para baixo. Se a sequência é realmente aleatória, a probabilidade de um valor alto ser seguido por outro ainda mais alto é muito menor que 0,5, e a probabilidade de um valor elevado ser seguido por um menor é maior que 0,5. As fórmulas, portanto, levam em conta esse aspecto de probabilidade “posicional” e o número esperado de repetições para cima e para baixo fica representado por (STEVENSON, 2001):

$$R = \frac{2N - 1}{3} \quad \text{Eq. (20)}$$

onde N é o total de observações. O desvio padrão do número de repetições para cima e para baixo é:

$$\sigma_R = \sqrt{\frac{(16N - 29)}{90}} \quad \text{Eq. (21)}$$

3.7 RESULTADOS

Os resultados obtidos através dos experimentos com os diversos métodos estão compilados a seguir e o arquivo utilizado para realização dos testes sofreu as transformações descritas no quadro abaixo:

Quadro 11: Características do arquivo de dados.

Características do arquivo
Soma dos saldos das variáveis STR-Clientes e CIP
Divisão dos valores por 1 milhão
Inclusão da média aritmética nos dias faltantes
Normalização dos dados
Exclusão da tendência
Exclusão da sazonalidade semanal
Exclusão de 1 componente senoidal na variável STR-Clientes (783,7 dias)
Exclusão de sazonalidades mensal (dias úteis e calendário)
Potência da série = 0,142292

Após a soma aritmética das variáveis STR e CIP e sua divisão por 10^{-6} ,

forma incluídos nos dias relativos aos feriados bancários, a média aritmética da base de dados. Em seguida normalizou-se os dados e extraiu-se a tendência linear, conforme já descrito anteriormente. Dando prosseguimento, extraiu-se a sazonalidade semanal através da subtração do valor médio de cada dia da semana, ou seja, para as segundas-feiras retirou-se dos respectivos valores a média de todas as segundas-feiras da base de dados, realizando-se o mesmo procedimento para as terças-feiras e assim sucessivamente.

Os ciclos, representados pelas componentes senoidais, foram excluídos utilizando-se a transformada de Fourier.

Por fim, as sazonalidades mensais dos dias úteis e dias calendários também foram suprimidas através do mesmo mecanismo usado para exclusão da sazonalidade semanal, sendo que, desta vez, as médias foram calculadas tendo em conta os dias úteis e dias calendário do mês.

3.7.1 Teste da autocorrelação

Os testes de autocorrelação consistem em analisar a relação entre variáveis defasadas e variáveis posteriores, com base em diferentes intervalos de defasagem e podem ser empregados para testar a existência de dependência ou independência de variáveis em uma série temporal (BRUNI, 2004).

Uma autocorrelação serial igual a zero significa que as mudanças de preços em períodos de tempos consecutivos não são correlacionáveis entre si e, dessa forma, as análises não podem se basear no passado para inferir padrões de comportamentos futuros (SILVA, 2004).

Estudos realizados por Queiroz (2004) apontam para a inexistência de autocorrelação serial no comportamento diário das reservas bancárias das instituições brasileiras. Cabe comentar que no trabalho foi utilizado o somatório das reservas de todas as instituições conjuntamente mas, ainda assim, remete-nos a hipótese deste comportamento se estender a cada banco individualmente.

A fim de avaliar as características da série STR + CIP foram calculadas as autocorrelações para vários períodos e com diversos atrasos, sempre visando localizar algum ciclo que pudesse auxiliar no treinamento da rede.

Análise de autocorrelação 01

O estudo iniciou-se com o cálculo da correlação de 1 até 40 atrasos, a partir do registro 401. Após o cálculo das autocorrelações, segue-se para o registro 441 e executa-se o procedimento novamente, até o final do arquivo.

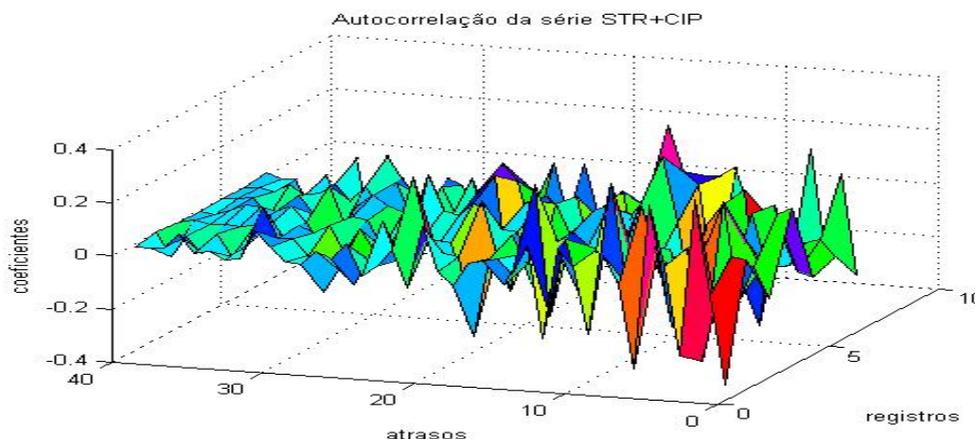


Gráfico 25: Autocorrelação da série STR+CIP de 1 até 40 atrasos.

Apesar de o arquivo já ter sido trabalhado, com a retirada das sazonalidades, ainda assim observa-se a existência de sazonalidades no gráfico representada por repetições ao longo do eixo dos registros. Vê-se, também, que a partir de 10 dias a correlação começa a não ter mais influência, com os coeficientes próximos de 0.

Outra característica presente na série diz respeito a grande quantidade de alternância nos sinais dos coeficientes. Se para um determinado atraso o coeficiente de autocorrelação calculado é positivo, no atraso seguinte ele se torna negativo. Isto acontece em função da alta frequência existente na base de dados trazendo, em consequência, uma dificuldade para o seu mapeamento.

Análise de autocorrelação 02

Neste segundo teste o cálculo da autocorrelação foi realizado para 20 atrasos e percorrendo a base registro a registro, ou seja, para o registro 1 calcula-se de 1 a 20 atrasos, em seguida faz-se o mesmo procedimento para o registro 2 e assim sucessivamente, até o término dos registros.

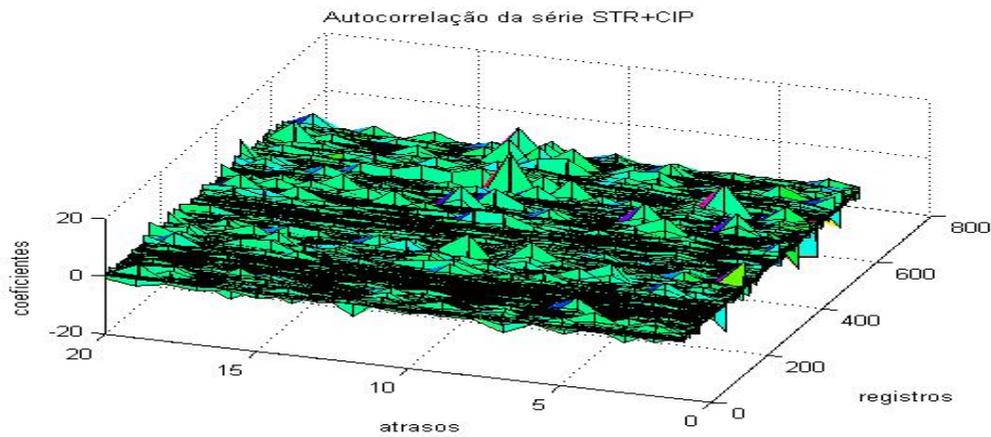


Gráfico 26: Autocorrelação da série STR+CIP – Todos os registros com atrasos de 1a 20.

Apesar da existência de pontos com coeficientes elevados, verifica-se que a maioria dos dados possuem baixa correlação. Nota-se, também, a inexistência de persistência nas correlações, ou seja, os valores não permanecem próximos, para a mesma quantidade de atrasos, ao se deslocar no tempo. Por exemplo: a correlação do registro 580 com o 587 (7 atrasos) corresponde a 4,592 e a de 581 com 588 (7 atrasos) é de -0,832. Ou seja, parece não existir ciclos bem definidos na base.

Para esta análise foi utilizada a seguinte equação para os cálculos da autocorrelação.

$$r(t, \tau) = \frac{s(t)s(t + \tau)}{\sigma_s^2} \quad \text{Eq. (22)}$$

onde $s(t)$ = valor do registro no instante t ;

τ = atrasos;

σ_s^2 = variância da série.

Este estudo busca obter graficamente uma visão da variação dos coeficientes ao longo do tempo e com isto detectar a existência de ciclos. Pode-se observar que a maioria dos dados encontram-se em vales com alguns poucos picos presentes. Como a escala do gráfico pode induzir a erros, em função da existência de alguns picos muito grandes, resolveu-se utilizar de filtro para amortecimento da série.

Análise de Autocorrelação 03

A fim de reduzir a escala dos coeficientes com o objetivo de melhor

representar os valores existentes nos vales verificados na análise anterior, foi gerado gráfico com a aplicação de média móvel, conforme representado pela equação abaixo.

$$r(\tau, t) = \frac{\frac{1}{t_2 - t_1} \sum_{t=t_1}^{t_2} s(t)s(t + \tau)}{\sigma_s^2} \quad \text{Eq. (23)}$$

onde $s(t)$ = valor do registro no instante t ;

t_1 = posição inicial.

t_2 = posição final.

τ = atrasos.

σ_s^2 = variância da série.

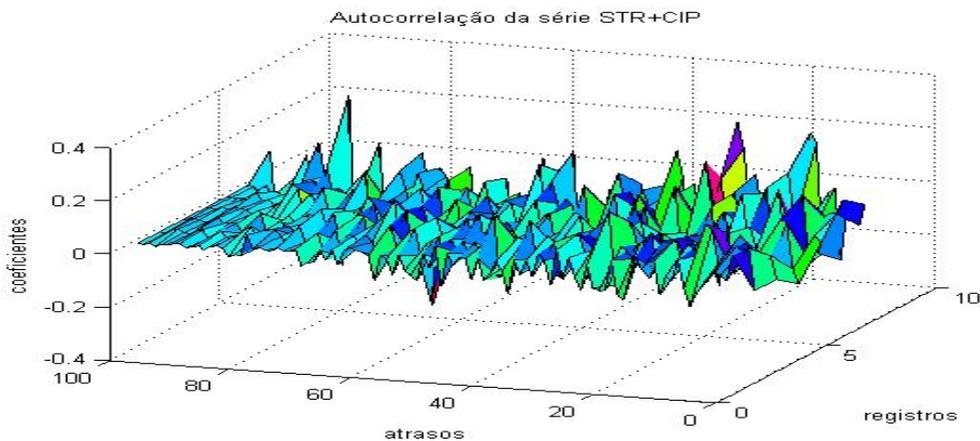


Gráfico 27: Autocorrelação da série STR+CIP – Com amortecimento média móvel.

Neste estudo foi utilizado uma média móvel de 100 posições e apesar da redução dos picos em relação ao gráfico da análise 02, ainda assim não se pode detectar persistência nas correlações da série. Em vista do exposto, optou-se pela criação de gráficos com variações da média móvel no amortecimento da série.

Análise de Autocorrelação 04

Enquanto no gráfico anterior a média móvel era fixa e sempre para a frente, aqui resolveu-se fazê-la computando-se x posições à frente e y posições atrás do ponto de referência. Os valores de x e y foram alterados de diversas formas indo de 0 a 10 para cada um dos lados. A figura que sintetiza este estudo está abaixo e representa as autocorrelações da série com uma média móvel de 7 valores, sendo 4 à frente e 2 para

trás, além, é claro, do elemento central.

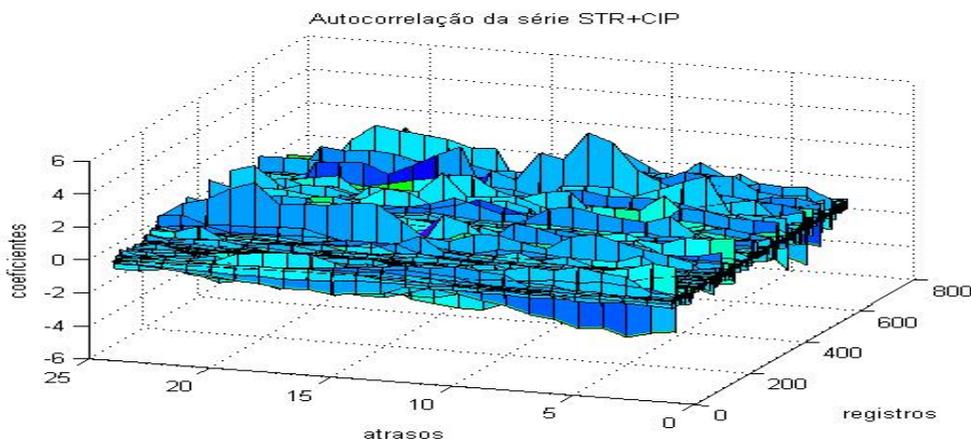


Gráfico 28: Autocorrelação da série STR+CIP – Com média móvel de 7 valores.

Para efeito de comparação, adiante tem-se os cálculos dos coeficientes sem a aplicação do amortecimento.

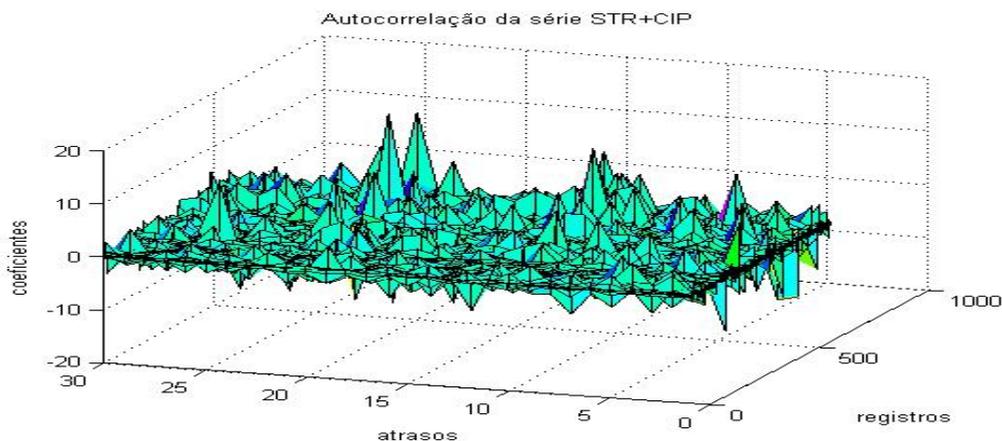


Gráfico 29: Autocorrelação da série STR+CIP.

Apesar das inúmeras tentativas para a detecção de correlações persistentes na base de dados, algumas delas apresentadas neste estudo, os gráficos não foram suficientemente claros para se obter um padrão que pudesse ser utilizado como auxílio no mapeamento dos dados. Novos estudos foram desenvolvidos com a utilização de outros recursos gráficos e de tabelas visando a descoberta de comportamentos repetitivos e importantes na série.

Análise de Autocorrelação 05

A representação gráfica de todos os registros da série bem como os diversos

atrasos implementados tornaria o estudo cansativo e pouca informação acrescentaria na tese, haja vista que após dezenas de análises acredita-se na inexistência de correlação periódica nos dados. Entretanto, a fim de apresentar a linha adotada que refuta a hipótese da existência de ciclos periódicos na série, foram gerados alguns gráficos mais significativos e apresentadas as conclusões obtidas.

Nesta análise, por exemplo, o gráfico foi obtido a partir do cálculo da correlação de todos os registros com 1 até 50 atrasos à frente. De posse desses valores calculou-se a média móvel de 5 correlações relativas aos atrasos e 3 correlações relativas aos registros, conforme exemplificado no gráfico abaixo.

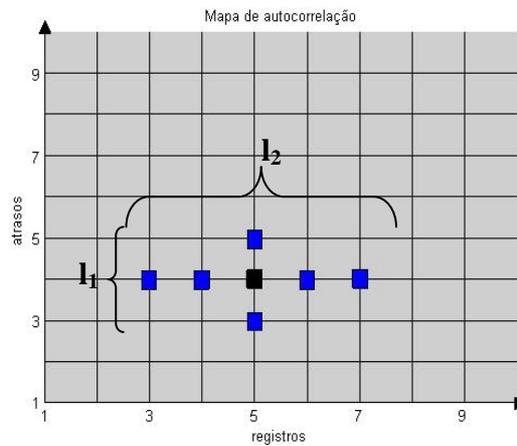


Gráfico 30: Gráfico ilustrativo da metodologia utilizada nos cálculos dos coeficientes de correlação.

A equação utilizada na geração do gráfico está apresentada a seguir.

$$r(t, \tau) = \frac{1}{l_1 l_2} \sum_{x=\frac{l_1-1}{2}}^{\frac{l_1-1}{2}} \sum_{y=\frac{l_2-1}{2}}^{\frac{l_2-1}{2}} r(x, y) \quad \text{Eq. (24)}$$

onde r = coeficiente de correlação.

l_1 = quantidade de atrasos.

l_2 = quantidade de registros.

Para o caso de se utilizar a média móvel com 3 posições nos atrasos (l_1) e 5 registros (l_2) temos:

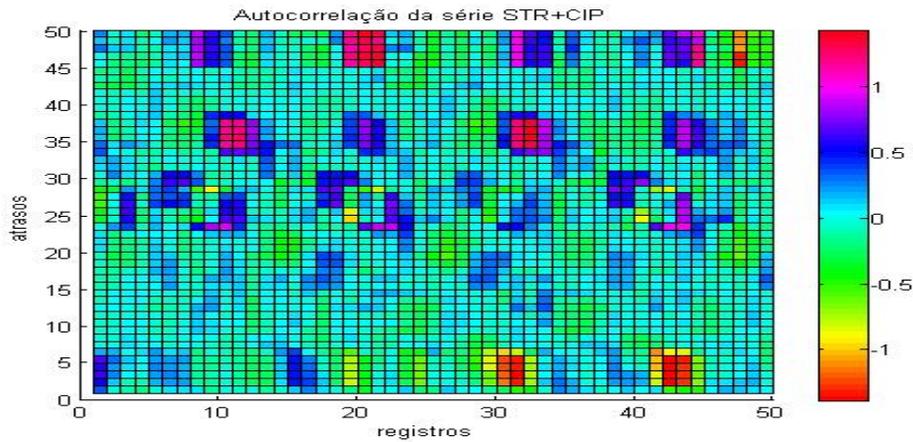


Gráfico 31: Autocorrelação da série STR + CIP para média de 3 posições no atraso e 5 registros.

Vê-se no gráfico que pouquíssimos pontos apresentam correlações significativas e, ainda assim, sem nenhuma regularidade. A maioria dos pontos encontram-se próximos a zero. Apesar de este gráfico haver sido gerado para o intervalo de registros entre 551 e 600, todos os demais intervalos foram plotados e apresentaram resultados similares.

Análise de Autocorrelação 06

Neste modelo foi introduzido o conceito de “variância instantânea” que pode ser entendida como a volatilidade dos dados, ou seja, o quanto ele se alterou durante um determinado período. Se os valores da métrica descrita na equação abaixo forem persistentes, deduz-se que há ciclos de volatilidade nos dados.

$$r(t, \tau) = \frac{s(t)s(t-\tau)}{\sqrt{\sigma_{s(t)}^2 \sigma_{s(t+\tau)}^2}} \quad \text{Eq. (25)}$$

Onde $s(t)$ = valor do registro no instante t .

τ = atrasos.

σ_s^2 = variância da série.

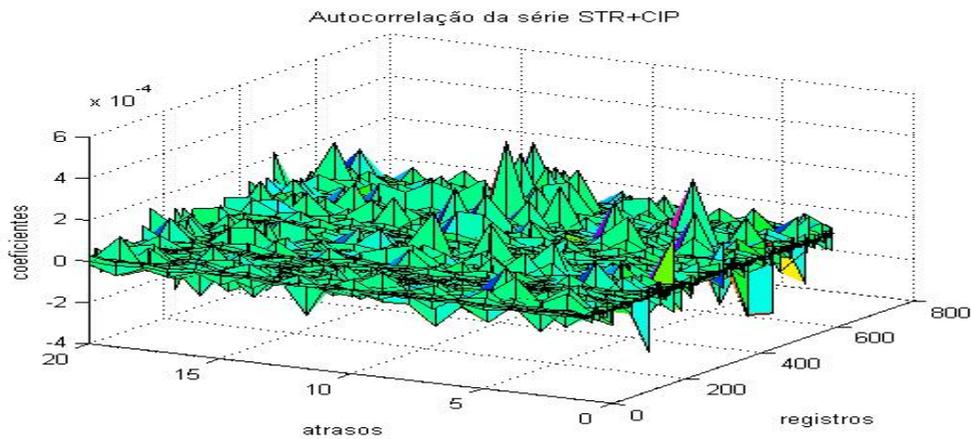


Gráfico 32: Autocorrelação da série STR + CIP com variância instantânea.

Ainda seguindo a mesma linha, calculou-se os coeficientes da série utilizando-se média móvel de 7 posições, sendo 3 para a frente e 3 para trás, além do registro central. O resultado pode ser visualizado abaixo.

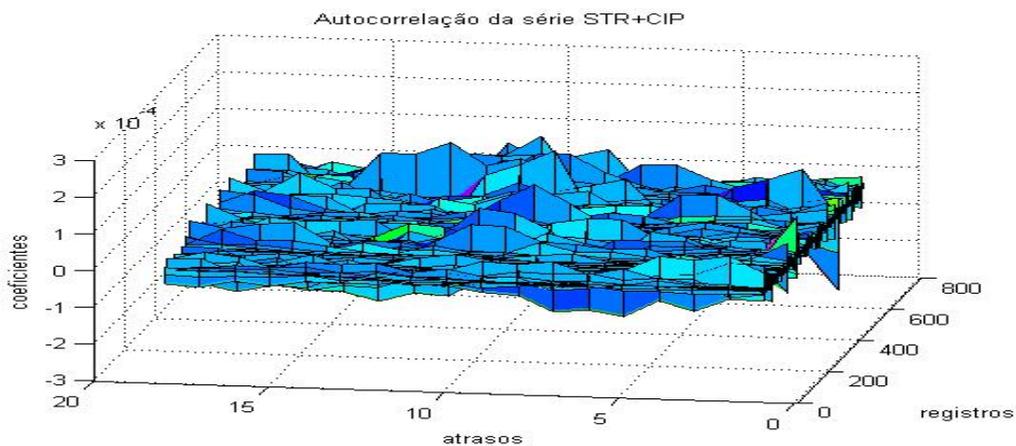


Gráfico 33: Autocorrelação da série STR + CIP com variância instantânea e média móvel de 5 posições.

Verifica-se nos gráficos acima a existência de alguma correlação na faixa de 5 e 10 atrasos o que já era de se esperar em função destes atrasos representarem 1 e 2 semanas de dias úteis. Entretanto, numa análise mais acurada verifica-se que este fenômeno só ocorre para alguns poucos dias. Seguindo o eixo x (registros) nota-se que os picos não se repetem em dias subsequentes, nem formam padrões para espaços regulares de dias. Este aspecto é melhor analisado no gráfico em 2D apresentado abaixo, onde pode-se observar a inexistência de repetições.

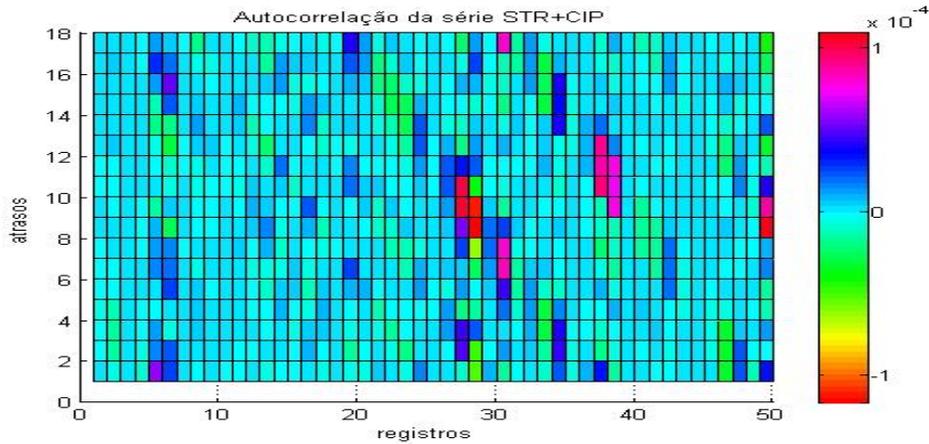


Gráfico 34: Autocorrelação da série STR + CIP com variância instantânea e média móvel de 5 posições.

Chatfield (CHATFIELD, 2004) define ruído branco (*white noise*) como uma sequência de variáveis aleatórias e identicamente distribuídas. A expressão deriva da física, particularmente da ótica, onde a luz branca corresponde a igual contribuição de todas as outras ondas (cores). Já ruído se relaciona a erro, excepcionalidade ou propriedades indesejáveis existentes em um sinal.

Estatisticamente falando, trata-se de um modelo que descreve o comportamento de um tipo de sinal que possui determinadas características, tais como a distribuição normal dos coeficientes de autocorrelação, média zero e variância igual a $1/n$, onde n representa o tamanho da amostra (BARTLETT *apud* BRUNI, 2004).

Quadro 12: Comparação entre os parâmetros de aleatoriedade para autocorrelação definidos por Bartlett e os obtidos.

	Média	Variância
Valor tabelado	0	0,001495
Análise 01	0,0121951	0,036733
Análise 02	0,019552	0,864217
Análise 03	0,000017	0,006205
Análise 04	0,017294	0,135187
	0,012449	0,930055
Análise 05	0,007027	0,055256
Análise 06	5,29 10e-7	6,33 10e-10
	5,78 10e-7	2,03 10e-10

No caso deste estudo, verifica-se a proximidade entre os resultados obtidos no cálculo das autocorrelações com os parâmetros estipulados por Bartlett, conforme pode ser visto na tabela acima, sugerindo, ainda que não enfaticamente, uma forte tendência para se considerar o resíduo da série como aleatório.

Análise de Autocorrelação 07

Braga e Calôba (2007) ao tratarem da previsão da polaridade dos índices do IBOVESPA necessitaram escolher dentre os 20 índices disponíveis quais os que seriam utilizados no problema e os respectivos atrasos que mais se correlacionavam com a saída da rede. Para tanto, para cada série foram calculados 20 atrasos e, em seguida, adicionado mais 10 registros e refeitos os cálculos, até que toda a série tivesse sido processada. Ao término, refez-se todo o procedimento no sentido inverso. Após a escolha das melhores séries e respectivos atrasos, os dados foram aplicados à rede e obteve-se resultados corretos em aproximadamente 70% dos registros da base de teste.

Adaptando a metodologia acima para este estudo, foi calculado o desvio padrão para cada dia do mês considerando média 0 e o resultado subtraído do desvio padrão normal. Estes cálculos foram realizados na ordem direta e inversa e para dias úteis e dias calendário. Ao término, foram gerados gráficos relativos ao comportamento dos resultados. As equações e metodologia adotadas estão descritas adiante.

O objetivo deste estudo é uma tentativa de estabelecer correlações entre os dias do mês (útil e calendário) a partir da variabilidade dos dados.

Quadro 13: Planilha para cálculo das métricas dos dias úteis em ordem direta.

	1º	2º	d (dias)	22º	23º	D ^f
1	x_1^1	x_2^1	x_{22}^1	x_{23}^1	
2	x_1^2	x_2^2	x_{22}^2	x_{23}^2	
.....	
m (mês)	x_d^m	
.....	
48	x_1^{48}	x_2^{48}	x_{22}^{48}	x_{23}^{48}	
M _d						
μ _d						
P _{0d}						
P _{1d}						
ΔP _d						

Onde:

x_d^m = valor normalizado do dia d, no mês m.

D^f = quantidade de dias faltantes, ou seja, para meses com 22 dias úteis esta coluna será marcada com 1, para 21 dias úteis, 2 e assim sucessivamente.

M_d = quantidade de eventos que formarão a média, ou seja, no caso de

existirem todos os 1^{os} dias úteis, o valor será 48.

$$\mu_d = \frac{1}{M_d} \sum_{m=1}^{48} x_d^m \quad \text{Eq. (26)}$$

$$P_{0d} = \left[\frac{1}{M_d} \sum_{m=1}^{48} (x_d^m)^2 \right]^{\frac{1}{2}} \quad \text{Eq. (27)}$$

$$P_{1d} = \left[\frac{1}{M_d} \sum_{m=1}^{48} (x_d^m - \mu_d)^2 \right]^{\frac{1}{2}} \quad \text{Eq. (28)}$$

$$\Delta P_d = P_{0d} - P_{1d} \quad \text{Eq. (29)}$$

Nos cálculos dos dias úteis na ordem direta, poderá não existir valores a partir do 18º dia útil. Já quando do cálculo da ordem inversa, tal situação acontecerá nos primeiros dias do mês pois a série estará deslocada iniciando-se no 23º dia útil, ou seja, na hipótese de o mês possuir apenas 20 dias úteis o 20º dia estará na coluna 23, o 19º dia útil na coluna 22 e assim sucessivamente. Com isto, os primeiros dias da não conterão valores. Situação semelhante ocorrerá nos cálculos dos dias calendário, onde só se encontra a série completa para até 28 dias.

O arquivo de trabalho foi tratado conforme descrito no quadro abaixo.

Quadro 14: Transformações no arquivo de dados.

Características do Arquivo de Dados
Divisão dos valores por 1 milhão
Redução dos valores para o intervalo [-1000, 2000]
Inclusão da média aritmética nos dias faltantes
Normalização dos dados
Exclusão da tendência
Exclusão da sazonalidade semanal
Exclusão de 6 componentes senoidais (períodos: 783,7, 21,8, 4,36, 3,63, 5,44 e 10,89 dias)
Exclusão da sazonalidade mensal (dias úteis e calendário)

O gráfico dos valores relativos aos dias úteis na ordem direta e inversa está apresentado abaixo.

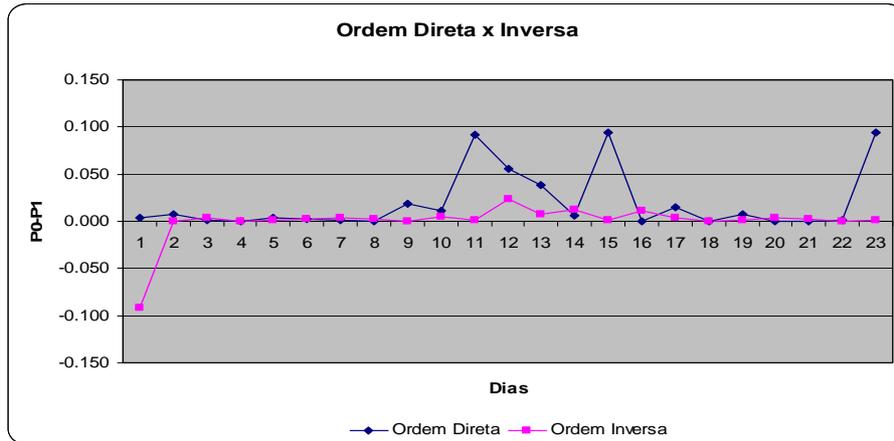


Gráfico 35: Análise dos dias úteis pela ordem direta x inversa.

Em seguida tem-se o gráfico dos dias calendário, também na ordem direta e inversa.

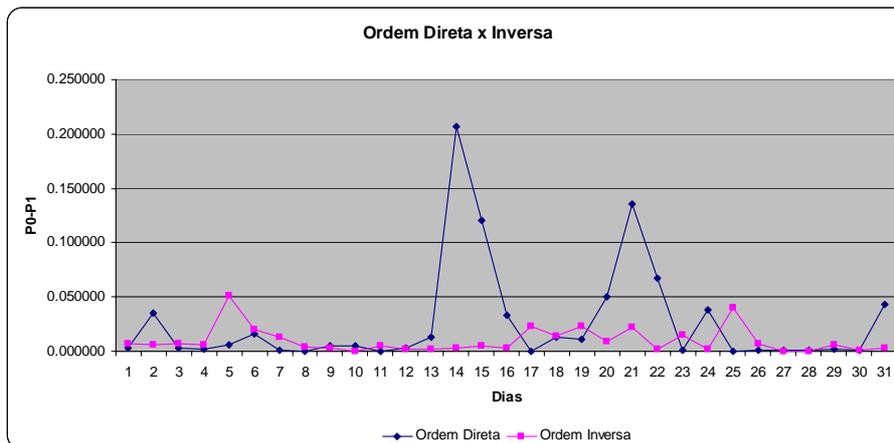


Gráfico 36: Análise dos dias calendário pela ordem direta x inversa.

3.7.2 Teste do qui-quadrado

Como mencionado anteriormente, um dos primeiros procedimentos para se avaliar a aleatoriedade de um conjunto de dados consiste em provar que o conjunto é uniforme nos intervalos, ou seja, que a distribuição dos dados não segue a curva normal de probabilidades.

De uma maneira geral pode-se dizer que dois grupos se comportam de forma semelhante se as diferenças entre as frequências observadas e as esperadas em cada categoria forem muito pequenas, próximas a zero.

O teste trabalha com duas hipóteses:

H_0 : não há associação entre os grupos

H1: há associação entre os grupos

Se, para um dado nível de significância, o valor do qui-quadrado for superior ao valor crítico da estatística do qui-quadrado (definido em tabela), rejeita-se a hipótese *H0*.

No caso das séries estudadas (STR-Clientes + CIP e série aleatória) foram utilizados vários intervalos de frequência no cálculo (5, 10, 20 e 50), ou seja, distribuiu-se o conjunto de dados em 5, 10, 20 e 50 grupos. Adotou-se o nível de significância igual a 0,05, como usual.

Todos os conjuntos da série sob estudo foram rejeitados, assim como os gerados pela função rand do Matlab, aceitos, conforme pode ser visto no quadro a seguir.

Quadro 15: Teste qui-quadrado para significância de 0,05 e várias frequências.

		STR + CIP		Valores aleatórios	
Freq.	Valor tabelado	Qui-quadrado	Situação registro	Qui-quadrado	Situação registro
5	9,49	1.641,33	rejeitado	4,05	aceito
10	16,92	2.105,21	rejeitado	8,35	aceito
20	30,14	2.204,37	rejeitado	24,00	aceito
50	66,34	2.274,67	rejeitado	104,49	rejeitado

Vê-se, portanto, que a série STR+CIP não pode ser considerada desassociada, baseado no estudo proposto por Karl Pearson, ou seja, para um nível de confiança de 95% os números referentes às sequências em estudo não são uniformemente distribuídos.

3.7.3 Teste de Kolmogorov-Smirnov

O princípio deste teste baseia-se na comparação da curva da frequência cumulativa dos dados, com a função de distribuição teórica em hipótese. Quando as duas curvas se sobrepõem à estatística de teste é calculada através da máxima diferença entre ambas. A magnitude da diferença estabelece-se probabilisticamente, segundo a lei de probabilidade dessa estatística, que se encontra tabelada. Em outras palavras, o teste de Kolmogorov-Smirnov mede a distância máxima entre os resultados de uma distribuição empírica a ser testada e os resultados associados à distribuição

hipoteticamente verdadeira. Assim, o teste avalia se duas amostras têm distribuições semelhantes, ou seja, se foram extraídas de uma mesma população. (MORCILLO, 2004).

O princípio básico para realização do teste é comparar as frequências acumuladas das duas amostras. Dessa forma, os seguintes passos devem ser adotados:

- Primeiramente ordenam-se as duas amostras.
- Constroem-se as distribuições de frequências acumuladas nos intervalos de classe de cada amostra.
- Calculam-se as diferenças entre as frequências acumuladas de cada amostra em cada um dos intervalos de classe. Este capítulo está estruturado sempre considerando as diferenças entre a primeira e a segunda amostra (A-B).
- Escolhe-se a maior diferença $[D_{máx}]$ que será comparada com $D_{crítico}$ (valor este tabelado).
- Se $D_{máx} \geq D_{crítico}$ rejeita-se a hipótese de igualdade das amostras.

Abaixo estão os resultados encontrados para o teste de Kolmogorov-Sminorv aplicado sobre os dados da série em estudo e sobre o arquivo gerado no Matlab.

Quadro 16: Teste de K-S com nível de significância 0,05.

	Valor tabelado	Cálculo K-S	Hipótese
STR + CIP	0,048286	0,258052	rejeitada
Aleatória	0,048286	0,500208	rejeitada

O teste de Kolmogorov-Smirnov implica em uma distribuição uniforme dos dados, característica esta necessária para considerar uma série aleatória. Entretanto, verifica-se que todas as séries foram rejeitadas, significando, portanto, que, baseado na distribuição Kolmogorov, nenhuma delas apresenta características de aleatoriedade.

Ao se comparar os testes do qui-quadrado e de Kolmogorov-Smirnov, verifica-se que ambos indicam se o gerador de aleatórios segue a distribuição pretendida. Entretanto, existem algumas diferenças entre eles e um exemplo é o fato de o teste de Kolmogorov-Smirnov se aplicar a distribuições contínuas, enquanto que o teste do qui-quadrado se aplica a distribuições discretas (dividindo-se o conjunto em

intervalos). Este último é muito sensível ao número de classes, conforme pode-se verificar no experimento anterior onde para os intervalos de 5, 10 e 20 a série aleatória foi aceita e para o intervalo de 50 rejeitada.

Outra diferença é a de que o teste do qui-quadrado necessita de grupos de dados, ao contrário do KS. Assim, a hipotética distribuição contínua do teste de Kolmogorov-Smirnov permite examinar o ajuste para toda a amostra, tendo uma avaliação mais completa da amostra. Sob esta ótica verifica-se que a série STR-CIP apresenta uma aderência ao modelo KS maior que a série aleatória do Matlab.

3.7.4 Teste de repetições de 2 tipos de observações

Quando os dados de uma série pertencem a somente 2 categorias (positivo/negativo, branco/preto, quente/frio etc) pode-se utilizar este teste a fim de se verificar a aleatoriedade da sequência. Para se tomar uma decisão neste sentido, é necessário se contar o número de repetições nos dados e compará-lo com o número esperado de repetições.

Quadro 17: Teste de repetições de 2 observações, com significâncias 0,05 e 0,01.

	Significância		STR + CIP	Aleatória
	0,05	0,01		
Repetições	$\pm 1,96$	$\pm 2,58$	-0,84	-0,78

Vê-se, no quadro acima, que os valores apurados nos testes de repetições ficaram dentro do intervalo para os níveis de significância 0,01 e 0,05, caracterizando, desta forma, a aleatoriedade dos dados.

3.7.5 Teste de repetições acima e abaixo da mediana

Quando os dados em estudo são numéricos, um método comum de teste da aleatoriedade de uma sequência focaliza quaisquer padrões que possam formar em relação à mediana amostral. Isto é, cada observação é rotulada como acima (A) ou abaixo (B) da mediana amostral e ignoram-se os valores que coincidem com a mediana (STEVENSON, 2001).

Após os cálculos efetuados para as séries sob estudo, obteve-se os seguintes resultados.

Quadro 18: Teste de repetições acima e abaixo da mediana, com significâncias 0,05 e 0,01.

	Significância		STR + CIP	Aleatória
	0,05	0,01		
Mediana	±1,96	±2,58	-0,57	-0,93

Os valores apurados nos testes de repetições acima e abaixo da mediana também ficaram dentro do intervalo para os níveis de significância 0,01 e 0,05 demonstrando que, para este teste, os dados são considerados aleatórios.

3.7.6 Teste de repetições acima e abaixo em relação ao anterior

Por último, tem-se o teste que utiliza as repetições para cima e para baixo como medida de aleatoriedade. Mede-se cada valor da sequência em relação ao valor anterior. Cada valor numérico é substituído por U (para cima) ou por D (para baixo), isto em relação ao valor anterior (STEVENSON, 2001).

Após os cálculos do teste obteve-se os seguintes resultados.

Quadro 19: Teste de repetições para cima e para baixo, com significâncias 0,05 e 0,01.

	Significância		STR + CIP	Aleatória
	0,05	0,01		
Cima/Baixo	±1,96	±2,58	-1,50	-0,03

Das séries analisadas tanto a STR-CIP como a gerada pelo Matlab seriam aceitas como aleatórias para todos os níveis de significância.

3.8 CONCLUSÕES

Este capítulo buscou verificar se o comportamento do resíduo da série STR + CIP e da série gerada pelo Matlab (para alguns testes) poderiam ser classificadas como aleatórias pois a simples observação delas no tempo evidencia padrões bastante irregulares. Assim, numa avaliação preliminar, apesar de o resíduo da série STR + CIP não ser caracterizado como aleatório pelos testes, o mesmo ocorrendo para a série gerada pelo Matlab, ambas parecem seguir um caminho quase-aleatório com alto grau de imprevisibilidade, sem quaisquer tendências pronunciadas, não indicando ajuste às principais formulações determinísticas.

Uma série verdadeiramente aleatória implica na inexistência de correlação entre os números, distribuição uniforme no espaço além de outras propriedades que foram descritas anteriormente. Dessa forma, a aplicação de um único teste para se avaliar o grau de aleatoriedade de uma série pode não determinar, tacitamente, a característica aleatória dos dados.

Em vista do exposto, optou-se por se fazerem 6 testes distintos (autocorrelação, qui-quadrado, Kolmogorov-Smirnov, além de 3 testes de repetições), de tal forma que, na hipótese de ser verificada aleatoriedade em todos os testes poderia-se considerar a série como possivelmente aleatória. Entretanto, constatou-se, através dos testes promovidos, que nenhuma das séries apresentaram características de aleatoriedade, pois todas elas, inclusive a gerada pelo Matlab, foram rejeitadas em pelo menos um dos testes de avaliação.

As análises e testes promovidos permitem tecer alguns comentários a respeito dos dados estudados.

- Teste de autocorrelação: Este teste procurou detectar a existência ou não de ciclos entre os valores no tempo. A série STR + CIP apresentou índices de variabilidade próximos de zero, demonstrando baixo grau de previsibilidade. Tiveram alguns casos em que o índice de variabilidade é alto para determinado dia sem, no entanto, permanecer alto nos dias subsequentes. Mesmo nos casos em que se encontra valores significativos em relação aos demais, a magnitude destes coeficientes é pequena o suficiente para se considerar aceitável a suposição de previsibilidade do comportamento da série.
- Teste do qui-quadrado: objetivou verificar o ajuste da distribuição à curva normal. Os resultados mostraram que para as frequências e nível de significância aplicadas, às distribuições das séries não se ajustam à distribuição Normal, a exceção dos primeiros níveis de frequência utilizado para a série gerada pelo Matlab. Por este teste a hipótese de normalidade não pode ser aceita.
- Teste de Kolmogorov-Smirnov: este teste também afastou a hipótese de as distribuições empíricas aproximarem-se significativamente da Normal

teórica, descartando a suposição de aleatoriedade. Neste caso, ambas as séries foram rejeitadas.

- Testes de repetições: aceitaram a hipótese de aleatoriedade para todos os níveis de significância aplicados e para todas as séries. Algumas observações podem ser formuladas para estes tipos de testes:
 - ◇ O teste de dois tipos de observações demonstrou que não se têm padrões de sequências bem caracterizados para os valores positivos e negativos da série;
 - ◇ O teste acima e abaixo da mediana também caracterizou os dados como aleatórios. Pode-se, com isto, afirmar que os conjuntos dos dados não apresentam ciclos bem definidos;
 - ◇ O teste acima e abaixo em relação ao anterior também apresentou índices de aleatoriedade para ambas as séries demonstrando a inexistência de tendências de crescimento ou de decrescimento das séries.

Constatou-se que as séries são constituídas basicamente pela componente “irregularidade”, que além de ser de alta frequência apresenta um elevado índice de dispersão (variância). Apesar de não haver sido constatada, inequivocamente, a aleatoriedade da série, a análise dos resultados mostrou que a irregularidade existente nas séries é muito forte.

Conforme já discutido, a existência de apenas um teste demonstrando um padrão não-aleatório nos dados já é suficiente para descartar a hipótese de aleatoriedade da série. Por outro lado, a aceitação da hipótese de aleatoriedade por alguns testes demonstra claramente que a série possui um comportamento não totalmente previsível e aponta na direção da dificuldade de se desenvolver modelos determinísticos de previsão baseados em características passadas dos dados.

4 MODELOS NEURAIIS

4.1 INTRODUÇÃO

Um dos principais instrumentos utilizados para se descrever um fenômeno é o modelo. Os modelos são versões simplificadas de algum problema ou situação da vida real. São usados para ilustrar certos aspectos da situação, evitando grande número de detalhes que talvez sejam irrelevantes para o problema (STEVENSON, 2001). Portanto, a construção de um modelo privilegiará as informações mais relevantes do problema em detrimento daquelas que, num primeiro momento, se mostrem menos importantes.

Modelos neurais são desenvolvidos de forma empírica, ou seja, partindo-se de um modelo básico vai-se calculando o erro médio quadrático (MSE – *Mean Squared Error*) dos conjuntos de treinamento (70% dos registros), validação (15% dos registros) e teste (15% dos registros) e ajustando os diversos parâmetros da rede até se chegar a uma solução ideal para o problema.

Dessa forma, o desenvolvimento dos modelos de previsão criados nesta tese passaram por várias etapas de refinamento, conforme será apresentado a seguir. Além disso, cada topologia de rede implementada foi executada 5 vezes sendo que o resultado apresentado corresponde a média aritmética dos experimentos. Cabe comentar, também, que o arquivo de entrada das redes é constituído pelo resíduo oriundo da extração das componentes determinísticas (tendência, sazonalidade e ciclo).

A fim de melhor visualizar as etapas de processamento implementadas nos dados, adiante tem-se um diagrama geral dos procedimentos. As particularidades acaso existentes serão discutidas quando da apresentação dos modelos.

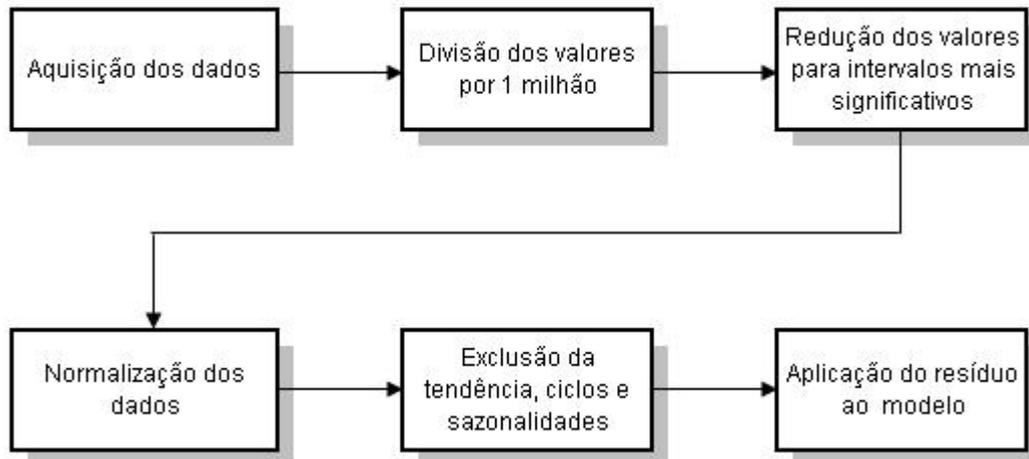


Figura 1: Diagrama de tratamento dos dados.

A fim de evitar o excesso de quadros no corpo da tese, as definições dos arquivos utilizados nas redes e as topologias das redes implementadas foram transportados para o Anexo 5.

4.2 PESQUISAS COM MODELOS NEURAI

Com o surgimento das redes neurais artificiais muitos problemas foram resolvidos ou melhor explicados do que com o uso de outras técnicas. Em função disso, houve na década de 90 uma grande utilização dos modelos neurais em praticamente todos os ramos da ciência. Encontram-se hoje artigos médicos, trabalhos em economia, astronomia, física nuclear, geografia, biologia, genética etc. utilizando modelos neurais para previsão, classificação e reconhecimento de padrões. Entretanto, alguns problemas mais complexos ou não respondem bem ou apresentam uma baixa performance nos resultados quando do uso dos modelos neurais fazendo com que a comunidade científica continue seus esforços visando descobrir novas soluções para os problemas.

Uma das classes de modelos mais difíceis de se mapear é aquela em que as séries possuem altas frequências nos dados. Vários pesquisadores vêm se dedicando à tentativa de se obter modelos robustos para este tipo de série. Congressos são realizados e artigos publicados apresentando, a cada dia, novidades para a modelagem deste tipo de série.

Allemão (ALLEMÃO, 2004) trabalhou com 3 séries envolvendo o comportamento da movimentação diária de numerário de agências bancárias, onde é

nítida a alta frequência dos dados. Apesar de as redes neurais artificiais terem apresentado desvios de menor magnitude quando comparados com aqueles obtidos pela aplicação dos métodos de previsão convencionais ainda assim os resultados não se mostraram satisfatórios a ponto de se poder utilizá-los na logística de distribuição de numerário.

Pasley e Austin (PASLEY; AUSTIN, 2004) também abordaram o problema de alta frequência dos dados. Em seu estudo os autores trabalharam com uma grande base de dados – aproximadamente 100 mil registros – referentes às taxas de câmbio do yen e do dólar americano e, para tanto, utilizaram de técnicas intensivas de previsão. Foi desenvolvido um modelo, denominado *fast k-NN*, onde se implementou uma rede neural binária baseada em matriz de correlação, através de um algoritmo próprio, chamado AURA (*Advanced Uncertainty Reasoning Architecture*) que é similar a um *perceptron feedforward* com uma única camada altamente conectada e com pesos binários. Foi considerado promissor o resultado alcançado de 56,9% de acerto.

Nguyen (NGUYEN; CHAN, 2004) trabalhando com redes neurais simples e múltiplas e modelo ARIMA obteve melhores resultados na previsão de curto prazo com o modelo de Box e Jenkins. Para previsões mais longas os melhores resultados foram originados das redes neurais múltiplas. Os dados tratados por Nguyen não podem ser classificados como de alta frequência pois trata-se de série relativa ao consumo de gás em localidade no Canadá onde a demanda se comporta de uma forma mais suave que aquela encontrada em algumas séries financeiras. Entretanto, chamou a atenção o fato de as redes não terem respondido bem no curto prazo demonstrando que, apesar de a série haver sido mapeada com baixo índice de erro (RMSE inferior a 0,06 nos dados de teste), a previsão não correspondeu às expectativas.

Taxa de erro baixa durante o treinamento da rede demonstra que as características da função foram devidamente mapeadas. Entretanto, durante as previsões a performance da rede pode não responder bem aos dados (ALLEMÃO, 2004) (NGUYEN; CHAN, 2004). É provável que este problema seja oriundo da pouca quantidade de dados ou alta frequência fazendo com que os padrões anteriores não se repitam gerando, em consequência, baixo índice de acertos.

Tentativas utilizando clusterização de dados antes de se promover a previsão foi utilizada por Lokmic e Smith (LOCKMIC; SMITH, 2000), que trabalharam com

uma das variáveis do fluxo de caixa de uma instituição financeira. O objetivo do trabalho era o de prever a apresentação de cheques pelos clientes para com isto obter a necessidade de caixa. Para tanto, na fase de pré-processamento, os cheques foram clusterizados em 3 grupos – valores baixos, médios e altos – utilizando-se uma rede de Kohonen e, posteriormente, com uma MLP (*MultiLayer Perceptron*) tentou obter os valores diários de apresentação dos cheques. Os resultados não foram satisfatórios pois, segundo os autores, além do número de variáveis envolvidas no problema ser muito limitado não se tem conhecimento dos padrões existentes nos grupos segmentados nem as características dos clientes que emitiram os cheques.

A quantidade diária de numerário em circulação representa um dos mais importantes fatores que influenciam a liquidez de mercados financeiros. Baseado neste conceito, Hlaváček *et al.* (HLAVÁČEK; CADA; HAKL, 2005) propuseram um modelo neural chamado *structured feedforward neural network* para tratar o assunto. Este modelo é derivado das redes com unidades de chaveamento apresentado originalmente por Bitzan (*apud* HLAVÁČEK; CADA; HAKL, 2005) e se propõe a dividir o espaço de entrada em conjuntos disjuntos (clusters) onde são aplicados métodos de regressão linear, processos estocásticos e redes neurais. Comparando este modelo com o ARIMA o autor encontrou uma ligeira melhora nos resultados oriundos das redes neurais.

O uso de comparações entre redes neurais e modelos auto-regressivos de Box e Jenkins vem sendo muito estudado. Ho *et al.* (HO; XIE; GOH, 2002) fizeram uma análise para o problema de falhas em sistemas de compressores em uma planta visando a detecção do melhor intervalo de tempo existente entre uma ocorrência e outra. A partir do momento em que existe uma confiança em um específico intervalo, há a possibilidade de se planejar a reposição da peça ou o seu conserto antes da parada da máquina, evitando-se, com isto, a ociosidade de mão-de-obra e redução da produtividade. A simulação trabalhou com 90 falhas críticas no período de 1968 e 1989.

Foram utilizadas, no estudo, redes do tipo recorrente e *feedforward*, sendo que as recorrentes obtiveram melhores resultados. Por sua vez, a rede recorrente quando comparada com o modelo auto-regressivo obteve performance similar, levando a conclusão que qualquer dos modelos podem ser aplicados para este tipo de problema.

Ainda seguindo esta linha de comparação entre modelos neurais e auto-

regressivos tem-se o trabalho de Zhang (ZHANG, 2003), onde é proposto um modelo híbrido, através do uso de redes neurais e ARIMA. O autor relaciona 3 razões para a adoção desta abordagem:

1. Geralmente encontra-se dificuldade em se determinar na prática se a série temporal é constituída por função linear ou não-linear ou se um particular método é mais efetivo que outro para a previsão, levando o pesquisador a testar diversos modelos;
2. No mundo real as séries temporais raramente são totalmente lineares ou não-lineares. Elas, frequentemente, contêm ambos os padrões. Neste caso nem o modelo neural nem o ARIMA são exclusivamente aderentes aos dados;
3. Em problemas reais frequentemente as séries apresentadas são complexas e um único modelo pode não capturar os diferentes padrões existentes tão bem como se houvesse um hibridismo associado.

Muitos estudos empíricos sugerem a combinação de vários modelos para se obter melhor performance nos resultados. Encontram-se junções de algoritmos genéticos com redes neurais, redes neurais com lógica *fuzzy*, etc. No artigo citado, Zhang resolveu trabalhar a parte linear dos dados através de modelo ARIMA e a parte não-linear via rede neural pois combinando diferentes modelos, diferentes aspectos existentes nos dados podem ser capturados.

Se considerarmos que uma série temporal é composta por uma estrutura linear de autocorrelação e uma componente não-linear, tem-se, então, duas componentes independentes a serem estimadas. Primeiro utiliza-se o modelo ARIMA para detecção da componente linear e o resíduo composto de relacionamento não-linear será tratado posteriormente com redes neurais artificiais.

Foram usadas três conhecidas bases de dados para os testes: a primeira constituída de 288 observações relativas ao período de 1700 à 1987 referentes as quantidades anuais de explosões solares. Esta série é considerada como sendo não-linear e não gaussiana; a segunda série contém 114 observações relativas às quantidades anuais, no período de 1821 à 1934, de captura de lincas no Canadá e, por último, a 3ª. série corresponde à taxa entre a libra esterlina e o dólar americano no período de 1980 a 1993 e é constituída de 731 registros semanais. Estas séries são muito utilizadas em

estudos estatísticos e de não-linearidade.

Em todas as séries os resultados se mostraram melhores utilizando-se o modelo híbrido ao invés dos modelos ARIMA e de redes neurais isoladamente. Cabe comentar que o ganho obtido na 3ª série entre as redes neurais e o modelo híbrido foi residual, notadamente para períodos maiores, de 6 e 12 meses a frente. Entretanto, ainda assim, houve melhoria. Quanto ao modelo ARIMA, individualmente a sua performance foi inferior tanto com relação às redes neurais como quando comparado ao modelo híbrido.

Gao *et al.* (GAO; KINOUCI; ITO; ZHAO, 2003) no artigo *Time Series Identifying and Modeling with Neural Networks* propõem a criação de um modelo neural para previsão da série TOPIX (*Tokyo Stock Price Index*). Para tanto foi utilizado o movimento diário no período de 1996-2001, sendo que os anos de 1996 a 1998 foram usados para treinamento e o resto para validação do modelo.

Segundo os autores, as séries contêm variados graus de ruído que foram tratados com métodos estatísticos clássicos para, posteriormente, treinar várias redes do tipo *feedforward* até se alcançar aquela que melhor os representasse. Aqui cabe comentar que os autores utilizaram um método desenvolvido por eles e denominado R-Método para definir a melhor arquitetura.

Para aumentar o grau de precisão nos resultados, o erro encontrado no treinamento, considerado como resíduo aleatório, foi utilizado como entrada de uma nova rede, com a mesma arquitetura da principal. Como o foco do artigo é apresentar alternativas de modelagem não existem resultados nem comparações com outros métodos de previsão. Também não são feitas simulações com outras séries. Entretanto, eles afirmam que todas as características da série foram capturadas através desta metodologia, corroborando o fato de redes neurais serem uma boa opção para o mapeamento de séries complexas e ruidosas.

Em outro artigo, Gao *et al.* (GAO; KINOUCI; ITO; ZHAO, 2005) abordaram o problema de se realizar previsões de uma série temporal referente ao mercado de ações da bolsa de Tóquio, chamado de TOPIX (*Tokyo Stock Price Index*). Segundo os autores o tratamento de séries temporais financeiras, notadamente àquelas relativas ao mercado acionário, carregam consigo uma dificuldade adicional que se trata do tamanho da série. Considerando, as questões políticas, o ambiente em que se insere

o mercado e outras variáveis exógenas que influenciam a formação do preço das ações, o uso de séries de períodos grandes não são ideais para representarem o passado, visando projeções do futuro.

Pode-se afirmar que o mesmo acontece com o fluxo de caixa das empresas, principalmente de bancos que são muito sensíveis à política do governo, conjunturas externas, inflação, reajustes de salários etc. Caberá, portanto, ao pesquisador dedicar espaço na definição do tamanho ideal da série. Esta proposta de tese trata de uma série com início em abril de 2002, quando da reestruturação do SPB, que ocasionou uma grande mudança em todo o Sistema Financeiro Nacional.

Em ‘*ANN for non-stationary time series*’ (KIM *et al.*, 2004) os autores destacam que séries não-estacionárias não são passíveis de modelagem, ou seja, há necessidade de se existir alguma regularidade a fim de aplicar os métodos de previsão clássicos. Estacionariedade de uma série é a independência relativamente à origem dos tempos. Um processo é dito estacionário se todas as distribuições permanecem as mesmas no tempo, ou seja, possuem a média e a variância constantes (MORETTIN, 1987).

Kim *et al.* (2004) basearam seus experimentos na série KOSPI (*Korea Stock Price Index*), notadamente não-estacionária, durante o período de 1994 a 2002. Deste conjunto foram extraídos três subconjuntos de dados, referentes aos anos de 1995, 1997 e 1999 e aplicados sobre eles 3 diferentes redes do tipo MLP. Os resultados se mostraram promissores levando o autor a afirmar que o uso de RNA é uma chave para o sucesso na análise de complexas séries temporais financeiras.

4.3 MODELOS NEURAI APLICADOS

As redes neurais artificiais são classificadas, segundo Simon Haykin (HAYKIN, 1999), em 3 categorias: *single-layer feedforward networks*, *multilayer feedforward networks* e *recurrent networks*. Nesta pesquisa utilizaram-se redes do tipo *multilayer feedforward*.

Não se tem ainda nenhuma fórmula capaz de prever qual a topologia a ser empregada visando à obtenção do melhor resultado para um determinado problema. Muitos estudos são feitos mostrando soluções para casos específicos, entretanto nenhum pesquisador conseguiu ainda definir parâmetros a serem adotados que resultem em uma

rede neural ótima, haja vista a enorme gama de variáveis a serem ajustadas, tais como: quantidade de camadas escondidas, quantidade de neurônios por camada, função de ativação, algoritmo de treinamento, quantidade de dados de entrada da rede, normalização de dados e medidas de performance entre outros.

Nesta tese foram utilizadas redes do tipo MLP – *MultiLayer Perceptron*, da classe de redes *multilayer feedforward networks*. Tal escolha se deveu ao fato de serem as mais utilizadas em problemas envolvendo previsão de séries temporais (ZHANG; PATUWO; HU, 1988) (BRAGA; CARVALHO; LUDEMIR, 1998).

Durante o andamento da pesquisa vários parâmetros foram alterados, visando à melhoria de resultados. Utilizou-se nas redes neurais implementadas, com maior frequência, 1 camada intermediária de neurônios do tipo tangente hiperbólica, pelo fato desta arquitetura ser um aproximador universal e ter se mostrado eficiente em diversos estudos pesquisados. O uso de 2 camadas também foi experimentado, pois as redes criadas desta forma também possuem a propriedade de aproximar qualquer função (CYBENKO *apud* BRAGA *et al.*, 1998) (KOLMOGOROV *apud* KÓVÁCS, 1996) (HORNIK *apud* ZHANG; PATUWO; HU, 1998).

Por fugir ao escopo desta tese, as arquiteturas e variações utilizadas nos modelos que foram rejeitados em função de não terem apresentado boa resposta serão comentadas, quando necessárias, somente com o intuito de conduzir a linha de raciocínio sem, no entanto, se ater aos detalhes.

O desempenho do treinamento das redes foi medido através do MSE (*Mean Squared Error*). Foram alterados, também, os parâmetros de momento, funções de treinamento, quantidade de camadas escondidas, mudanças nos fatores de incremento e decremento das taxas de aprendizado e número de épocas visando à melhoria da performance da rede.

Outra métrica utilizada para avaliar o desempenho foi desenvolvida em conjunto com os analistas do banco. As previsões realizadas atualmente pelo banco permitem uma margem de erro de R\$ 200 milhões no conjunto das variáveis do fluxo da movimentação das reservas bancárias. Considerando que este valor se aplica a todo o universo, optou-se por se calcular os percentuais de acertos para erros até R\$ 50, R\$ 100, R\$ 150 e R\$ 200 milhões.

Por fim, como forma de acompanhamento das transformações realizadas foi

utilizado o conceito de potência apresentado no Capítulo 2, quando da discussão a respeito da decomposição da série. A potência dos dados, definida como a média do somatório dos registros ao quadrado, permeou todo o trabalho. A cada passo realizado avaliava-se a potência com o intuito de verificar se houve aumento (inserção de ruído) ou redução do valor (exclusão de informação). Caso apresentasse aumento, a linha era, normalmente, descartada, a não ser que indícios apontassem para a continuação. Quando se verificava redução, dava-se prosseguimento ao estudo, partindo-se para outra etapa.

Algumas características das redes são válidas em todos os experimentos e, por isto, não serão apresentadas a cada modelo. Em todos os modelos, por exemplo, foi utilizado o algoritmo de treinamento *backpropagation* e as redes possuíam apenas 1 único neurônio linear na camada de saída para representar a variável. Os pares entrada-saída foram montados com um intervalo de 2 dias, a fim de representar a previsão para $t + 2$. Aqui cabe esclarecer que por impossibilidade operacional não há interesse de se tratar a previsão para $t + 1$ pois os valores do fluxo de caixa são compilados apenas ao final do dia e a previsão necessariamente tem que ser informada para o departamento responsável pela negociação financeira do banco no início da manhã. Ou seja, até às 9h a previsão tem que estar definida a fim de que se possam fazer todas as transações financeiras necessárias ao equilíbrio do caixa do banco. Entretanto, só ao final do dia é que se consegue compilar o movimento do dia anterior, pois é necessário realizar o processamento de milhões de lançamentos e contabilização do resultado em centenas de variáveis. Dessa forma, há necessidade de se realizar a previsão para, no mínimo, $t + 2$.

Antes de iniciar a apresentação dos modelos neurais implementados, cabe comentar que o arquivo de entrada da rede sofreu na etapa de pré-processamento várias transformações. Dentre elas, talvez a mais importante para o propósito desta tese seja a exclusão das componentes determinísticas.

O resíduo ou irregularidade encontrado após as transformações será o objeto de estudo. Conforme comentado, não raro houve a necessidade de se retornar ao ponto inicial e refazer todo o pré-processamento, às vezes em função de aumento da base de dados, outras a fim de se alterar as etapas com inclusão ou exclusão de técnicas para redução da informação existente na base. Ou seja, baseado nos resultados da modelagem neural do resíduo novas abordagens foram utilizadas na etapa de pré-

processamento com o intuito de melhorar a performance do modelo.

Extração de Componentes Determinísticas 01

Após a extração da parte determinística da série, apresentada no quadro abaixo, o resíduo resultante pode ser considerado como o erro de previsão. Assim, a sua reconstituição através da inserção dos valores subtraídos fornece o erro na mesma escala dos dados originais.

Quadro 20: Transformações da base de dados.

Transformações da base de dados
Divisão dos valores por 1 milhão
Redução dos valores para o intervalo [-1000, 2000]
Normalização dos dados
Exclusão da tendência
Exclusão da primeira componente de Fourier, referente ao período de 756 dias
Exclusão das semanas incompletas e com início e fim em meses diferentes
Potência da série = 0,219079

Com a retirada das semanas incompletas e das semanas que se iniciam em um mês e terminam em outro, a base original com 756 registros, referente ao período de 01/04/2003 a 31/03/2006, ficou reduzida a 575 registros.

Adiante tem-se a quantidade de registros (Qtd.) e percentuais (%) em relação ao total da base cujos resíduos foram inferiores ou iguais aos limites de 50, 100, 150 e 200 milhões.

Quadro 21: Estatística de acertos após extração da parte determinística da série.

Até R\$ 50 milhões		Até R\$ 100 milhões		Até R\$ 150 milhões		Até R\$ 200 milhões	
Qtd.	%	Qtd.	%	Qtd.	%	Qtd.	%
64	11,1	137	23,8	211	36,7	265	46,1

Modelo Neural 01

O arquivo utilizado neste experimento, oriundo das transformações descritas no Anexo 5, foi reduzido para 465 registros quando da preparação dos pares entrada-saída (Treinamento = 307, Validação = 79 e Teste = 79). Isto aconteceu em razão da definição de que os pares deveriam estar, necessariamente, em sequência, ou seja, com a exclusão de algumas semanas da base nem sempre ocorria do par estar sequencialmente ordenado e, nestes casos, os registros eram retirados. A seguir tem-se os resultados.

Quadro 22: Performance do modelo 01.

Número de neurônios	Até R\$ 50 milhões		Até R\$ 100 milhões		Até R\$ 150 milhões		Até R\$ 200 milhões	
	Qtd.	%	Qtd.	%	Qtd.	%	Qtd.	%
1	7,4	9,4	16,2	20,5	25,2	31,9	34,6	43,8
2	7,8	9,9	16,6	21,0	26,4	33,4	35,0	44,3
5	8,4	10,6	18,4	23,3	27,0	34,2	33,4	42,3
10	9,6	12,2	16,4	20,8	25,0	31,6	31,8	40,3

Verifica-se no quadro acima que o uso de rede neural, nesta situação, não se mostrou apropriado. Ao contrário, vê-se que antes da sua utilização os resultados haviam alcançado, somente com a extração das componentes determinísticas, índices de previsão superiores.

Cabe comentar que o aumento do número de neurônios não apresentou melhora significativa nos resultados, assim como também não se vislumbrou melhora com o aumento das épocas de treinamento que foram aumentadas durante o treinamento, alcançando o limite de 10.000 épocas.

Para efeito de análise do desempenho da rede, após a etapa de treinamento foi calculada a diferença entre os valores de entrada e saída dos registros de treinamento, a fim de verificar o seu desempenho quanto ao mapeamento dos dados. Abaixo tem-se o quadro com os resultados alcançados e nota-se que a performance é similar àquela alcançada no teste, quando deveria ser superior, haja vista que na etapa de treinamento é apresentada à rede tanto os valores de entrada como de saída.

Quadro 23: Performance dos dados de treinamento do modelo 01.

Número de neurônios	Até R\$ 50 milhões		Até R\$ 100 milhões		Até R\$ 150 milhões		Até R\$ 200 milhões	
	Qtd.	%	Qtd.	%	Qtd.	%	Qtd.	%
1	38,0	12,4	74,8	24,4	107,6	35,0	134,4	43,8
2	38,6	12,6	75,4	24,6	109,6	35,7	134,8	43,9
5	40,8	13,3	74,4	24,2	109,4	35,6	137,4	44,8
10	41,8	13,6	78,0	25,4	110,8	36,1	138,0	45,0

Extração de Componentes Determinísticas 02

Este pré-tratamento diferiu do anterior somente pela extração da sazonalidade semanal da série, ou seja, antes de se criar os conjuntos de treinamento, validação e teste calculou-se o valor médio de cada dia da semana – segundas-feiras, terças-feiras etc – e subtraiu-se essa média dos dias correspondentes na base de dados.

Após este cálculo, foram criados os conjuntos para utilização na rede neural.

Similar ao procedimento anterior, calculou-se a parte determinística extraída da série através da reconstituição do resíduo (erro). Os resultados podem ser avaliados adiante.

Quadro 24: Estatística de acertos após extração da parte determinística da série.

Até R\$ 50 milhões		Até R\$ 100 milhões		Até R\$ 150 milhões		Até R\$ 200 milhões	
Qtd.	%	Qtd.	%	Qtd.	%	Qtd.	%
72	12,5	149	25,9	219	38,1	275	47,8

Nota-se que houve melhora em relação ao estudo anterior quando a sazonalidade semanal não havia sido extraída. A potência da série que anteriormente era de $P_{ant} = 0,219079$ passou para $P_{atual} = 0,212572$, confirmando o acerto da decisão.

Modelo Neural 02

Ainda seguindo o mesmo conceito aplicado ao modelo neural 01, aqui também foram utilizados pares entrada-saída sequenciais e as semanas incompletas ou com início e fim em meses diferentes foram retiradas do conjunto. Os resultados da aplicação do modelo estão descritos abaixo e as transformações dos dados encontram-se no Anexo 5.

Quadro 25: Performance do modelo 02.

Número de neurônios	Até R\$ 50 milhões		Até R\$ 100 milhões		Até R\$ 150 milhões		Até R\$ 200 milhões	
	Qtd.	%	Qtd.	%	Qtd.	%	Qtd.	%
1	10,0	12,7	19,2	24,3	26,8	33,9	32,8	41,5
2	9,2	11,6	17,8	22,5	25,2	31,9	31,8	40,3
5	8,8	11,1	17,2	21,8	26,0	32,9	32,4	41,0
10	8,6	10,9	17,2	21,8	24,2	30,6	30,4	38,5

Apesar da melhora da potência obtida com a retirada da sazonalidade semanal, o mesmo não se refletiu nos resultados da rede. O aumento das quantidades e percentuais de acerto, quando existentes, é estatisticamente insignificante. Além disso, não se pode deixar de comentar o fato de estes resultados serem inferiores aqueles alcançados somente com a extração das componentes determinísticas. Isto sugere, conforme discutido no capítulo 3, que a série possui uma forte componente irregular cujo mapeamento é de difícil solução.

Modelo Neural 03

Considerando a dinâmica existente no segmento bancário, onde o comportamento geral do movimento se altera constantemente em função de fatores internos e externos, optou-se pela criação dos conjuntos de treinamento, validação e testes em ordem aleatória e não sequencialmente como utilizado nos modelos anteriores, que poderia, em função dos dados utilizados, gerar um treinamento com característica distinta do conjunto de teste. A aleatoriedade da ordem dos conjuntos busca refletir as alterações verificadas nos dados no decorrer do tempo e fornecer subsídios para um melhor desempenho da rede.

Para a realização deste experimento foi utilizado o mesmo arquivo de dados do modelo anterior e os resultados estão compilados abaixo.

Quadro 26: Performance do modelo 03.

Número de neurônios	Até R\$ 50 milhões		Até R\$ 100 milhões		Até R\$ 150 milhões		Até R\$ 200 milhões	
	Qtd.	%	Qtd.	%	Qtd.	%	Qtd.	%
1	8,8	11,1	20,6	26,1	28,0	35,4	36,2	45,8
2	11,4	14,4	19,4	24,6	31,8	40,3	38,8	49,1
5	10,8	13,7	21,0	26,6	32,4	41,0	38,6	48,9
10	9,4	11,9	20,4	25,8	31,4	39,7	41,0	51,9

A abordagem utilizada se mostrou adequada, apresentando um ganho em relação ao modelo anterior.

As performances medidas através do MSE tem permanecido estáveis nos três conjuntos de dados durante todos os experimentos (vide quadro de performance deste modelo abaixo). Cabe relatar que apesar de haver sido escolhido para apresentar os resultados apenas os modelos com 1, 2, 5 e 10 neurônios na camada escondida, experimentos com até 50 neurônios foram realizados e nenhum ganho significativo foi observado, conforme pode ser visto no quadro, bem como na representação gráfica do treinamento de uma das redes com 1 neurônio na camada escondida.

Quadro 27: MSE de treinamento, validação e teste da rede 03.

Número de neurônios	Treinamento	Validação	Teste
1	0,227229	0,184675	0,259082
2	0,228364	0,180575	0,266795
5	0,224024	0,181777	0,264779
10	0,221392	0,185445	0,267193

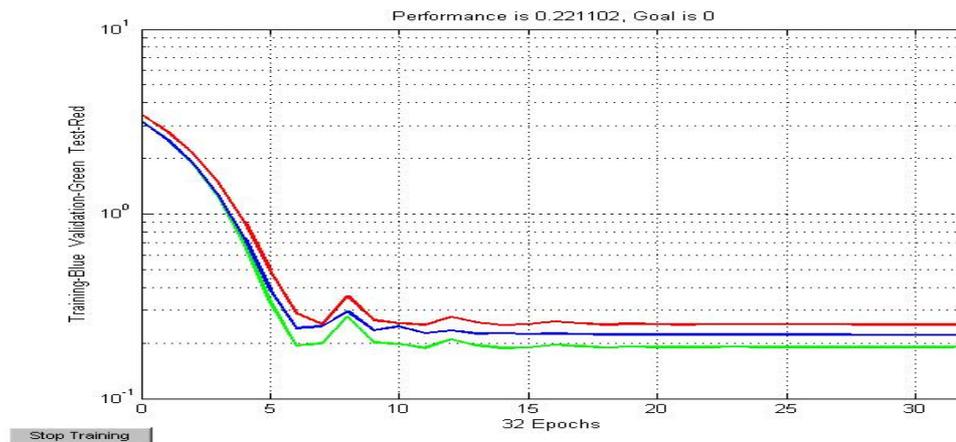


Gráfico 37: Performance do modelo com 1 neurônio na camada escondida.

A definição da quantidade ideal de neurônios em uma rede neural é um assunto controverso em que os pesquisadores ainda não conseguiram chegar a um denominador comum. Várias propostas são feitas e, entretanto, não há consenso entre as diversas sugestões apresentadas.

Para citar alguns exemplos, têm-se, nos casos de apenas 1 camada escondida, as sugestões de Lippmann e Hecht-Nielsen de se usar “ $2n + 1$ ” neurônios, onde n corresponde ao número de variáveis de entrada, já Wong propõe “ $2n$ ” neurônios, Tang avaliza o uso de “ n ” como o ideal e Kang “ $n/2$ ” neurônios. Todos relacionados no artigo de Zhang (ZHANG; PATUWO; HU, 1998).

Zhang *et al.* (1998) afirmam que o caminho mais comum para se determinar o número de neurônios é via experimentos ou por tentativa e erro. Cichocki segue a mesma linha e sugere a criação de redes superdimensionadas (CICHOCKI; UNBEHAUEN, 1995) com a subsequente eliminação dos neurônios em excesso, através do processo de poda.

A escolha do número de neurônios passa, fundamentalmente, pelo empirismo e análise do processo, visando não subdimensionar a rede, o que ocasionaria dificuldades de convergência, nem superdimensioná-la, levando-a a memorizar os padrões de treinamento (*overfitting*) ao invés de extrair as características gerais que permitirão a generalização ou reconhecimento dos padrões não vistos durante o treinamento.

Modelo Neural 04

Uma rede neural artificial do tipo *feedforward* é composta basicamente de camadas de neurônios interconectados. Nesta estrutura está presente todo o conhecimento da rede, ou seja, é na matriz de pesos, modificada durante o treinamento, que a rede retém o conhecimento sobre a função de mapeamento dos dados de entrada nas saídas desejadas.

O processo realizado por cada nó da rede é definido pela combinação dos processamentos realizados pelos nós da camada anterior que estão conectados a ele. As funções implementadas definem como é realizada a divisão do espaço de decisão.

Em 1957 Kolmogorov demonstrou que funções contínuas de 3 variáveis podem ser representadas como superposições e composições de funções de 1 variável (NASCIMENTO; YONEYAMA, 2002) (KOVÁCS, 1996). Baseado nesta teoria Cybenko (*apud* ZHANG; PATUWO; HU, 1998), em 1989, demonstrou que redes neurais com uma camada de neurônios com funções de transferência não lineares e uma camada com funções de transferência lineares são capazes de aproximar qualquer função contínua. Acrescentando-se uma camada adicional pode-se aproximar qualquer função matemática com a acurácia desejada (BRAGA; CARVALHO; LUDEMIR, 2007) (NASCIMENTO; YONEYAMA, 2002) (ZHANG; PATUWO; HU, 1998). Assim, ainda segundo Cybenko (*apud* ZHANG; PATUWO; HU, 1998):

“...that a network never needs more than two hidden layers to solve most problems including forecasting.”

Uma boa aproximação da função não significa, necessariamente, uma boa previsão. A fase de treinamento pode mapear perfeitamente a função mas se não houver padrões nos dados o processo de previsão será de baixa qualidade.

Este experimento tem por objetivo verificar se a utilização de uma única camada pode apresentar resultados mais consistentes que o uso de várias camadas. Tal premissa se baseia na hipótese de que a obtenção de retas formadas por combinações lineares das entradas pode, na média, apresentar melhores resultados no mapeamento dos dados do que a tentativa de tratamento não-linear.

A rede 04 utilizou o mesmo arquivo de dados construído para o experimento 03. A diferença desta rede para a anterior é que neste modelo não existe camada escondida, mas tão-somente os nós de entrada e a camada de saída com apenas 1

neurônio linear. Além disso, foi excluída, também, a representação binária do dia da semana.

Seguindo procedimento adotado anteriormente, aqui, ao invés de se criar os conjuntos de treinamento, validação e teste em ordem sequencial, optou-se pela criação com ordem aleatória. A fim de tornar os resultados mais robustos, foram criados 3 conjuntos de cada tipo (treinamento, validação e teste) e cada conjunto aplicado 5 vezes na rede. Os resultados, referentes à média das execuções de cada conjunto, estão descritos abaixo.

Quadro 28: Performance do modelo 06.

Número de neurônios	Até R\$ 50 milhões		Até R\$ 100 milhões		Até R\$ 150 milhões		Até R\$ 200 milhões	
	Qtd.	%	Qtd.	%	Qtd.	%	Qtd.	%
1	13,4	17,0	19,6	24,8	32,6	41,3	39,8	50,4
1	10,8	13,7	22,2	28,1	30,8	39,0	38,4	48,6
1	7,6	9,6	17,0	21,5	26,2	33,2	32,8	41,5

Verifica-se que com apenas um neurônio na camada de saída a performance geral da rede é similar as anteriores, demonstrando que a complexidade na arquitetura da rede talvez não seja o melhor caminho para a melhoria dos resultados devendo-se buscar novas abordagens, notadamente no tratamento dos dados ou uso de séries auxiliares.

4.4 TRATAMENTO DOS DIAS DE FERIADOS

As análises realizadas trouxeram conhecimento dos registros utilizados e conduziram os estudos para a adoção de novas estratégias no tratamento da base de dados, no tocante a criação do arquivo de entrada-saída.

Considerando que a ausência de dados decorrentes de feriados pode descaracterizar o comportamento dos dias da semana, optou-se por completar a série com o valor médio da série original. Dessa forma, 28 novos registros com a média aritmética foram incluídos e aboliu-se a retirada das semanas com início e fim em meses diferentes. O tratamento da série consistiu em:

- Divisão dos valores por 1 milhão.
- Alteração dos limites inferior (-1.000) e superior (2.000).
- Inclusão da média aritmética nos dias faltantes.

- Normalização dos dados através da subtração da média e divisão por 2 desvios-padrão.
- Retirada da tendência linear.
- Retirada da sazonalidade semanal (subtração da média do dia da semana).
- Retirada de dois ciclos senoidais referentes às componentes com período de 783,7 dias e 21,8 dias.

Outro procedimento adotado neste pré-tratamento foi a exclusão de mais uma componente senoidal, referente ao período de 21,8 dias. Este período, detectado na análise do espectro da transformada de Fourier corresponde a 1 mês, considerando os dias úteis, e é razoável inferir que exista alguma ciclicidade nos dados.

Após a realização das transformações mencionadas, foi criado o gráfico de autocorrelação a fim de se localizar as relações existentes com o passado.

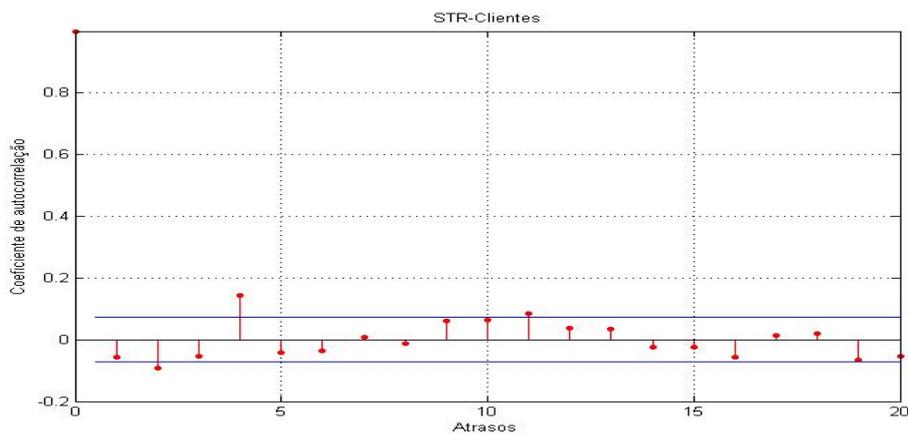


Gráfico 38: Autocorrelação da STR-Clientes após retirada de 2 ciclos senoidais.

Verifica-se que dentre todos os atrasos mapeados aqueles relativos a 2 e 4 dias são os que apresentam os maiores índices de correlações.

Extração de Componentes Determinísticas 03

O tratamento dos dados para utilização nas próximas redes divergiu dos anteriores pela inserção de 28 novos registros relativos aos dias de feriados e a extração de uma 2ª componente senoidal, escolhida através da análise do espectro de Fourier.

O cálculo da parte determinística extraída da série através da reconstituição do resíduo (erro) encontra-se no quadro abaixo.

Quadro 29: Estatística de acertos após extração da parte determinística da série.

Até R\$ 50 milhões		Até R\$ 100 milhões		Até R\$ 150 milhões		Até R\$ 200 milhões	
Qtd.	%	Qtd	%	Qtd.	%	Qtd.	%
97	12,4	188	24,0	274	34,9	357	45,5

A potência da série que anteriormente era de $P_{ant} = 0,212572$ passou para $P_{atual} = 0,187556$.

O percentual de acertos teve uma ligeira redução o que pode ser explicado pela entrada de 208 novos registros. Como a maioria destes dados são oriundos das semanas incompletas e daquelas cujo início e fim estão em meses diferentes é possível imaginar que são dados com altas frequências. Os primeiros em função de serem sobrecarregados pelos feriados e os últimos em razão do início do mês onde normalmente o movimento é maior.

Modelo Neural 05

Baseado no conhecimento adquirido aplicou-se à rede o arquivo tratado conforme descrito no Anexo 5. Os resultados seguem a ordem de número de neurônios e utilização de arquivo com e sem a representação binária do dia da semana.

Quadro 30: Performance do modelo 05.

Número de neurônios	Até R\$ 50 milhões		Até R\$ 100 milhões		Até R\$ 150 milhões		Até R\$ 200 milhões	
	Qtd.	%	Qtd	%	Qtd.	%	Qtd.	%
1	19,0	14,6	32,6	25,1	46,2	35,5	59,8	46,0
1	17,6	13,5	33,6	25,8	48,4	37,2	61,6	47,4
5	20,8	16,0	35,2	27,1	46,8	36,0	60,2	46,3
5	16,0	12,3	30,6	23,5	43,6	33,5	56,8	43,7

Vê-se que não houve ganho de performance com as alterações implementadas, nem mesmo o aumento do número de camadas e de neurônios influenciaram os resultados.

Modelo Neural 06

Como os atrasos verificados no mapa de autocorrelação não se mostraram adequados para o mapeamento dos dados, optou-se pela verificação de limites superiores aos atrasos de 20 posições. Assim, foram calculados índices até 30 atrasos e

notou-se uma significativa correlação com o 22º dia, conforme gráfico abaixo, o que motivou a criação de novo modelo contemplando esta característica.

Análises superiores a 30 atrasos foram realizadas mas não apresentaram índices passíveis de estudo. Além disso, a literatura mostra que com o passar do tempo as relações tendem a diminuir, o que foi comprovado no estudo de autocorrelação apresentado no capítulo 3.

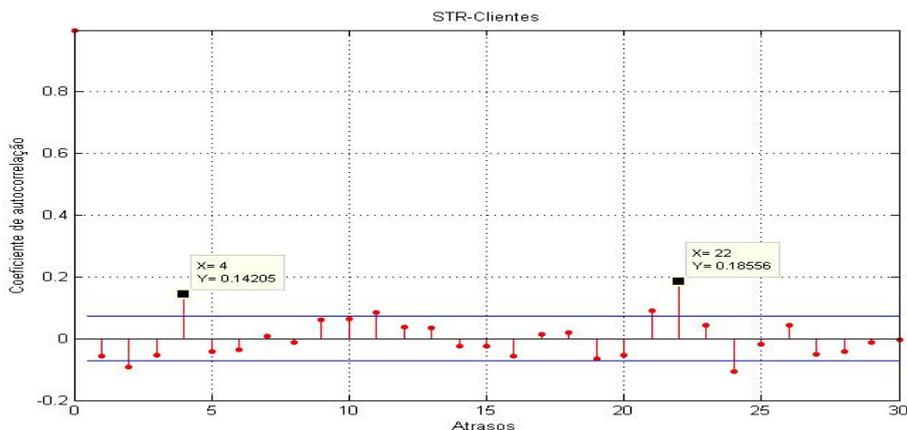


Gráfico 39: Autocorrelação da STR-Clientes após retirada de 2 ciclos senoidais.

O quadro abaixo apresenta os resultados seguindo a ordem de número de neurônios e utilização de arquivo com e sem a representação binária do dia da semana.

Quadro 31: Performance do modelo 06.

Número de neurônios	Até R\$ 50 milhões		Até R\$ 100 milhões		Até R\$ 150 milhões		Até R\$ 200 milhões	
	Qtd.	%	Qtd.	%	Qtd.	%	Qtd.	%
1	16,6	14,6	32,8	28,8	44,6	39,1	54,8	48,1
1	13,8	12,1	33,2	29,1	47,0	41,2	57,8	50,7
5	19,8	17,4	33,2	29,1	46,0	40,4	55,0	48,2
5	15,2	13,3	30,0	26,3	44,6	39,1	55,8	48,9

As alterações promovidas não se mostraram adequadas ao problema. Foram testadas redes com 1 e 5 neurônios na camada de entrada visando a obtenção de resultados melhores. Entretanto, verificando o quadro de acertos, vê-se que a exceção de alguns picos não muito significativos a maioria dos resultados manteve-se constante para todos os limites e com qualquer quantidade de neurônios.

4.5 TRATAMENTO DOS CICLOS SENOIDAIS

Considerando os baixos índices de performances alcançados até o momento, optou-se, antes de se treinar novas redes, pela exclusão de outros ciclos senoidais. Tal escolha se deu em função de haverem sido detectadas elevadas componentes de períodos significativos, conforme pode ser verificado no gráfico a seguir.

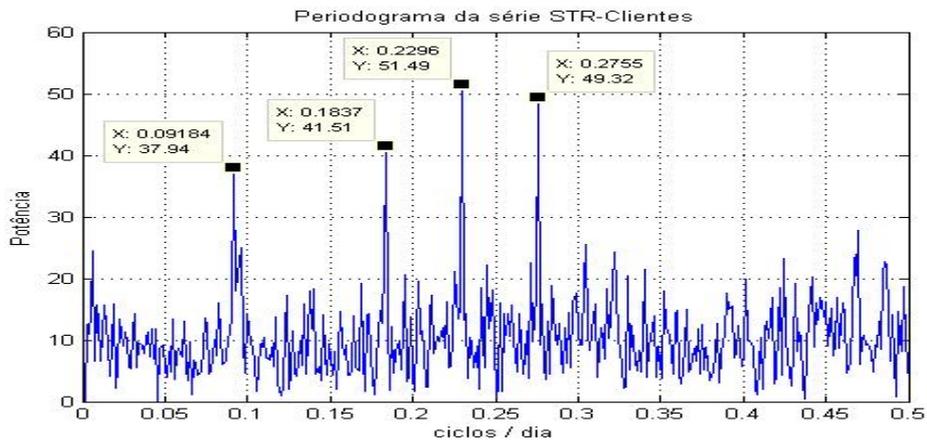


Gráfico 40: Representação da série STR-Clientes.

Cabe registrar que o periodograma acima já está com as duas primeiras componentes excluídas (783,7 dias e 21,8 dias). O novo arquivo que será utilizado nos próximos experimentos terão excluídas as componentes referentes aos períodos de 4,36 dias, 3,63 dias, 5,44 dias e 10,89 dias, correspondentes às coordenadas (0,2296, 51349), (0,2756, 49,32), (0,1837, 41,51) e (0,09184, 37,94) respectivamente. A seguir tem-se o periodograma da série após a extração destas 4 componentes.

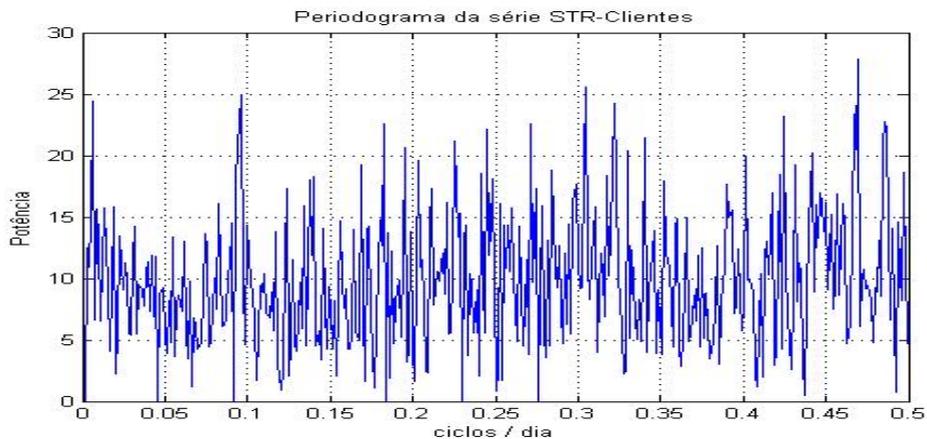


Gráfico 41: Periodograma da STR-Clientes após extração de mais 4 componentes.

Extração de Componentes Determinísticas 04

A potência da série que, anteriormente, encontrava-se em $P_{ant} = 0,187556$ quando da extração da 2ª componente senoidal, passou para $P_3 = 0,178930$, $P_4 = 0,171015$, $P_5 = 0,165408$ e $P_6 = 0,160724$. Esta redução demonstra que a série ainda possuía componentes cíclicas passíveis de remoção. Estudos subsequentes de retirada de novas senóides apresentaram diminuição residual na potência série não motivando novas exclusões.

Adiante tem-se o cálculo da parte determinística extraída da série através da reconstituição do resíduo.

Quadro 32: Estatística de acertos após extração da parte determinística da série.

Até R\$ 50 milhões		Até R\$ 100 milhões		Até R\$ 150 milhões		Até R\$ 200 milhões	
Qtd.	%	Qtd.	%	Qtd.	%	Qtd.	%
100	12,8	195	24,9	280	35,7	366	46,7

Cabe comentar que o ganho com a extração de mais 4 componentes foi, à primeira vista, insignificante, o que poderia induzir a pesquisa a não considerar esta etapa. Entretanto, dada a dificuldade existente no mapeamento e consequente previsão dos dados, acredita-se que o ganho, ainda que de dimensões reduzidas vale para efeitos práticos, pois o acerto de mais 9 dias de previsão (diferença entre o cálculo obtido na Extração de Componentes Determinísticas 03 e esta) dentro dos limites de R\$ 200 milhões é extremamente significativo para a empresa.

Posteriormente promoveu-se o estudo da autocorrelação do resíduo da série (vide gráfico abaixo) obtendo-se três atrasos com índices acima do intervalo de confiança em $t - 1$, $t - 4$ e $t - 9$. Dentre estes, foi utilizado no próximo experimento somente $t - 4$ pois para o caso de $t - 1$ não há condições de contemplá-lo no modelo em função de restrições operacionais já explicitadas e para $t - 9$ verificou-se que além do índice do atraso não ser de valor significativo ele não possui nenhuma relação conhecida, ou seja, não representa múltiplos de semanas.

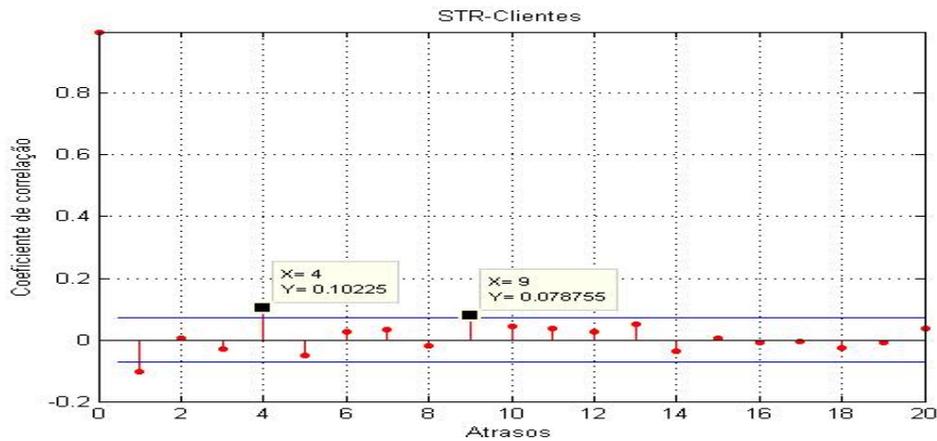


Gráfico 42: Autocorrelograma da STR-Clientes após extração de mais 4 componentes.

Modelo Neural 07

Este modelo procurou inserir as características mapeadas no estudo dos ciclos senoidais, ou seja, no arquivo de entrada foram excluídas as 6 principais componentes senoidais e trabalhou-se com atraso de 4 dias. As demais características tratadas no pré-processamento estão descritas no Anexo 5.

Quadro 33: Performance do modelo 07.

Número de neurônios	Até R\$ 50 milhões		Até R\$ 100 milhões		Até R\$ 150 milhões		Até R\$ 200 milhões	
	Qtd.	%	Qtd.	%	Qtd.	%	Qtd.	%
1	18,0	13,6	34,8	26,4	49,6	37,6	62,2	47,1
1	18,8	14,2	37,8	28,6	52,8	40,0	63,0	47,7
5	16,6	12,6	33,2	25,2	50,0	37,9	62,4	47,3
5	17,0	12,9	38,6	29,2	54,4	41,2	63,4	48,0

A retirada das componentes determinísticas tende a deixar o resíduo resultante cada vez mais estocástico pois representa a não linearidade dos dados. O uso das redes neurais vêm ao encontro deste problema, tentando mapear esta aleatoriedade. Entretanto, os resultados obtidos até o momento mostram que as abordagens utilizadas ainda não conseguiram captar o comportamento da série. Provavelmente em função da inexistência de padrões que se repetam no tempo.

4.6 AVALIAÇÃO DA MÉDIA DOS DIAS ÚTEIS

Média dos Dias 11, 12, 13, 15 e 23

Em adição as análises realizadas até o momento, fez-se um estudo a respeito das médias dos dias úteis, ou seja, verificou-se os valores médios para o 1º dia útil do mês, 2º dia útil do mês e assim respectivamente, até o 23º dia útil. O gráfico representativo das médias encontra-se abaixo.

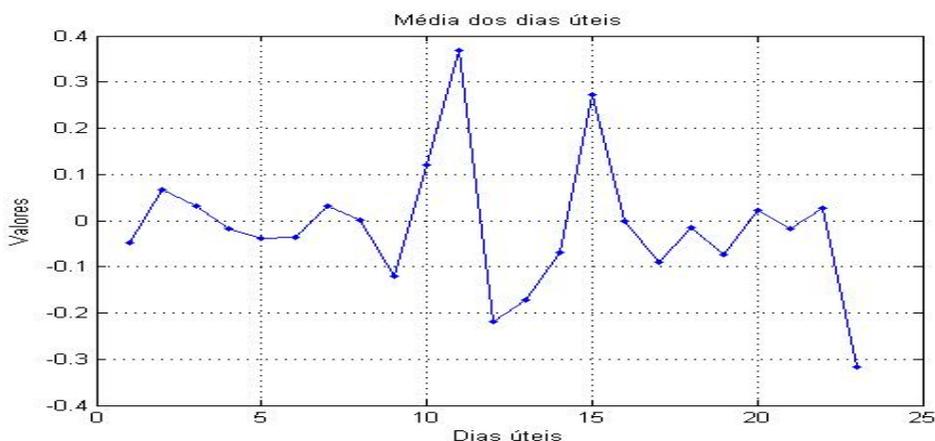


Gráfico 43: Média dos dias úteis.

Verifica-se que os valores relativos aos dias 11, 12, 13, 15 e 23 estão divergentes em relação aos demais que oscilam na faixa aproximada de $[-0,1, 0,1]$. Assim, optou-se por subtrair a média destes dias visando compactar todo o conjunto no intervalo aproximado de $[-0,1, 0,1]$ e com isso facilitar o mapeamento da série. O novo gráfico ficou, então, com a seguinte característica.

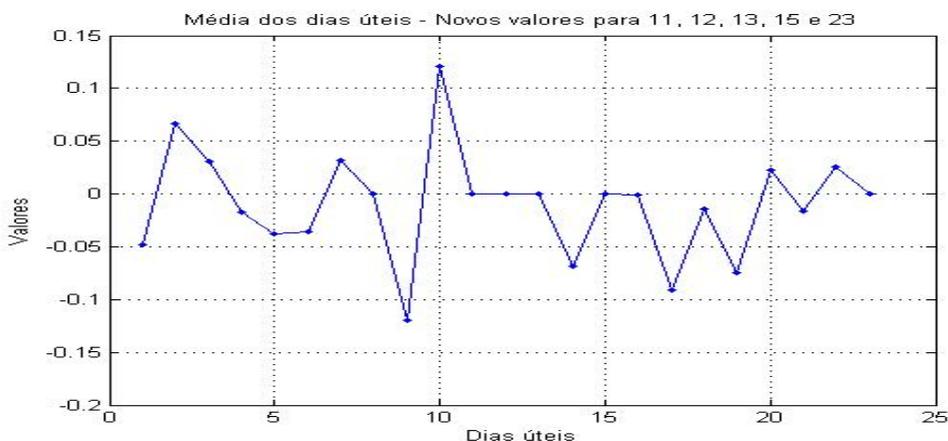


Gráfico 44: Dias úteis com subtração da média dos dias 11, 12, 13, 15 e 23.

Com o adoção do procedimento descrito acima, novo conjunto de dados foi criado com potência $P = 0,146864$.

Extração de Componentes Determinísticas 05

Adiante tem-se o cálculo da parte determinística extraída da série através da reconstituição do resíduo.

Quadro 34: Estatística de acertos após extração da parte determinística da série.

Até R\$ 50 milhões		Até R\$ 100 milhões		Até R\$ 150 milhões		Até R\$ 200 milhões	
Qtd.	%	Qtd	%	Qtd.	%	Qtd.	%
105	13,4	195	24,9	284	36,2	367	46,8

Vê-se com este procedimento que apesar de reduzido houve um ganho tanto na potência quanto no aumento da quantidade de registros dentro das faixas estipuladas, demonstrando que ainda existe espaço para extração de sazonalidades na série.

Modelo Neural 08

Este modelo é semelhante ao Modelo Neural 07 com a diferença que, neste caso, o arquivo utilizado teve a média dos dias 11, 12, 13, 15 e 23 subtraída da série original.

Os resultados obtidos encontram-se a seguir.

Quadro 35: Performance do modelo 08.

Número de neurônios	Até R\$ 50 milhões		Até R\$ 100 milhões		Até R\$ 150 milhões		Até R\$ 200 milhões	
	Qtd.	%	Qtd	%	Qtd.	%	Qtd.	%
1	17,2	13,0	34,8	26,4	49,0	37,1	59,0	44,7
1	17,8	13,5	32,8	24,8	49,0	37,1	59,8	45,3
5	16,0	12,1	34,0	25,8	49,0	37,1	60,0	45,5
5	15,6	11,8	32,0	24,2	46,2	35,0	60,2	45,6

Apesar de o modelo não ter respondido adequadamente aos dados, com performance inferior, inclusive, a do modelo 07, ainda assim a extração da média para alguns dias apresentou, no tocante à parte determinística, um bom resultado, direcionando a pesquisa por este caminho.

Média dos Dias Úteis

Em seguida optou-se pela subtração da média de todos os dias úteis, a exceção do 23º dia útil pelo fato de só existirem 4 registros na base relativos a este dia. Após este procedimento calculou-se a potência obtendo o valor de $P = 0,144764$.

A fim de se avaliar o comportamento da série após todas as etapas, gerou-se o gráfico abaixo.

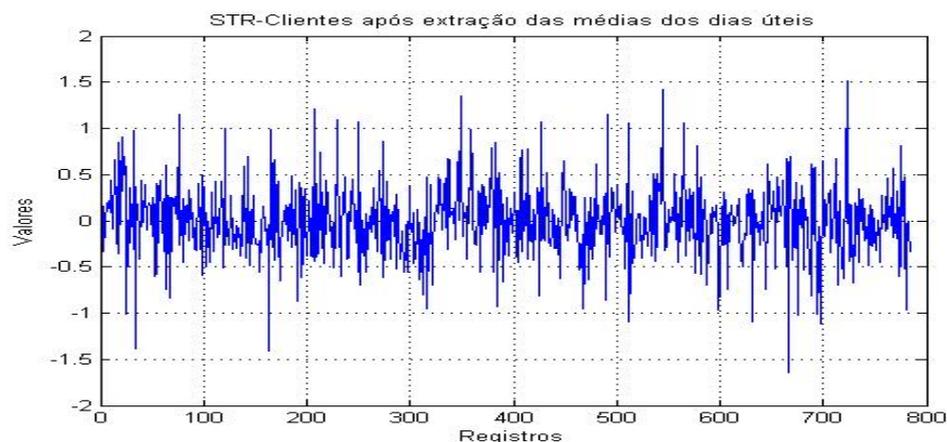


Gráfico 45: Série após subtração da média.

Visualmente se pode observar a inexistência de tendência, ciclos ou sazonalidades que, acaso existentes, devem ser de pequena magnitude, o que já vem sendo constatado nos últimos estudos.

Extração de Componentes Determinísticas 06

Para avaliar a performance da previsão baseado tão-somente nas extrações determinísticas, criou-se o quadro abaixo após a reconstituição do resíduo.

Quadro 36: Estatística de acertos após extração da parte determinística da série.

Até R\$ 50 milhões		Até R\$ 100 milhões		Até R\$ 150 milhões		Até R\$ 200 milhões	
Qtd.	%	Qtd.	%	Qtd.	%	Qtd.	%
109	13,9	193	24,6	290	37,0	380	48,5

Novamente percebe-se um pequeno ganho em relação ao experimento anterior. Em seguida os arquivos gerados nas fases 5 e 6 da Extração de Componentes Determinísticas serão aplicados em um modelo neural com vistas a avaliar a performance para estes conjuntos de dados. Após estas alterações foi criado um novo

correlograma, que pode ser visto abaixo.

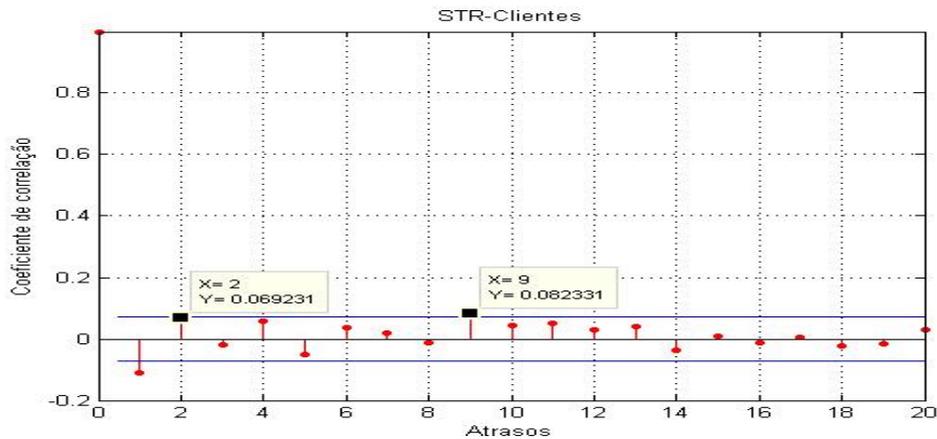


Gráfico 46: Correlograma da série STR-Clientes.

Verifica-se que o atraso de 4 dias que era significativo no gráfico anterior teve seu índice de correlação reduzido e os atrasos de 2 e 9 dias tiveram seus índices ampliados.

Modelo Neural 09

O modelo 09 segue as mesmas diretrizes do experimento anterior. Entretanto, baseado no gráfico de autocorrelação acima, trabalhou-se com 3 tipos de entradas.

- Atraso de 2 dias e com e sem a representação binária do dia da semana e sem a representação binária do dia da semana.
- Atraso de 9 dias e com e sem a representação binária do dia da semana.
- Atrasos de 2 e 9 dias conjuntamente e com e sem a representação binária do dia da semana.

No Anexo 5 encontra-se todas as características do arquivo e a topologia da rede aplicada.

O quadro abaixo apresenta os resultados seguindo a ordem de número de neurônios e com e sem representação binária do dia da semana.

Quadro 37: Performance do modelo 09.

Número de neurônios	Até R\$ 50 milhões		Até R\$ 100 milhões		Até R\$ 150 milhões		Até R\$ 200 milhões	
	Qtd.	%	Qtd.	%	Qtd.	%	Qtd.	%
1	20,4	15,5	38,8	29,4	54,2	41,1	66,8	50,6
1	22,2	16,8	40,4	30,6	53,0	40,2	64,8	49,1
5	22,2	16,8	38,0	28,8	53,2	40,3	66,6	50,5
5	23,4	17,7	41,8	31,7	56,6	42,9	64,8	49,1
1	17,8	13,7	33,0	25,4	49,4	38,0	63,6	48,9
1	20,0	15,4	32,2	24,8	50,0	38,5	63,2	48,6
5	16,6	12,8	31,4	24,2	47,0	36,2	64,2	49,4
5	18,0	13,8	32,8	25,2	50,0	38,5	63,6	48,9
1	16,8	12,9	35,0	26,9	52,8	40,6	65,0	50,0
1	16,8	12,9	34,0	26,2	53,0	40,8	63,6	48,9
5	16,8	12,9	34,0	26,2	52,6	40,5	66,2	50,9
5	17,8	13,7	35,2	27,1	49,4	38,0	62,0	47,7

Verificou-se que a performance das redes permanece estável, não tendo sido observadas melhorias significativas apesar das diversas alterações realizadas.

4.7 ANÁLISE GRÁFICA DO ERRO

Chatfield (CHATFIELD, 2004) afirma que a primeira e mais importante análise a ser efetuada em uma série temporal é o estudo gráfico dos dados, denominado *time plot*. Através dos gráficos visualizasse importantes características da série que em uma análise numérica dificilmente se perceberia. Extrapolando este conceito, o estudo gráfico do erro também pode trazer informações importantes para o desenvolvimento dos modelos.

No gráfico abaixo, produzido com o arquivo tratado para o modelo neural 09, encontram-se no eixo das abscissas os valores originais dos dados e no eixo das coordenadas o respectivo erro de previsão. Na situação ideal todos os dados estariam sobrepostos na linha horizontal representativa de erro 0. Não é o que se observa. Grande parte dos registros está substancialmente afastada da origem.

A maior concentração dos dados originais são de valores positivos (1º e 4º quadrantes) com 617 registros correspondendo a 78,7% do total da base de 784 registros. Já para os erros, a distribuição é uniforme, com 391 registros positivos (49,9%) e 393 negativos (50,1%).

Fenômeno idêntico acontece quando se separam os registros originais cujos

erros ultrapassaram o limite de 200 milhões. Neste caso, os originais se distribuem em 305 positivos (75,5%) e 99 negativos (24,5%). Os erros acima de 200 milhões obtiveram distribuição idêntica àquela ocorrida quando do cálculo realizado com o total da base, com 201 positivos (49,8%) e 203 negativos (50,2%).

O desbalanceamento da base pode ter influenciado para que não se obtivesse bons resultados. A teoria de redes neurais recomenda que sejam apresentados à rede quantidades semelhantes de características, a fim de não viciá-la. Não foi o que ocorreu. Os registros positivos representam $\frac{3}{4}$ do total e esta particularidade pode ter prejudicado o desempenho.

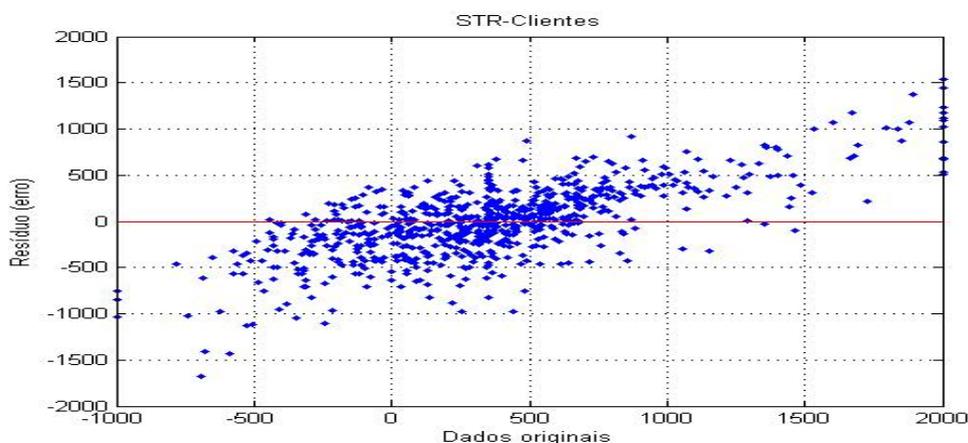


Gráfico 47: Dados originais x Erro.

Na próxima visão temos um gráfico entre os valores determinísticos extraídos da série, no eixo das abscissas e os valores dos respectivos erros, representados pelo eixo das coordenadas.

O gráfico mostra que para grandes valores de x os respectivos erros também o são, ou seja, nestes casos as equações aplicadas não conseguem extrair muita informação. Isto pode ser um fator de dificuldade no mapeamento das redes neurais pois a característica da rede de possuir alta frequência já produz, por si só, uma complexidade no mapeamento. Se, aliado a isto, ainda existir elevadas amplitudes, pode tornar extremamente difícil a solução do problema.

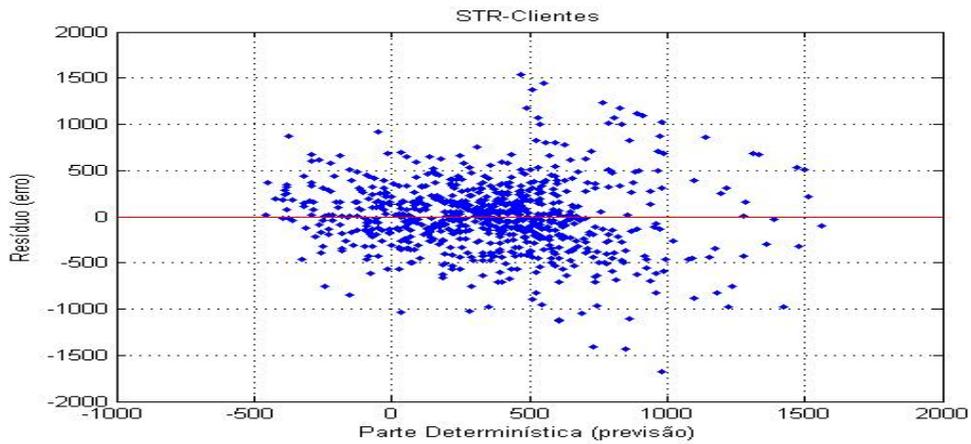


Gráfico 48: Parte determinística x Erro.

A fim de avaliar a independência entre os valores residuais da série, foi criado o gráfico de autocorrelação abaixo.

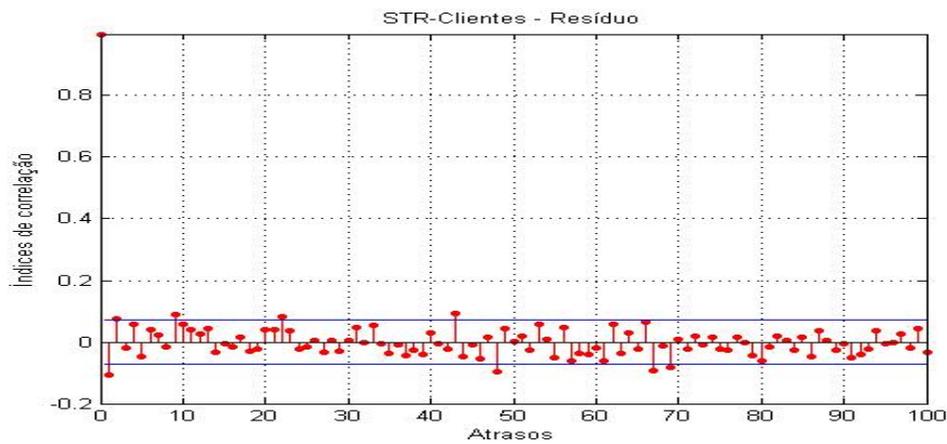


Gráfico 49: Função de autocorrelação do resíduo da série STR-Clientes.

Vê-se que a exceção de alguns poucos pontos, os índices da série estão bem próximos de 0, com os limites de confiança no intervalo $[-0,0714, 0,0714]$. Este resultado nos remete à hipótese de tratar-se de um ruído branco, conforme definido no capítulo 3. A fim de se avaliar a distribuição dos coeficientes, foi criado o histograma apresentado a seguir.

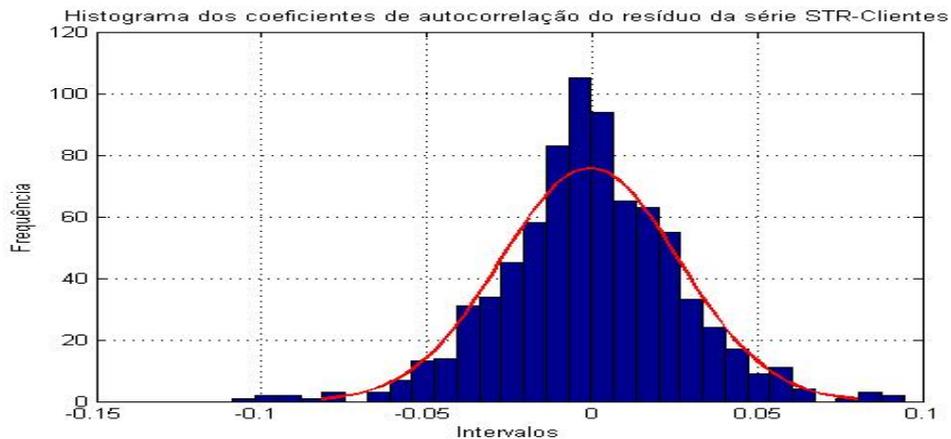


Gráfico 50: Histograma dos coeficientes de autocorrelação do resíduo da série STR-Clientes.

O histograma demonstra que a distribuição dos dados está bem próxima da normal, caracterizando, também, uma das premissas para se considerar o resíduo da série como um ruído branco. Entretanto, ao tempo em que os gráficos auxiliam nas análises é o estudo quantitativo que corrobora as hipóteses e, neste caso, foram calculadas a média e variância dos dados, obtendo-se os valores abaixo.

Quadro 38: Comparação entre os valores definidos por Bartlett para um ruído branco e os calculados.

	Média	Variância
Valor tabelado	0	0,001276
Valor calculado	0,000638	0,002003

Os números mostram que caracterizar a série como um ruído branco talvez seja precipitado mas considerá-la próxima desta característica é totalmente factível. O uso de outras séries trabalhando em conjunto talvez possa caminhar para uma melhoria nos resultados.

4.8 CONCLUSÃO

As técnicas abordadas neste capítulo foram responsáveis por grande parte do esforço dedicado a este trabalho de pesquisa. A complexidade da série sob estudo tornou não trivial a construção e ajustes dos parâmetros do modelo neural. Ao contrário, após várias tentativas em se obter um bom desempenho nos treinamentos voltava-se para a fase de pré-processamento com o intuito de avaliar e utilizar novos métodos para tratamento dos dados visando melhorar os resultados.

Bishop (BISHOP, 1996) apresenta uma excelente citação para mostrar a importância do tratamento dos dados, quando diz:

“Outro caminho através do qual a performance da rede pode ser melhorada, algumas vezes dramaticamente, é incorporando conhecimento prévio aos dados. Tal conhecimento se refere a relevantes informações que podem ser utilizadas no desenvolvimento da solução, melhorando o treinamento dos dados.”

Apesar da importância do tema, poucos autores ligados à construção de redes neurais artificiais dedicam espaço em seus livros para discorrer sobre técnicas de pré e pós-processamento, tendo que se recorrer a literatura de outras áreas, notadamente estatística (BUSSAB; MORETTIN, 2001) (LEVINE; BERESON; STEPHAN, 2000) e *data mining* (LITTLE; RUBIN, 1987) (HAIR; ANDERSON *et al.*, 1998) (HAN; KAMBER, 2001) (PYLE, 1999).

Os resultados obtidos com o uso das RNA não confirmaram a hipótese da sua superioridade no tratamento de dados não-lineares e ruidosos. Apesar de alguns modelos neurais haverem conseguido melhorar o índice de assertividade dos dados, ainda assim as redes que conseguiram tal feito o fizeram por uma margem muito reduzida para considerá-las eficazes. Como exemplo temos o quadro abaixo onde vemos uma compilação dos resultados alcançados na extração das componentes e, ao lado, os melhores resultados obtidos com cada modelo de rede. Vê-se que os modelos 3 a 7 e o 9 obtiveram um pequeno ganho no tratamento do arquivo da série apesar de, para os propósitos a que se destinam, este ganho não ser desprezível.

Quadro 39: Comparação da performance dos modelos.

Extração de componentes determinísticas		Modelos neurais	
1	46,1%	1	44,3%
2	47,8%	2	41,5%
		3	51,9%
		4	50,4%
3	45,5%	5	47,4%
		6	50,7%
4	46,7%	7	48,0%
5	46,8%	8	45,6%
6	48,5%	9	50,9%

Cabe comentar que a comparação acima apresenta uma importante incoerência quando se fala dos conjuntos que foram utilizados em cada etapa: a extração das componentes determinísticas trabalhou com a série original, quando ainda detinha muita informação, já os modelos neurais trataram o ruído oriundo dessa fase. Não foi feita nenhuma análise comparativa entre os métodos estatísticos clássicos e os neurais para esta etapa do trabalho, até porque foge ao escopo da tese. O que se discute aqui é que ao se chegar na fronteira da utilização dos métodos convencionais, supôs-se que as RNA pudessem levar os resultados a patamares mais significativos do que os alcançados. Não foi o que se verificou até o momento.

A seção 2 deste capítulo já antevia as dificuldades a serem superadas. Muitos autores (PASLEY; AUSTIN, 2004) (LOCKMIC; SMITH, 2000) (HLAVÁČEK; CADA; HAKL, 2005) trabalharam com RNA em séries financeiras de altas frequências e obtiveram resultados similares ao alcançado nos experimentos apresentados neste capítulo, demonstrando a dificuldade em se mapear este tipo de dado.

5 OUTRAS ABORDAGENS NEURAIAS

5.1 INTRODUÇÃO

As tentativas de se prever, com redes neurais artificiais, o comportamento do resíduo oriundo da extração das componentes determinísticas da série STR-Clientes, utilizando-se unicamente os registros relativos à própria série, não se mostraram eficazes. Tal fato decorre da não descoberta, ou inexistência, de correlações com o passado que pudessem auxiliar na previsão deste resíduo, ou seja, não foram observados padrões nos dados.

Nos estudos realizados detectou-se um comportamento com características de aleatoriedade, ou bem próximo disto. Apesar de não se poder afirmar tacitamente tratar-se de um ruído branco, os cálculos e análises promovidas apontam fortemente para este caminho.

Entretanto, não obstante as redes neurais não terem trazido ganhos na performance dos índices de previsão obtidos com as técnicas estatísticas clássicas de decomposição de série, ainda assim existe espaço para melhoraria dos resultados através de novos tratamentos nos dados, inserção de outras séries no modelo de forma a auxiliar a série original ou, até mesmo, aumento do número de registros.

Trabalho da pesquisadora Lam (2004) aborda o uso de redes do tipo MLP para utilização em previsão financeira e, para tanto, a autora utilizou 16 séries financeiras e 11 séries macroeconômicas para prever o comportamento de 364 empresas. A autora apresenta diversos artigos onde os resultados na previsão de séries temporais são, de alguma forma, ruins. Ela defende a tese de que mais importante do que o uso de modelos neurais complexos, modelos estatísticos convencionais etc é a escolha das variáveis que farão parte do conjunto de dados a ser aplicado ao modelo.

No artigo *An Empirical Analysis of Data Requirements for Financial Forecasting with Neural Networks*, Walczak (2001) utilizou várias séries referentes à taxa de câmbio entre o dólar e outras moedas em um período de 22 anos e 4 meses,

visando uma previsão de um dia à frente ($d + 1$). Deste total foram separados 21 anos e 9 meses para treinamento e o restante para teste.

O primeiro item discutido pelo autor refere-se à quantidade de dados no treinamento da rede. Apesar de haver elencado diversos pesquisadores que corroboram a idéia de que quanto mais dados melhor para se realizar a previsão de uma série temporal, Walczak demonstra que os melhores resultados foram obtidos com períodos entre 1 e 6 anos, ou seja, com uma quantidade de registros bem inferior aquela disponível para o trabalho.

A utilização de dados externos para auxílio da modelagem também foi descartada pelo pesquisador que já havia tratado este tipo de série com mais 18 séries adicionais e concluído que o uso de série homogênea (sem utilização de informações externas) produz melhor performance.

Vê-se, portanto, a inexistência de consenso quando se trata de discutir a melhor direção a ser seguida, pois a quantidade de abordagens que podem ser adotadas impede o projetista dos modelos de caminhar em um único sentido.

Encontram-se neste capítulo os estudos mais significativos, dentre os vários realizados, que foram conduzidos visando o aumento assertividade do modelo de previsão.

5.2 TRATAMENTO DAS DEMAIS VARIÁVEIS INTEGRANTES DO FLUXO DE CAIXA

Dando prosseguimento aos estudos, optou-se pelo uso de outras variáveis para auxiliar a principal – STR-Clientes – nos modelos de previsão. Para tanto, foram tratadas as demais séries integrantes da reserva bancária, no período de 01/04/2003 a 31/03/2006.

Quadro 40: Estatística básica das variáveis componentes da movimentação da reserva bancária.

	Média	Mínimo	Máximo	Desvio
01. Numerário	-186,17	-1.537,99	600,74	289,11
02. Compensação	-57,82	-656,30	581,93	124,07
03. Arrecadação via STN / STR	-626,76	-4.530,80	5,47	746,83
04. Ordens Bancárias	491,47	-1.025,48	5.064,11	769,47
05. Câmbio	-14,36	-1.016,81	1.063,73	235,02
06. Compulsório	1,60	-341,24	229,05	51,82
07. SELIC	-10,53	-6.792,34	6.378,11	1.703,03
08. CETIP	-20,77	-1.411,88	1.497,00	152,83
09. Tesouro Nacional - Outras	425,54	-976,74	7.534,41	937,90
11. Lançamentos BACEN – SLB	4,94	-1.336,94	540,93	65,93
13. BMF - Derivativos	-0,44	-88,79	115,28	10,58
14. CBLC	3,55	-1.386,78	2.008,33	127,87
16. STR – IF	35,72	-418,54	922,00	113,16
17. STR - Clientes	362,25	-1,150,90	4.256,37	543,22
18. CIP	-400,33	-1.470,49	1.145,66	329,92

Após o cálculo das estatísticas básicas de cada componente foi realizado, a exemplo do tratamento da variável STR-Clientes, o cálculo da frequência dos valores em faixas e posterior a redução da quantidade de faixas, trazendo para o patamar mais significativo os valores muito altos ou muito baixos. Os quadros das distribuições de frequências encontram-se no Anexo 6.

Por fim, as variáveis 10. TECBAN, 12. Redesconto e 15. BMF–Ativos não foram tratadas em função de possuírem valores insignificantes (TECBAN) ou não terem quantidade significativa de registros com movimento (Redesconto e BMF-Ativos).

Foi realizado, também, o cálculo de correlação cruzada entre as variáveis com o objetivo de verificar a existência de padrões de similaridade entre duas ou mais séries.

Cabe registrar que para realização da correlação cruzada, as variáveis auxiliares passaram pelo mesmo tratamento dispensado à variável alvo, ou seja, após a agregação dos valores dispersos para patamares mais significativos os registros faltantes, relativos aos feriados, foram completados com a média aritmética da respectiva série. Em seguida, foi realizada a normalização com a subtração do valor médio e divisão por 2 desvios-padrão. Após este procedimento, extraiu-se a tendência linear e a sazonalidade semanal e, por fim, os ciclos senoidais, através da transformada de Fourier. Cada variável teve a quantidade de ciclos senoidais extraídos com base no seu comportamento, ou seja, não existiu um número fixo de ciclos a ser retirado. Para

cada caso foram avaliadas as componentes mais significativas de cada uma delas e procedeu-se a respectiva exclusão. Nos arquivos com os dados de Câmbio, Compulsório, SLB, STR-IF e CIP foram extraídas 9 componentes, nos de Compensação e Tesouro Nacional 8 componentes, SELIC e BMF-Derivativos, 7 componentes, STN/STR, CETIP e STR-Clientes, 6 componentes, Numerário, Ordens Bancárias e CBLC 4 componentes.

Quadro 41: Correlação cruzada entre as variáveis do fluxo de caixa.

	Num	Compe	STN/STR	OB	Câmbio	Compul	Selic
Num	1,000	0,060	0,051	-0,061	-0,006	-0,022	-0,043
Compe	0,060	1,000	0,098	0,002	-0,016	0,014	-0,015
STN/STR	0,051	0,098	1,000	0,014	0,004	0,027	-0,157
OB	-0,061	0,002	0,014	1,000	0,009	-0,045	-0,248
Câmbio	-0,006	-0,016	0,004	0,009	1,000	0,017	-0,089
Compul	-0,022	0,014	0,027	-0,045	0,017	1,000	-0,051
Selic	-0,043	-0,015	-0,157	-0,248	-0,089	-0,051	1,000
Cetip	0,034	-0,057	0,060	0,014	0,027	0,065	-0,074
TN	-0,154	-0,023	-0,218	-0,053	-0,039	0,021	-0,050
SLB	-0,017	0,047	0,012	0,022	0,002	-0,095	-0,058
BMF	0,030	-0,051	0,052	0,017	-0,030	0,091	0,032
CBLC	-0,060	-0,043	-0,041	0,031	0,028	0,011	0,019
STR-IF	0,065	-0,295	-0,030	-0,079	0,028	0,056	-0,056
STR-Clí	0,007	-0,105	0,093	-0,007	-0,055	-0,013	-0,137
CIP	0,128	-0,032	0,063	0,051	-0,029	0,038	-0,135

	Cetip	TN	SLB	BMF	CBLC	STR-IF	STR-Clí	CIP
Num	0,034	-0,154	-0,017	0,030	-0,060	0,065	0,007	0,128
Compe	-0,057	-0,023	0,047	-0,051	-0,043	-0,295	-0,105	-0,032
STN/STR	0,060	-0,218	0,012	0,052	-0,041	-0,030	0,093	0,063
OB	0,014	-0,053	0,022	0,017	0,031	-0,079	-0,007	0,051
Câmbio	0,027	-0,039	0,002	-0,030	0,028	0,028	-0,055	-0,029
Compul	0,065	0,021	-0,095	0,091	0,011	0,056	-0,013	0,038
Selic	-0,074	-0,050	-0,058	0,032	0,019	-0,056	-0,137	-0,135
Cetip	1,000	-0,004	0,053	0,027	-0,007	0,006	0,059	-0,039
TN	-0,004	1,000	0,065	-0,057	0,045	0,071	0,012	-0,016
SLB	0,053	0,065	1,000	0,020	-0,010	-0,046	0,053	-0,025
BMF	0,027	-0,057	0,020	1,000	0,013	0,043	0,000	-0,014
CBLC	-0,007	0,045	-0,010	0,013	1,000	0,048	-0,112	-0,072
STR-IF	0,006	0,071	-0,046	0,043	0,048	1,000	0,076	0,031
STR-Clí	0,059	0,012	0,053	0,000	-0,112	0,076	1,000	-0,024
CIP	-0,039	-0,016	-0,025	-0,014	-0,072	0,031	-0,024	1,000

Verifica-se, no quadro acima, que os coeficientes de correlação entre as variáveis estão próximos de zero, o que denota a quase inexistência de correlação entre elas e, conseqüentemente, uma dificuldade em utilizá-las como auxílio na previsão.

Dentro deste contexto, as correlações mais significativas, ou seja, aquelas cujo valor absoluto é superior a 0,1 estão destacadas para melhor visualização.

Em seguida, calculou-se as correlações entre a variável alvo (STR-Clientes) e as demais para até 20 atrasos, registrando-se os valores mais significativos no quadro abaixo e os respectivos gráficos no Anexo 7.

Quadro 42: Variáveis auxiliares e atrasos.

Variável	Atrasos	Coef. correlação
Numerário	16	0,0883
Compensação	10	0,0731
STN / STR	14	-0,1300
Ordens Bancárias	3	-0,1004
Câmbio	13	0,0844
Compulsório	19	-0,0793
SELIC	13	-0,0777
CETIP	13	-0,0555
TN	10 e 14	0,1808 e 0,1503
SLB	20	0,0955
BMF	19	-0,0843
CBLC	4	-0,0769
STR-IF	10	-0,0667
CIP	9	-0,1060

Modelo Neural 10

A rede 10 foi construída para prever a variável STR-Clientes com o uso de variáveis auxiliares componentes da movimentação de reservas bancárias. Dessa forma, com base no estudo de correlação cruzada apresentado anteriormente, foram definidas quais variáveis seriam utilizadas e qual atraso seria implementado.

Nem todas as variáveis foram utilizadas neste modelo. As variáveis CETIP e STR-IF foram excluídas pois os índices de correlação cruzada com a variável STR-Clientes estão aquém do intervalo de confiança. Após esta eliminação, sobraram 12 variáveis que foram, uma a uma, trabalhadas junto com a STR-Clientes. Cada variável auxiliar, juntamente com a variável alvo, gerou um arquivo de entrada na rede que foi treinada com 1 e 5 neurônios e com e sem representação binária do dia da semana.

O quadro abaixo apresenta os resultados para cada variável.

Quadro 43: Performance do modelo 10.

Variável	Número de neurônios	Até R\$ 50 milhões		Até R\$ 100 milhões		Até R\$ 150 milhões		Até R\$ 200 milhões	
		Qtd.	%	Qtd.	%	Qtd.	%	Qtd.	%
Numeração	1	19,8	15,7	35,2	27,9	50,8	40,3	63,6	50,5
	5	21,2	16,8	37,6	29,8	52,2	41,4	65,0	51,6
Compen-sação	1	19,8	15,0	37,2	28,2	52,2	39,5	68,4	51,8
	5	22,0	16,7	36,2	27,4	51,8	39,2	67,4	51,1
STN / STR	1	18,6	13,4	34,4	24,7	50,4	36,3	64,8	46,6
	5	18,8	13,5	35,4	25,5	52,2	37,6	65,8	47,3
Ordens Bancárias	1	23,8	17,8	40,2	30,0	54,8	40,9	64,0	47,8
	5	22,6	16,9	40,0	29,9	54,2	40,4	67,0	50,0
Câmbio	1	15,6	12,7	33,4	27,2	47,4	38,5	61,0	49,6
	5	15,2	12,4	30,4	24,7	47,2	38,4	60,8	49,4
Compul-sório	1	15,6	12,1	32,0	24,8	45,4	35,2	59,2	45,9
	5	15,6	12,1	33,4	25,9	49,4	38,3	61,6	47,8
SELIC	1	20,8	16,3	41,0	32,0	58,6	45,8	69,4	54,2
	5	21,4	16,7	41,0	32,0	57,6	45,0	68,4	53,4
Tesouro Nacional	1	14,6	12,0	29,8	24,4	45,0	36,9	59,2	48,5
	5	16,2	13,3	33,2	27,2	48,6	39,8	59,4	48,7
SLB	1	18,0	14,6	37,0	30,1	52,2	42,4	64,6	52,5
	5	18,4	15,0	36,6	29,8	52,0	42,3	63,0	51,2
BMF	1	16,6	12,0	33,0	23,9	46,2	33,5	58,4	42,3
	5	17,4	12,6	30,6	22,2	46,2	33,5	60,4	43,8
CBLC	1	20,4	15,5	36,4	27,6	52,8	40,0	64,8	49,1
	5	21,2	16,1	35,4	26,8	51,2	38,8	67,0	50,8
CIP	1	17,2	12,9	36,4	27,4	54,2	40,8	69,2	52,0
	5	18,8	14,1	38,0	28,6	54,0	40,6	67,8	51,0

Dentre todas as séries utilizadas, verifica-se que somente as variáveis SELIC, SLB e CIP apresentaram resultados superiores aos alcançados até o momento, quando se trabalhou exclusivamente com a STR-Clientes (modelo neural 02). As demais obtiveram performance similares ou inferiores. Considerando os baixos índices de correlações alcançados, já esperava-se rendimento deficiente do modelo.

Chama atenção o fato de os melhores resultados não terem sido alcançados com os melhores índices de correlação. As variáveis TN e STN/STR foram as que apresentaram as maiores correlações e, no entanto, ao serem aplicadas à rede não responderam eficazmente apesar da existência de alguma correlação entre as variáveis. Isto significa que o padrão não se repete por toda a série mas, tão-somente, para alguns intervalos específicos. Na análise de autocorrelação apresentada no capítulo 2 vê-se, claramente, a existência de períodos com significativas correlações seguida de períodos

com inexistência de correlação.

Modelo Neural 11

Baseado nos resultados obtidos no experimento anterior, optou-se pela criação de um novo modelo utilizando como entrada da rede, juntamente com a STR-Clientes, as variáveis que melhores resultados apresentaram, ou seja, a SELIC, SLB e CIP.

Os resultados estão compilados no quadro abaixo.

Quadro 44: Performance do modelo 011.

Número de neurônios	Até R\$ 50 milhões		Até R\$ 100 milhões		Até R\$ 150 milhões		Até R\$ 200 milhões	
	Qtd.	%	Qtd.	%	Qtd.	%	Qtd.	%
1	20,2	16,6	35,8	29,3	50,0	41,0	61,6	50,5
5	20,4	16,7	36,2	29,7	49,8	40,8	62,4	51,1
10	21,6	17,7	36,6	30,0	53,0	43,4	64,2	52,6
20	19,4	15,9	39,2	32,1	53,4	43,8	63,4	52,0

Verifica-se que o uso de variáveis integrantes da movimentação da reserva bancária conjuntamente não está agregando melhoria no modelo. Procurou-se, também, neste modelo, acrescentar mais neurônios com vistas a aumentar os graus de liberdade da rede e, com isto, facilitar o mapeamento. A tentativa não se mostrou adequada haja vista que os resultados caminharam no sentido de redução da performance.

Modelo Neural 12

Neste modelo todas as variáveis definidas como auxiliares foram utilizadas em conjunto com a STR-Clientes, no arquivo entrada-saída.

Quadro 45: Performance do modelo 012.

Número de neurônios	Até R\$ 50 milhões		Até R\$ 100 milhões		Até R\$ 150 milhões		Até R\$ 200 milhões	
	Qtd.	%	Qtd.	%	Qtd.	%	Qtd.	%
1	17,0	13,9	33,2	27,2	44,2	36,2	57,4	47,0
5	17,6	14,4	31,8	26,1	43,4	35,6	56,0	45,9

Nos resultados apresentados acima vê-se que a utilização de todas as variáveis como entrada da rede degradou a performance. Apesar da existência de alguma correlação entre as variáveis auxiliares e a principal, o relacionamento entre elas

não foi suficientemente forte para caracterizar um padrão de comportamento repetitivo ao longo do tempo que pudesse ser mapeado e projetado no futuro.

Extração de Sazonalidades

Considerando que a extração da sazonalidade mensal dos dias do calendário e dos dias úteis na variável STR-Clientes reduziu mais fortemente a potência da série do que a retirada de diversos ciclos senoidais através da transformada de Fourier, resolveu-se adotar o mesmo procedimento para as demais variáveis. Entretanto, como a primeira componente, abrangendo todo o período (784 dias) é comum a todas as variáveis, optou-se pela sua exclusão, antes de se retirar a sazonalidade mensal. Os cálculos de cada etapa estão apresentados a seguir.

Quadro 46: Evolução da potência das séries.

	Sem sazo semanal	Sem 1 comp. Fourier	Sem sazo mensal - calend	Sem sazo mensal - útil
Numerário	0,169647	0,084190	0,068450	0,067692
Compensação	0,165178	0,153296	0,131437	0,128582
STN/STR	0,244228	0,207926	0,087515	0,079404
OB	0,234879	0,207804	0,153373	0,149513
Câmbio	0,245300	0,230807	0,226583	0,224018
Compulsório	0,242896	0,217731	0,212362	0,204603
SELIC	0,247618	0,242850	0,221732	0,217996
CETIP	0,242078	0,236396	0,206581	0,203087
TN	0,244991	0,217131	0,075976	0,065403
SLB	0,243302	0,239248	0,229998	0,225430
BMF	0,246484	0,241333	0,236819	0,235051
CBLC	0,248465	0,244700	0,242111	0,237346
STR IF	0,185112	0,174867	0,158530	0,156906
STR Clientes	0,207513	0,197253	0,121352	0,117430
CIP	0,175957	0,158237	0,136934	0,134930

Considerando que valores significativos de correlação podem ser um facilitador na modelagem dos dados, calculou-se os índices de correlação entre cada variável auxiliar e a STR-Clientes a fim de se definir os atrasos, acaso existentes, em que haja alguma relação entre as séries. Assim, foi criado o quadro abaixo onde se apresenta os atrasos mais evidentes das diversas variáveis e que serão utilizados como entrada da rede neural.

Quadro 47: Atrasos significativos entre as var. auxiliares e STR-Clientes.

	Atrasos
Numerário	0
Compensação	5
STN/STR	11 e 20
OB	0
Câmbio	0
Compulsório	0
SELIC	14
CETIP	5 e 12
TN	11
SLB	16
BMF	13
CBLC	0
STR IF	0
CIP	2, 3, 6 e 8

Apesar de a variável STN/STR possuir uma correlação para um atraso de 20, optou-se por não utilizá-lo em função de estar muito distante da origem.

As variáveis Numerário, Ordens Bancárias, Câmbio, Compulsório, CBLC e STR-IF não apresentaram correlações com a STR-Clientes passíveis de uso.

Modelo Neural 13

Após tratar todas as variáveis individualmente com a STR-Clientes, criou-se uma rede com a aplicação conjunta de todas elas com a principal.

O quadro abaixo apresenta os resultados seguindo a ordem de número de neurônios e utilização de arquivo com e sem a representação binária do dia da semana.

Quadro 48: Performance do modelo 013.

Número de neurônios	Variável	Até R\$ 50 milhões		Até R\$ 100 milhões		Até R\$ 150 milhões		Até R\$ 200 milhões	
		Qtd.	%	Qtd.	%	Qtd.	%	Qtd.	%
1	Compe	20,0	15,5	39,0	30,2	54,6	42,3	69,4	53,8
5		21,4	16,6	40,2	31,2	57,0	44,2	72,0	55,8
1		23,0	17,8	42,2	32,7	55,0	42,6	70,6	54,7
5		20,4	15,8	40,0	31,0	54,0	41,9	70,6	54,7
1	STN / STR	20,0	15,9	39,4	31,3	56,2	44,6	69,4	55,1
5		18,0	14,3	34,0	27,0	52,2	41,4	63,6	50,5
1		23,8	18,9	39,8	31,6	57,2	45,4	68,0	54,0
5		17,6	14,0	35,8	28,4	52,4	41,6	66,2	52,5

1	SELIC	19,2	15,4	36,6	29,3	55,8	44,6	66,8	53,4
5		19,4	15,5	35,0	28,0	51,8	41,4	68,0	54,4
1		16,8	13,4	36,8	29,4	54,0	43,2	66,8	53,4
5		18,6	14,9	34,2	27,4	52,0	41,6	67,2	53,8
1	CETIP	18,4	14,6	35,4	28,1	54,6	43,3	66,6	52,9
5		20,6	16,3	40,2	31,9	54,4	43,2	65,8	52,2
1		16,0	12,7	39,6	31,4	54,4	43,2	67,0	53,2
5		18,0	14,3	35,0	27,8	52,2	41,4	66,0	52,4
1	TN	19,0	15,1	38,2	30,3	52,6	41,7	66,6	52,9
5		18,8	14,9	37,6	29,8	55,4	44,0	67,8	53,8
1		19,6	15,6	40,6	32,2	55,2	43,8	68,2	54,1
5		19,8	15,7	37,8	30,0	53,8	42,7	68,4	54,3
1	SLB	17,4	14,0	38,2	30,8	54,8	44,2	67,4	54,4
5		18,0	14,5	38,4	31,0	56,8	45,8	69,2	55,8
1		18,8	15,2	39,4	31,8	54,0	43,5	68,6	55,3
5		17,6	14,2	36,6	29,5	51,4	41,5	68,0	54,8
1	BMF	18,4	14,8	38,0	30,6	56,8	45,8	68,8	55,5
5		17,4	14,0	37,8	30,5	56,4	45,5	68,8	55,5
1		22,2	17,9	39,2	31,6	55,4	44,7	69,6	56,1
5		18,4	14,8	39,4	31,8	54,6	44,0	69,6	56,1
1	CIP	21,0	15,9	40,2	30,5	56,6	42,9	71,6	54,2
5		21,8	16,5	37,4	28,3	56,4	42,7	70,2	53,2
1		20,2	15,3	37,6	28,5	58,0	43,9	73,0	55,3
5		18,2	13,8	35,6	27,0	56,6	42,9	71,6	54,2
1	Todas as variáveis	17,4	14,0	38,2	30,8	55,6	44,8	66,0	53,2
5		20,2	16,3	36,0	29,0	50,0	40,3	61,2	49,4
1		20,2	16,3	39,2	31,6	54,4	43,9	64,4	51,9
5		19,2	15,5	37,6	30,3	52,8	42,6	64,0	51,6

A performance dos resultados encontra-se nos gráficos abaixo, cujo eixo das abscissas contém os valores e o das ordenadas o percentual médio de acertos. A linha contínua representa o percentual de acertos, enquanto a pontilhada o de erros.

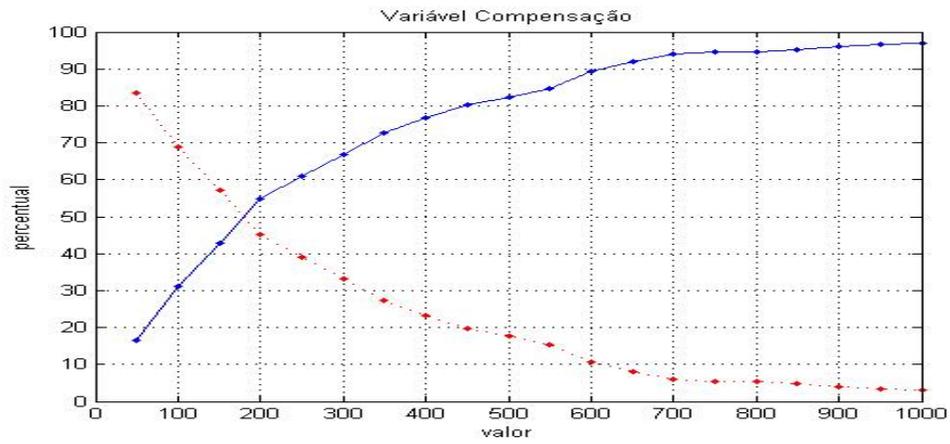


Gráfico 51: Representação do valor e percentual de acertos e erros da variável compensação.

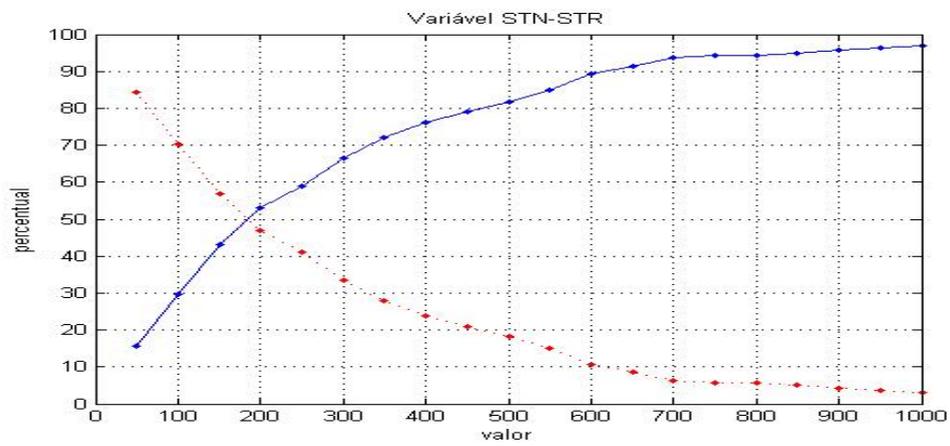


Gráfico 52: Representação do valor e percentuais de acertos e erros da variável STN-STR.

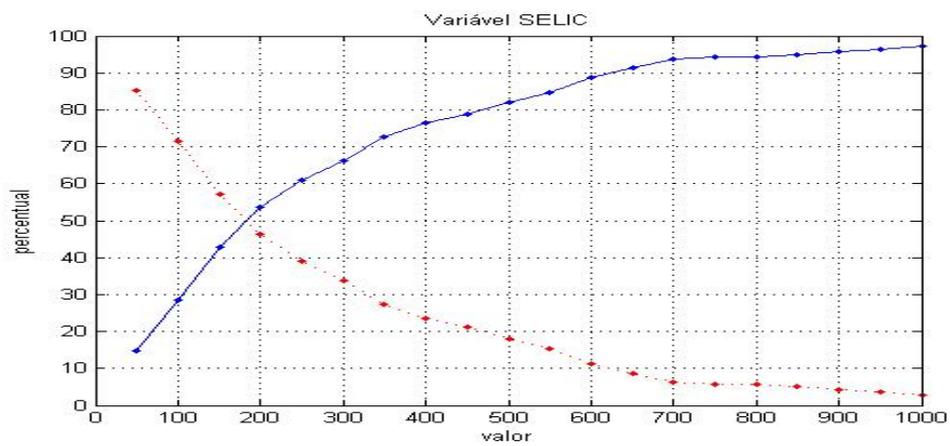


Gráfico 53: Representação do valor e percentuais de acertos e erros da variável SELIC.

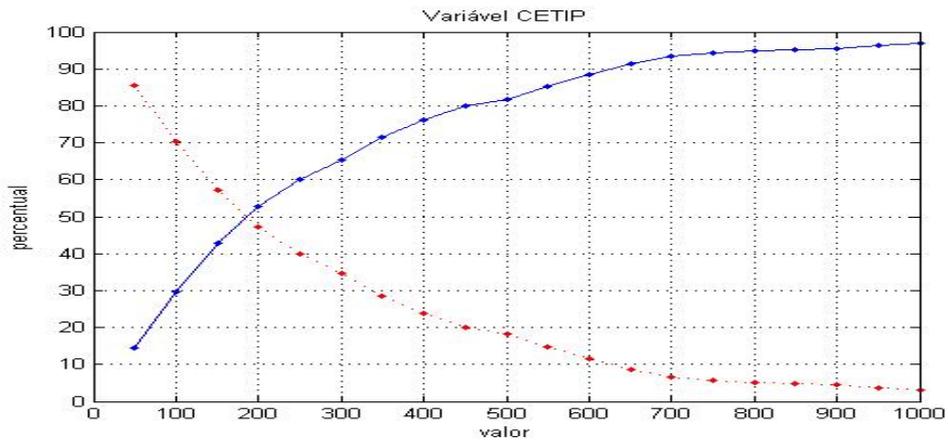


Gráfico 54: Representação do valor e percentuais de acertos e erros da variável CETIP.

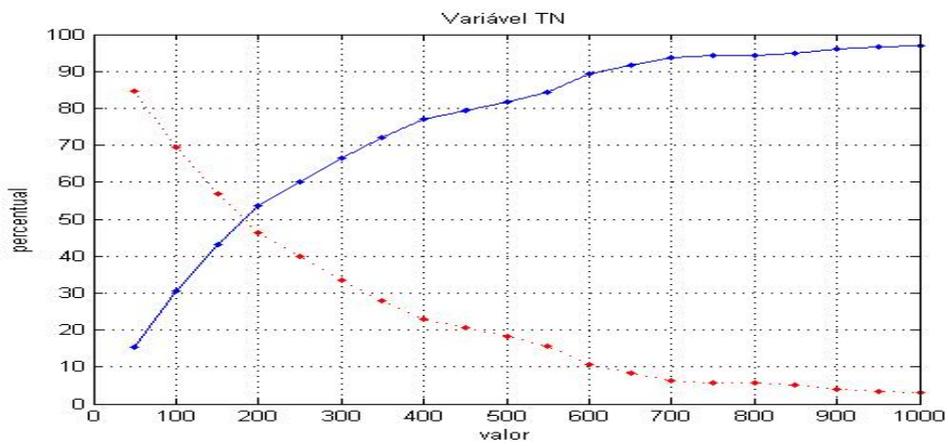


Gráfico 55: Representação do valor e percentuais de acertos e erros da variável TN.

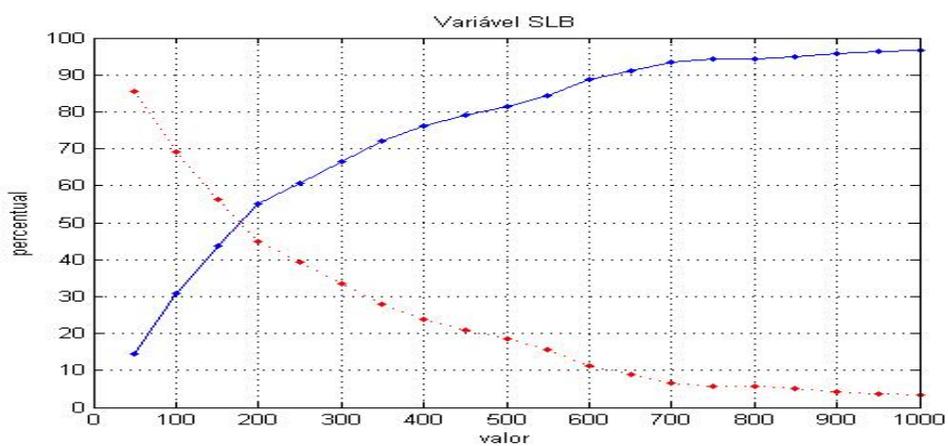


Gráfico 56: Representação do valor e percentuais de acertos e erros da variável SLB.

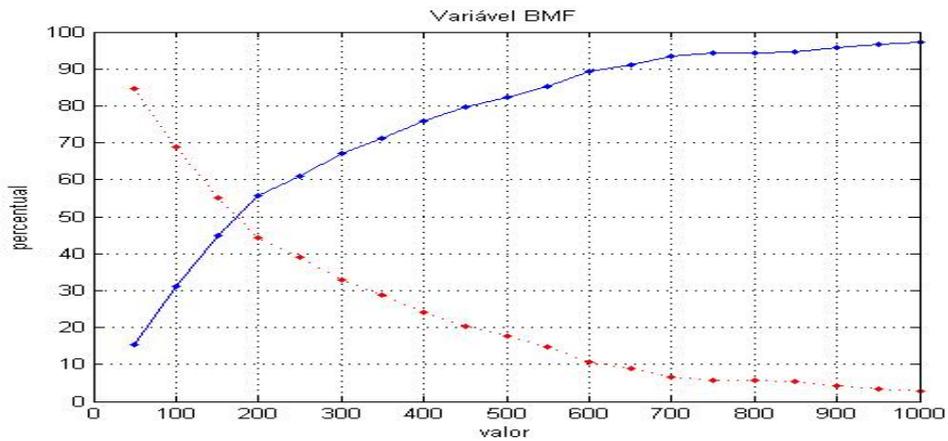


Gráfico 57: Representação do valor e percentuais de acertos e erros da variável BMF.

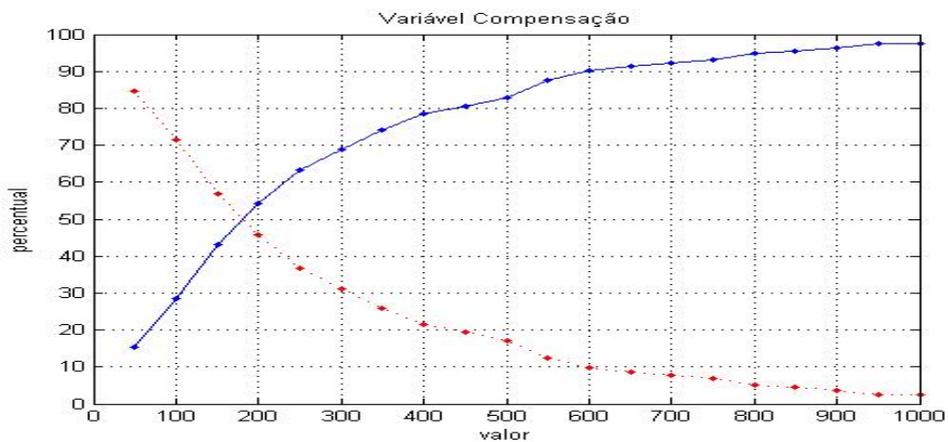


Gráfico 58: Representação do valor e percentuais de acertos e erros da variável CIP.

Analisando os gráficos, verifica-se que a característica das curvas são praticamente as mesmas para todas as variáveis, ou seja, não existe uma preferência no uso de qualquer variável, o que foi corroborado pelos dados numéricos apresentados na tabela acima.

Em uma observação mais atenta detecta-se a existência de três segmentos de reta bem definidos: o primeiro, com a maior inclinação, ocorre até a faixa de R\$ 200 milhões, o segundo, até R\$ 600 milhões e o último para valores superiores a R\$ 600 milhões. Isto demonstra que acima de 200 milhões, o crescimento do erro vai reduzindo. Ou seja, apesar da existência de erros muito grandes eles são poucos. Na média, a quantidade de erros no intervalo entre [200, 600] se situa em 35,0% e acima de R\$ 600 milhões, 10,9%. Nos gráficos pode-se verificar as relações entre acertos e erros da previsão.

Esta análise induz a se avaliar o comportamento da utilização de rede neural para prever unicamente os valores acima de R\$ 200 milhões. É possível que devido a grande amplitude dos dados, a rede não esteja conseguindo alcançar os picos, haja vista que a existência, também, de elevada frequência, faz com que seja necessário tratar um pico positivo em um instante t e, em um instante $t + 1$, um pico negativo, dificultando o mapeamento.

Modelo Neural 14

Baseado no experimento anterior, optou-se pela criação de rede neural utilizando apenas os registros superiores a R\$ 200 milhões de reais. Tal premissa decorreu do fato de haver sido verificado nos estudos anteriores que os maiores erros encontravam-se nos registros de maiores valores. Além disso, no modelo neural 13, concluiu-se que talvez alguma parte do erro seja relativa a dificuldade da rede mapear os valores mais elevados. Assim, decidiu-se pela criação de um modelo que previsse somente os valores maiores que este patamar.

Neste modelo gerou-se novo gráfico de autocorrelação sendo que só foram utilizados os registros de valores superiores a 200 milhões. Dessa forma, como pode ser observado no gráfico abaixo, o atraso com maior índice é referente a 5 dias, apesar de haver autocorrelação superior ao limite de confiança para atrasos de 12 e 20. Entretanto, optou-se por trabalhar somente com 5 dias que corresponde a uma semana.

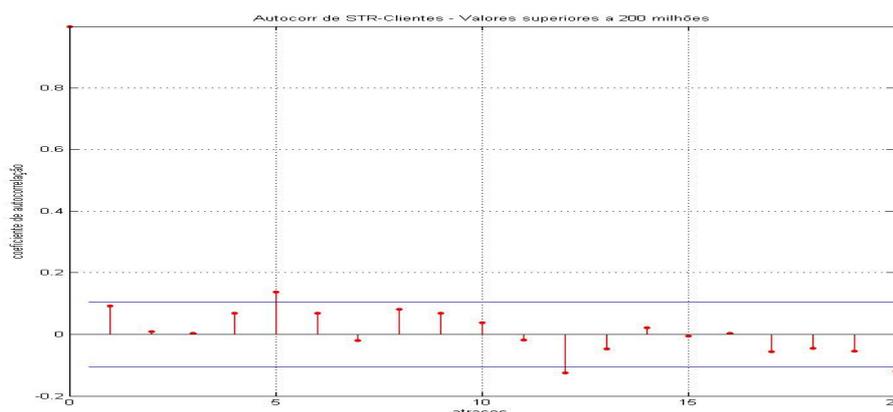


Gráfico 59: Autocorrelação da STR-Clientes para valores superiores a R\$ 200 milhões

Após o treinamento e validação da rede foi aplicado no modelo o conjunto entrada-saída de teste obtendo-se os resultados abaixo.

Quadro 49: Performance da rede 018.

Número de neurônios	Até R\$ 50 milhões		Até R\$ 100 milhões		Até R\$ 150 milhões		Até R\$ 200 milhões	
	Qtd.	%	Qtd.	%	Qtd.	%	Qtd.	%
1	0,0	0,0	1,0	0,4	14,4	5,7	30,4	12,0
5	3,2	1,3	7,6	3,0	15,6	6,1	34,2	13,5
1	3,4	1,3	10,2	4,0	26,2	10,3	42,2	16,6
5	2,2	0,9	11,4	4,5	24,4	9,6	40,2	15,8

A hipótese de facilidade no mapeamento dos valores mais elevados não se mostrou correta. Pelo contrário, a performance do modelo para estes valores foi extremamente baixa. Atribui-se esse péssimo desempenho, novamente, a amplitude e s inflexões existentes na série o que torna o modelo extremamente difícil de se mapear.

5.3 VARIÁVEIS EXTERNAS

É senso comum imaginar a influência de variáveis econômicas no ambiente empresarial. Notadamente no segmento bancário, as despesas e faturamento de uma empresa são passíveis de forte influência oriunda dos contextos econômico, político e social do país. As políticas de geração de empregos, o pagamento de 13º salário e aumentos salariais refletem em um acréscimo no montante de dinheiro em circulação na economia que se traduz no aumento de transações comerciais, inversões, empréstimos etc. Por outro lado, o aumento da alíquota do compulsório dos bancos promovem um enxugamento da liquidez com a consequente redução da atividade econômica. Todos esses fatores, dentre diversos outros, influenciam de forma significativa a movimentação bancária.

Baseado nesta premissa, promoveu-se estudos visando a utilização de variáveis econômicas que pudessem auxiliar no mapeamento da série STR-Clientes. Para tanto foram analisados índices, taxas e indicadores econômicos existentes no sítio do IPEA-Instituto de Pesquisa Econômica Aplicada (www.ipea.gov.br), órgão vinculado à Secretaria de Assuntos Estratégicos do Ministério do Planejamento Brasileiro. Tais indicadores são agrupados, por periodicidade, em 4 faixas: diário, mensal, trimestral e anual. Para o propósito desta tese, apenas a faixa diária faz sentido pois as demais não acompanham a dinâmica necessária da previsão desejada, ou seja, é preciso que o índice esteja disponível diariamente ao início da manhã para que se possa utilizá-lo no modelo. Ainda assim, mesmo os índices diários disponíveis não

necessariamente disponibilizam seus valores referentes à $t - 1$. A maioria só é apresentada com defasagens de $t - 3$ (importações e exportações em dólares americanos) ou até maiores, como no caso da taxa de juros over, DI-Depósito Interbancário, preço do petróleo bruto e outros.

Dessa forma, dentre as dezenas de indicadores disponíveis, verificou-se que apenas a taxa de câmbio do dólar americano agrega as condições necessárias para utilização, ou seja, tem-se a garantia de obtê-la com o resultado em $t - 1$ e é de fácil aquisição. Além disso, a moeda norte-americana é atrelada a diversos contratos e transações comerciais no Brasil, o que não acontece com os índices de ações, tanto brasileiros como internacionais.

Modelo Neural 15

Este modelo se caracterizou pelo uso da cotação do dólar como arquivo auxiliar de entrada da rede. Para tanto, apurou-se os valores da moeda norte-americana no mesmo período da STR-Clientes e realizou-se um pré-processamento semelhante ao aplicado na variável principal, conforme descrito abaixo.

Quadro 50: Arquivo de dados para a rede 13.

Características do tratamento do arquivo de dólar
Inclusão da média aritmética nos dias faltantes
Normalização dos dados
Exclusão da tendência
Exclusão da sazonalidade semanal
Exclusão de 4 componente senoidais

Juntamente com o arquivo acima, utilizou-se o arquivo da STR-Clientes após as transformações descritas no Anexo 5. Posteriormente, calculou-se a correlação cruzada entre as variáveis obtendo-se os atrasos de $t - 2$ a $t - 9$ com alguma correlação.

Os resultados podem ser vistos abaixo.

Quadro 51: Performance do modelo 013.

Número de neurônios	Até R\$ 50 milhões		Até R\$ 100 milhões		Até R\$ 150 milhões		Até R\$ 200 milhões	
	Qtd.	%	Qtd.	%	Qtd.	%	Qtd.	%
1	13,3	10,3	29,8	23,1	42,2	32,7	56,1	43,5
2	14,1	10,9	31,0	24,0	45,2	35,0	58,4	45,3
5	15,8	12,2	27,6	21,4	40,8	31,6	57,1	44,3
10	14,9	11,6	27,1	21,0	42,3	32,8	55,9	43,3

Vê-se, na compilação acima, que o uso da cotação do dólar americano para auxílio na previsão da variável degradou o resultado em relação aos modelos anteriores. Tal fato, pode ser atribuído a inexistência de correlações significativas entre as variáveis fazendo com que o comportamento do dólar dificultasse o mapeamento da STR. Cabe destacar que os índices de correlação para todos os atrasos foram muito baixos, não excedendo, em nenhum caso, a 0,05.

5.4 AGREGAÇÃO DE VARIÁVEIS

Quando da reestruturação do SPB, em abril de 2002, o Banco sob estudo criou 2 variáveis, denominadas Sistema de Transferência de Reservas – Clientes e Câmara Interbancária de Pagamentos – CIP, através das quais se realizam a movimentação de recursos dos clientes com outros bancos. Na ocasião a diferença existente entre as duas se limitava ao valor que era permitido cada uma operar, ou seja, na variável CIP só podia se realizar lançamentos até R\$ 5.000,00, enquanto a STR-Clientes era destinada aos valores superiores. Com o tempo, ambas as variáveis começaram a trabalhar com qualquer quantia. Atualmente, estas duas variáveis se confundem, tanto na operacionalização de suas rubricas como na aceitação de qualquer montante.

Tendo vista o exposto, optou-se pela unificação das duas variáveis, trabalhando com uma única série formada pela soma aritmética.

Adiante tem-se a estatística básica das variáveis individualmente e em conjunto para o período de 01/04/2003 a 31/03/2006. Diferente dos demais tratamentos, neste caso não se restringiu os valores a nenhum intervalo.

Quadro 52: Estatística das variáveis.

	Valores em R\$ milhão					
	Média	Mediana	Mínimo	Máximo	Variância	Desvio
STR - Clientes	362,25	331,07	-1.150,90	4.256,37	295.091,86	543,22
CIP	-400,33	-392,94	-1.470,49	1.145,66	108.848,75	329,92
STR – Clientes + CIP	-38,09	-124,27	-1.710,57	4.524,83	330.941,00	575,27

As transformações realizadas na série podem ser vistas adiante.

Gráfico 60: Transformações da série STR-Clientes e CIP.

Transformações da série
Soma dos saldos das variáveis STR-Clientes e CIP
Divisão dos valores por 1 milhão
Inclusão da média aritmética nos dias faltantes
Normalização dos dados
Exclusão da tendência
Exclusão da sazonalidade semanal
Exclusão de 1 componente senoidal na variável STR-Clientes (783,7 dias)
Exclusão de sazonalidades mensal (dias úteis e calendário)
Potência da série = 0,142292

Posteriormente foram criados gráficos dos valores da série STR-Clientes e CIP, conjuntamente, sendo acompanhado pelos de histograma e autocorrelação.

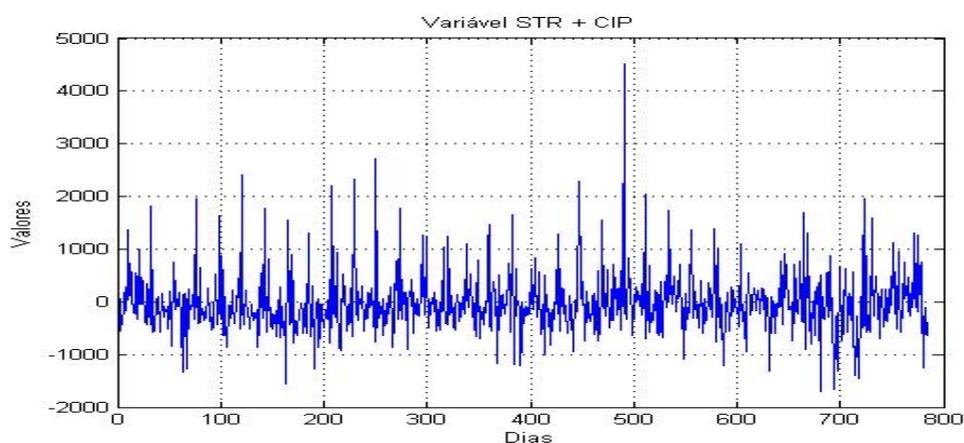


Gráfico 61: Série STR-Clientes e CIP.

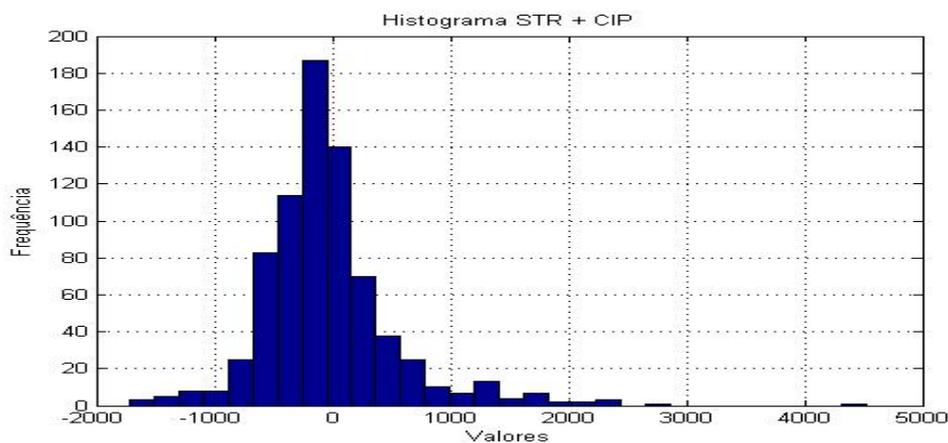


Gráfico 62: Histograma da série STR-Clientes e CIP.

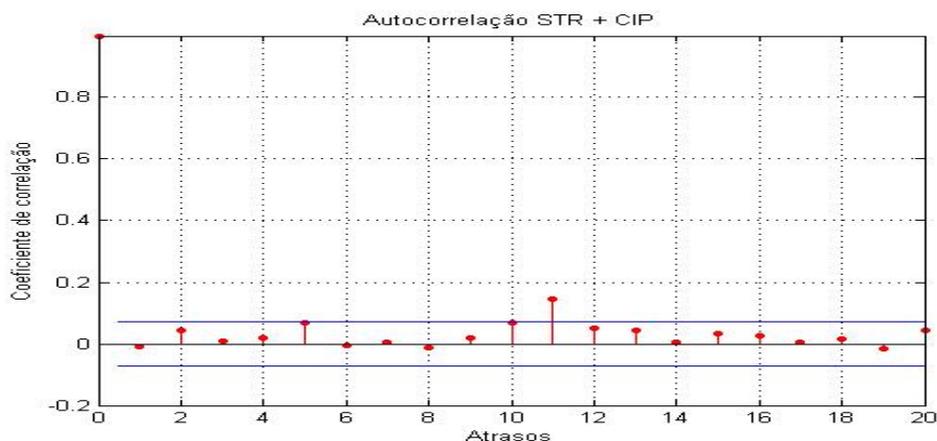


Gráfico 63: Autocorrelação da STR-Clientes e CIP para 20 atrasos.

O comportamento apresentado no primeiro gráfico, é semelhante aquele já observado para a variável STR-Clientes individualmente, ou seja, a sua característica é de um ruído branco.

Fica evidente no histograma a distribuição quase normal dos dados, apesar da existência de assimetria à direita, caracterizada por uma queda mais acentuada neste lado do que à esquerda.

Por fim, a função de autocorrelação mostra que praticamente inexistente qualquer relação para atrasos até 20 dias úteis (aproximadamente 1 mês), a exceção do 11º dia e, mesmo assim, com baixo índice (0,14563).

A performance de acertos da série após a extração das componentes determinísticas está tabulada abaixo.

Quadro 53: Performance de acertos.

Até R\$ 50 milhões		Até R\$ 100 milhões		Até R\$ 150 milhões		Até R\$ 200 milhões	
Qtd.	%	Qtd.	%	Qtd.	%	Qtd.	%
101	12,9	198	25,3	276	35,2	366	46,7

Pelos resultados alcançados verifica-se que a união das duas variáveis acrescentou ruído à série final, reduzindo os índices de performance já alcançados.

Modelo Neural 16

Baseado na análise descrita anteriormente, foi criado um modelo com a entrada exclusiva do arquivo STR-Clientes + CIP e atraso de 11 dias, referente a mais

significativa autocorrelação dos dados, conforme pode ser verificado no acima.

A definição do pré-processamento encontra-se no Anexo 5 e os resultados obtidos, seguindo a ordem de número de neurônios e utilização de arquivo com e sem representação binária do dia da semana, estão a seguir. Além disso, estão representados, também, experimentos com arquivo em ordem sequencial e aleatória.

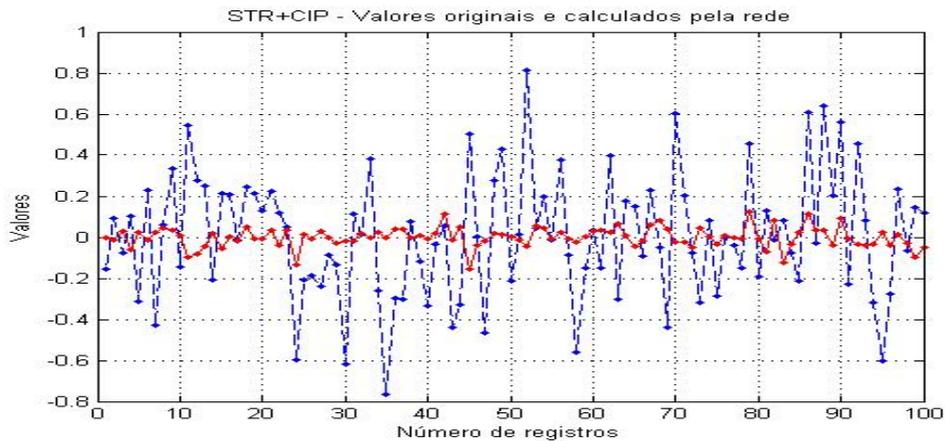
Quadro 54: Performance de acertos.

	Número de neurônios	Até R\$ 50 milhões		Até R\$ 100 milhões		Até R\$ 150 milhões		Até R\$ 200 milhões	
		Qtd.	%	Qtd.	%	Qtd.	%	Qtd.	%
Arq. seq. com bin.	1	10,2	8,8	19,0	16,4	26,6	22,9	39,2	33,8
	5	9,4	8,1	20,0	17,2	29,4	25,3	40,6	35,0
Arq. seq. sem bin.	1	12,0	10,3	20,2	17,4	26,4	22,8	41,6	35,9
	5	10,8	9,3	19,0	16,4	27,2	23,4	39,6	34,1
Arq. aleat. com bin.	1	13,0	11,2	27,4	23,6	40,2	34,7	53,0	45,7
	5	12,6	10,9	25,4	21,9	38,8	33,4	51,0	44,0
Arq. aleat. sem bin.	1	11,6	10,0	26,4	22,8	40,0	34,5	54,8	47,2
	5	13,6	11,7	28,8	24,8	40,6	35,0	51,2	44,1

Nenhuma das redes testadas respondeu bem aos dados. Nem mesmo houve, neste experimento, uma melhoria em relação a extração da parte determinística, onde foi alcançado o patamar de 46,8%, para o limite de R\$ 200 milhões. Aqui, apenas uma rede superou este patamar e, mesmo assim, por um percentual mínimo que pode-se considerar como erro estatístico.

Efetivamente os dados estão se mostrando com características aleatórias, pois todas as tentativas e estratégias utilizadas não estão apresentando bons resultados.

O treinamento realizado pelas redes não está conseguindo capturar as características dos dados. Abaixo encontra-se um gráfico com os 100 primeiros registros de treinamento (linha pontilhada) e os respectivos valores calculados pela rede no treinamento (linha contínua). Verifica-se que a saída da rede não consegue mapear os dados.



Quadro 55: Valores originais e calculados pela rede.

Vários experimentos foram realizados e os resultados sempre se mostraram ruins. Foram testadas redes com 1 e 2 camadas ocultas. Tentou-se saturar a rede com neurônios e utilizou-se critérios de parada onde se deixou a rede treinando por milhares de épocas. Nenhum dos resultados foi considerado satisfatório.

Modelo Neural 17

Similar ao modelo neural 13, aqui projetou-se uma rede de cuja entrada participaram todas as variáveis auxiliares em conjunto com a série STR+CIP. Estudos de autocorrelação para definição dos atrasos a serem aplicados forneceram subsídios para a criação do arquivo entrada-saída. O pré-tratamento dos dados encontra-se no Anexo 5.

Quadro 56: Atrasos significativos entre var. auxiliares e STR-Clientes e CIP.

	Atrasos
Numerário	15 e 16
Compensação	0
STN/STR	9 e 11
OB	13
Câmbio	7 e 13
Compulsório	0
SELIC	17
CETIP	0
TN	6 e 7
SLB	10
BMF	0
CBLC	0
STR-IF	11
STR + CIP	5

O quadro abaixo apresenta os resultados seguindo a ordem de número de neurônios e utilização de arquivo com e sem representação binária do dia da semana e sem representação binária do dia da semana. Além disso, estão representados, também, experimentos com arquivo em ordem sequencial de entrada de dados e aleatória.

Quadro 57: Performance de acertos.

	Número de neurônios	Até R\$ 50 milhões		Até R\$ 100 milhões		Até R\$ 150 milhões		Até R\$ 200 milhões	
		Qtd.	%	Qtd.	%	Qtd.	%	Qtd.	%
Arq. seq. com bin.	1	10,2	8,9	18,2	15,8	27,6	24,0	39,0	33,9
	5	8,6	7,5	18,0	15,7	29,4	25,6	37,6	32,7
Arq. seq. sem bin.	1	8,6	7,5	18,8	16,3	28,4	24,7	38,0	33,0
	5	8,8	7,7	19,4	16,9	28,0	24,3	38,0	33,0
Arq. aleat. com bin.	1	13,8	12,0	25,6	22,3	37,6	32,7	48,6	42,3
	5	16,8	14,6	28,8	25,0	40,6	35,3	50,2	43,7
Arq. aleat. sem bin.	1	12,4	10,8	24,8	21,6	37,6	32,7	51,0	44,3
	5	10,6	9,2	23,4	20,3	37,0	32,2	46,4	40,3

Verifica-se que a estratégia de se utilizar as variáveis auxiliares para auxílio da rede não se mostrou factível, obtendo-se resultados inferiores àqueles alcançados somente com a série univariada.

Redes com Atraso no Tempo

O tempo constitui um elemento essencial no processo de aprendizagem e

pode ser contínuo ou discreto. Independentemente da sua forma, o tempo é uma entidade ordenada que é básica para muitas tarefas encontradas na prática, como mercado de ações, processamento de sinais, evolução das vendas de um produto etc. É através da incorporação do tempo na operação de uma rede neural que ela é capacitada a seguir as variações estatísticas em processos não-estacionários como flutuações em preços do mercado de *commodities*, consumo de energia elétrica e fluxo de caixa de uma empresa, dentre outros.

A computação convencional assim como a maioria dos algoritmos de treinamento de redes neurais não estão aptos a tratar automaticamente a temporalidade dos dados. Muitos algoritmos de treinamento são incapazes de implementar mapeamentos dinâmicos como, por exemplo, o algoritmo *backpropagation*, que pode apenas aprender mapeamentos estáticos. Para que uma rede neural seja dinâmica, ela deve ter “memória” e uma forma simples de inserir memória de curto prazo na estrutura de uma rede neural é através de atrasos de tempo, que podem ser implementados a nível sináptico dentro da rede ou na camada de entrada da rede (HAYKIN, 2001).

Um artifício muito utilizado para o processamento temporal através destas redes envolve o uso de janelas de tempo, onde a entrada da rede utiliza trechos dos dados temporais como se eles formassem padrões estatísticos (BRAGA; CARVALHO; LUDEMIR, 1998).

Esta tese utilizou, em todos os experimentos envolvendo redes do tipo MLP, uma janela de tempo, visando simular um processamento dinâmico. Foram testadas várias janelas a fim de se verificar o tamanho ideal para a entrada de dados. Entretanto, apesar das diversas tentativas não se alcançou resultados satisfatórios em nenhuma das aplicações com MLP e, por isto, optou-se pelo uso de redes que implementassem na sua arquitetura o atraso no tempo. Segundo Rocha *et al* (2007), as redes com capacidade de representar internamente situações temporais são classificadas em dois grupos:

1. Redes com dinâmica interna: os valores passados da entrada e da saída não são re-introduzidos na entrada da rede, o que significa que nenhum conhecimento a priori é utilizado. A dinâmica é inserida pelas chamadas memórias de contexto e se dá através da realimentação das camadas escondidas. Um exemplo deste tipo de arquitetura é a rede Elman, Jordan e Hopfield.

2. Redes com dinâmica externa: utilizam os atrasos de tempo da entrada e/ou realimentação da camada de saída para a camada de entrada. Um exemplo deste tipo é a rede TDNN (*Time Delay Neural Network*), FIR (*Finite Impulse Response*) e NARX (*Nonlinear Autoregressive with Exogenous Input*).

Modelo Neural 18

Este modelo utilizou uma arquitetura baseada nas redes Elman (1990), que permite a criação de modelos que levam em consideração aspectos temporais e comportamentos dinâmicos, onde a saída de uma unidade depende de seu estado em um tempo anterior. Ao contrário das redes *feedforward* este tipo de rede, também conhecida como recorrentes ou redes com memória, possuem laços de realimentação, ou seja, a saída de um neurônio pode ser a entrada para outro de uma camada precedente ou, no caso de auto-realimentação, para o próprio neurônio. As redes recorrentes não possuem organização rígida e seus neurônios têm a liberdade para se conectarem a qualquer outro (HAYKIN, 2001) e esta característica lhe dá a capacidade de implementar a temporalidade na sua própria arquitetura.

Esta rede, além das unidades de entrada, intermediárias e de saída, possui unidades de contexto. As unidades de entrada e saída interagem com o ambiente externo, enquanto as unidades intermediárias e de contexto não o fazem. As unidades intermediárias podem ter funções de ativação lineares ou não lineares e as unidades de contexto são usadas apenas para memorizar as ativações anteriores das unidades intermediárias e podem ser consideradas como atrasos no tempo (BRAGA; CARVALHO; LUDEMIR, 1998). As unidades de contexto são importantes quando o processo de aprendizagem dos exemplos ocorre ao longo do tempo, ou seja, sempre que os valores passados da rede influenciam os valores presentes.

A definição do pré-processamento dos dados e a topologia de rede implementada estão no Anexo 5.

Os resultados obtidos, seguindo a ordem de número de neurônios e arquivos com e sem representação binária do dia da semana, encontram-se a seguir.

Quadro 58: Performance de acertos.

Número de neurônios	Até R\$ 50 milhões		Até R\$ 100 milhões		Até R\$ 150 milhões		Até R\$ 200 milhões	
	Qtd.	%	Qtd.	%	Qtd.	%	Qtd.	%
1	7,4	6,0	20,0	16,3	28,0	22,8	38,4	31,2
2	9,4	7,6	19,0	15,4	27,6	22,4	39,2	31,9
5	8,4	6,8	18,4	15,0	27,8	22,6	37,2	30,2
10	8,6	7,0	18,6	15,1	29,0	23,6	39,0	31,7
1	8,8	7,2	20,0	16,3	25,0	20,3	38,0	30,9
2	9,4	7,6	18,2	14,8	26,8	21,8	39,2	31,9
5	8,8	7,2	19,0	15,4	26,6	21,6	39,4	32,0
10	7,6	6,2	20,0	16,3	25,6	20,8	39,2	31,9

Contrariando as expectativas, o modelo Elman não foi superior aos modelos MLP em nenhum caso. Nem mesmo em relação a extração da parte determinística houve algum ganho. Ao contrário, este modelo introduziu ruído na série, pois reduziu a performance já obtida em experimentos anteriores.

Fica claro, também, que o uso de variáveis binárias para auxiliar o mapeamento não vêm apresentando melhoria, ou seja, devido as características aleatórias da série, o fato de informarmos à rede tratar-se de uma segunda-feira, terça-feira ou outro dia qualquer, não melhora o seu desempenho.

Modelo Neural 19

Este modelo utilizou uma rede do tipo TDNN para mapear os dados. Trata-se de uma rede do tipo *feedforward* de múltiplas camadas cujos neurônios das camadas intermediária e de saída são replicados ao longo do tempo (HAYKIN, 2001). A vantagem dessa rede em relação às demais é que a TDNN pode proporcionar memória a todos os neurônios das camadas escondidas e de saída, enquanto no janelamento só se consegue introduzir memória na primeira camada escondida.

A topologia da rede TDNN é basicamente a de uma MLP onde para cada conexão pode existir uma memória associada. As TDNN apresentam como entrada para a rede não apenas a entrada atual mas um vetor com a entrada atual e as últimas T entradas anteriores. Em resumo, as TDNN são redes estáticas que aprendem não com uma entrada, mas com uma sequência de entradas (presente e passada) (ROCHA, 2007).

No Anexo 5 tem-se a definição do pré-tratamento dos dados assim como a topologia da rede. Os resultados alcançados estão a seguir.

Quadro 59: Performance de acertos.

Número de neurônios	Até R\$ 50 milhões		Até R\$ 100 milhões		Até R\$ 150 milhões		Até R\$ 200 milhões	
	Qtd.	%	Qtd.	%	Qtd.	%	Qtd.	%
1	8,4	6,8	19,8	16,1	25,8	21,0	38,6	31,4
2	8,4	6,8	20,2	16,4	25,0	20,3	38,8	31,5
5	8,6	7,0	18,0	14,6	25,4	20,7	39,6	32,2
10	7,8	6,3	19,4	15,8	25,6	20,8	39,4	32,0

O modelo TDNN também não apresentou resultados satisfatórios. Cabe comentar que a flexibilidade de implementação de atrasos na rede auxiliou na realização de vários experimentos. Atrasos fixos e variando tanto na camada intermediária como na de saída foram testados mas os resultados apontaram para o descarte do modelo.

5.5 CONCLUSÃO

Os experimentos realizados até o momento se mostraram ineficazes para o objetivo definido de apresentar índices de performance de acerto superiores a 80%, para um patamar de R\$ 200 milhões. Entretanto, apesar do baixo desempenho ainda assim os resultados obtidos pelos modelos têm sido superiores àqueles alcançados pelo banco, conforme quadro abaixo em que se tem compilados a quantidade e percentual de dias cujos erros de previsões foram iguais ou inferiores aos limites estipulados na primeira coluna. Esta estatística se refere a variável STR-Clientes individualmente, durante o ano de 2007 e para erros até 1 bilhão de reais.

Quadro 60: Quantidade e percentual de dias com erro de previsão para a variável STR-Clientes.

	Quantidade de dias	Percentual de acertos		Quantidade de dias	Percentual de acertos
x <= 50	26	10,4	x <= 550	187	75,1
x <= 100	49	19,7	x <= 600	195	78,3
x <= 150	65	26,1	x <= 650	202	81,1
x <= 200	87	34,9	x <= 700	204	81,9
x <= 250	108	43,4	x <= 750	209	83,9
x <= 300	130	52,2	x <= 800	214	85,9
x <= 350	146	58,6	x <= 850	221	88,8
x <= 400	159	63,9	x <= 900	226	90,8
x <= 450	169	67,9	x <= 950	233	93,6
x <= 500	182	73,1	x <= 1000	237	95,2

Abaixo tem-se quadro semelhante ao anterior com a estatística para as

variáveis STR-Clientes e CIP, conjuntamente.

Quadro 61: Quantidade e percentual de dias com erro de previsão para as variáveis STR-Clientes e CIP conjuntamente.

	Quantidade de dias	Percentual de acertos		Quantidade de dias	Percentual de acertos
x <= 50	21	8,4	x <= 550	180	72,3
x <= 100	44	17,7	x <= 600	197	79,1
x <= 150	60	24,1	x <= 650	203	81,5
x <= 200	79	31,7	x <= 700	208	83,5
x <= 250	100	40,2	x <= 750	214	85,9
x <= 300	111	44,6	x <= 800	217	87,1
x <= 350	120	48,2	x <= 850	222	89,2
x <= 400	137	55,0	x <= 900	226	90,8
x <= 450	155	62,2	x <= 950	229	92,0
x <= 500	173	69,5	x <= 1000	232	93,2

Trata-se, conforme já comentado anteriormente, de problema extremamente complexo que a empresa vem buscando há anos melhoria nas previsões, sem, no entanto, alcançar índices considerados satisfatórios.

Apesar de representar um valor significativo, a meta de R\$ 200 milhões como patamar para se considerar uma previsão como correta se torna pequena ao compará-la com o movimento realizado. No ano de 2007, por exemplo, o movimento médio diário das variáveis STR-Clientes e CIP conjuntamente, em valores absolutos, foi da ordem de R\$ 573 milhões, com valor máximo de R\$ 2,17 bilhões e mínimo de R\$ 3,19 bilhões.

Além disso os erros são tratados de forma absoluta, ou seja, pela métrica definida (corte em R\$ 200 milhões) um erro de, por exemplo, R\$ 201 milhões tem o mesmo peso de um erro de R\$ 1 bilhão, o que, obviamente, para efeitos práticos não tem a mesma importância. Há necessidade de se avaliar o uso de pesos na métrica de performance para punir menos os erros próximos do patamar definido e aumentar a punição para os maiores erros.

A extração das componentes determinísticas da série foi trabalhada cuidadosamente deixando como resultado um resíduo cujas características inviabilizaram qualquer outro tratamento convencional. O resíduo aplicado à rede, quando apresentava ganho era pequeno, conforme pode ser verificado nos modelos, mas, ainda assim, permitiu um aumento na melhoria da previsão, sem contudo poder considerar os resultados como ideais.

6 ESTRATÉGIAS NÃO NEURAIAS E AVALIAÇÕES

6.1 INTRODUÇÃO

O uso de diversos modelos para tratar de um mesmo assunto é uma prática que vem se consolidando não só pela facilidade em se utilizar os modelos através de *softwares* comerciais e acadêmicos como pela quantidade de modelos existentes, o que leva o pesquisador a buscar a melhor alternativa para representar o seu problema.

Tashaya-Temizel e Ahmad (2005) testaram 9 séries com modelos distintos de redes neurais (TDNN) e ARIMA. Posteriormente, foram calculadas as previsões para um modelo híbrido ARIMA-NN, onde a parte linear era tratada no algoritmo de Box-Jenkins e o resíduo aplicado à rede. O resultado mostrou que 5 das 9 séries responderam melhor ao modelo híbrido, e as 4 restantes obtiveram melhores resultados com modelos neurais individuais.

Bao *et al.* (2007) fizeram uma comparação de performance da previsão da densidade da madeira utilizando uma rede neural MLP, outra MLP treinada com o *backpropagation* otimizado pelo algoritmo de swarm (PSO – *Particle Swarm Optimization*) e um modelo ARIMA. A densidade da madeira é a mais importante característica para se determinar a qualidade do produto e está ligada a diversas propriedades, tais como: dureza, taxa de encolhimento e resistência. Os resultados da pesquisa mostraram que a rede neural treinada com o algoritmo otimizado por PSO apresentou melhores resultados que os demais.

Apesar de a grande maioria dos trabalhos de comparação de performance utilizarem redes neurais e modelos ARIMA, encontramos outros que extrapolam esta fronteira. Bouqata *et al.* (2000) fizeram um comparativo entre modelos neurais, Lógica Difusa (*Fuzzy Logic*) e ARIMA para a série de consumo privado da economia espanhola, obtendo melhores resultados com o uso do modelo difuso.

Nesta tese, optou-se, fundamentalmente, pela utilização de redes neurais artificiais no mapeamento dos dados, haja vista a hipótese do resíduo final, após a

extração das componentes determinísticas, possuir características de não linearidade. Entretanto, após várias análises verificou-se que a maior parte do resíduo é composto por ruído branco, que se caracteriza pela aleatoriedade dos dados, e, por consequência, sem expectativas de se conseguir melhoraria nos índices alcançados.

Apesar das características do resíduo já mencionadas, cabe verificar a qualidade dos resultados com outros modelos, notadamente o ARIMA por ser um dos mais utilizados em finanças, assim como avaliar a utilização de outras estratégias para alcançar o objetivo proposto que é o desenvolvimento de um modelo de auxílio a tomada de decisões. Sob esta ótica, a Lógica Difusa pode fornecer boas aproximações da série, através da introdução de incertezas no modelo.

6.2 MODELO ARIMA - AUTOREGRESSIVE INTEGRATED MOVING AVERAGE

O modelo ARIMA, proposto por Box e Jenkins (BOX; JENKINS; REINSEL, 1994) faz parte da classe mais geral de modelos de previsão de séries temporais e podem ser aplicados em séries estacionárias ou não.

A sigla ARIMA em português significa auto-regressivo, integrado e médias móveis. Auto-regressivo corresponde à defasagem da série transformada, o termo médias móveis às defasagens dos erros aleatórios e o termo integrado ao processo de diferenciação da série original a fim de torná-la estacionária.

A condição de estacionariedade que é introduzida no modelo como forma simplificadora implica em considerar constantes a média e a variância da série temporal $\{Z_1, Z_2, \dots, Z_t\}$ e a covariância entre Z_t e Z_{t+k} dependente apenas do intervalo k .

A classe de modelos ARIMA é descrita matematicamente como (ABELÉM, 1994)

$$Z_k = \phi_1 Z_{k-1} + \phi_2 Z_{k-2} + \dots + \phi_p Z_{k-p} + \varepsilon_k + \theta_1 \varepsilon_{k-1} + \theta_2 \varepsilon_{k-2} + \dots + \theta_q \varepsilon_{k-q} \quad \text{Eq. (30)}$$

Onde $\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-q}$ são os q erros passados das séries. As constantes $\phi_1, \phi_2, \dots, \phi_p$ são os parâmetros auto-regressivos do modelo, enquanto $\theta_1, \theta_2, \dots, \theta_q$, são os parâmetros das médias móveis.

O modelo ARIMA possui 3 parâmetros necessários que são passados à série e representados como ARIMA(p,d,q) onde p é o número de termos auto-regressivos, d é

o número de diferenças não sazonais e q é o número de termos médias móveis. Todos os parâmetros são inteiros e iguais ou maiores que zero.

É importante ressaltar que se $p = 0$ o modelo resultante será somente constituído de médias móveis, no caso de $d = 0$ temos que a série original já é estacionária, não necessitando de se realizar nenhuma diferenciação nos dados; por último, a hipótese de $q = 0$ significa que o modelo é puramente regressivo.

A construção de um modelo ARIMA se constitui de 3 etapas bem definidas (BOX; JENKINS; REINSEL, 1994) (MONTGOMERY; JOHNSON; GARDINER, 1990):

- Identificação do modelo;
- Estimação; e
- Diagnóstico.

Alguns autores, a exemplo de Bowerman (BOWERMAN; O'CONNELL, 1979) consideram que a construção do modelo é realizada em 4 passos – identificação, estimação, diagnóstico e previsão. Entretanto, a fase de previsão não pode ser considerada como integrante da construção do modelo pois o mesmo já foi desenvolvido nos passos anteriores. A previsão corresponde tão-somente à aplicação do modelo. Adotou-se, para efeito desta tese, somente os 3 passos mais aceitos como sendo os responsáveis pela fase de construção do modelo.

A identificação do modelo consiste na tentativa de se encontrar os parâmetros p , d e q que representem satisfatoriamente a série histórica em questão. Esta é a etapa considerada como sendo a mais crítica do método (MORETTIN; TOLOI, 1987).

Inicialmente, a fim de se tornar a série estacionária, diferencia-se a série Z_k tantas vezes quanto necessárias. O número de diferenciações corresponde ao parâmetro d . Após esta etapa resta descobrir quais os melhores valores de p e q . Tais parâmetros são obtidos através dos cálculos das autocorrelações teóricas e autocorrelações parciais estimadas (correlações internas entre membros da série separados por intervalos de tempo constantes) cujos comportamentos devem imitar o das respectivas quantidades teóricas.

A segunda etapa de construção do modelo está relacionada com a estimação. Nesta fase já se tem identificado um modelo provisório para a série que foi obtida no

passo 1. Calcula-se, então, os valores estimados dos coeficientes autoregressivos e/ou médias móveis e ajusta-se o modelo para a série em questão. Para realizar tal ajuste a maioria dos autores sugere o uso do método de mínimos quadrados. Entretanto, Morettin (MORETTIN; TOLOI, 1987) trabalha com o cálculo de máxima verossimilhança para a realização de ajustes do modelo.

Finalmente, tem-se a fase de diagnóstico onde através do modelo construído nas etapas anteriores busca-se verificar se ele representa ou não corretamente os dados. A verificação é feita, na maioria das vezes, através da análise dos resíduos gerados pelo modelo construído.

Ao se ajustar um modelo ARIMA procura-se representar toda a estrutura de dependências dos dados existentes na série. Assim, a existência de resíduos no modelo não deve ter nenhuma relação com o tempo pois, caso exista, significa que o modelo construído não foi suficientemente abrangente. Nestes casos, deve-se retornar ao primeiro passo e iniciar novamente o processo de construção.

Após a conclusão com sucesso das etapas anteriores chega-se ao objetivo do trabalho que é o de realizar previsões k passos à frente baseado nos valores existentes na série sob estudo.

As previsões são obtidas diretamente a partir da equação do modelo construído nos passos anteriores. Seja $Z_k(j)$ a previsão de Z_{k+j} obtida no instante k . A equação de previsão é apenas a obtida da própria expressão do modelo para Z_t , escrita para o instante de interesse $t + j$.

6.2.1 Aplicação do Modelo ARIMA

O arquivo utilizado neste experimento, oriundo das transformações descritas no Anexo 5, consistiu em 675 registros iniciais para ajuste e os 109 registros finais para teste do resultado da previsão gerada pelo modelo. Cabe destacar que os cálculos foram feitos com auxílio do *software* forecast.pro.

A fim de se verificar o aumento de performance, acaso existente, com o uso do modelo ARIMA é necessário se calcular quanto da parte determinística é responsável pela previsão. Para tanto, a reconstituição do resíduo apresentou os seguintes percentuais de acerto para um universo de 109 registros relativos ao período de 01/11/2005 a 31/03/2006.

Quadro 62: Estatística de acertos após extração da parte determinística da série.

Até R\$ 50 milhões		Até R\$ 100 milhões		Até R\$ 150 milhões		Até R\$ 200 milhões	
Qtd.	%	Qtd.	%	Qtd.	%	Qtd.	%
19	17,4	38	34,9	56	51,4	72	66,1

Os resultados da aplicação do modelo estão descritos abaixo.

Quadro 63: Performance dos modelos ARIMA.

Modelo	Até R\$ 50 milhões		Até R\$ 100 milhões		Até R\$ 150 milhões		Até R\$ 200 milhões	
	Qtd.	%	Qtd.	%	Qtd.	%	Qtd.	%
ARIMA(1,1,1)	17	15,6	37	33,9	59	54,1	77	70,6
ARIMA(1,0,3)	15	13,8	36	33,0	60	55,0	74	67,9
ARIMA(1,0,2)	20	18,3	41	37,6	59	54,1	72	66,1
ARIMA(2,2,3)	10	9,2	21	19,3	32	29,4	43	39,4
ARIMA(3,0,3)	18	16,5	39	35,8	56	51,4	71	65,1

Dentre os modelos apresentados, verifica-se que os 3 primeiros são os que melhores resultados apresentam apesar de não terem obtido aumento significativo em relação a extração pura e simples da parte determinística. Entretanto, demonstra que pela simplicidade de utilização do modelo bem como o baixo custo computacional de implementação, no geral os resultados se mostram melhores que os alcançados com o uso de redes neurais artificiais.

O gráfico abaixo apresenta na parte de cima a série residual a ser prevista e na parte de baixo as previsões obtidas com o modelo ARIMA(1,1,1), considerado o melhor modelo, apesar de ter sido superior aos demais somente no patamar de R\$ 200 milhões. O modelo ARIMA(1,0,2) não obteve a melhor solução para o patamar de R\$ 200 milhões, entretanto alcançou desempenho superior ao demais nos limites de R\$ 50 e R\$ 100 milhões. Conclui-se que ambos são adequados ao propósito de geração de previsão.

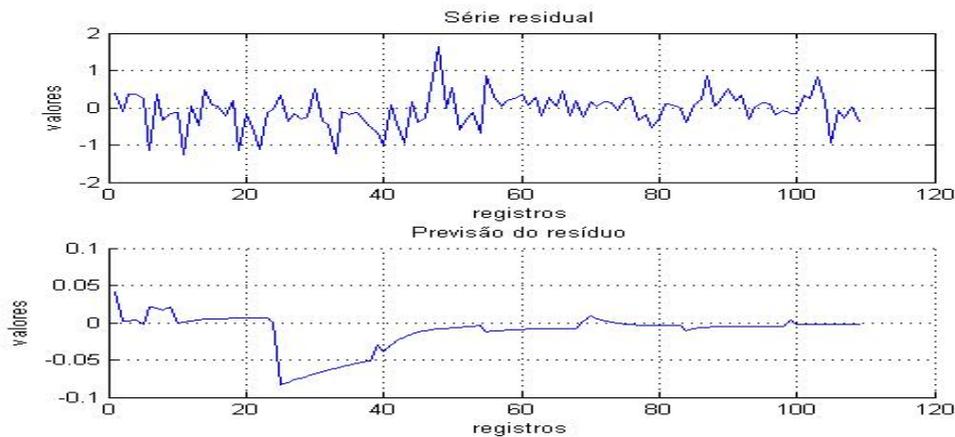


Gráfico 64: Série residual e previsão.

Destaca-se o fato da escala entre os gráficos ser da ordem de 10^2 , ou seja, os valores da previsão são, em média, 100 vezes inferiores aos valores dos resíduos, significando que o modelo ARIMA conseguiu extrair da série apenas uma pequena parcela. Isto pode ser quantificado através do cálculo da potência que antes da aplicação do modelo valia $Pot = 0,210606$ e, posteriormente, $Pot = 0,207099$.

Conclui-se, portanto, que pouco de linearidade passível de extração através do ARIMA ainda persiste na série.

6.3 MODELO DIFUSO

A Lógica Difusa, também conhecida como Lógica Nebulosa ou *Logic Fuzzy*, é uma forma elegante e eficaz para solucionar problemas, pois sistemas construídos com base nesta teoria são mais compactos e baratos que aqueles que utilizam soluções convencionais. Além disso, esse paradigma possui a capacidade de processar dados inexatos ou parcialmente corretos. De fato, os sistemas difusos vêm sendo utilizados em diversos tipos de aplicações comerciais, inclusive na área de previsão de séries temporais (KUO, R.J.; XUE, K.C., 1999) (SHEEN, J.N., 2004).

Aristóteles (384 - 322 a.C.) estabeleceu um conjunto de regras para que conclusões pudessem ser aceitas logicamente. O emprego da lógica de Aristóteles leva a uma linha de raciocínio baseado em premissas e conclusões. Desde então a lógica ocidental tem sido binária, isto é, uma declaração é falsa ou verdadeira, não podendo ser ao mesmo tempo parcialmente verdadeira ou falsa (TAKEMURA, 2008).

A lógica de Aristóteles trata as sentenças classificando-as como verdadeiras

ou falsas. Não obstante, muitas das experiências humanas não podem ser classificadas simplesmente como verdadeiras ou falsas. Este tipo de resposta, ou seja, um sim ou não a estas questões é, na maioria das vezes, incompleto. Na verdade, entre a certeza de ser e a certeza de não ser, existem infinitos graus de incerteza. Esta imperfeição intrínseca à informação representada numa linguagem natural, tem sido tratada matematicamente no passado com o uso da teoria das probabilidades.

A Lógica Difusa viola as suposições Aristotélicas. O conceito de dualidade, estabelecendo que algo pode e deve coexistir com o seu oposto, faz a Lógica Difusa parecer natural. Além disso, ela tem se mostrado mais adequada para tratar imperfeições da informação do que a teoria das probabilidades. De forma mais objetiva e preliminar, pode-se definir Lógica Difusa como sendo uma ferramenta capaz de capturar informações vagas, em geral descritas em uma linguagem natural e convertê-las para um formato numérico, de fácil manipulação pelos computadores de hoje em dia.

A Teoria Difusa surgiu com a publicação do artigo intitulado “*Fuzzy Sets*” por Lotfi A. Zadeh, em 1965. Neste trabalho, ele introduziu o conceito de conjunto difuso, cujas fronteiras não são bem demarcadas ou precisas, em contraste com os conjuntos tradicionais, nitidamente definidos e que possuem fronteiras precisas. (ZADEH, 1965). Este trabalho é apresentado, também, como uma abordagem alternativa para o tratamento de incertezas pois ele mostra que em muitos aspectos essa teoria se parece com a teoria da probabilidade – o uso do intervalo unitário e o fato de que ambos os conceitos descrevem incertezas, apesar de o tipo de incerteza tratado pela Estatística diferir daquele tratado pela Lógica Difusa (OLIVEIRA, 1999).

A Lógica Difusa é baseada na Teoria dos Conjuntos Difusos e esta é uma generalização da Teoria dos Conjuntos tradicionais. A Teoria Nebulosa foi desenvolvida para resolver os paradoxos gerados à partir da classificação hermética de que um elemento pertence ou não a um conjunto, ou seja, uma proposição lógica clássica tem apenas dois extremos: ou é verdadeira ou falsa. Entretanto, na lógica nebulosa, uma premissa varia em grau de verdade de 0 a 1, o que leva a premissa a ser parcialmente verdadeira ou parcialmente falsa.

Com a incorporação do conceito de “grau de verdade”, a Teoria dos Conjuntos Difusos estende a Teoria dos Conjuntos tradicionais. Os grupos são

rotulados qualitativamente (usando termos linguísticos, tais como: alto, frio, pequeno, perto etc.) e os elementos destes conjuntos são caracterizados variando o grau de pertinência (valor que indica o grau em que um elemento pertence a um conjunto).

Teoria dos Conjuntos Tradicionais

A teoria dos conjuntos difusos é, em grande parte, uma extensão da teoria dos conjuntos tradicionais. Dessa forma, cabe apresentar uma rápida revisão dos conceitos desta teoria.

Existem três formas através do qual um conjunto A pode ser definido:

1. Um conjunto A cujo os membros são a_1, a_2, a_3 é geralmente definido por:

$$A = \{ a_1, a_2, a_3 \} \quad \text{Eq. (31)}$$

Este tipo de definição é aplicado somente a conjuntos finitos.

2. Um conjunto também pode ser representado por:

$$A = \{x|P(x)\} \quad \text{Eq. (32)}$$

onde o símbolo $|$ denota a frase “tal que,” e $P(x)$ designa a proposição da forma “ x tem a propriedade P ”. Isto é, A é definido por esta notação como o conjunto de todos elementos de X para qual a proposição $P(x)$ é verdadeira.

3. Por fim, um conjunto pode ser definido por uma função, geralmente chamada de função característica, que declara quais elementos de X são membros do conjunto e quais não são. Então, nestes casos, um conjunto A é definido por sua função característica, γ_A representada a seguir.

$$\gamma_A(x) = \begin{cases} 1 & \text{para } x \in A \\ 0 & \text{para } x \notin A \end{cases} \quad \text{Eq. (33)}$$

Caso um conjunto não contenha nenhum membro ele é chamado de conjunto vazio e representado por \emptyset . Alguns conjuntos tradicionais podem ser vistos no quadro adiante.

Quadro 64: Exemplos de conjuntos tradicionais.

Conjunto	Elementos	Descrição
Z	$\{\dots, -2, -1, 0, 1, 2, \dots\}$	Conjunto dos números inteiros
N	$\{1, 2, 3, \dots\}$	Conjunto dos números naturais
Z_0	$\{0, 1, 2, \dots\}$	Conjunto dos números inteiros não negativos
R		Conjunto dos números reais
R^+		Conjunto dos números reais não negativos

Dentre as operações utilizadas na teoria dos Conjuntos Tradicionais pode-se destacar a *união* representado por \cup , *intersecção* representado por \cap e o *complemento* representado por \neg .

A união dos conjuntos A e B é denotada por:

$$A \cup B = \{x | x \in A \text{ ou } x \in B\} \quad \text{Eq. (34)}$$

A intersecção de dois conjuntos A e B é denotada por:

$$A \cap B = \{x | x \in A \text{ e } x \in B\} \quad \text{Eq. (35)}$$

O complemento representa os elementos de um universo que não fazem parte de um conjunto:

$$\neg A = \{x | x \in X \text{ e } x \notin A\} \quad \text{Eq. (36)}$$

Teoria dos Conjuntos Difusos

Um conjunto difuso é definido em um universo de discurso (conjunto base) X , e caracterizado pela sua função de pertinência:

$$A : X \rightarrow [0,1] \quad \text{Eq. (37)}$$

onde $A(x)$ representa o grau com que x pertence a A e expressa a extensão com que x se enquadra na categoria representada por A .

Uma função de pertinência particular pode ser visualizada por meio da equação abaixo. Verifica-se que esta função é triangular e as variáveis a , b e c são parâmetros da função.

$$\mu(x) = \begin{cases} \frac{x-a}{b-a} & \text{se } x \in [a, b) \\ \frac{c-x}{c-b} & \text{se } x \in [b, c] \\ 0 & \text{caso contrário} \end{cases} \quad \text{Eq. (38)}$$

Conforme definido anteriormente, a teoria dos conjuntos difusos é uma extensão da teoria dos conjuntos tradicionais. Assim, as principais operações e relações entre conjuntos difusos são definidas como extensão das operações e relações tradicionais, como pode ser visto no quadro abaixo, onde A e B denotam conjuntos difusos sobre um conjunto base X e $A(x)$ e $B(x)$ representam os graus de pertinência de x nos conjuntos difusos A e B respectivamente.

Quadro 65: Operações e relações com conjuntos difusos.

Nº	Operação	Representação	Natureza
1	Complemento	$\neg A(x) = 1 - A(x)$	Operação
2	Diferença	$(A \neq B)$ se $A(x) \neq B(x)$ para pelo menos um elemento de $x \in X$	Relação
3	Igualdade	$(A = B)$ se $A(x) = B(x)$ para todo $x \in X$	Relação
4	Inclusão	$(A \subseteq B)$ se $A(x) \leq B(x)$ para todo $x \in X$	Relação
5	Intersecção	$A \cap B = A(x) \cap B(x) = \min [A(x), B(x)]$	Operação
6	União	$A \cup B = A(x) \cup B(x) = \max [A(x), B(x)]$	Operação

Sistemas Difusos

Um sistema difuso representa uma função (geralmente não-linear) entre duas ou mais variáveis através da descrição linguística da relação entre elas. O conhecimento para esta descrição pode envolver conceitos vagos e/ou imprecisos, representados por sentenças nebulosas do tipo:

$$x \text{ é } A \quad \text{Eq. (39)}$$

Um sistema difuso é então um conjunto de regras do tipo se-então que mapeia entradas em saídas. Estas regras definem regiões no espaço entrada x saída ($X \times Y$). Um sistema difuso $F: X \rightarrow Y$ aproxima a função $f: X \rightarrow Y$ cobrindo seu gráfico com fragmentos ou remendos que se sobrepõem (*patches*). Cada regra determina um remendo. A aproximação da função melhora a medida que os fragmentos aumentam em

número e diminuem em tamanho como é mostrado na figura abaixo.

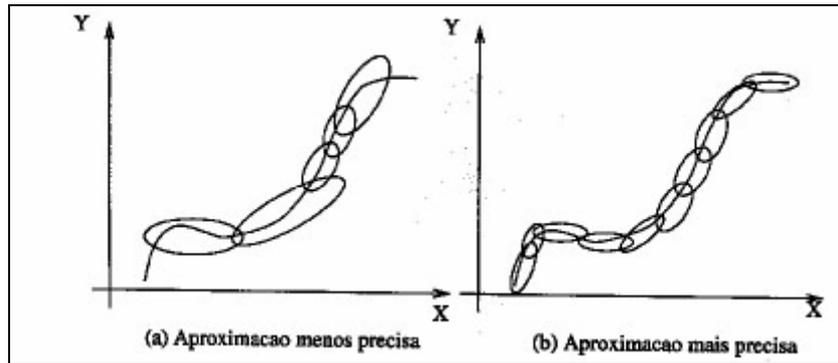


Figura 2: Gráficos mostrando como aumentar a precisão da aproximação nebulosa.

Regras Nebulosas

Uma regra nebulosa é uma sentença nebulosa que expressa o conhecimento da relação entre duas ou mais variáveis através de uma relação entre os conjuntos difusos associados a cada sentença. Por exemplo, o conjunto difuso A pode ser ligado ao conjunto difuso B pela regra $A \rightarrow B$, escrita como:

$$\text{Se } x \text{ é } A \text{ então } y \text{ é } B \quad \text{Eq. (40)}$$

Um modelo linguístico relacionando as variáveis x e y pode ser escrito como um conjunto de N regras que relacionam os termos $A_i, i = 1 \dots n$ e $B_j, j = 1 \dots m$, por regras $A_i \rightarrow B_j$. Cada combinação (i, j) representa uma regra k escrita como:

$$\text{Se } x \text{ é } A_i \text{ então } y \text{ é } B_j, k = 1 \dots N \quad \text{Eq. (41)}$$

Modelos Difusos

Mamdani propôs na década de 70 um método de inferência difuso que foi por muitos anos um padrão para a utilização dos conceitos da Lógica Nebulosa em processamento de conhecimento. As regras de produção em um modelo de Mamdani possuem relações nebulosas tanto em seus antecedentes como em seus consequentes.

Por exemplo

SE Temperatura = AumentandoMuito

E Pressão = MuitoAlta

ENTÃO PerigoDeExplosão = Grande

Como a maioria das aplicações de interesse possui sistemas convencionais

de aquisição e atuação baseados em grandezas numéricas, o modelo de Mamdani inclui módulos de interface que transformam as variáveis nebulosas geradas em variáveis numéricas proporcionais, adequadas para os sistemas de atuação existentes. Abaixo tem um modelo esquemático de um diagrama de processamento difuso utilizando o modelo de inferência de Mamdani [3].

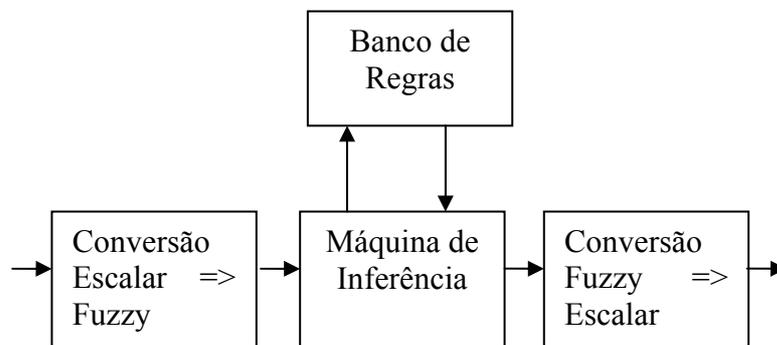


Figura 3: : Diagrama de processamento difuso.

Na década de 80, foi proposta por Takagi, Sugeno e Kang uma estrutura de inferência também baseada na teoria dos conjuntos difusos. Esta estrutura é denominada sistema de inferência de Sugeno, modelo inferência difuso paramétrico ou simplesmente modelo TSK. Os autores demonstraram que estes modelos funcionam como bons aproximadores para sistemas que podem ser completa ou satisfatoriamente representados apenas por meio de suas relações de entrada e saída.

Similarmente ao modelo de Mamdani, os modelos difusos TSK são também baseados na utilização de uma base de regras condicionais de inferência. Porém, no modelo TSK os consequentes das regras, em vez de serem formados por relações nebulosas, compõem-se de equações paramétricas relacionando as entradas e as saídas do processo. Adiante se tem um exemplo de uma regra genérica de um modelo TSK (REZENDE, 2003).

$$\begin{array}{ll}
 \text{SE} & x_1 = A_i \\
 \text{E} & x_2 = A_j \\
 \text{ENTÃO } y & = \psi(x_1, x_2)
 \end{array}$$

6.3.1 Aplicação do Modelo Difuso

Para efeito de comparação com outros paradigmas discutidos nesta tese,

foram criados modelos difusos para verificar a performance da série de resíduos.

O arquivo utilizado para este experimento foi tratado conforme os procedimentos descritos no Anexo 5.

Foram gerados, no total, 3 (três) modelos difusos a fim de se verificar qual o que melhor responderia aos dados. Para tanto, os dias úteis da semana foram divididos em 3 conjuntos, os dias do mês em 3 e 5 conjuntos e os dias úteis do mês em 3 e 5 conjuntos. Os meses do ano não foram *fuzzificados* sendo que cada mês correspondeu foi codificado com um valor entre 1 e 12, iniciando-se com janeiro. Os dias do mês e os dias úteis do mês *fuzzificados* de duas formas cada um e combinados entre si. Assim, ficou-se com os seguintes modelos:

- Modelo 1: dias da semana divididos em 3 conjuntos; dias do mês divididos em 3 conjuntos, dias úteis do mês em 3 conjuntos e meses do ano correspondentes a 12 conjuntos. Neste modelo foram geradas 324 regras (3 x 3 x 3 x 12).
- Modelo 2: dias da semana divididos em 3 conjuntos; dias do mês divididos em 3 conjuntos, dias úteis do mês em 5 conjuntos e meses do ano correspondentes a 12 conjuntos. Neste modelo foram geradas 540 regras (3 x 3 x 5 x 12).
- Modelo 3: dias da semana divididos em 3 conjuntos; dias do mês divididos em 5 conjuntos, dias úteis do mês em 5 conjuntos e meses do ano correspondentes a 12 conjuntos. Neste modelo foram geradas 900 regras (3 x 5 x 5 x 12).

Em todos os modelos foi utilizado uma janela de 5 posições.

Os resultados da aplicação do modelo estão descritos abaixo.

Quadro 66: Performance dos modelos difusos.

Modelo	Até R\$ 50 milhões		Até R\$ 100 milhões		Até R\$ 150 milhões		Até R\$ 200 milhões	
	Qtd.	%	Qtd.	%	Qtd.	%	Qtd.	%
Modelo 01	27	22,5	40	33,3	54	45,0	57	47,5
Modelo 02	14	11,7	26	21,7	39	32,5	49	40,8
Modelo 03	10	8,3	17	14,2	26	21,7	39	32,5

Os valores apresentados não se mostraram satisfatórios apresentando índices de acerto baixos.

Chamou atenção o fato de o primeiro modelo ser o mais genérico de todos e, apesar disto, ter conseguido o melhor resultado. Isto demonstra que os modelos 2 e 3 estão superparametrizados, ou seja, a divisão do universo de discurso em mais conjuntos fez com que se perdesse em generalidade, acarretando índices mais baixos de acertos no conjunto de teste.

A fim de se aumentar as taxas de acerto, um caminho a ser adotado seria o de se aumentar a generalização dos modelos, ou seja, trabalhar com menos regras. Entretanto, esta premissa não garante o êxito do experimento pois um excesso de generalização pode levar a uma baixa representação dos dados, influenciando negativamente nos resultados.

A seguir tem-se o gráfico dos resíduos e das previsões alcançadas no modelo 01, a fim de se avaliar graficamente a performance obtida.

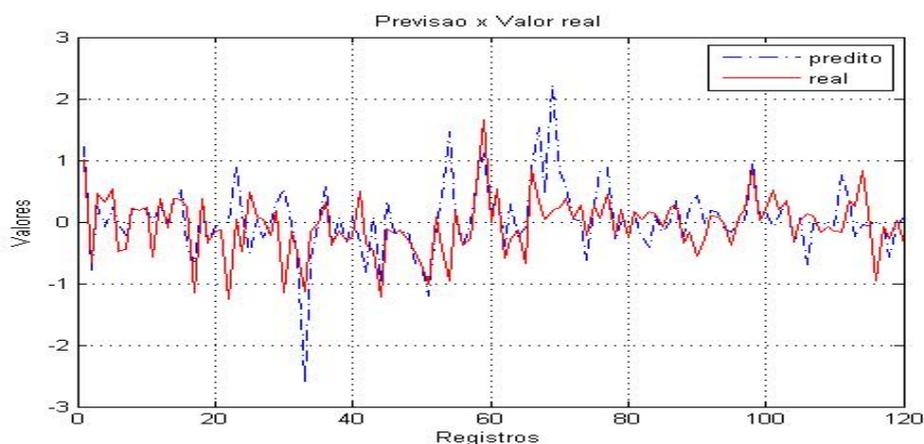


Gráfico 65: Modelo difuso 01. Previsão x resíduo.

Vê-se que apesar de o modelo haver captado de forma geral o comportamento da curva (linha tracejada), ainda assim muito dos dados não foram devidamente representados, notadamente com relação aos picos existentes na série.

6.4 AVALIAÇÃO DOS MODELOS

No decorrer desta pesquisa muitas foram as alterações realizadas, principalmente no tocante a aquisição de dados. Como o modelo projetado objetiva o uso imediato pelo banco, a aquisição de registros passou por constantes acréscimos, sempre visando ter a posição mais atualizada.

Conforme já discutido, a dinâmica existente no setor bancário é uma das

mais rápidas existentes. Enquanto uma indústria leva meses, ou até anos para implementar um novo produto ou modificar algum bem já existente, no segmento bancário ocorre em poucos meses, às vezes até em dias. Isto faz com que toda a série histórica seja alterada de forma drástica. Além disso, tem-se que levar em consideração, também, as fusões e aquisições realizadas pelos bancos, onde os ativos do banco adquirido são incorporados imediatamente na diversas rubricas contábeis da instituição compradora, modificando o perfil destas contas.

Estes problemas não foram tratados no decorrer da pesquisa, ou seja, os modelos já implementados e testados não foram modificados a cada aquisição nova de dados. Além disso, com o constante aprendizado a respeito das suas características, os registros mais antigos, quando apresentavam distorções no seu comportamento, foram sendo retirados com a chegada de novos. Exemplo desta atitude foi a exclusão do primeiro ano da série que, conforme pode ser verificado no estudo promovido no Capítulo 2, apresentava características divergentes dos demais anos, em razão do reflexo oriundo da reestruturação do SPB. Posteriormente, houve a unificação das variáveis STR-Clientes e CIP em função de procedimentos operacionais que o banco passou a adotar, trabalhando, a grosso modo, de forma idêntica com ambas as variáveis.

Em função dessas premissas, a comparação entre os diversos paradigmas utilizados nesta tese pode induzir a pesquisa a tirar conclusões equivocadas sobre qual o melhor caminho a se adotar no cálculo dos valores previstos. Assim, optou-se por separar os dados relativos ao período de 01/04/2006 a 31/03/2007 (248 registros) para realização de testes nos modelos já treinados e a devida comparação dos resultados. Escolheu-se, para cada paradigma utilizado o modelo que melhor o representasse e, em seguida, aplicou-se a série de testes, após o tratamento das componentes determinísticas descritas abaixo.

6.4.1 Extração de Componentes Determinísticas

A primeira etapa consistiu no pré-tratamento dos dados, através das seguintes transformações.

Quadro 67: Transformações no arquivo de dados para aplicação nos modelos ARIMA, difuso e ICA.

Características do arquivo (Período 01/04/2006 a 31/03/2007)
Soma dos saldos das variáveis STR-Clientes e CIP
Divisão dos valores por 1 milhão
Inclusão da média aritmética nos dias faltantes
Normalização dos dados
Exclusão da tendência
Exclusão da sazonalidade semanal
Exclusão de 1 componente senoidal na variável STR-Clientes (783,7 dias)
Exclusão de sazonalidades mensal (dias úteis e calendário)
Potência da série = 0,137203

Os resultados apurados após a extração das componentes determinísticas estão compilados no quadro abaixo, onde se pode ver a quantidade de registros e percentuais em relação ao total da base cujos resíduos foram inferiores ou iguais aos limites de 50, 100, 150 e 200 milhões.

Quadro 68: Estatística de acertos após extração da parte determinística da série.

Até R\$ 50 milhões		Até R\$ 100 milhões		Até R\$ 150 milhões		Até R\$ 200 milhões	
Qtd.	%	Qtd.	%	Qtd.	%	Qtd.	%
23	9,3	46	18,5	61	24,6	79	31,9

Os resultados se mostraram inferiores aqueles já alcançados em experimentos anteriores. Isto pode ser explicado pelo fato de a série utilizada para testes haver sido modificada com os parâmetros definidos na série de treinamento, ou seja, a exclusão da tendência, das sazonalidades semanais e mensais foram calculadas baseada nos índices de treinamento.

Além disso, tem-se a questão do tamanho da série. Por se tratar de um movimento de 1 ano, certamente as características utilizadas na extração das componentes na massa de treinamento ficaram defasadas. Este modelo teria que passar por uma atualização para utilizá-lo nos dados de teste. A título de exemplo, se fossem feitos os cálculos para somente os 50 primeiros registros da base de testes, o resultado seria de 40% de acertos para o patamar de R\$ 200 milhões.

Abaixo tem-se o gráfico da série e da previsão calculada. Para efeito de melhor visualização, optou-se por dividir os dados em 2 áreas, onde na parte superior tem-se metade dos registros. A linha pontilhada representa os dados a serem previstos e a contínua a previsão obtida.

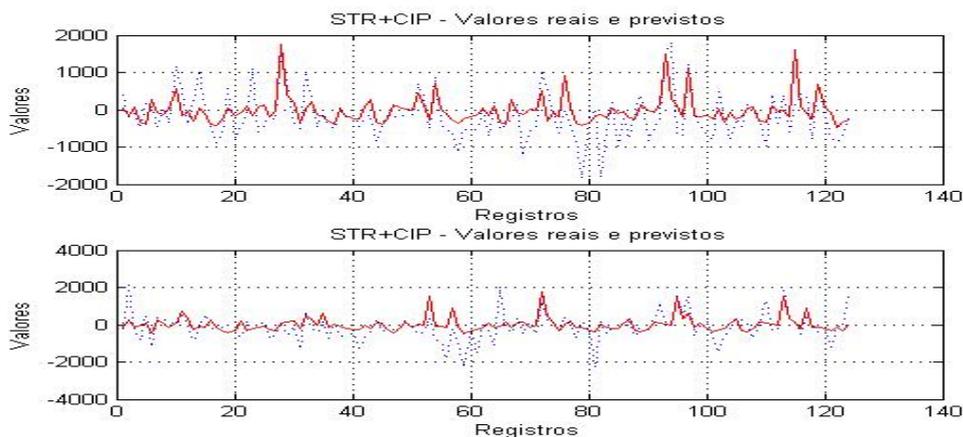


Gráfico 66: Série STR+CIP e sua previsão, após extração de componentes determinísticas.

6.4.2 Modelo Neural

Para os cálculos da previsão utilizando redes neurais, os dados foram aplicados no modelo neural 16 com 1 neurônio na camada escondida e sem representação binária do dia da semana, que foi o que obteve melhores performances durante o desenvolvimento. Cabe comentar que em função do atraso de 11 posições utilizado nos pares entrada-saída da rede, o conjunto de teste foi reduzido para 237 registros.

Os resultados apurados estão no quadro abaixo.

Quadro 69: Estatística de acertos utilizando modelo neural para os dados de teste.

Até R\$ 50 milhões		Até R\$ 100 milhões		Até R\$ 150 milhões		Até R\$ 200 milhões	
Qtd.	%	Qtd.	%	Qtd.	%	Qtd.	%
28	11,8	53	22,4	81	34,2	109	46,0

Vê-se que o ganho percentual obtido foi residual e, mesmo assim, só nos últimos limites (até R\$ 150 e R\$ 200). Entretanto, apesar deste pequeno ganho, a quantidade de registros corretos é inferior àquelas obtidas com a extração puramente das componentes determinísticas. Estes resultados demonstram que a aplicação de redes neurais nesta situação não se mostrou eficiente.

6.4.3 Modelo ARIMA

Considerando o desempenho semelhante ocorrido com os modelos ARIMA(1,1,1) e ARIMA(1,0,2) optou-se pela realização deste teste com ambos.

Quadro 70: Performance dos modelos ARIMA.

Modelo	Até R\$ 50 milhões		Até R\$ 100 milhões		Até R\$ 150 milhões		Até R\$ 200 milhões	
	Qtd.	%	Qtd	%	Qtd.	%	Qtd.	%
ARIMA(1,1,1)	36	14,5	74	29,8	103	41,5	132	53,2
ARIMA(1,0,2)	42	16,9	74	29,8	102	41,1	134	54,0

Ratificando os resultados anteriores, aqui os dois modelos apresentaram muita semelhança, com resultados praticamente idênticos.

Os modelos ARIMA tem demonstrado um melhor desempenho em relação aos demais. A razão disso pode ser atribuída ao fato de o modelo mapear os dados linearmente o que, se por um lado, não se obtém grandes ajustes por outro extrai, na média, melhores soluções que os demais algoritmos não-lineares.

6.4.4 Modelo Difuso

Para o teste do modelo difuso utilizou-se os parâmetros do Modelo 1 que apesar de ser o mais simples entre todos os testados, foi o que apresentou os melhores resultados.

- Modelo 1: dias da semana divididos em 3 conjuntos; dias do mês divididos em 3 conjuntos, dias úteis do mês em 3 conjuntos e meses do ano correspondentes a 12 conjuntos. Total de 324 regras (3 x 3 x 3 x 12).

Os resultados da aplicação do modelo estão descritos abaixo.

Quadro 71: Performance do modelo difuso.

Até R\$ 50 milhões		Até R\$ 100 milhões		Até R\$ 150 milhões		Até R\$ 200 milhões	
Qtd.	%	Qtd	%	Qtd.	%	Qtd.	%
44	18,0	71	29,1	87	35,7	110	45,1

Houve, como já se esperava, uma pequena redução quando comparado com a performance dos testes anteriores. Entretanto, em relação aos índices alcançados após a extração pura e simples das componentes determinísticas, houve um significativo acréscimo.

6.5 CONCLUSÕES

A despeito de todas as tentativas feitas e modelos utilizados, não se conseguiu chegar a resultado ideal, apesar de haver tido uma significativa melhora em relação aos modelos existentes atualmente na empresa, passando de uma performance de acertos da ordem de 35% (Quadro 60) no patamar de R\$ 200 milhões, para percentuais entre 45% e 54%, dependendo do modelo utilizado (vide quadro abaixo). Ou seja, obteve-se um ganho de 28,6% para o pior modelo e 54,3% para o melhor que, traduzindo em dias, representa um acréscimo de 25 e 48 dias de acertos durante um ano (utilizou-se para cálculo o ano de 2007, cujos dias úteis foram 249).

A título de comparação, segue quadro comparativo analítico e gráfico dos diversos modelos aplicados na solução do problema.

Quadro 72: Performance dos diversos modelos testados.

Modelo	Até R\$ 50 milhões		Até R\$ 100 milhões		Até R\$ 150 milhões		Até R\$ 200 milhões	
	Qtd.	%	Qtd.	%	Qtd.	%	Qtd.	%
Modelo do Banco	21	8,4	44	17,7	60	24,1	79	31,7
Extração de comp.	23	9,3	46	18,5	61	24,6	79	31,9
Redes Neurais	28	11,8	53	22,4	81	34,2	109	46,0
ARIMA(1,1,1)	36	14,5	74	29,8	103	41,5	132	53,2
ARIMA(1,0,2)	42	16,9	74	29,8	102	41,1	134	54,0
Lógica Difusa	44	18,0	71	29,1	87	35,7	110	45,1

No gráfico abaixo a curva representativa do modelo ARIMA corresponde à média dos modelos ARIMA(1,1,1) e ARIMA(1,0,2).

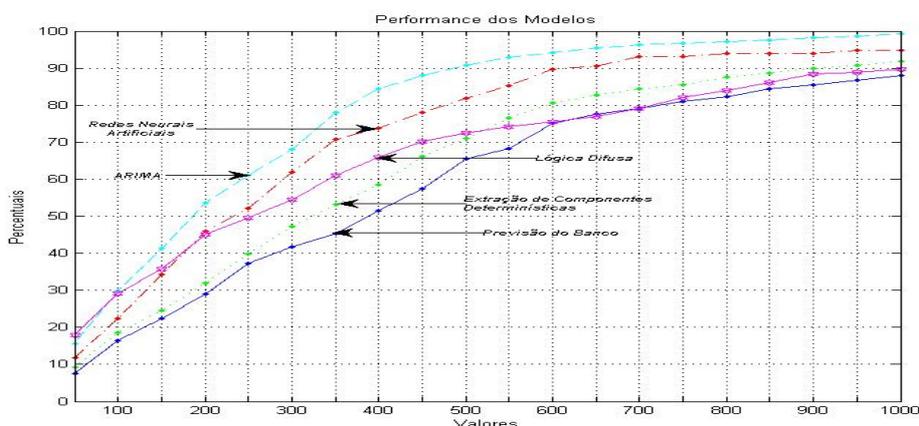


Gráfico 67: Performance de todos os modelos.

7 CONCLUSÃO

7.1 OBJETIVOS

O objetivo principal deste trabalho foi o de mapear o comportamento das principais séries componentes da movimentação da reserva bancária de um grande banco brasileiro, com vistas a criação um modelo computacional para a previsão destas variáveis em um horizonte de 2 dias à frente.

Procurou-se inicialmente entender o comportamento das séries com o objetivo de identificar a existência de algum padrão ou classes de padrões. Estudos de correlação, análises do comportamento dos feriados e da relação entre os dias da semana e do mês foram realizadas visando a detecção de padrões ou repetições que auxiliassem no desenvolvimento do modelo. Em seguida, trabalhou-se com a exclusão das componentes determinísticas utilizando, para tanto, a aplicação de técnicas da estatística clássica, tais como a decomposição das séries e transformações para o domínio da frequência.

Constatou-se que as séries são constituídas basicamente pela componente irregularidade, que além de ser de alta frequência apresenta um elevado índice de dispersão (variância). Em vista desta particularidade estudos sobre aleatoriedade foram realizados, onde se buscou a comprovação deste fato. Apesar de não haver sido constatada, inequivocamente, a aleatoriedade da série, a análise dos resultados mostrou que a irregularidade existente nas séries é muito forte.

Há de se comentar que o uso das redes neurais artificiais para o mapeamento de funções complexas e não-lineares é hoje um dos mais aceitos na comunidade científica. E, dentre os modelos de redes, as MLP são as mais utilizadas, sendo referenciadas na maioria dos trabalhos de previsão de séries temporais quando se trata de modelos não clássicos. Entretanto, como os resultados não alcançaram os índices desejados, outras arquiteturas de redes neurais foram implementadas, visando a obtenção de melhoria na performance.

O uso de redes neurais depende fortemente de dois fatores críticos: o ajuste dos diversos parâmetros que influenciam o desempenho da rede e o tratamento dispensado ao dados que serão processados.

No primeiro caso, a experimentação prática e experiência do projetista são cruciais para um bom desempenho. Não foram minimizados os esforços para se alcançarem os objetivos. Além da execução de milhares de modelos com ajustes das variáveis intrínsecas ao algoritmo foram realizadas exaustivas análises dos resultados, buscando descobrir métodos a serem utilizados para melhoria dos resultados.

Quanto ao segundo fator crítico, este também foi foco da tese já que foram realizadas dezenas de estudos a respeito do tratamento dos dados de entrada. Análises de correlações, formas de normalização, redução do universo das variáveis componentes do fluxo de caixa e uso de variáveis externas para auxílio no mapeamento da série foram alguns dos itens testados e avaliados.

Outros paradigmas também foram implementados, a exemplo de modelos difusos e ARIMA. Os resultados serviram para embasamento da argumentação de tratar-se de séries complexas, cujas pesquisas para a solução do problema não podem se restringir a este trabalho.

Além das técnicas descritas nesta tese, foi estudado também o uso de ICA (*Independent Components Analysis*) no tratamento dos dados. Trata-se de uma técnica cujo propósito é o descobrimento das fontes originais de vários sinais misturados. Como as variáveis STR-Clientes e CIP possuem características similares e, em alguns casos, são operacionalizadas conjuntamente, levantou-se a hipótese de os dados estarem misturados, ou seja, características da STR incluídas na CIP e vice-versa. Dessa forma, com o uso do algoritmo de ICA procurou-se obter os dados representativos de cada variável mais limpos. Após este procedimento, os conjuntos foram aplicados às redes buscando a obtenção de uma melhor performance. Os resultados não foram satisfatórios, assim como também não trouxe melhoria o uso de todas as séries no algoritmo. Em função da baixa performance obtida, os resultados não foram apresentados.

Alguns autores, a exemplo de Hyärinen *et al.* (2001) e Kiviloto e Oja (1998) vêm utilizando o ICA para previsão de séries temporais obtendo bons resultados. Destaca-se, entretanto, que em ambos os casos as séries são referentes ao preço de

commodities que diferem substancialmente das séries financeiras utilizadas nesta tese.

7.2 PRINCIPAIS DIFICULDADES ENCONTRADAS

Apesar de a previsão da movimentação de reservas ser um problema inerente a todo o segmento bancário quase nenhuma literatura sobre o assunto está disponível. Os estudos de fluxo de caixa existentes focam nas empresas comerciais e trabalham basicamente com os valores de entrada e saída de dinheiro sem se aterem aos diversos desdobramentos integrantes deste fluxo.

O movimento do fluxo de caixa das empresas não bancárias têm um comportamento diferente de um banco. A grosso modo, uma empresa não bancária tem como receitas tão-somente os valores relativos a comercialização de mercadorias ou serviços e, como despesas, o pagamento de faturas, impostos, despesas trabalhistas etc. Todos esses itens, de algum modo, possuem algum grau de previsão. Seguindo a classificação apresentada no capítulo 2, poderia-se enquadrar a maioria deles como Fluxo Certo ou Previsível.

Não é o que se observa nos bancos onde o produto comercializado – dinheiro – é muito sensível as conjunturas econômicas, as negociações realizadas, ao tipo de produto ofertado e, até mesmo, as características pessoais do cliente. O sistema bancário brasileiro é extremamente ágil e interligado, facilitando sobremaneira as mudanças de uma instituição para outra. Com isto, o problema da previsão passa, também, pela capacidade de se analisar as características gerais do comportamento da base de clientes da instituição.

O dinamismo que afeta o sistema bancário também é um fator de dificuldade quando se deseja mapear a sua movimentação. Produtos com características distintas são lançados frequentemente sempre que se observa uma tendência ou um segmento consumidor. A facilidade em se criar e operacionalizar um produto bancário faz com que estes produtos sejam inseridos no portfólio dos bancos com uma rapidez difícil de ser encontrada nos outros estratos da sociedade. Com isto, receitas que eram contabilizadas em determinadas contas passam para outra em questão de meses ou, até mesmo, dias. A série histórica que vinha apresentando um determinado comportamento de repente começa a apresentar características diferentes das conhecidas, dificultando a

análise e, conseqüentemente, o seu mapeamento.

A inexistência de outras variáveis que possuam correlação com as variáveis principais, notadamente a STR-Clientes e, posteriormente, STR+CIP, também foi um fator que pode ter contribuído para a pesquisa não ter alcançado melhores resultados.

O uso da métrica fixa de R\$ 200 milhões para avaliação da performance da rede também não se mostrou adequada, necessitando de uma revisão neste critério a fim de melhor representar o erro pois, atualmente, qualquer valor superior a este patamar, independente da distância que se encontra, é considerado erro e este erro tem o mesmo peso quando excede o patamar em, digamos, R\$ 1 milhão ou em R\$ 1 bilhão.

A impossibilidade de se trabalhar com o horizonte de $t + 1$ no cálculo da previsão pode ser considerada também um fator de redução de performance. É sabido que quanto maior o intervalo entre a data e a previsão, piores são os resultados. Entretanto, esta é uma limitação operacional imposta pelo problema e sem solução à curto prazo. Os dados para utilização no modelo só estão disponíveis em horário que inviabiliza sua utilização no modelo para o dia seguinte pois as decisões a serem tomadas baseadas nas projeções dos valores das variáveis já tiveram que ser feitas, restando, portanto, a opção de se trabalhar com projeção para $t + 2$.

Por fim, o tratamento dos *outliers* também não apresentou resultados satisfatórios. Não se detectou regularidades nem padrões nestes dados. Os registros, nestas situações, apesar de parecerem espúrios em função da aparente divergência com os demais, foram conferidos e checados com a fonte e correspondem efetivamente ao movimento.

A ocorrência de valores significativamente superiores à média geral dificulta o seu mapeamento, além de produzir reflexos nos registros seguintes. Apesar deste problema, estes registros não puderam ser excluídos por não serem espúrios, nem se conseguiu tratá-los por não apresentarem nenhum padrão de comportamento detectável. Gráficamente, observou-se que nestes casos a rede tentava mapear o valor, não alcançando o seu pico e, quando o registro seguinte tinha valor muito inferior, a rede não conseguia alcançá-lo, ficando geralmente no meio do caminho. Assim, tanto o primeiro registro como o seguinte apresentavam erro de previsão muito alto. Como a série possui alta frequência e elevada amplitude, esta característica degradou os resultados. A título de exemplificação, abaixo tem-se o gráfico referente a 50 registros

do conjunto de treinamento do modelo neural 16 e verifica-se na comparação entre os valores reais (linha pontilhada) e os mapeados (linha contínua), o problema comentado, notadamente na faixa situada entre os registros de números 15 a 30.

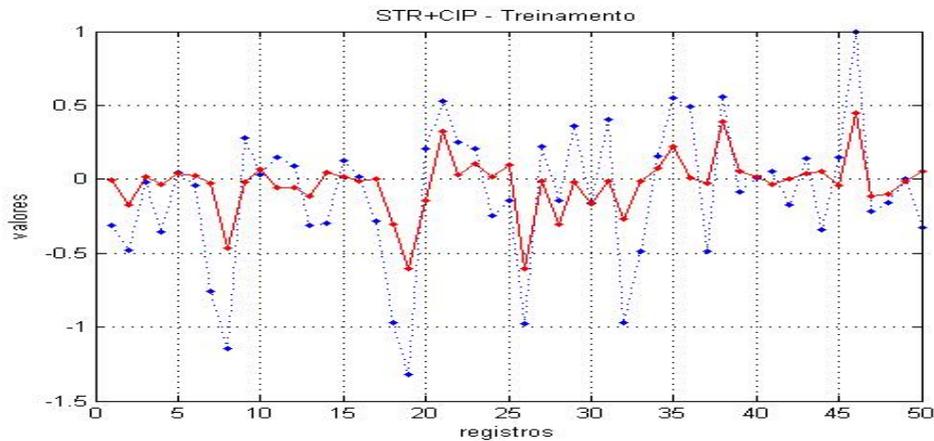


Gráfico 68: Mapeamento dos dados de treinamento.

7.3 TRABALHOS FUTUROS

Em termos de trabalhos futuros, muito ainda há a se fazer pois o problema se mostrou mais complexo do que se supunha inicialmente. Abaixo têm-se as principais linhas a serem abordadas em estudos futuros.

- Tratamento dos dados – levantar e avaliar a aplicação de novas formas de transformação e normalização dos dados de maneira a melhor evidenciar os possíveis padrões existentes.
- Aplicar novos paradigmas aos dados, buscando melhoria nos índices de acerto de previsão.
- Análise multivariável – tentar identificar e obter outras variáveis dentro do contexto do problema que possam ajudar na explicação do fenômeno.
- Métricas de desempenho – conceber uma métrica para avaliar o desempenho dos modelos, que não penalize de forma desigual erros de previsão elevados e erros baixos e que melhor reflita a realidade do problema e a expectativa do usuário.

A construção do modelo neural foi a que consumiu a maior parte do tempo de pesquisa em função do fato de ser considerado, atualmente, o paradigma mais poderoso para se aproximar uma função. Assim, várias redes foram construídas durante

a fase de pesquisa bem como muitos parâmetros de ajuste foram aplicados. Entretanto, o modelo ARIMA, utilizado para comparação, apresentou resultados superiores aos modelos neurais e difusos. Isto leva a proposição de se realizar, no futuro, um aprofundamento no estudo dos modelos lineares.

8 ANEXOS

1. VARIÁVEIS DO FLUXO DE CAIXA

Quadro 73: Relação das variáveis integrantes do fluxo de caixa classificadas.

Variável	Subitens	Tipo de Fluxo
Numerário	Depósito – Instituição Financeira	Fluxo Certo
	Saque – Instituição Financeira	Fluxo Certo
	Custodiante – crédito de tarifa	Fluxo Previsível
Compensação	Resultado noturno – Instituição Financeira	Fluxo Não Previsível
	Resultado noturno – BACEN	Fluxo Não Previsível
	Resultado noturno – BNDES (até 01/11/02)	Fluxo Certo
	Resultado diurno – Instituição Financeira	Fluxo Não Previsível
	Resultado diurno – BACEN	Fluxo Não Previsível
	Resultado diurno – BNDES (até 01/11/02)	Fluxo Certo
	Devolução do pré-depósito	Fluxo Certo
	Pré-depósito – do dia	Fluxo Certo
	Pré-depósito – ajuste da noturna	Fluxo Não Previsível
	Pré-depósito – ajuste da diurna	Fluxo Não Previsível
Arrecadação via STN / STR	Tributos Federais – DARF	Fluxo Previsível
	INSS – Arrecadação	Fluxo Previsível
	Tributos Municipais	Fluxo Previsível
	PIS / PASEP	Fluxo Não Previsível
	Tributos Estaduais (ICMS)	Fluxo Previsível
	FGTS – Entrada	Fluxo Previsível
	FGTS – Saída	Fluxo Previsível
	FGTS – Repasse Arrecadação	Fluxo Certo
	FGTS – Lei 8.727 CEF	Fluxo Certo
	FCVS / SFH	Fluxo Não Previsível
	Contribuição Sindical	Fluxo Não Previsível
Ordens Bancárias	Outros Bancos	Fluxo Previsível
	Instituição Financeira – Crédito	Fluxo Previsível
	Instituição Financeira – Débito	Fluxo Previsível
	Folha de Pagamento	Fluxo Previsível
	On-line pagamento (OP 15)	Fluxo Previsível
	On-line pagamento (OP 16)	Fluxo Não Previsível
	On-line pagamento (OP 29)	Fluxo Não Previsível
	ITR / ISO / SIMPLES / FUNCHEQUE (OP 51)	Fluxo Certo
	DIV-TDA / SIBR / ELET-DIV SID (OP 51)	Fluxo Certo
	PROEX / AGF / EGF / RECOOP (OP 51)	Fluxo Certo
	Lei 8.727 (OP 51)	Fluxo Certo
	FPEM / ICMS - DESON / ANP (OP 51)	Fluxo Certo
	On-line pagamento (OP 52)	Fluxo Não Previsível

Câmbio	PCAM380 via STR – Instituição Financeira – Entrada	Fluxo Certo
	PCAM380 via STR – Instituição Financeira – Saída	Fluxo Certo
	PCAM380 via STR – BNDES	Fluxo Certo
	PCAM383 BM&F CÂMBIO: operações NET - Interbancário	Fluxo Certo
	PCAM 383 BM&F CÂMBIO: Aporte garantias financeiras	Fluxo Certo
	PCAM 383 BM&F CÂMBIO: Devolução, rendimento ou resgate de garantias	Fluxo Certo
	PCAM 300 via LDL - CLEARING BANK	Fluxo Certo
Compulsório	Operações de câmbio	Fluxo Certo
	Encaixe da poupança	Fluxo Certo
	Poupex	Fluxo Certo
	Recursos à vista	Fluxo Certo
	Exigibilidade adicional	Fluxo Certo
SELIC	IF Compromissada: Doadas / Tomadas – BACEN	Fluxo Certo
	IF Compromissada: Tomadas – Mercado	Fluxo Certo
	IF Compromissada: Doadas – Mercado	Fluxo Certo
	IF Definitivas: Compras/Vendas – BACEN	Fluxo Certo
	IF Definitivas: Vendas – Mercado	Fluxo Certo
	IF Definitivas: Compras – Mercado	Fluxo Certo
	Cliente 2	Fluxo Certo
	Cliente específico	Fluxo Certo
	IF -BI Definitiva – Mercado	Fluxo Certo
	PREVI Definitiva – Mercado	Fluxo Certo
	Fundo do Município Rio Definitiva – Mercado	Fluxo Certo
	IF DTVM Definitiva – Mercado	Fluxo Certo
	Operações Instituição Financeira – Resgate	Fluxo Certo
	Operações Instituição Financeira – Amortização	Fluxo Certo
	Operações Instituição Financeira – Juros	Fluxo Certo
	Operações IF -BI – Resgate	Fluxo Certo
	Operações PREVI – Resgate	Fluxo Certo
	Operações Fundo Município Rio – Resgate	Fluxo Certo
	Operações IF DTVM, Seguridade e Extra-mercado - Amortização	Fluxo Certo
	Repasse de Impostos	Fluxo Certo
	Mercado - Revendas: Instituição Financeira	Fluxo Certo
	Mercado - Recompras: Instituição Financeira	Fluxo Certo
	Mercado - Revendas/Recompras: IF BI	Fluxo Certo
	Mercado - Revendas/Recompras: IF-PREVI	Fluxo Certo
	Mercado - Revendas/Recompras: Fundo Município Rio	Fluxo Certo
	Mercado - Revendas/Recompras: DTVM	Fluxo Certo
	Limite financeiro para não liquidante	Fluxo Certo
Custos de utilização SELIC	Fluxo Certo	
CETIP	Operações NET/Sistema CET/SND (entrada)	Fluxo Não Previsível
	Operações NET/Sistema CET/SND (saída)	Fluxo Não Previsível
	Liquidação valor bruto / Sistema CET / Sistema Nacional de Debêntures.	Fluxo Não Previsível
	Recebimento de eventos (resgate de papéis)	Fluxo Não Previsível
	Aporte de Garantias Financeiras	Fluxo Não Previsível
Devolução, Rendimentos ou Resgate de Garantias	Fluxo Não Previsível	

Tesouro Nacional - Outras	INSS - Pagamento de Benefícios	Fluxo certo
	INSS - Pagamento de Tarifas	Fluxo Não Previsível
	INSS - Pagamentos Outros	Fluxo Não Previsível
	INSS - Transferências Diversas	Fluxo certo
	Secretaria da Receita Federal – Restituição de IR	Fluxo certo
	Transferência. da conta única para reserva – Diversos	Fluxo Previsível
	Transferência de reserva para conta única – Diversos	Fluxo Previsível
	Pagamentos e/ou devoluções de recursos do TN	Fluxo Não Previsível
	Pagamentos de funcionalismo público	Fluxo certo
TECBAN	Movimento do dia	Fluxo Previsível
	Aporte/Devolução de garantias financeiras	Fluxo Não Previsível
	Devolução, rendimentos ou resgate de garantias	Fluxo Não Previsível
Lançamentos BACEN – SLB	Lançamentos Diversos - Regionais/FOPAG (Crédito)	Fluxo Não Previsível
	Lançamentos Diversos - Dívida Rio/Centrus/FGC (Débito)	Fluxo Não Previsível
	Lançamentos Diversos - Lei 8727 (Débito)	Fluxo certo
	Lançamentos Diversos - BID 541 (Débito)	Fluxo Não Previsível
	BACEN - Compe Devedora	Fluxo Não Previsível
	BACEN - Compe Credora	Fluxo Não Previsível
Redesconto	Redesconto intradia	Fluxo Certo
	Pagamento de redesconto	Fluxo Certo
BM&F Derivativos	Operações NET/Yeld-IFDsvm/Ajuste	Fluxo Certo
	Pagamento de Eventos	Fluxo Certo
	Recebimento de Eventos	Fluxo Certo
	Aporte de Garantias Financeiras	Fluxo Certo
	Devolução, rendimentos ou resgate de garantias	Fluxo Certo
CBLC	Operações NET/IMC/Ações	Fluxo Certo
	Pagamento de Eventos/AEB/(ID)Eventos	Fluxo Certo
	Recebimento de Eventos/IMC/Eventos	Fluxo Certo
	Tarifa de Custódia	Fluxo Certo
	Aporte de Garantias Financeiras	Fluxo Certo
	Devolução, rendimentos ou resgate de garantias	Fluxo Certo
BM&F Ativos	Op. NET/Sist. CET/SND Sistema Nacional de Debêntures	Fluxo Certo
	Aporte de Garantias Financeiras	Fluxo Certo
	Devolução, rendimentos ou resgate de garantias	Fluxo Certo

Sistema Transferência Reservas - Instituições Financeiras	Pagamento de Benefícios	Fluxo Previsível
	Operação de Cessão de Créditos	Fluxo Não Previsível
	Cartões de Crédito	Fluxo Previsível
	Cartões de Débito	Fluxo Previsível
	Pagamento de funcionalismo público	Fluxo Previsível
	Pagamento de fornecedor público	Fluxo Previsível
	FGC - Fundo Garantidor de Crédito	Fluxo certo
	Repasse FNDE - Fundo Nacional de Desenvolvimento Educacional	Fluxo Não Previsível
	Acerto de Compensação	Fluxo Não Previsível
	Taxa serviço CCF	Fluxo Não Previsível
	Repasses da Lei 8727	Fluxo certo
	Repasses de valores do BNDES	Fluxo Não Previsível
	Repasses demais bancos de desenvolvimento	Fluxo Não Previsível
	Liquidação de compromissos junto ao BNDES	Fluxo Não Previsível
	Liquidação compromissos junto demais bancos desenv.	Fluxo Não Previsível
	Transferência de recursos ao governo federal	Fluxo Não Previsível
	Repasses do governo federal	Fluxo Não Previsível
	Resgate de cotas do FDS - Fundo Desenvolvimento Social	Fluxo Não Previsível
	Movimentação de Valores - Fundo de Aval	Fluxo Não Previsível
	Pagamento a corretoras	Fluxo Não Previsível
Outros	Fluxo Não Previsível	
Sistema Transferência Reservas - Clientes	Transferência a débito outra inst. Financeira-entrada	Fluxo Não Previsível
	Transferência a crédito outra inst. Financeira-saída	Fluxo Não Previsível
	Transferência conta cliente para IF-entrada	Fluxo Não Previsível
	Transferência conta cliente para IF-saída	Fluxo Não Previsível
	Transferência IF para conta cliente-entrada	Fluxo Não Previsível
	Transferência IF para conta cliente-saída	Fluxo Não Previsível
	Transferência entre contas de diferentes titulares-entrada	Fluxo Não Previsível
	Transferência entre contas de diferentes titulares-saída	Fluxo Não Previsível
	Transferência entre contas de mesma titularidade-entrada	Fluxo Não Previsível
	Transferência entre contas de mesma titularidade-saída	Fluxo Não Previsível
	Transferência para devolução indevida-entrada	Fluxo Não Previsível
	Transferência para devolução indevida-saída	Fluxo Não Previsível
	Transferência para depósito judicial-entrada	Fluxo Não Previsível
	Transferência para depósito judicial-saída	Fluxo Não Previsível
	Transferência reservas bancárias para liquidação de documento com código barras-entrada	Fluxo Não Previsível
	Transf. RB para liq. documento com código barras-saída	Fluxo Não Previsível
CIP	Depósito operacional / Transferência entre I.F. / Pagamento de benefícios / Movimento operacional	Fluxo Não Previsível
	Resultado Financeiro / Transferência do resultado líquido	Fluxo Não Previsível
	Ajuste dos pagamentos compensados	Fluxo Não Previsível

2. VARIÁVEIS DO FLUXO DE CAIXA COM REGRAS DE AGREGAÇÃO

Quadro 74: Aplicação das regras às variáveis.

Variável	Subitens	Tipo de Fluxo	Regra Aplicada
Numerário	Depósito – IF	FC	Regra 02 - Agregação - FC
	Saque – IF	FC	
Compensação ⁽¹⁾	Resultado noturno – BNDES	FC	Regra 02 - Agregação - FC
	Resultado diurno – BNDES	FC	
	Devolução do pré-depósito	FC	
	Pré-depósito – do dia	FC	
	Pré-depósito – ajuste da noturna	FNP	
	Pré-depósito – ajuste da diurna	FNP	
Arrecadação via STN / STR	FGTS – Repasse Arrecadação	FC	Regra 01 - Sem movimentação
	FGTS – Lei 8.727 CEF	FC	
	FCVS / SFH	FNP	
	Tributos Federais – DARF	FP	Regra 03 - Agregação - FP
	INSS – Arrecadação	FP	
	Tributos Municipais	FP	
	Tributos Estaduais (ICMS)	FP	
	FGTS – Entrada	FP	
	FGTS – Saída	FP	
Ordens Bancárias ⁽²⁾	DIV-TDA / SIBR / ELET-DIV SID	FC	Regra 01 - Sem movimentação
	PROEX / AGF / EGF / RECOOP	FC	
	ITR / ISO / SIMPLES / FUNCHEQUE	FC	Regra 02 - Agregação - FC
	Lei 8.727	FC	
	FPEM / ICMS - DESON / ANP	FC	
	Outros Bancos	FP	Regra 03 - Agregação - FP
	IF – Crédito	FP	
	IF – Débito	FP	
	Folha de Pagamento	FP	
	On-line pagamento (OP 15)	FP	
	On-line pagamento (OP 16)	FNP	
	On-line pagamento (OP 29)	FNP	
	On-line pagamento (OP 52)	FNP	
	Variável	Subitens	Tipo de Fluxo
Câmbio	IF – Entrada	FC	Regra 02 - Agregação - FC
	IF – Saída	FC	
	BNDES	FC	
	Interbancário	FC	
	Aporte garantias financeiras	FC	
	Devolução, rendimento ou resgate	FC	
	CLEARING BANK	FC	

Compulsório	Operações de câmbio	FC	Regra 02 - Agregação - FC
	Encaixe da poupança	FC	
	Poupex	FC	
	Recursos à vista	FC	
	Exigibilidade adicional	FC	
SELIC	IF Compr.: Doadas / Tomadas – BC	FC	Regra 02 - Agregação - FC
	IF Compr.: Tomadas – Mercado	FC	
	IF Compr.: Doadas – Mercado	FC	
	IF Definitivas: C&V – BACEN	FC	
	IF Definitivas: Vendas – Mercado	FC	
	IF Definitivas: Compras – Mercado	FC	
	Cliente 2	FC	
	Cliente específico	FC	
	IF-BI Definitiva – Mercado	FC	
	PREVI Definitiva – Mercado	FC	
	Fundo do Mun. Rio – Mercado	FC	
	IF DTVM Definitiva – Mercado	FC	
	Operações IF – Resgate	FC	
	Operações IF – Amortização	FC	
	Operações IF – Juros	FC	
	Operações IF-BI – Resgate	FC	
	Operações PREVI – Resgate	FC	
	Op. Fundo Mun. Rio – Resgate	FC	
	IF DTVM, Seguridade, Extra-mercado	FC	
	Repasse de Impostos	FC	
	Mercado - Revendas: IF	FC	
	Mercado - Recompras: IF	FC	
	Mercado - IF BI	FC	
	Mercado - IF-PREVI	FC	
Mercado - Fundo Município Rio	FC		
Mercado - DTVM	FC		
Limite financeiro para não liquidante	FC		
Custos de utilização SELIC	FC		
CETIP	Aporte de Garantias Financeiras	FNP	Regra 01 - Sem movimentação
Tesouro Nacional – Outras	INSS - Pagamento de Benefícios	FC	Regra 02 - Agregação - FC
	INSS - Transferências Diversas	FC	
	SRF – Restituição de IR	FC	
	Pagamentos de funcis público	FC	Regra 03 - Agregação - FP
	Transf. da conta única para reserva	FP	
	Transf. de reserva para conta única	FP	
TECBAN	Aporte/Devolução de garantias	FNP	Regra 01 - Sem movimentação
	Garantias - Devol., Rend., Resgate	FNP	
Lançamentos BACEN / SLB	Lei 8727 (Débito)	FC	Regra 01 - Sem movimentação
	BID 541 (Débito)	FNP	
	Dívida Rio/Centrus/FGC (Débito)	FNP	Regra 03 - Agregação - FP

Variável	Subitens	Tipo de Fluxo	Regra Aplicada
Redesconto	Redesconto intradia	FC	Regra 01 - Sem movimentação
	Pagamento de redesconto	FC	
BM&F Derivativos	Operações NET/Yeld-IFDsvm/Ajuste	FC	Regra 02 - Agregação - FC
	Pagamento de Eventos	FC	
	Recebimento de Eventos	FC	
	Aporte de Garantias Financeiras	FC	
	Garantias - Devol., Rend., Resgate	FC	
CBLC	Operações NET/IMC/Ações	FC	Regra 02 - Agregação - FC
	Pagamentos Eventos/AEB/(ID)	FC	
	Recebimento Eventos/IMC	FC	
	Tarifa de Custódia	FC	
	Aporte de Garantias Financeiras	FC	
	Garantias - Devol., Rend., Resgate	FC	
BM&F Ativos	Op. NET/Sist. CET/SND	FC	Regra 02 - Agregação - FC
	Aporte de Garantias Financeiras	FC	
	Garantias - Devol., Rend., Resgate	FC	
Sistema Transferência Reservas - Instituições Financeiras	Resgate de cotas do FDS	FNP	Regra 01 - Sem movimentação
	Repasse FNDE	FNP	
	FGC - Fundo Garantidor de Crédito	FC	Regra 02 - Agregação - FC
	Repases da Lei 8727	FC	
	Pagamento de Benefícios	FP	Regra 03 - Agregação - FP
	Cartões de Crédito	FP	
	Cartões de Débito	FP	
	Pagamento de funcionalismo público	FP	
Pagamento de fornecedor público	FP		

- (1) Compensação (regra 02) – Os campos Pré-Depósito – Ajuste da Noturna e Pré-Depósito – Ajuste da Diurna individualmente são considerados como de Fluxo Não Previsível. Entretanto, quando agregados com o campo Pré-Depósito – Do Dia possuem características de Fluxo Certo, ou seja, não se tem a previsão individual destas colunas, mas o somatório dos três campos fornecem um valor exato da movimentação.
- (2) Ordens Bancárias (regra 03) – Os subitens agregados nesta regra apesar de alguns serem de característica Não Previsível, quando agrupados são previsíveis, ou seja, o banco consegue prever o valor deste subitem, mas não das variáveis que o compõem.

3. VARIÁVEIS APÓS REDUÇÃO DA DIMENSIONALIDADE – 1ª ETAPA

Quadro 75: Base de dados após 1ª Etapa de redução da dimensionalidade.

Variável	Subítem	Tipo de Fluxo
Numerário	Agregação – FC	Fluxo Certo
	Custodiante – crédito de tarifa	Fluxo Previsível
Compensação	Resultado noturno – Instituição Financeira	Fluxo Não Previsível
	Resultado noturno – BACEN	Fluxo Não Previsível
	Resultado diurno – Instituição Financeira	Fluxo Não Previsível
	Resultado diurno – BACEN	Fluxo Não Previsível
	Agregação – FC	Fluxo Certo
Arrecadação via STN / STR	PIS / PASEP	Fluxo Não Previsível
	Contribuição Sindical	Fluxo Não Previsível
	Agregação – FP	Fluxo Previsível
Ordens Bancárias	Agregação – FP	Fluxo Previsível
Câmbio	Agregação – FC	Fluxo Certo
Compulsório	Agregação – FC	Fluxo Certo
SELIC	Agregação – FC	Fluxo Certo
CETIP	Operações NET/Sistema CET/SND (entrada)	Fluxo Não Previsível
	Operações NET/Sistema CET/SND (saida)	Fluxo Não Previsível
	Liquidação valor bruto / Sist. CET / SND	Fluxo Não Previsível
	Recebimento de eventos (resgate de papéis)	Fluxo Não Previsível
	Devolução, rendimentos ou resgate de garantias	Fluxo Não Previsível
Tesouro Nacional - Outras	INSS - Pagamento de Tarifas	Fluxo Não Previsível
	INSS - Pagamentos Outros	Fluxo Não Previsível
	Pagamento e/ou devoluções de recursos do TN	Fluxo Não Previsível
	Agregação – FC	Fluxo Certo
	Agregação – FP	Fluxo Previsível
TECBAN	Agregação – FP	Fluxo Previsível
Lançamentos BACEN – SLB	Lançamentos Diversos - Regionais/FOPAG (Crédito)	Fluxo Não Previsível
	Lançamentos Diversos - Dívida Rio/Centrus/FGC (Débito)	Fluxo Não Previsível
	Agregação – FP	Fluxo Previsível
BM&F Derivativos	Agregação – FC	Fluxo Certo
CBLC	Agregação – FC	Fluxo Certo
BM&F Ativos	Agregação – FC	Fluxo Certo
STR - IF	Operação de Cessão de Créditos	Fluxo Não Previsível
	Acerto de Compensação	Fluxo Não Previsível
	Taxa serviço CCF	Fluxo Não Previsível
	Repasse de valores do BNDES	Fluxo Não Previsível
	Repasse demais bancos de desenvolvimento	Fluxo Não Previsível
	Liquidação de compromissos junto ao BNDES	Fluxo Não Previsível
	Liquidação compromissos junto demais bancos	Fluxo Não Previsível
	Transferência de recursos ao governo federal	Fluxo Não Previsível
	Repasse do governo federal	Fluxo Não Previsível
	Movimentação de Valores - Fundo de Aval	Fluxo Não Previsível
	Pagamento a corretoras	Fluxo Não Previsível
	Outros	Fluxo Não Previsível
	Agregação – FC	Fluxo Certo
	Agregação – FP	Fluxo Previsível

Sstema Transferência Reservas - Clientes	Transferência a débito outra IF – entrada	Fluxo Não Previsível
	Transferência a crédito outra IF – saída	Fluxo Não Previsível
	Transferência conta cliente para IF – entrada	Fluxo Não Previsível
	Transferência conta cliente para IF – saída	Fluxo Não Previsível
	Transferência IF para conta cliente – entrada	Fluxo Não Previsível
	Transferência IF para conta cliente – saída	Fluxo Não Previsível
	Transferência entre contas de diferentes titulares – entrada	Fluxo Não Previsível
	Transferência entre contas de diferentes titulares – saída	Fluxo Não Previsível
	Transferência entre contas de mesma titularidade – entrada	Fluxo Não Previsível
	Transferência entre contas de mesma titularidade – saída	Fluxo Não Previsível
	Transferência para devolução indevida – entrada	Fluxo Não Previsível
	Transferência para devolução indevida – saída	Fluxo Não Previsível
	Transferência para depósito judicial – entrada	Fluxo Não Previsível
	Transferência para depósito judicial – saída	Fluxo Não Previsível
	Transferência RB para liq. doc. c/ código barras – entrada	Fluxo Não Previsível
	Transferência RB para liq. doc. c/ código barras – saída	Fluxo Não Previsível
CIP	Dep. Oper. / Transf. entre I.F. / Pagto. Benef. / Mov. Oper.	Fluxo Não Previsível
	Resultado Financeiro/Transferência do resultado líquido	Fluxo Não Previsível
	Ajuste das Pags Compensadas	Fluxo Não Previsível

4. VARIÁVEIS APÓS REDUÇÃO DA DIMENSIONALIDADE – 2ª ETAPA

Quadro 76: Variáveis de FC e FP após a 2ª etapa de redução de dimensionalidade.

Variável	Subitens	Tipo de Fluxo
Numerário	Agregação – FP	Fluxo Previsível
Arrecadação via STN / STR	Agregação - FP	Fluxo Previsível
Ordens Bancárias	Agregação – FP	Fluxo Previsível
Câmbio	Agregação – FC	Fluxo Certo
Compulsório	Agregação – FC	Fluxo Certo
SELIC	Agregação – FC	Fluxo Certo
Tesouro Nacional - Outras	Agregação - FP	Fluxo Previsível
TECBAN	Agregação – FP	Fluxo Previsível
BM&F – Derivativos	Agregação – FC	Fluxo Certo
CBLC	Agregação – FC	Fluxo Certo
BM&F – Ativos	Agregação – FC	Fluxo Certo

Quadro 77: Variáveis de FNP após a 2ª. etapa de redução de dimensionalidade.

Variável	Subitens	Tipo de Fluxo
Compensação	Resultado noturno – Instituição Financeira	Fluxo Não Previsível
	Resultado noturno – BACEN	Fluxo Não Previsível
	Resultado diurno – Instituição Financeira	Fluxo Não Previsível
	Resultado diurno – BACEN	Fluxo Não Previsível
	Agregação – FC	Fluxo Certo
CETIP	Operações NET/Sistema CET/SND	Fluxo Não Previsível
	Liquidação valor bruto / Sist. CET / SND	Fluxo Não Previsível
	Recebimento de eventos (resgate de papéis)	Fluxo Não Previsível
	Devolução, rendimentos ou resgate de garantias	Fluxo Não Previsível
Lançamentos BACEN – SLB	Lançamentos Diversos - Regionais/FOPAG (Crédito)	Fluxo Não Previsível
	BACEN - Compe	Fluxo Não Previsível
	Agregação – FP	Fluxo Previsível
STR - IF	Operação de Cessão de Créditos	Fluxo Não Previsível
	Acerto de Compensação	Fluxo Não Previsível
	Taxa serviço CCF	Fluxo Não Previsível
	Repases de valores do BNDES	Fluxo Não Previsível
	Repases demais bancos de desenvolvimento	Fluxo Não Previsível
	Liquidação de compromissos junto ao BNDES	Fluxo Não Previsível
	Liquidação compromissos junto demais bancos	Fluxo Não Previsível
	Transferência de recursos ao governo federal	Fluxo Não Previsível
	Repases do governo federal	Fluxo Não Previsível
	Movimentação de Valores - Fundo de Aval	Fluxo Não Previsível
	Pagamento a corretoras	Fluxo Não Previsível
	Outros	Fluxo Não Previsível
	Agregação – FC	Fluxo Certo
	Agregação – FP	Fluxo Previsível

Variável	Subitens	Tipo de Fluxo
Sstema Transferência Reservas - Clientes	Transferência a débito outra IF	Fluxo Não Previsível
	Transferência conta cliente para IF	Fluxo Não Previsível
	Transferência conta cliente para IF	Fluxo Não Previsível
	Transferência entre contas de diferentes titulares	Fluxo Não Previsível
	Transferência entre contas de mesma titularidade	Fluxo Não Previsível
	Transferência para devolução indevida	Fluxo Não Previsível
	Transferência para depósito judicial	Fluxo Não Previsível
	Transferência RB para liq. documento com código barras	Fluxo Não Previsível
CIP	Dep. Oper. / Transf. entre I.F. / Pagto. Benef. / Mov. Oper.	Fluxo Não Previsível
	Resultado Financeiro/Transferência do resultado líquido	Fluxo Não Previsível
	Ajuste das Pags Compensadas	Fluxo Não Previsível

5. CARACTERÍSTICAS DOS ARQUIVOS DE DADOS E TOPOLOGIAS DE REDES

Quadro 78: Arquivo de dados para a rede 01.

Características do arquivo (modelo 01)	
Divisão dos valores por 1 milhão	
Redução dos valores para o intervalo [-1000, 2000]	
Normalização dos dados	
Exclusão da tendência	
Exclusão de 1 componente senoidal (período de 755,9 dias)	
Exclusão das semanas incompletas e com início e fim em meses diferentes	
Potência da série = 0,219079	

Quadro 79: Topologia da rede 01.

Topologia da rede (modelo 01)	
Entrada	7 registros referentes aos dias $t - 4$, $t - 6$ e representação binária do dia da semana (1-de-n)
Função de propagação da camada escondida	Tangente hiperbólica
Quantidade de neurônios na entrada	1, 2, 5 e 10

Quadro 80: Arquivo de dados para as redes 02, 03 e 04.

Características do arquivo (modelos 02, 03 e 04)	
Divisão dos valores por 1 milhão	
Redução dos valores para o intervalo [-1000, 2000]	
Normalização dos dados	
Exclusão da tendência	
Exclusão de 1 componente senoidal (período de 755,9 dias)	
Exclusão das semanas incompletas e com início e fim em meses diferentes	
Exclusão da sazonalidade semanal	
Potência da série = 0,212572	

Quadro 81: Topologias das redes 02 e 03.

Topologia das redes (modelos 02 e 03)	
Entrada	7 registros referentes aos dias $t - 4$, $t - 6$ e representação binária do dia da semana (1-de-n)
Função de propagação da camada escondida	Tangente hiperbólica
Quantidade de neurônios na entrada	1, 2, 5 e 10
Observações	Escolha dos pares entrada-saída de forma aleatória no modelo 03

Quadro 82: Topologia da rede 04.

Topologia da rede (modelo 04)	
Entrada	2 registros referentes aos dias $t - 4$ e $t - 6$
Função de propagação da camada escondida	Linear
Quantidade de neurônios na entrada	1
Observações	Escolha dos pares entrada-saída de forma aleatória

Quadro 83: Arquivo de dados para as redes 05 e 06.

Características do arquivo (modelos 05 e 06)
Divisão dos valores por 1 milhão
Redução dos valores para o intervalo $[-1000, 2000]$
Inclusão da média aritmética nos dias faltantes
Normalização dos dados
Exclusão da tendência
Exclusão da sazonalidade semanal
Exclusão de 2 componentes senoidais (períodos de 783,7 e 21,8 dias)
Potência da série = 0,187556

Quadro 84: Topologia da rede 05.

Topologia da rede (modelo 05)	
Entrada 01	2 registros referentes aos dias $t - 4$ e $t - 6$
Entrada 02	7 registros referentes aos dias $t - 4$, $t - 6$ e representação binária do dia da semana (1-de-n)
Função de propagação da camada escondida	Tangente hiperbólica
Quantidade de neurônios na entrada	1 e 5
Observações	Escolha dos pares entrada-saída de forma aleatória

Quadro 85: Topologia da rede 06.

Topologia da rede (modelo 06)	
Entrada 01	2 registros referentes aos dias $t - 4$ e $t - 22$
Entrada 02	7 registros referentes aos dias $t - 4$, $t - 22$ e representação binária do dia da semana (1-de-n)
Função de propagação da camada escondida	Tangente hiperbólica
Quantidade de neurônios na entrada	1 e 5
Observações	Escolha dos pares entrada-saída de forma sequencial

Quadro 86: Arquivo de dados para a rede 07.

Características do arquivo (modelo 07)	
Divisão dos valores por 1 milhão	
Redução dos valores para o intervalo [-1000, 2000]	
Inclusão da média aritmética nos dias faltantes	
Normalização dos dados	
Exclusão da tendência	
Exclusão da sazonalidade semanal	
Exclusão de 6 comp. senoidais (períodos de 783,7, 21,8, 4,4, 3,6, 5,4 e 10,9 dias)	
Potência da série = 0,160724	

Quadro 87: Arquivo de dados para a rede 08.

Características do arquivo (modelo 08)	
Divisão dos valores por 1 milhão	
Redução dos valores para o intervalo [-1000, 2000]	
Inclusão da média aritmética nos dias faltantes	
Normalização dos dados	
Exclusão da tendência	
Exclusão da sazonalidade semanal	
Exclusão de 6 comp. senoidais (períodos de 783,7, 21,8, 4,4, 3,6, 5,4 e 10,9 dias)	
Exclusão da média dos dias úteis 11, 12, 13, 15 e 23	
Potência da série = 0,146864	

Quadro 88: Topologia da redes 07 e 08.

Topologia das redes (modelos 07 e 08)	
Entrada 01	1 registro referente ao dia t - 4
Entrada 02	6 registros referentes ao dia t - 4 e representação binária do dia da semana (1-de-n)
Função de propagação da camada escondida	Tangente hiperbólica
Quantidade de neurônios na entrada	1 e 5
Observações	Escolha dos pares entrada-saída de forma aleatória

Quadro 89: Arquivo de dados para a rede 09.

Características do arquivo (modelo 09)	
Divisão dos valores por 1 milhão	
Redução dos valores para o intervalo [-1000, 2000]	
Inclusão da média aritmética nos dias faltantes	
Normalização dos dados	
Exclusão da tendência	
Exclusão da sazonalidade semanal	
Excl.de 6 comp. senoidais (períodos: 783,7, 21,8, 4,4, 3,6, 5,4 e 10,9 dias)	
Exclusão da sazonalidade mensal (dias úteis)	
Potência da série = 0,144764	

Quadro 90: Topologia da rede 09.

Topologia da rede (modelo 09)	
Entrada 01	1 registro referente ao dia $t - 2$
Entrada 02	6 registros referentes ao dia $t - 2$ e representação binária do dia da semana (1-de-n)
Entrada 03	1 registro referente ao dia $t - 9$
Entrada 04	6 registros referentes ao dia $t - 9$ e representação binária do dia da semana (1-de-n)
Entrada 05	2 registros referentes aos dias $t - 2$ e $t - 9$
Entrada 06	7 registros referentes aos dias $t - 2$ e $t - 9$ e representação binária do dia da semana (1-de-n)
Função de propagação da camada escondida	Tangente hiperbólica
Quantidade de neurônios na entrada	1 e 5
Observações	Escolha dos pares entrada-saída de forma aleatória

Quadro 91: Arquivo de dados para as redes 10, 11, 12 e 13.

Características do arquivo (modelos 10, 11, 12 e 13)	
Divisão dos valores por 1 milhão	
Redução dos valores para o intervalo apropriado para cada variável	
Inclusão da média aritmética nos dias faltantes	
Normalização dos dados	
Exclusão da tendência para cada variável	
Exclusão da sazonalidade semanal para cada variável	
Exclusão de componentes senoidais para cada variável	
Exclusão da sazonalidade mensal para cada variável (dias úteis e calendário)	

Quadro 92: Topologia da rede 10.

Topologia da rede (modelo 10)	
Entrada	2 registros referentes ao dia $t - 2$ da variável principal e o respectivo registro atrasado das variáveis auxiliares. No caso da variável TN houve um acréscimo de 1 entrada.
Função de propagação da camada escondida	Tangente hiperbólica
Quantidade de neurônios na entrada	1 e 5
Observações	Escolha dos pares entrada-saída de forma aleatória

Quadro 93: Topologia da rede 11.

Topologia da rede (modelo 11)	
Entrada	4 registros referentes ao dia $t - 2$ da variável principal e os respectivos registros atrasados das variáveis auxiliares (SELIC, SLB e CIP).
Função de propagação da camada escondida	Tangente hiperbólica
Quantidade de neurônios na entrada	1, 5, 10 e 20
Observações	Escolha dos pares entrada-saída de forma aleatória

Quadro 94: Topologia da rede 12.

Topologia da rede (modelo 12)	
Entrada	14 registros referentes ao dia $t - 2$ da variável principal e o respectivo registro atrasado da variável auxiliar. No caso da variável TN houve um acréscimo de 1 entrada.
Função de propagação da camada escondida	Tangente hiperbólica
Quantidade de neurônios na entrada	1 e 5
Observações	Escolha dos pares entrada-saída de forma aleatória

Quadro 95: Topologia da rede 13.

Topologia da rede (modelo 13)	
Entrada 01	2 registros, sendo 1 correspondente as variáveis Compensação, STN/STR, SELIC, TN, SLB e BMF e o outro correspondente a variável alvo. Esta quantidade de entradas se refere ao modelo neural sem representação binária do dia da semana. No caso em que foi inserido a representação binária, a quantidade de entradas passou para 7 elementos. Para a variável CETIP houve mais um registro por conta da existência de 2 atrasos com correlação cruzada significativa ficando, portanto, o arquivo de entrada com 3 elementos no caso de não haver representação binária e 8 elementos quando da representação binária do dia da semana. Para a variável CIP a entrada foi constituída de 5 elementos, referentes aos atrasos dos dias $t - 2$, $t - 3$, $t - 6$ e $t - 8$ do arquivo auxiliar e $t - 2$ da variável alvo para a rede sem representação binária do dia da semana e 10 elementos referentes aos atrasos descritos acima e a representação binária do dia da semana.
Entrada 02	13 elementos referentes aos atrasos dos diversos arquivos juntos para a rede sem representação binária do dia da semana e 18 registros referentes aos atrasos dos diversos arquivos e a representação binária do dia da semana.
Função de propagação da camada escondida	Tangente hiperbólica
Quantidade de neurônios na entrada	1, 2, 5 e 10
Observações	Escolha dos pares entrada-saída de forma aleatória

Quadro 96: Arquivo de dados para a rede 14.

Características do arquivo (modelo 14) – Acima de R\$ 200 milhões
Divisão dos valores por 1 milhão
Redução dos valores para o intervalo [-1000, 2000]
Inclusão da média aritmética nos dias faltantes
Normalização dos dados
Exclusão da tendência
Exclusão da sazonalidade semanal
Exclusão de 1 componente senoidal (períodos de 783,7 dias)
Exclusão da sazonalidade mensal (dias úteis e calendário)
Potência da série = 0,239157

Quadro 97: Topologia da rede 14.

Topologia da rede (modelo 14)	
Entrada 01	1 registro referente ao dia $t - 5$
Entrada 02	6 registros referentes aos dias $t - 5$ e representação binária do dia da semana (1-de-n)
Função de propagação da camada escondida	Tangente hiperbólica
Quantidade de neurônios na entrada	1 e 5

Quadro 98: Arquivo de dados para a rede 15.

Características do arquivo (modelo 15) - Dólar
Divisão dos valores por 1 milhão
Redução dos valores para o intervalo [-1000, 2000]
Inclusão da média aritmética nos dias faltantes
Normalização dos dados
Exclusão da tendência
Exclusão da sazonalidade semanal
Excl.de 6 comp. senoidais (períodos: 783,7, 21,8, 4,4, 3,6, 5,4 e 10,9 dias)
Exclusão da sazonalidade mensal
Excl.de 3 comp. senoidais (períodos: 2,1, 10,5 e 156,8 dias)
Potência da série = 0,138248

Quadro 99: Topologia da rede 15.

Topologia da rede (modelo 15)	
Entrada	9 registros referentes ao dia $t - 2$ da variável principal e os respectivos atrasos da variável auxiliar ($t - 2$ a $t - 9$).
Função de propagação da camada escondida	Tangente hiperbólica
Quantidade de neurônios na entrada	1, 2, 5 e 10
Observações	Escolha dos pares entrada-saída de forma aleatória

Quadro 100: Arquivo de dados para as redes 16, 17, 18 e 19.

Características do arquivo (modelo 16, 17, 18 e 19)
Soma dos saldos das variáveis STR-Clientes e CIP
Divisão dos valores por 1 milhão
Inclusão da média aritmética nos dias faltantes
Normalização dos dados
Exclusão da tendência
Exclusão da sazonalidade semanal
Exclusão de 1 componente senoidal na variável STR-Clientes (783,7 dias)
Exclusão de sazonalidades mensal (dias úteis e calendário)
Potência da série = 0,142292

Quadro 101: Topologia da rede 16.

Topologia da rede (modelo 16)	
Entrada 01	1 registro referente ao dia $t - 11$
Entrada 02	6 registros referentes aos dias $t - 11$ e representação binária do dia da semana (1-de-n)
Função de propagação da camada escondida	Tangente hiperbólica
Quantidade de neurônios na entrada	1 e 5
Observações	Escolha dos pares entrada-saída de forma sequencial e aleatória Pares com e sem representação binária do dia da semana

Quadro 102: Topologia da rede 17.

Topologia da rede (modelo 17)	
Entrada	13 elementos para as variáveis STR_CIP, Numerário, STN-STR, OB, Câmbio, SELIC, TN, SLB, STR-IF referentes aos atrasos definidos no quadro acima para o caso de entradas sem representação do dia da semana e 18 elementos com a representação do dia da semana binário.
Função de propagação da camada escondida	Tangente hiperbólica
Quantidade de neurônios na entrada	1 e 5
Observações	Escolha dos pares entrada-saída de forma aleatória Pares com e sem representação binária do dia da semana

Quadro 103: Topologia da rede 18.

Topologia da rede (modelo 18)	
Entrada 01	1 registro referente ao dia $t - 2$
Função de propagação da camada escondida	Tangente hiperbólica
Quantidade de neurônios na entrada	1, 2, 5 e 10
Observações	Escolha dos pares entrada-saída de forma sequencial Pares com e sem representação binária do dia da semana

Quadro 104: Topologia da rede 19.

Topologia da rede (modelo 19)	
Entrada 01	1 registro referente ao dia t - 2
Função de propagação da camada escondida	Tangente hiperbólica
Quantidade de neurônios na entrada	1, 2, 5 e 10
Observações	Escolha dos pares entrada-saída de forma sequencial Pares com e sem representação binária do dia da semana Atrasos na 1ª camada escondida variando de 1 a 5 Atrasos na camada de saída variando de 1 a 3

Quadro 105: Transformações no arquivo de dados para aplicação nos modelos ARIMA, difuso e ICA.

Características do arquivo (Modelos ARIMA, Difuso e ICA)
Soma dos saldos das variáveis STR-Clientes e CIP
Divisão dos valores por 1 milhão
Inclusão da média aritmética nos dias faltantes
Normalização dos dados
Exclusão da tendência
Exclusão da sazonalidade semanal
Exclusão de 1 componente senoidal na variável STR-Clientes (783,7 dias)
Exclusão de sazonalidades mensal (dias úteis e calendário)
Potência da série = 0,137203

6. QUADROS DE FREQUÊNCIAS DAS VARIÁVEIS

As linhas em destaque correspondem aos valores cujos intervalos foram modificados. Os valores agregados representam menos de 1% do total de registros.

Quadro 106: Distribuição de frequência da variável Numerário.

01. Numerário	Qtd.	Acumulado	Percentual	Percentual acumulado
-2000,0 < x <= -1500,0	1	1	0,1323	0,1323
-1500,0 < x <= -1000,0	9	10	1,1905	1,3228
-1000,0 < x <= -500,00	91	101	12,03704	13,35979
-500,00 < x <= 0,0000	410	511	54,2328	67,5926
0,0000 < x <= 500,000	244	755	32,2751	99,8677
500,000 < x <= 1000,00	1	756	0,13228	100

Quadro 107: Distribuição de frequência da variável Compensação.

02. Compensação	Qtd.	Acumulado.	Percentual	Percentual acumulado
-800,00 < x <= -600,00	1	1	0,1323	0,1323
-600,00 < x <= -400,00	2	3	0,2646	0,3968
-400,00 < x <= -200,00	79	82	10,4497	10,8466
-200,00 < x <= 0,0000	421	503	55,6878	66,5344
0,0000 < x <= 200,000	237	740	31,3492	97,8836
200,000 < x <= 400,000	15	755	1,9841	99,8677
400,000 < x <= 600,000	1	756	0,1323	100

Quadro 108: Distribuição de frequência da variável STN/STR.

03. Arrec. via STN / STR	Qtd.	Cumul.	Percentual	Percentual acumulado
-5000,0 < x <= -4000,0	2	2	0,2646	0,2646
-4000,0 < x <= -3000,0	12	14	1,5873	1,8519
-3000,0 < x <= -2000,0	45	59	5,952	7,8042
-2000,0 < x <= -1000,0	78	137	10,3175	18,1217
-1000,0 < x <= 0,0000	618	755	81,7460	99,8677
0,0000 < x <= 1000,00	1	756	0,1323	100

Quadro 109: Distribuição de frequência da variável Ordens Bancárias.

04. Ordens Bancárias	Qtd.	Cumul.	Percentual	Percentual acumulado
-2000,0 < x <= -1000,0	1	1	0,1323	0,1323
-1000,0 < x <= 0,0000	114	115	15,0794	15,2116
0,0000 < x <= 1000,00	518	633	68,5183	83,7302
1000,00 < x <= 2000,00	84	717	11,1111	94,8413
2000,00 < x <= 3000,00	21	738	2,7778	97,6190
3000,00 < x <= 4000,00	12	750	1,5873	99,2063
4000,00 < x <= 5000,00	5	755	0,6614	99,8677
5000,00 < x <= 6000,00	1	756	0,1323	100

Quadro 110: Distribuição de frequência da variável Câmbio.

05. Câmbio	Qtd.	Cumul.	Percentual	Percentual acumulado
-1500,0 < x <= -1000,0	1	1	0,1323	0,1323
-1000,0 < x <= -500,00	29	30	3,8360	3,9683
-500,00 < x <= 0,0000	345	375	45,6349	49,6032
0,0000 < x <= 500,000	371	746	49,0741	98,6772
500,000 < x <= 1000,00	9	755	1,1905	99,8677
1000,00 < x <= 1500,00	1	756	0,1323	100

Quadro 111: Distribuição de frequência da variável Compulsório.

06. Compulsório	Qtd.	Cumul.	Percentual	Percentual acumulado
-400,00 < x <= -300,00	1	1	0,1323	0,1323
-300,00 < x <= -200,00	10	11	1,3228	1,4550
-200,00 < x <= -100,00	21	32	2,7778	4,2328
-100,00 < x <= 0,0000	86	118	11,3757	15,6085
0,0000 < x <= 100,000	616	734	81,4815	97,0899
100,000 < x <= 200,000	18	752	2,3810	99,4709
200,000 < x <= 300,000	4	756	0,5291	100

Quadro 112: Distribuição de frequência da variável SELIC.

07. SELIC	Qtd.	Cumul.	Percentual	Percentual acumulado
-8000,0 < x <= -6000,0	2	2	0,2646	0,2646
-6000,0 < x <= -4000,0	13	15	1,7196	1,9841
-4000,0 < x <= -2000,0	63	78	8,3333	10,3175
-2000,0 < x <= 0,0000	291	369	38,4921	48,8095
0,0000 < x <= 2000,00	318	687	42,0635	90,8730
2000,00 < x <= 4000,00	59	746	7,8042	98,6772
4000,00 < x <= 6000,00	9	755	1,1905	99,8677
6000,00 < x <= 8000,00	1	756	0,1323	100

Quadro 113: Distribuição de frequência da variável CETIP.

08. CETIP	Qtd.	Cumul.	Percentual	Percentual acumulado
-1500,0 < x <= -1000,0	2	2	0,2646	0,2646
-1000,0 < x <= -500,00	6	8	0,7937	1,0582
-500,00 < x <= 0,0000	423	431	55,9524	57,0106
0,0000 < x <= 500,000	322	753	42,5926	99,6031
500,000 < x <= 1000,00	2	755	0,2646	99,8677
1000,00 < x <= 1500,00	1	756	0,1323	100

Quadro 114: Distribuição de frequência da variável Tesouro Nacional - Outras.

09. Tesouro Nac. - Outras	Qtd.	Cumul.	Percentual	Percentual acumulado
-2000,0 < x <= 0,0000	140	140	18,5185	18,5185
0,0000 < x <= 2000,00	579	719	76,5873	95,1058
2000,00 < x <= 4000,00	17	736	2,2487	97,3545
4000,00 < x <= 6000,00	18	754	2,3810	99,7354
6000,00 < x <= 8000,00	2	756	0,264550265	100

Quadro 115: Distribuição de frequência da variável Lançamentos Bacen - SLB.

11. Lanç. BACEN – SLB	Qtd.	Cumul.	Percentual	Percentual acumulado
-1500,0 < x <= -1000,0	1	1	0,1323	0,1323
-1000,0 < x <= -500,00	0	1	0	0,1323
-500,00 < x <= 0,0000	227	228	30,0265	30,1587
0,0000 < x <= 500,000	527	755	69,7090	99,8677
500,000 < x <= 1000,00	1	756	0,1323	100

Quadro 116: Distribuição de frequência da variável BMF-Derivativos.

13. BMF - Derivativos	Qtd.	Cumul.	Percentual	Percentual acumulado
-100,00 < x <= -50,000	1	1	0,1323	0,1323
-50,000 < x <= 0,0000	398	399	52,6455	52,7778
0,0000 < x <= 50,0000	355	754	46,9577	99,7354
50,0000 < x <= 100,000	1	755	0,1323	99,8677
100,000 < x <= 150,000	1	756	0,1323	100

Quadro 117: Distribuição de frequência da variável CBLC.

14. CBLC	Qtd.	Cumul.	Percentual	Percentual acumulado
-1500,0 < x <= -1000,0	3	3	0,3968	0,3968
-1000,0 < x <= -500,00	2	5	0,26455	0,6614
-500,00 < x <= 0,0000	329	334	43,5185	44,1799
0,0000 < x <= 500,000	419	753	55,4233	99,6032
500,000 < x <= 1000,00	1	754	0,1323	99,7354
1000,00 < x <= 1500,00	1	755	0,1323	99,8677
1500,00 < x <= 2000,00	0	755	0	99,8677
2000,00 < x <= 2500,00	1	756	0,1323	100

Quadro 118: Distribuição de frequência da variável STR-IF.

16. STR – IF	Qtd.	Cumul.	Percentual	Percentual acumulado
-600,00 < x <= -400,00	1	1	0,1323	0,1323
-400,00 < x <= -200,00	1	2	0,1323	0,2646
-200,00 < x <= 0,0000	465	467	61,5079	61,7725
0,0000 < x <= 200,000	227	694	30,0265	91,7989
200,000 < x <= 400,000	50	744	6,6138	98,4127
400,000 < x <= 600,000	9	753	1,1905	99,6032
600,000 < x <= 800,000	1	754	0,1323	99,7354
800,000 < x <= 1000,00	2	756	0,2646	100

Quadro 119: Distribuição de frequência da variável STR-Clientes.

17. STR – Clientes	Qtd.	Cumul.	Percentual	Percentual acumulado
-2000,0 < x <= -1000,0	3	3	0,3968	0,3968
-1000,0 < x <= 0,0000	164	167	21,6931	22,0899
0,0000 < x <= 1000,00	522	689	69,0476	91,1376
1000,00 < x <= 2000,00	54	743	7,1429	98,2804
2000,00 < x <= 3000,00	12	755	1,5873	99,8677
3000,00 < x <= 4000,00	0	755	0	99,8677
4000,00 < x <= 5000,00	1	756	0,1323	100

Quadro 120: Distribuição de frequência da variável CIP.

18. CIP	Qtd.	Cumul.	Percentual	Percentual acumulado
-1500,0 < x <= -1000,0	28	28	3,7037	3,7037
-1000,0 < x <= -500,00	260	288	34,3915	38,0952
-500,00 < x <= 0,0000	397	685	52,5132	90,6085
0,0000 < x <= 500,000	70	755	9,2593	99,8677
500,000 < x <= 1000,00	0	755	0	99,8677
1000,00 < x <= 1500,00	1	756	0,1323	100

7. CÁLCULOS DE CORRELAÇÕES

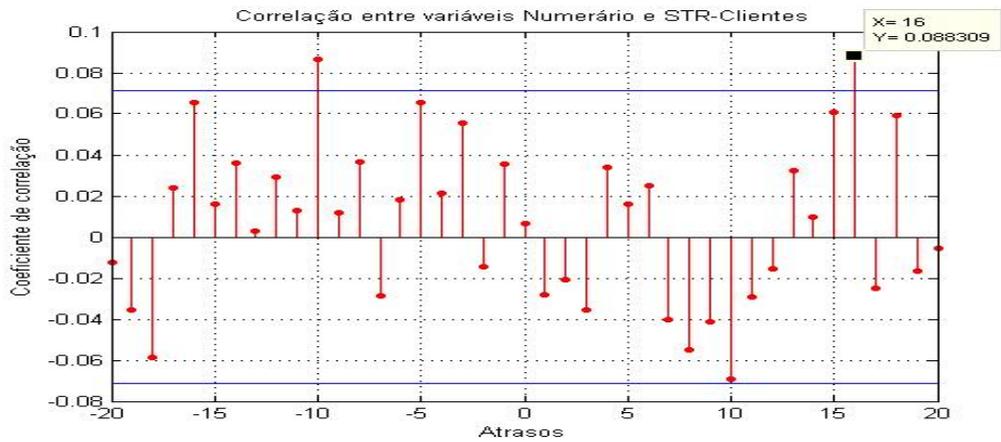


Gráfico 69: Correlação entre variáveis Numerário e STR-Clientes.

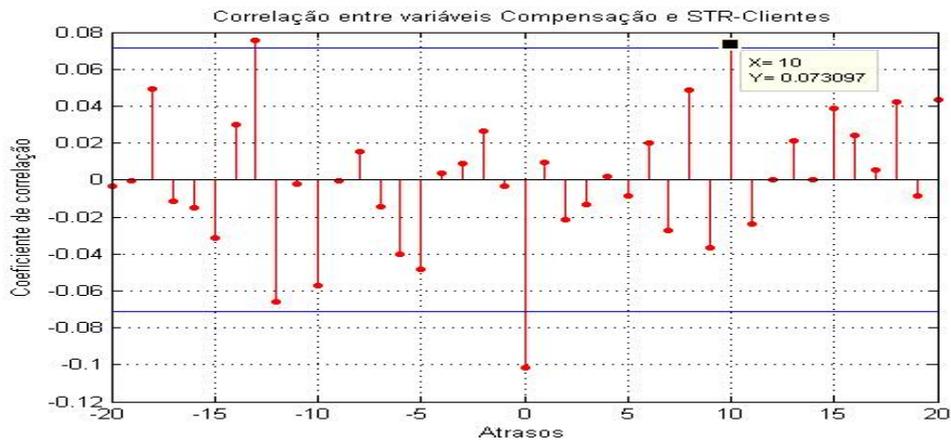


Gráfico 70: Correlação entre variáveis Compensação e STR-Clientes.

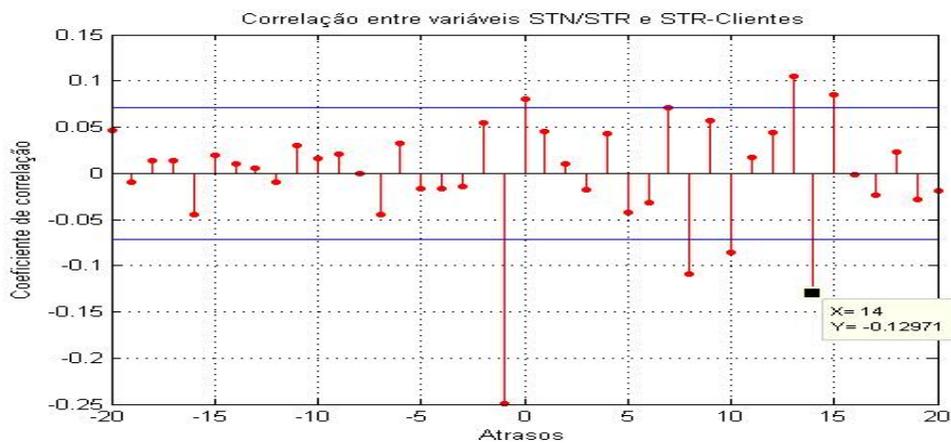


Gráfico 71: Correlação entre variáveis STN/STR e STR-Clientes.

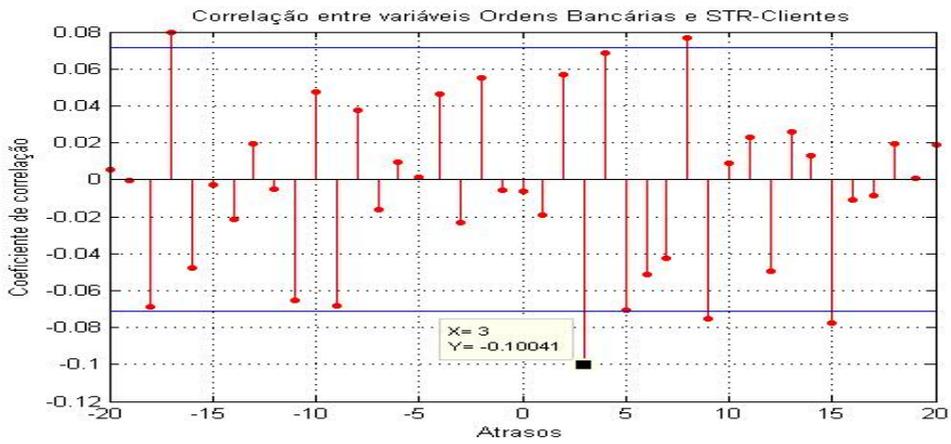


Gráfico 72: Correlação entre variáveis Ordens Bancárias e STR-Clientes.

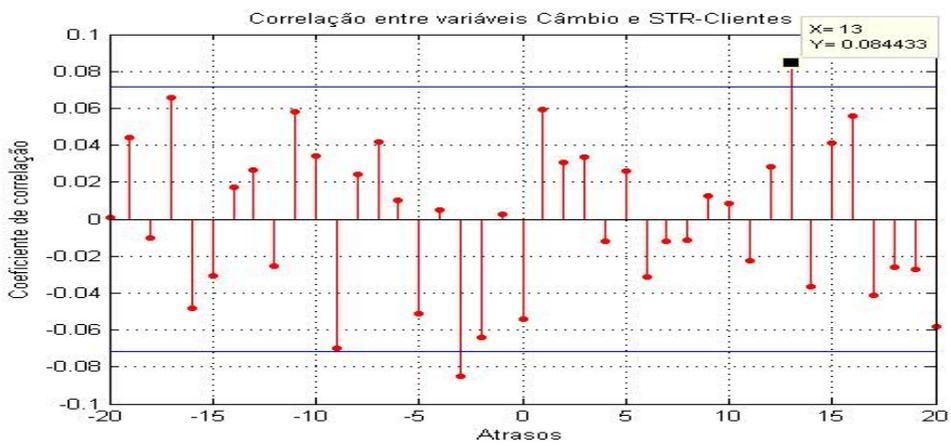


Gráfico 73: Correlação entre variáveis Câmbio e STR-Clientes.

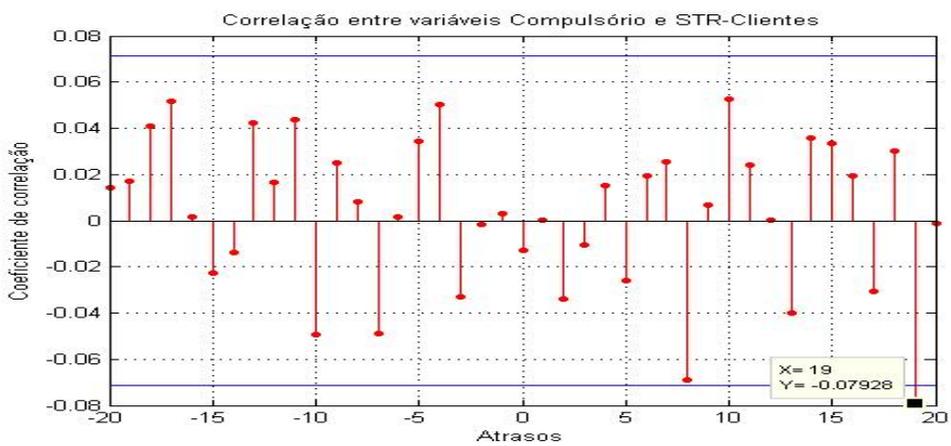


Gráfico 74: Correlação entre variáveis Compulsório e STR-Clientes.

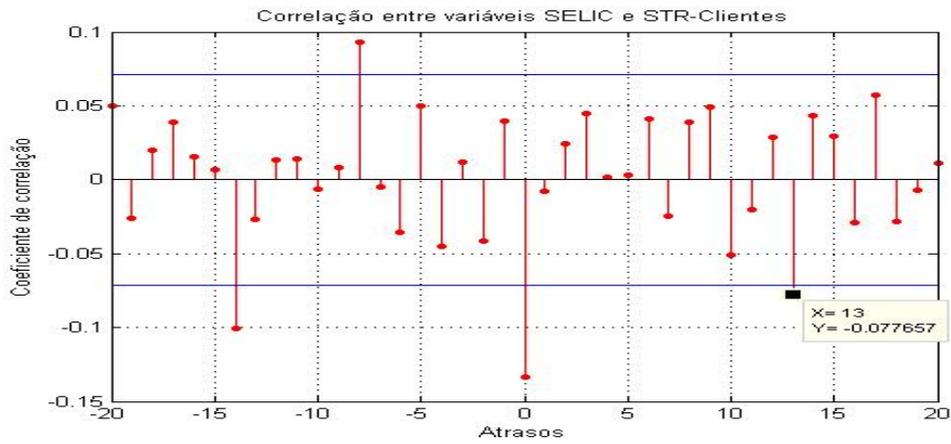


Gráfico 75: Correlação entre variáveis SELIC e STR-Clientes.

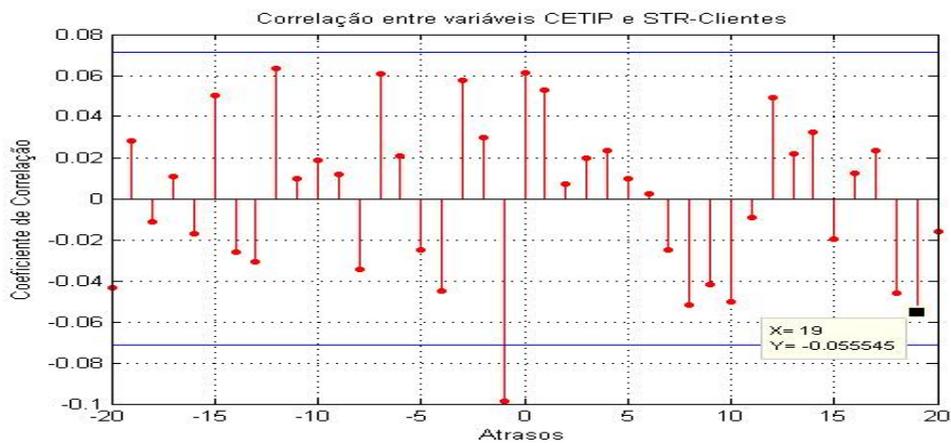


Gráfico 76: Correlação entre variáveis CETIP e STR-Clientes.



Gráfico 77: Correlação entre variáveis Tesouro Nacional e STR-Clientes.

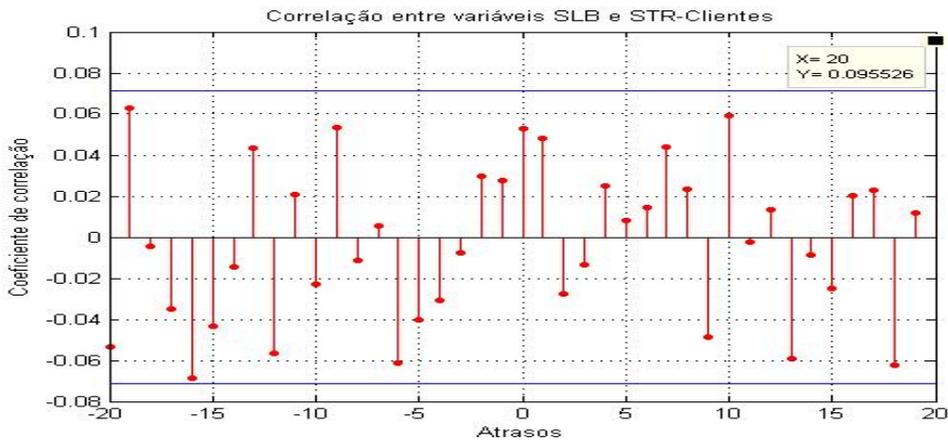


Gráfico 78: Correlação entre variáveis SLB e STR-Clientes.

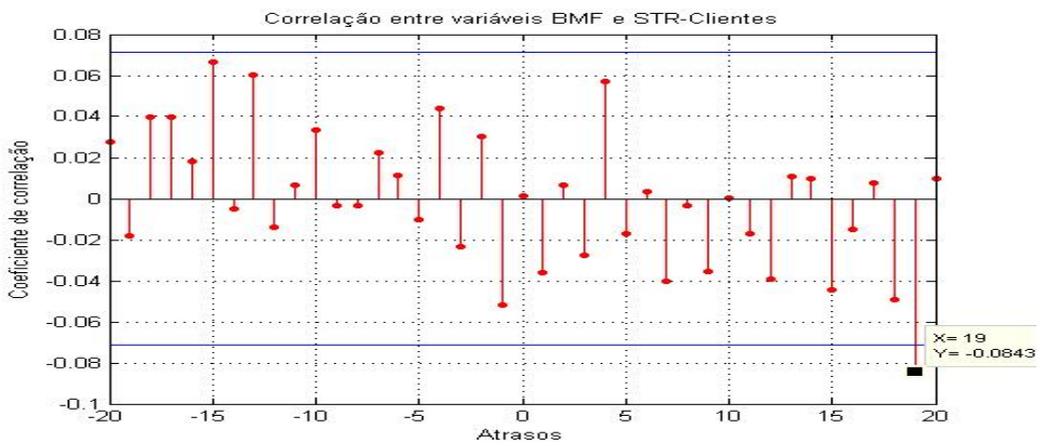


Gráfico 79: Correlação entre variáveis BMF e STR-Clientes.

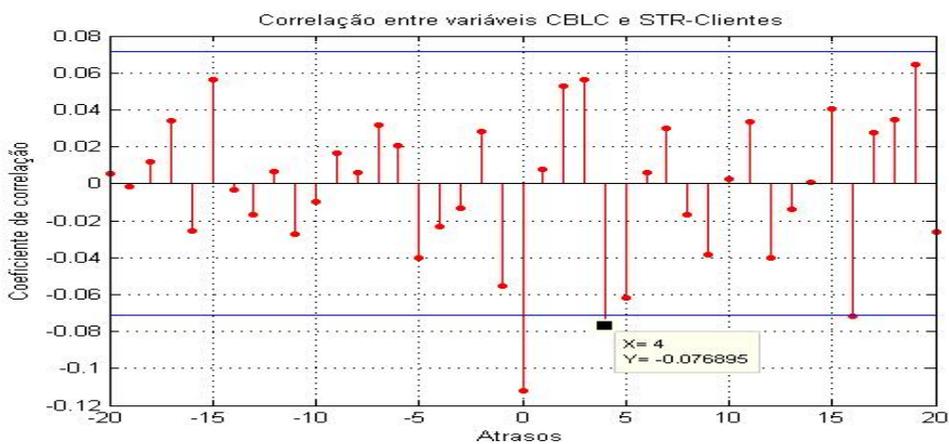


Gráfico 80: Correlação entre variáveis CBLC e STR-Clientes.

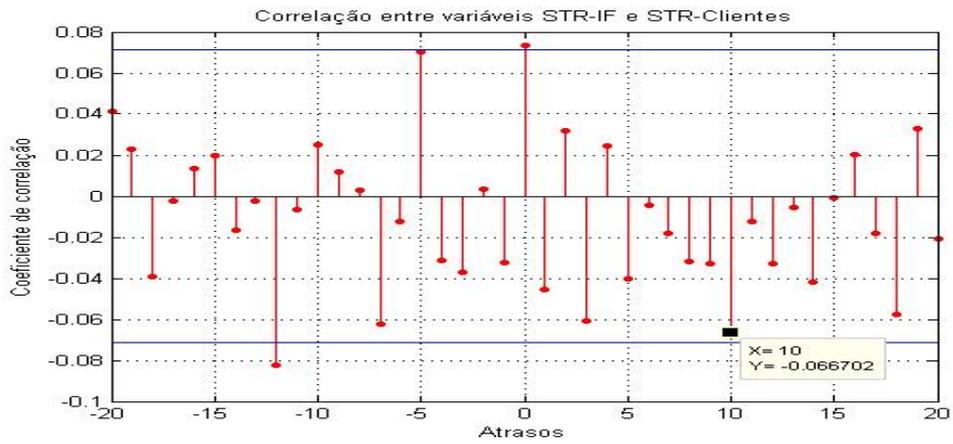


Gráfico 81: Correlação entre variáveis STR-IF e STR-Clientes.

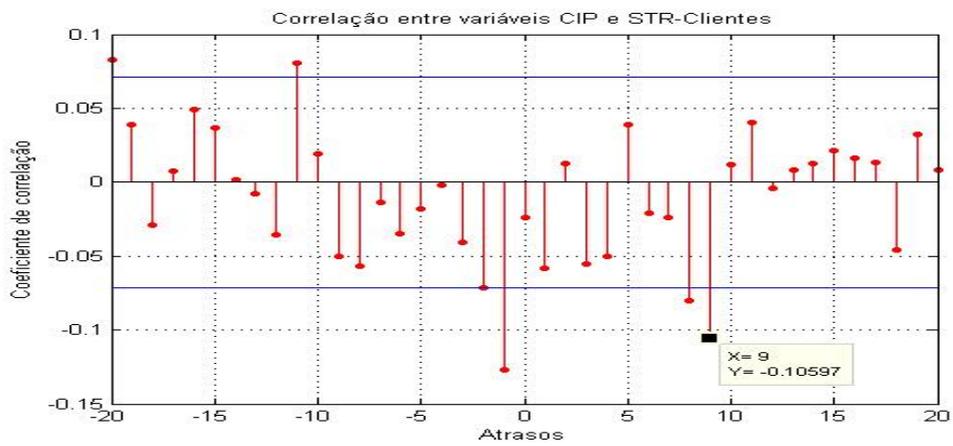


Gráfico 82: Correlação entre variáveis CIP e STR-Clientes.

9 REFERÊNCIAS

- ABELÉM, A. J. G. **Redes Neurais Artificiais na Previsão de Séries Temporais**. 1994. 108 p. Dissertação (Mestrado em Engenharia Elétrica) – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro.
- AFIF, A. A.; AZEN, J. P. **Statistical Analysis – A Computer Oriented Approach**. 2a. ed. New York: Academic Press, Inc. 1979. 442 p.
- ALLEMÃO, M. A. F. **Redes Neurais Aplicadas à Previsão de Demanda de Numerário em Agências Bancárias**. Dissertação (Mestrado em Informática) – Departamento de Ciência da Computação, Universidade Federal do Rio de Janeiro, Rio de Janeiro. 2004.
- BACHA, H.; MEYER, W. **A Neural Network Architecture for Load Forecasting**. IEEE International Joint Conference on Neural Networks, 2, p. 442-447. 1992. 6 p.
- BASILÉIA. **Sound Practices for Managing Liquidity in Banking Organisations**. Basileia, Suíça. 2000. 24 p.
- BISHOP, C. M. **Neural Networks for Pattern Recognition**. Oxford: Oxford University Press Inc., 1996. 482 p.
- BOUQATA, B; BENSALID, A.; PALLIAM, R.; GOMEZ SKARMETA, A. F. **Time Series Prediction using Crisp and Fuzz Neural Networks: A Comparative Study**. Computational Intelligence for Financial Engineering. p.170-173. 2000. 4 p.
- BOWERMAN, B. L.; O'CONNELL, R. T. **Forecasting & Time Series**. Belmont: Duxbury Press., 1979. 481 p.
- BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C. **Time Analysis Forecasting and Control**. New Jersey: Prentice Hall, 1994. 598 p.
- BRAGA, A. P.; CARVALHO, A. P. L.; LUDEMIR, T. B. **Fundamentos de Redes Neurais Artificiais – 11ª Escola de Computação**. Rio de Janeiro: Imprinta Gráfica e Editora Ltda., 1998. 246 p.

_____. **Redes Neurais Artificiais: Teoria e Aplicações**. Rio de Janeiro: LTC – Livros Técnicos e Científicos Editora S. A., 2007. 226 p.

BRAGA, P. C. A., CALÔBA, L. P. **Ibovespa Forecasting Using Hybrid Models**. Proceedings of the Computational Economics & Financial and Industrial Systems (IFAC-CEFIS 2007), Istanbul, Turkey, vol. -, pp. 1-8, October 9-11, 2007.

BRUNI, A. L. **A Eficiência Informacional do Mercado de ADRS Brasileiros: Uma Análise com Testes de Autocorrelação, Raiz Unitária e Cointegração**. Gestão & Planejamento. Revista do Programa de Pós-Graduação em Administração. Ano 5, N° 9 (Jan/Jun). ISSN 1516-9103. Salvador, 2004. 79 p.

BUSSAB, W.; MORETTIN, P. A. **Métodos Quantitativos – Estatística Básica**. 4ª ed. São Paulo: Saraiva S/A Livreros Editores, 2001. 321 p.

CHATFIELD, C. **The Analysis of Time Sereies – An Introduction**. 6ª ed. Florida: CRC Press LLC, 2004, 333 p.

CICHOCKI, A.; UNBERHAUEN, R. **Neural Networks for Optimization and Signal Processing**. New York: John Wiley & Sons, 1995. 526 p.

DALBELLO, L. **A Relevância do Fluxo de Caixa como Ferramenta de Gestão Financeira para Avaliação da Liquidez e Capacidade de Financiamento de Empresas**. Dissertação (Mestrado em Engenharia de Produção) – Universidade Federal de Santa Catarina – UFSC, Santa Catarina, 1999, 181 p.

EBECKEN, N. F. F. **Resumo de Probabilidade e Estatística**. Apostila do curso “Reconhecimento de Padrões”. Rio de Janeiro: COPPE/UFRJ, 2001. 60 p.

ELMAN, J. L. **Finding Structure in Time**. Cognitive Science, 14, 179-211. 1990, 28 p.

FARIA, A. N. **Organização de Empresas**. Distribuidora Record de Serviços de Imprensa Ltda. – vol 1, Rio de Janeiro, 1967, 260 p.

FAZENDA, Ministério da. **Tesouro Nacional - Glossário**. Disponível em http://www.tesouro.fazenda.gov.br/tesouro_direto/servicos/glossario/glossario_1.asp. Acessado em 28/12/2005.

FERNANDES, L. G. L.; NAVAU, P. O.; PORTUGAL, M. S., **Previsão de Séries de Tempo: Redes Neurais Artificiais e Modelos Estruturais**. XVII Encontro Brasileiro de Econometria, Salvador, 1996, v.26, n.2, p. 253-276.

GAO, D.; KINOUCI, Y.; ITO, K.; ZHAO, X. **Time Series Identifying and Modeling with Neural Networks**. IEEE 0-7803-7898/03. 2003., 6 p.

_____. **Neural networks for event extraction from time series: a back propagation algorithm approach**. Future Generation Computer Systems 21 (2005) 1096-1105 0167-739X doi: 10.1016/j.future.2004.03.009. Elsevier. 2005, 10 p.

HAIR JR., J. F.; ANDERSON, R. E. *et al.* **Multivariate Data Analysis** 5. ed. New Jersey: Prentice Hall, 1998. 730 p.

HAN, J.; KAMBER, M. **Data Mining: Concepts and Techniques**. San Francisco: Morgan Kaufmann Publishers, 2001. 550 p.

HAYKIN, S. **Neural Networks - A Comprehensive Foundation** 2. ed. New Jersey: Prentice Hall, 1999. 842 p.

_____. **Redes Neurais – Princípios e Prática**. 2. ed. Porto Alegre: Bookman, 2001. 900 p.

HLAVÁČEK, M.; CADA, J.; HAKL, F. **The Application of Structured Feedforward Neural Networks to the Modelling of the Daily Series of Currency in Circulation**. ICNC 2005, LNCS 3610, 1234-1246, Springer-Verlag, Berlin, 2005, 13 p.

HO, S.L.; XIE, M.; GOH, T.N. **A Comparative Study of Neural Network and Box-Jenkins ARIMA Modeling in Time Series Prediction**. Computer & Industrial Engineering 42 (2002) 371-375, 5 p.

HYVÄRINEN, Aapo; KARHUNEN, Juha, OJA, Erkki. **Independent Component Analysis**. Ed. John Willey & Sons, Inc. New York. 2001. 481 p.

KIM, Yoon; OH, Kyong J.; KIM, Chiho; DO, Jong D. **Artificial neural networks for non-stationary time series**. Neurocomputing 61 (2004) 439-447. doi: 10.1016/J.neucom.2004.04.002, 2004, 9 p.

KIVILUOTO, Kimmo; OJA, Erkki. **Independent Component Analysis for Parallel Financial Time Series**. ICONIP 98. Kitayushu. Japan. Oct. 1998, 895-898. 4 p.

KOVÁCS, Z. L. **Redes Neurais Artificiais: Fundamentos e Aplicações** 2ª ed. São Paulo: Edição Acadêmica, 1996. 174 p.

KUO, R.J.; XUE K.C. **Fuzzy Neural Networks with Application to Sales**

- Forecasting.** Fuzzy Sets and Systems 108, 123-143. 1999. 21 p.
- LAM, M. **Neural network techniques for financial performance prediction: integrating fundamental and technical analysis.** Decision Support Systems 37 (2004) 567-581. 0167-923/03. doi:10.1016/S0167-9236(03)00088-5. 2005, 15 p.
- LEVINE, D. M.; BERESON, M. L.; STEPHAN, D. **Estatística: Teoria e Aplicações Usando o Microsoft Excel em Português.** Rio de Janeiro: LTC - Livros Técnicos e Científicos Editora S.A., 2000. 812 p.
- LIMA, E. C. P.; SAGRILO, L. V. S. **Confiabilidade Estrutural.** Rio de Janeiro: Curso MSO – Programa de Engenharia Civil – UFRJ/COPPE. Disponível em <http://www.oceanica.ufrj.br/MSO/downloads/ceoff/>. Acesso em 23/05/2003.
- LITTLE, R. J. A.; RUBIN, D. B. **Statistical Analysis with Missing Data.** New York: John Wiley & Sons, 1987. 278 p.
- LOCKMIC, L.; SMITH, K. A. **Cash Flow Forecasting using Supervised and Unsupervised Neural Networks.** IEEE 0-7695-0619-4/00, 2000. 5 p.
- LOYOLA, G. **Regulação Prudencial e Mercado de Câmbio – Aprimorando o Mercado de Câmbio Brasileiro.** Disponível em <http://www.bmf.com.br/2004/pages/imprensa1/destaques/2003/dezembro/pdf/GustavoLoyola.pdf>. São Paulo, 04/12/2003. 12 p. Acessado em 02/04/2006.
- MONTGOMERY, D. C.; JOHNSON, L. A.; GARDINER, J. S. **Forecasting & Time Series Analysis** 2ª ed. New York: McGraw-Hill Inc., 1990. 381 p.
- MORETTIN, P. A.; TOLOI, C. M. C. **Previsão de Séries Temporais.** 2. ed. São Paulo: Atual Editora Ltda, 1987. 439 p.
- MORETTIN, P. A. **Econometria Financeira – Um Curso em Séries Temporais Financeiras.** São Paulo: Editora Edgard Blucher Ltda, 2008. 336 p.
- MORCILLO, A. M. **Teste do qui-quadrado (X^2).** Disponível em http://www.fcm.unicamp.br/centros/ciped/mp639/teste%20Qui_quadrado.pdf. Acessado em 02/04/2004.
- _____. **Teste de Kolmogorov-Smirnov.** Disponível em <http://www.fcm.unicamp.br/centros/ciped/mp639/Teste%20de%20Kolmogorov.pdf>. Acessado em 13/04/2004.

- MORETTIN, P. A.; TOLOI, Clélia M. C. **Previsão de Séries Temporais**. 2ª ed. São Paulo: Atual Editora Ltda, 1987. 439 p.
- NASCIMENTO JR., C. L.; YONEYAMA, T.. **Inteligência Artificial em Controle e Automação**. 1ª reimpressão. São Paulo: Editora Edgard Blucher Ltda., 2002. 218 p.
- NETER, J.; WASSERMAN, W.; WHITMORE, G. A.. **Applied Statistics**. 4. ed. Boston: Simon & Schuster, Inc., 1992. 992 p.
- NGUYEN, H. H., CHAN, C. W. **Multiple neural networks for a long term time series forecast**. *Neural Comput & Applic* 13:90-98 doi 10.1007/s00521-003-0390-z, 2004, 9 p.
- OLIVEIRA JR., H. A. **Lógica Difusa – Aspectos práticos e aplicações**. Ed. Interciência Ltda. 1999. 192 p.
- PARKINSON, K. L.; KALLBERG, J. G. **Corporate Liquidity, Business, One Irwin**, New York, 1993.
- PASLEY, A.; AUSTIN, J. **Distribution forecasting of high frequency time series**. *Decision Support Systems*. 37 (2004) 501-513. doi: 10.1016/S0167-923(03)00083-6. 2004, 13 p.
- PYLE, D. **Data Preparation for Data Mining**, Morgan Kaufmann Publishers, Inc, San Francisco, 1999. 535 p.
- QUEIROZ, M. F. **Comportamento Diário do Mercado Brasileiro de Reservas Bancárias – Nível e Volatilidade – Implicações na Política Monetária**. Proceedings of the 32th Brazilian Economics Meeting. RePEc:anp:en2004:096. 2004.
- QUEJI, L. M. **Modelo de Fluxo de Caixa Prospectado para Pequenas Empresas Comerciais à Luz do Ciclo de Vida**. Dissertação (Mestrado em Engenharia de Produção) – Universidade Federal de Santa Catarina – UFSC, Santa Catarina, 2002, 158 p.
- REZENDE, S. O. (Organizadora) **Sistemas Inteligentes – Fundamentos e Aplicações**. Ed. Manole, São Paulo, 2003. 525 p.
- ROCHA, D. M., **Redes Neurais para Modelagem de Sistemas Estruturais Offshore Dinâmicos Não-Lineares com Histerese**. Tese de D.Sc., COPPE/UFRJ, Rio de Janeiro, RJ, Brasil, 2007.

ROCHA, D. M.; KAYSER JÚNIOR, D. L.; EBECKEN, N. F. F.; CALÔBA, L. P. **Redes Neurais para Modelagem de Sistemas Estruturais Offshore Dinâmicos Não Lineares com Histerese**. Cadernos do IME – Série Estatística – vol. 23. Universidade do Estado do Rio de Janeiro – UERJ, Rio de Janeiro, 2007, 56 p.

SHEEN, J.N. **Fuzzy Financial Profitability Analyses of Demand Side Management Alternatives from Participant Perspective**. Information Sciences 169 (2005) 329-364. Elsevier Science Direct. 2004. 36 p.

SILVA, M. A. V. R. **A Hipótese da Eficiência do Mercado Acionário e Algumas Anomalias Encontradas**. Disponível em <http://www.unitau.br/nupes/hipotese.htm>. Acessado em 19/03/2004.

SILVEIRA, M. **Finanças – Noções Gerais**. Disponível em http://www.rits.org.br/gestao_teste/ge_testes/ge_mat01_financtxt0.cfm. Acessado em 10/10/2008.

SOUSA, A. F.; BARROS, L. A. B. C. **Propriedades Estatísticas dos Fluxos de Caixa e Modelos de Gerenciamento de Caixa**. Caderno de Pesquisas em Administração, São Paulo, v. 01, n° 12, 2° trim./2000, São Paulo, SP.

SPIEGEL, M. R. **Estatística**. Rio de Janeiro: SEDEGRA – Sociedade Editora e Gráfica Ltda, 1961. 580 p.

STEVENSON, W. J. **Estatística Aplicada à Administração**. São Paulo: Harbra Ltda., 2001. 495 p.

TAKEMURA, R. Y. **Controle Inteligente – Tutorial**. Universidade Estadual de Maringá. Disponível em http://www.din.uem.br/ia/control/fuz_prin.htm. Acessado em 10/09/2008.

TASKAYA-TEMIZEL, T.; AHMAD, K. **Are ARIMA Neural Network Hybrids Better than Single Models?**. International Joint Conference on Neural Networks. Montreal, Canadá, vol. 5. p. 3192-3197. 2005. 6 p.

VILLALBA, G. B.; SOUSA, A. F. **Modelos de Administração de Caixa – Análise Empírica**. V SEMEAD, Ensaio Finanças, junho 2001. Internet. <http://www.ead.fea.usp.br/semead/5semead/Finan%20as/Modelos%20de%20Administracao%20de%20Caixa.pdf>. Acesso em: 26 jul. 2004.

WALCZAK, S. **An Empirical Analysis of Data Requirements for Financial Forecasting with Neural Networks**. Journal of Management Information Systems. Spring 2001. Vol. 17. n° 4. p. 203-222. 2001. 20 p.

ZADEH, L. A. **Fuzzy Sets**. Information and Control, n.8, p. 338-353. 1965. 16 p.

ZANETI Jr., L. A.; ALMEIDA, F. C. **Exploração do Uso de Redes Neurais na Previsão do Comportamento de Ativos Financeiros**. III SEMEAD, Ensaio Finanças. Internet. <http://www.fia.com.br/labfin/pesquisa/artigos/arquivos/235.pdf>. Acesso em: 26 jul. 2004.

ZDANOWICZ, J. E. **Fluxo de Caixa: uma decisão de planejamento e controle financeiros**. Porto Alegre, D. C. Luzzatto Editores Ltda., 1989, 270 p.

ZHANG, G. P. **Time series forecasting usign a hybrid ARIMA and neural network model**. Neurocomputing 50 (2003) 159-175, 17 p.

ZHANG, G.; PATUWO, B. Eddy; HU, Michael Y. **Forecasting with artificial neural networks: The state of the art**. International Journal of Forecasting. 14 (1998) 35-62, 28 p.

Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)